

INFORMATION TECHNOLOGY AND INTELLIGENT  
TRANSPORTATION SYSTEMS

# Frontiers in Artificial Intelligence and Applications

The book series Frontiers in Artificial Intelligence and Applications (FAIA) covers all aspects of theoretical and applied Artificial Intelligence research in the form of monographs, doctoral dissertations, textbooks, handbooks and proceedings volumes.

The FAIA series contains several sub-series, including 'Information Modelling and Knowledge Bases' and 'Knowledge-Based Intelligent Engineering Systems'. It also includes the biennial European Conference on Artificial Intelligence (ECAI) proceedings volumes, and other EurAI (European Association for Artificial Intelligence, formerly ECCAI) sponsored publications. An editorial panel of internationally well-known scholars is appointed to provide a high quality selection.

Series Editors:

J. Breuker, N. Guarino, J.N. Kok, J. Liu, R. López de Mántaras,  
R. Mizoguchi, M. Musen, S.K. Pal and N. Zhong

## Volume 296

*Recently published in this series*

- Vol. 295. J. Mizera-Pietraszko, R. Rodriguez Jorge, D.M. Almazo Pérez and P. Pichappan (Eds.), *Advances in Digital Technologies – Proceedings of the 8th International Conference on Applications of Digital Information and Web Technologies ICADIWT 2017*
- Vol. 294. F. Bex and S. Villata (Eds.), *Legal Knowledge and Information Systems – JURIX 2016: The Twenty-Ninth Annual Conference*
- Vol. 293. S. Sun, A.J. Tallón-Ballesteros, D.S. Pamučar and F. Liu (Eds.), *Fuzzy Systems and Data Mining II – Proceedings of FSDM 2016*
- Vol. 292. H. Jaakkola, B. Thalheim, Y. Kiyoki and N. Yoshida (Eds.), *Information Modelling and Knowledge Bases XXVIII*
- Vol. 291. G. Arnicans, V. Arnicane, J. Borzovs and L. Niedrite (Eds.), *Databases and Information Systems IX – Selected Papers from the Twelfth International Baltic Conference, DB&IS 2016*
- Vol. 290. J. Seibt, M. Nørskov and S. Schack Andersen (Eds.), *What Social Robots Can and Should Do – Proceedings of Robophilosophy 2016 / TRANSOR 2016*
- Vol. 289. I. Skadiņa and R. Rozis (Eds.), *Human Language Technologies – The Baltic Perspective – Proceedings of the Seventh International Conference Baltic HLT 2016*
- Vol. 288. À. Nebot, X. Binefa and R. López de Mántaras (Eds.), *Artificial Intelligence Research and Development – Proceedings of the 19th International Conference of the Catalan Association for Artificial Intelligence, Barcelona, Catalonia, Spain, October 19–21, 2016*

ISSN 0922-6389 (print)  
ISSN 1879-8314 (online)

# Information Technology and Intelligent Transportation Systems

Proceedings of the 2nd International Conference on Information  
Technology and Intelligent Transportation Systems (ITITS 2017),  
Xi'an, China, June 10, 2017

Edited by

**Prof. Valentina Emilia Balas, Ph.D.**

*Department of Automation and Applied Informatics, Faculty of Engineering,  
"Aurel Vlaicu" University of Arad, Arad, Romania/IEEE Senior Member*

**Professor Lakhmi C. Jain**

*Faculty of Science, Technology and Mathematics, University of Canberra,  
Canberra, Australia*

**Professor Xiangmo Zhao**

*School of Information Engineering, Chang'an University, Xi'an, P.R. China*

and

**Professor Fuqian Shi**

*College of Information and Engineering, Wenzhou Medical University,  
P.R. China*

**IOS**  
Press

Amsterdam • Berlin • Washington, DC

© 2017 The authors and IOS Press.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without prior written permission from the publisher.

ISBN 978-1-61499-784-9 (print)

ISBN 978-1-61499-785-6 (online)

Library of Congress Control Number: 2017948688

*Publisher*

IOS Press BV

Nieuwe Hemweg 6B

1013 BG Amsterdam

Netherlands

fax: +31 20 687 0019

e-mail: [order@iospress.nl](mailto:order@iospress.nl)

*For book sales in the USA and Canada:*

IOS Press, Inc.

6751 Tepper Drive

Clifton, VA 20124

USA

Tel.: +1 703 830 6300

Fax: +1 703 830 2300

[sales@iospress.com](mailto:sales@iospress.com)

LEGAL NOTICE

The publisher is not responsible for the use which might be made of the following information.

PRINTED IN THE NETHERLANDS

# Preface

The 2nd International Conference on Information Technology and Intelligent Transportation Systems (ITITS 2017) was held in Xi'an on June 10–11, 2017, and provided a platform for professionals and researchers from industry and academia to present and discuss recent advances in the field of information technology and intelligent transportation systems. Intelligent transport systems vary in technologies applied, from basic management systems to more advanced application systems. Information technology plays a large role in intelligent transportation systems, including wireless communication; computational technologies; floating car data/cellular data: sensing technologies and video vehicle detection. The technology of intelligent transportation systems includes theoretic and applied topics such as emergency vehicle notification systems; automatic road enforcement; collision avoidance systems and some cooperative systems.

The conference will foster cooperation among organizations and researchers involved in these fields, invite well-known professors world-wide to further explore these topics, and discuss in-depth technical issues with the presenters. All papers were reviewed by 3–4 referees and the program chairs of the conference committee made a selection based on the score of each paper. This year, ITITS 2017 received more than 200 papers from 4 countries and 65 papers were accepted finally. ITITS 2017 was sponsored by Shaanxi Computer Society, Chang'an University and co-sponsored by Xi'an University of Technology, Northwestern Polytechnical University, CAS, Shaanxi Sirui Advanced Materials CO. LTD and Special Aircraft Engineering Research Institute.

We would like to thank the authors of papers and the reviewers for the expertise and contributions. The assistance provided by the conference committee members and all those involved in the organization of conference is gratefully acknowledged.

**The Editors:**

Prof. Valentina Emilia Balas, Ph.D.  
“Aurel Vlaicu” University of Arad  
Faculty of Engineering  
Department of Automation and Applied Informatics  
IEEE Senior Member  
77 B-dul Revolutiei  
310130 Arad  
Romania  
E-mail: [balas@drbalas.ro](mailto:balas@drbalas.ro); [valentina.balas@uav.ro](mailto:valentina.balas@uav.ro)

Professor Lakhmi C. Jain  
Faculty of Science, Technology and Mathematics  
University of Canberra  
Canberra  
ACT 2617  
Australia  
E-mail: [jainlakhmi@gmail.com](mailto:jainlakhmi@gmail.com)

Professor Xiangmo Zhao  
Chang’an University  
School of Information Engineering  
Nan Er Huan Zhong Duan  
Xi’an 710064  
P.R. China

Professor Fuqian Shi  
College of Information and Engineering  
Wenzhou Medical University, P.R. China  
Address: 7B-219, Chashan Campus  
Wenzhou Medical University  
Wenzhou City, 325035, P.R. China  
Email: [fuqian.shi@qq.com](mailto:fuqian.shi@qq.com)

# Contents

Preface	v
An Optimized Method for Web Service Discovery Based on Semantics <i>Xiaoyan Tang, Dahui Luo and Juan Huan</i>	1
A Preprocessing Algorithm Before Conflict Detection Based on Hybrid Surveillance <i>Yu-bo Wang, Hai-jun Wang, Wei-dong Jiao, Bo Zhou and Xiao-yun Shen</i>	8
Combined Bayesian Network-Based Recognition of Lane Changing Behavior <i>Yan Shan, Chang Wang, Juan Gao, Dingbo Song, Aisheng He and Ruibin Zhang</i>	21
Development of an Aircraft Turnaround Time Estimation Model Based on Discrete Time Simulation <i>Haiqiang Wang, Min Wang and Yang Wu</i>	29
Research on Wireless Mesh Network's Application of Position Information Sharing Service <i>Xiao-tao Xu, Cai-wang Guo and Shi Wang</i>	37
Powertrain Parameter Matching and Control Strategy of Electric Drive Chassis for Extended-Range Electric Vehicle <i>Jing-bo Zhao, Hai-mei Liu and Shao-yi Bei</i>	47
Studies on Gravimeter Leveling Algorithm Based on Mixed Sensitivity Robust Control Theory <i>Hai-miao Wang, Yong-yuan Qin, Xiang-wei Qiao and Zhao-fei Zhang</i>	58
An Approach of Telecommunication System for Vehicles and Infrastructures <i>Jiawei Wang, Nan Xu, Zhiyuan Liao, Cong Fei, Jiaheng Li and Wenpeng Zeng</i>	66
The Tracker with Online Training Based on the TLD Algorithm <i>Yongfeng Qi and Peng Zhang</i>	76
Research on the Construction of Network Security Attack and Defense Range System in Power Monitoring System <i>Wei Mingxin</i>	82
A Fuel Saving and Safety Driving Guidance System Applied on the Commercial Vehicle Combining GPS and GIS <i>Liu Hao and Hai Yuanquan</i>	91
Study on Passenger Flow Analysis and Prediction Method of the Public Transport Operation Passenger Line of the Adjacent City <i>Yiming Zhou and Qingge Pang</i>	100

Research on Frequencies Change Rules of Single-Damaged Simply Supported Beam <i>Zhaoguang Hu</i>	108
Effect of Vehicle Combinations on the Instability of Heterogeneous Urban Traffic Flow <i>Cheng Ye, Juan Xu, Shangzhi Xu, Yeqing Qian and Zhipeng Li</i>	117
Design and Implementation of an APP-Based Intelligent Service System <i>Jia Bao, Jun Tao, Chao Wen and Jingwen Zhang</i>	125
Traffic Incident Severity Detection Model Based on Catastrophe Theory <i>Hongwei Li, Sulan Li, Hongwei Zhu, Xing Zhao and Xiaoli Zhang</i>	134
A Method of Constructing Scenario Information Model Based on Power Operation Scene <i>Zhansheng Hou, Lin Peng, Haiyun Han, Min Xu, Gang Wang, He Wang and Qiang Zhou</i>	142
Research on Location Planning of Urban Rail Transit Emergency Facility Considering Vulnerability Node <i>Lei Zhu and Ning Zhang</i>	150
Channel Estimation in OFDM System Based on LS and FEC Iteration <i>Wenyan Chen, Ying Huang, Xiang Zou, Wenbo Fan and Jianxin Guo</i>	160
Container Terminal Operation System in Emission Control Area Using Modelling Based on Multi Agents <i>Jue Hou</i>	168
Flight Delay Propagation Analysis Based on the Mechanism of the Susceptible-Infected-Susceptible Model <i>Weiwei Wu, Haoyu Zhang, Jinfu Zhu and Frank Witlox</i>	180
Research on Constant Power Charging and Discharging of Battery Based on LCL Filter <i>Mengtao Huang and Jiamei Zhao</i>	193
An Improved Quantum Evolutionary Algorithm for High Dimension Knapsack Problem <i>Li Hao, Ma Lei and Qin Na</i>	203
Objects Detection Based on Multi-BING Feature Model <i>Su Yue, Chen Mianshu, Sang Aijun, Li Xiaoni and Li Mengying</i>	212
Print Defect Detection Based on Image Processing <i>Mengtao Huang and Qinyao Li</i>	222
New HRM Practices and Innovation Performance; The Moderating Role of Information Technology Ambidexterity <i>Abdul Waheed, Xiaoming Miao, Naveed Ahmad, Salma Waheed and Abdul Majeed</i>	228
The Droplets Constant Culturing Analyzer Based on Microfluidic Chip Control <i>Qiao Xinyong, Cai Qiang, Huang Cancan and Wang Zhenhua</i>	238



Research on UAV Online Trajectory Planning Algorithm Based on C/FD-GMRES Method	244
<i>Bin Fang, Jiang Yong Yang, Yan Wang and Shuo Xu</i>	
Time-Varying Formation Control and Collision Avoidance for Unmanned Aerial Vehicles Based on Position Estimation	255
<i>Yujie Feng, Qing Wang and Chaoyang Dong</i>	
An Improved Map-Matching Method Based on Hidden Markov Model	266
<i>Yang Linjian, Zhao Xiangmo, Zhang Wei, Meng Fanlin, Cheng Xiaodong and An Yisheng</i>	
Study on Application of Wavelet Transform Reconstruction Acceleration in Vehicle Vibration Displacement Measurement	275
<i>Feng Xiao-feng, Fang Bin and Liu Jiang-hong</i>	
dSPACE Based HIL Simulation Platform for Missile Control System	283
<i>Yun Chen, Qinghua Ma, Gen Wang and Haochun Miao</i>	
Toward Intelligent Traffic Light Control with Quality-of-Service Provisioning	290
<i>Lei Miao and Lijian Xu</i>	
$H_{\infty}$ Anti-Interference Control of EPS System for Distributed Driving Electric Vehicle	299
<i>Zhang Ni, Jing-bo Zhao and Xin Fan</i>	
Research on Vehicle Navigation System Based on Low-Cost Sensors	311
<i>Jiankui Du, Kanghua Tang and Xiaoping Hu</i>	
Fault Diagnosis Method for High-Speed Train Lateral Damper Based on Variational Mode Decomposition and Multiscale Entropy	321
<i>Changxi Li, Xiantai Gou, Xiu Li and Weidong Jin</i>	
The Influence of Using XBRL on Quoted Company's Information Quality	331
<i>Yu-chen Yang, Yan-li Chen, Lei Wang and Chen Tang</i>	
Big Data Analysis for Output of Chinese Research Papers Based on SCI	343
<i>Li Yang and Fang Xu</i>	
ZLCC: Vehicle Detection and Fine-Grained Classification Based on Deep Network Responses and Hierarchical Learning	350
<i>Chen Joya and Shunxi Li</i>	
Health Care Management Using Knowledge Management and Information Technology	361
<i>Jawad Karamat, Shurong Tong, Abdul Waheed and Kashif Mahmood</i>	
The Applied Research of Image Dehazing Algorithms in License Plate Location	371
<i>Wang Rui and Wang Guoyu</i>	
Application of Grey Correlation Matching Model on Informatizational Assessment of Ports	379
<i>Min Zhang</i>	

The Systematic Design of Road Traffic Signal Recognition Based on the Integration of ZigBee with Active FRID <i>Ziyi Liang and Ping Xu</i>	388
Research on Path Planning of Automatic Handling Robot <i>Yuanbin Hou, Liu He, Chunfeng Song, Yanyan Zhang and Dong Li</i>	396
Metro Intelligent Riding System Based on Dynamic Hybrid People Identification <i>Yuanbin Hou, Liu He, Yun Bai and Chen Li</i>	404
Energy Management Strategy for a Hybrid Vehicular Power System Based on Haar Wavelet Transform <i>Bo Li, Shaoyi Bei, Jingbo Zhao, Ting Zhou, Zhang Ni and Qing Liu</i>	413
Design and Development of Chongqing Traffic Weather Service System Based on WebGIS <i>Shigang Han, Baicheng Xia and Fanhua Min</i>	424
PCA Based Improved Face Recognition System <i>Fayaz Ali Dharejo, Munsif Ali Jatoi, Zongbo Hao and Majid Ali Tunio</i>	429
60-GHz 15 × 20 Flat Array Antenna <i>Jiangmei Tang, Bin Yao and Qinhong Zheng</i>	441
Digital Image Watermarking Tools: State-of-the-Art <i>Surekha Borra, Lakshmi H.R., Nilanjan Dey, Amira S. Ashour and Fuqian Shi</i>	450
Fault Diagnosis of Mine Main Fan Based on Improved Particle Swarm Wavelet Neural Network <i>Yong Yang, Xiaodan Yuan and Chenchen Cui</i>	460
Statistical Image Analysis Based Automated Leaves Classification <i>Manar Bati Al-Otaibi, Amira S. Ashour, Nilanjan Dey, Rahaf Abdullah Al Quthami, Asrar Abdullah Al-Nufaei and Fuqian Shi</i>	469
Fractional Dimension Parameters Modelling and Optimization Based on Pulverized Coal Image <i>Zheng Wang and Mei Wang</i>	480
Modified Cuckoo Search Based Neural Networks for Forest Types Classification <i>Sankhadeep Chatterjee, Nilanjan Dey, Soumya Sen, Amira S. Ashour, Simon James Fong and Fuqian Shi</i>	490
Monocular Vision Based Object Recognition and Tracking for Intelligent Robot <i>Qian Gao, Lilian Zhang, Xiaoping Hu and Xiaofeng Wang</i>	499
Mice Liver Cirrhosis Microscopic Image Analysis Using Gray Level Co-Occurrence Matrix and Support Vector Machines <i>Yu Wang, Luying Cao, Nilanjan Dey, Amira S. Ashour and Fuqian Shi</i>	509

The Overview of Chinese Cooperative Intelligent Transportation System Vehicular Communication Application Layer Specification and Data Exchange Standard	516
<i>Yijia Feng, Dazhi He, Lei Niu, Ming Yang and Yunfeng Guan</i>	
Algorithm Design of GNSS/INS Integrated Navigation for Vehicle Location in Cities	527
<i>Han Wang, Maosong Wang and Kun Wen</i>	
An Approach to Identify Travel Hotspots at Night Based on Mobile Phone Data	537
<i>Guanchen Dai, Linchao Li, Jian Zhang, Xiaoli Zhang, Yi Lei, Zhongyue Yang and Qiuying Peng</i>	
Online Road Boundary Detection for Sparse 3D Laser	548
<i>Pengpeng Sun, Xiangmo Zhao, Ya Sun and Haigen Min</i>	
Application and Structure Design of Polyvinylidene Fluoride Membranes in Road Energy Collection	558
<i>Jian Xiao, Yufei Zhang, Xiang Zou, Pengfei Li, Hongyang Gen and Linlin Huang</i>	
Research on Setting and Benefit Evaluation of Integrated Waiting Area Based on VISSIM Platform	565
<i>XinChao Chen, Si Qin, Jian Zhang and Hong Pan</i>	
Implementation and Adoption of E-HRM in Small and Medium Enterprises of Pakistan	573
<i>Abdul Waheed, Salma Waheed, Jawad Karamat, Naveed Ahmad and Abdul Majeed</i>	
The Relationship Between Product Modularization and Open Innovation Based on Sports Industry	583
<i>Min Zhang and Xun Wen</i>	
Study on Prediction of the Shanghai Composite Index Based on EMD and NARX Neural Network	590
<i>Yan Xiu and Xinye Chen</i>	
Subject Index	597
Author Index	601

This page intentionally left blank

# An Optimized Method for Web Service Discovery Based on Semantics

Xiaoyan Tang <sup>a,1</sup>, Dahui Luo <sup>a</sup> and Juan Huan <sup>b</sup>

<sup>a</sup>*Department of Computer Science, Changzhou College of Information Technology, Changzhou, Jiangsu, 213164, China*

*tangdeyan@163.com, dahuiluo@gmail.com*

<sup>b</sup>*School of Information Engineering, Changzhou University, Changzhou, Jiangsu, 213164, China*

*huanjuan@cczu.edu.cn*

**Abstract.** The lack of web services semantic description leads to inadequate services search results from services registries. We propose a novel approach to semantically cluster web services, and group the similar web services based on the services similarity matrix so as to improve services discovery and composition. The proposed approach uses WordNet-Similarity to compute the similarity of terms of service name, description, I/O parameters, and then re-weights the services vector to decide the similarity of services. The method based on web service semantics we introduced has an acceptable result.

**Keywords.** web service, semantics, similarity matrix, WordNet, service discovery, clustering

## 1. Introduction

Web service is a distributed, modularized application component, which is available in Web environment <sup>[1]</sup>. Web service discovery is one of common applications and aims at selecting appropriate multiple web services to integrate into one system for achieving more complicated functions.

The lack of web services semantic description leads to inadequate services search results from services registries. We propose a novel approach to semantically cluster web services and divide similar web services into different groups so that services discovery and composition will be improved. Firstly, the proposed approach extracts the web services names, description, I/O parameters from WSDL and then preprocesses this information, including stop-words removal and term stemming, and gets a bags of words. Secondly, we model web services to weight vectors according to TF-IDF (Term Frequency-Inverse Document Frequency) to create document matrix. Thirdly, we use WordNet::Similarity <sup>[2]</sup> to compute the similarity of terms, and then re-weight the services vector to decide the similarity of services. Lastly, the web services are clustered based on services similarity matrix and grouped into reasonable clusters.

---

<sup>1</sup> Xiaoyan Tang. Associate-professor. Master's degree in computer application. Research interest: service computing, semantic Web, data mining. E-mail: tangdeyan@163.com.

In Section 2, we propose a way of computing web services similarity between based on semantics. Then we propose the services clustering algorithm to form service community. In Section 4, the experimental evaluation is discussed. In section 5, some related works are commented. In the last Section, we conclude our paper and discuss our planning for future work.

## 2. Web Services Similarity Computation Based on Semantics

We propose a novel four-phase method of web services clustering based on semantics. Figure 1 shows an overview of this approach.

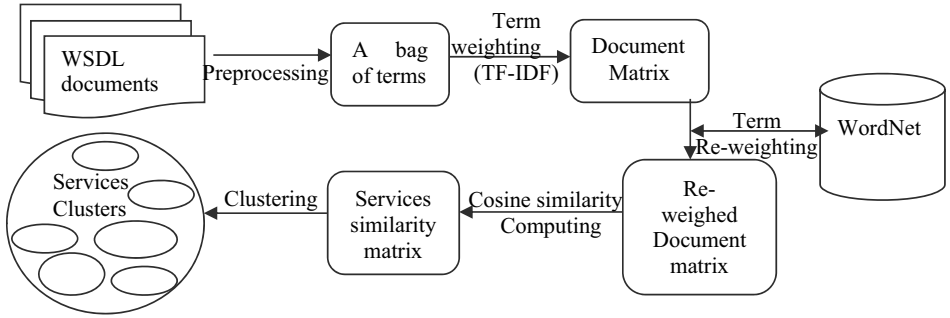


Figure 1. Services clustering based on semantics

### 2.1. Services Document Preprocessing

We parse the WSDL files of web services and get the information including service names, descriptions, I/O parameters, and extract relevant terms from above information using text mining techniques. The detailed process includes splitting the sentences to terms, eliminating stop words which contains request, response, soap, post and other service-related words, stemming the terms. After parsing and preprocessing, we get a bag of terms, assuming  $t$  terms totally.

### 2.2. Document Vector Representation

As described in 2.1, we model web services documents as  $t$ -dimensional vectors and every distinct word contained in the document corresponds to each element of the vector. We use TF-IDF ranking (Salton & Buckley)<sup>[3]</sup> to model web services and assign a weight to each term in the vector, which expresses the importance of every term within the document.

Within document  $j$ ,  $W_{ij}$  represents the importance of word  $i$ :

$$w_{ij} = tf_{ij} * idf_i = tf_{ij} * \log_2\left(\frac{N}{df_i}\right) \quad (1)$$

Where,  $tf_{ij}$  and  $idf_i$  represent term  $i$ 's frequency, term  $i$ 's inverse-document-frequency within document  $j$ , respectively.  $N$  is the amount of collected service documents, the actual number of documents which contain term  $i$  is denoted as  $df_i$ .

When web services are modelled into vectors, we then use Cosine similarity measure. Let  $a, b$  be two web service vectors. The vector inner product computes their similarity according to the Vector Space Model (VSM) [4].

$$\text{sim}(a, b) = a \cdot b = \frac{\sum_{i=1}^t w_{ia} \cdot w_{ib}}{\sqrt{\sum_{i=1}^t w_{ia}^2} \sqrt{\sum_{i=1}^t w_{ib}^2}} \quad (2)$$

Where,  $w_{ia}, w_{ib}$  are the weights of term  $i$  in the two vectors.

Due to the lack of semantics consideration, this modeling method needs to be improved. As we know, sometimes, closely related documents may not contain same terms. We often use different words and phrases to express semantically similar concepts when we organize documents and queries. So it is not effective to compare documents directly based on word-based VSM.

As an example, there are five web service  $s_1, s_2, \dots, s_5$ , after step 2.1 the terms are as follows  $A, \text{fruit}, \text{banana}, \text{apple}, B, C$ . Assuming the five document vector are shown in Table 1 according to formula (1).

**Table 1.** Services document vector based on TF-IDF

	A	fruit	pear	apple	B	C
$s_1$	0	0.5	0	0	0.5	0
$s_2$	0	0	0.8	0	0	1
$s_3$	1	0.1	0.2	0.3	0	0
$s_4$	1	0	0	0	0.5	1
$s_5$	1	0	0	0	1	1

The similarity matrix of web services is shown in Table 2.

**Table 2.** Services similarity matrix

	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$
$s_1$	0.00	0.00	0.07	0.24	0.41
$s_2$	0.00	0.00	0.12	0.52	0.45
$s_3$	0.07	0.12	0.00	0.62	0.54
$s_4$	0.24	0.52	0.62	0.00	0.96
$s_5$	0.41	0.45	0.54	0.96	0.00

It shows that the similarity between  $s_1$  and  $s_2$  is 0, but actually,  $s_1$  and  $s_2$  is similar, because terms fruit and banana have certain similarity.

We propose a method to re-weight the terms taking terms semantics into consideration based on WordNet.

### 2.3. Term Re-weighting

WordNet is an English lexical database to establish terms semantic relations. Nouns, verbs, adjectives and adverbs are grouped into synonym sets [5]. WordNet::Similarity provided by Ted Pedersen et al, is designed to compute semantics relatedness of words using Perl, [6].

Since documents are syntactically modelled into term vectors  $T = \{t_1, t_2, \dots, t_n\}$ , after words semantics relatedness computation, the weight  $w_{ij}$  should be adjusted to  $w'_{ij}$ , considering term  $i$  may semantically related with terms  $i'$  within document  $T$ .

$$w'_{ij} = w_{ij} + \frac{1}{n} \sum_{\substack{i \neq i' \text{ and } i' \in T \\ \text{SemSim}(i, i') \geq \tau}} w_{ij} \text{SemSim}(i, i') \quad (3)$$

Where  $n$  is the number of words  $i'$  which meets the requirement  $\text{im}(i, i') \geq \tau$  and  $i \neq i'$  and  $i' \in T$ ,  $\tau$  is a user defined threshold,  $\text{SemSim}(i, i')$  denotes the semantic similarity of  $i$  and  $i'$  calculated using WordNet::Similarity. The more semantically

similar the terms are within the documents, the higher weights the formula assign. The similarity between a re-weighted document vectors  $a$  and  $b$  is computed as

$$\text{sim}(a, b) = a \cdot b = \frac{\sum_{i=1}^t w_{ia}' * w_{ib}'}{\sqrt{\sum_{i=1}^t w_{ia}'^2} \sqrt{\sum_{i=1}^t w_{ib}'^2}} \quad (4)$$

Where,  $w_{ia}'$  and  $w_{ib}'$  are the re-weighted weights of  $w_i$  term taking semantics into account in the document vector  $a$  and  $b$  respectively. The closer similarity there are between documents vector  $a$  and  $b$ , the higher similarity score they have. The following example illustrates the vector term re-weighting process. The semantic similarity between concepts can be calculated based on the measure proposed by Lin<sup>[7]</sup> in the WordNet.

SemSim("fruit", "apple")=0.7712

SemSim("fruit", "pear")=0.8314

SemSim("apple", "pear")=0.9172

After re-weighting the services terms, the new five document vector are shown in Table 3 according to formula (3).

**Table 3.** Five services document vector based on re-weighted TF-IDF

	A	fruit	pear	apple	B	C
$s_1$	0	0.50	0.21	0.19	0.5	0
$s_2$	0	0.33	0.80	0.37	0	1
$s_3$	1	0.30	0.38	0.43	0	0
$s_4$	1	0	0	0	0.5	1
$s_5$	1	0	0	0	1	1

The semantic similarity matrix of services is as follows according to formula (4):

**Table 4.** Semantic services similarity matrix

	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$
$s_1$	0.00	0.39	0.34	0.22	0.38
$s_2$	0.39	0.00	0.34	0.22	0.38
$s_3$	0.34	0.34	0.00	0.22	0.38
$s_4$	0.22	0.22	0.22	0.00	0.38
$s_5$	0.38	0.38	0.38	0.38	0.00

From Table 4, we can see that the similarity between  $s_1$  and  $s_2$  is 0.39 rather than 0. The similarity between  $s_1$  and  $s_3$  is 0.34 instead of 0.07.

Expanding the comparison with a threshold  $\tau$  will increase the weight of term in the description. The lower threshold values  $\tau$  (e.g.,  $\tau = 0.1$ ) we set, the more likely many new terms are introduced and the topic of the query is diffused. We can learn appropriate threshold values by training. In this paper, we set  $\tau = 0.7$  (Only very similar terms will be expanded in the query).

### 3. Services Clustering

#### 3.1. Web Service Model

A web services collection can be set

$$S = \{s_1, s_2, \dots, s_m\},$$

The model of web service is

$$s = \{sd, sn, inps\{pn\}, outps\{pn\}\}$$



Where,  $sd$  is textual description of the web service,  $sn$  is name of the web service,  $inps$  is set of inputs of the web service,  $outps$  is set of outputs of the web service, which is shown in Figure 2.

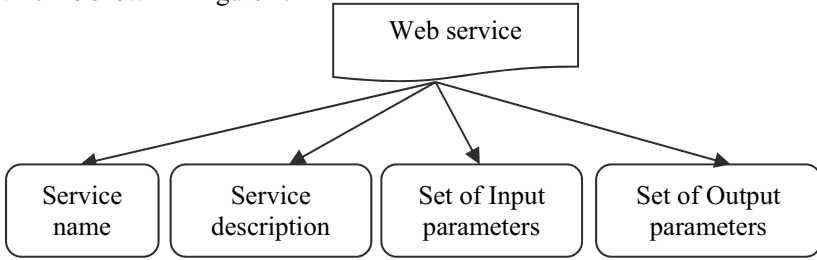


Figure 2. Services Model

Get services description collection is  $SD = \{sd_1, sd_2, \dots, sd_m\}$ , services names collection is  $SN = \{sn_1, sn_2, \dots, sn_m\}$ , services inputs collection is  $INPS = \{inps_1\{pn, c\}, inps_2\{pn, c\}, \dots, inps_m\{pn, c\}\}$ , services outputs collection is  $OUTPS = \{outps_1\{pn, c\}, outps_2\{pn, c\}, \dots, outps_m\{pn, c\}\}$ .

$sd, sn, inps\{pn, c\}$  and  $outps\{pn, c\}$  can be deduced to term vectors, each distinct word contained in the document or set is represented as each element in the vector.

$$\begin{aligned} \text{sim}(s_i, s_j) = & \\ u_1 \text{sim}(sd_i, sd_j) + u_2 \text{sim}(sn_i, sn_j) + u_3 \text{sim}(inps_i\{pn, c\}, inps_j\{pn, c\}) + & \\ u_4 \text{sim}(outps_i\{pn, c\}, outps_j\{pn, c\}) & \end{aligned} \quad (5)$$

$$\text{Where } \sum_{k=1}^4 u_k = 1.$$

### 3.2. Clustering Problem Statement

Clustering is an unsupervised learning method, which organizes a data set into a number of groups/ clusters [8]. Hierarchical clustering groups web services over a variety of scales by creating a cluster tree or dendrogram. A bottom-up algorithm of hierarchical clustering can be applied to service clustering based on semantic similarity. After clustering, services can be mapped into different clusters that are loosely coupled and highly-cohesive [9]. Then the efficiency of service discovery can be improved based on service clusters. The algorithm for service clustering is described in detail as Algorithm 1.

Algorithm 1: Service Clustering

Input: All services in service register, the threshold for service similarity  $\sigma$ .

Output: The set of service clusters  $SC = \{SC_1, SC_2, \dots, SC_k\}$

(1): Local  $SC = \emptyset$  ;

(2) :For each service  $s_i$  that does not belong to any service cluster, build a new cluster  $\{S\}$ .

(3):  $SC = SC \cup \{SC_i\}$ ; // add this cluster to SC

(4): For each service  $s_j$  that do not belong to any service clusters

(5):  $\text{sim}(s_i, s_j)$  // compare  $s_i$  with all the other services that also do not belong to any clusters

(6): If  $\text{sim}(s_i, s_j) \gg \sigma$  then add  $s_j$  to  $\{SC_i\}$

(7): Repeat the steps of 2 and 3;

(8): Return SC.

#### 4. Empirical Evaluation

We implemented service discovery using the following two methods for our experiments: (1) traditional TF-IDF based service discovery; (2) semantics and clustering based service discovery based on OWLS-TC version 4.0 as test collection which contains 1083 web services described using owl-s. Then we use WordNet::similarity to compute the word semantic similarity. According to formula (5),  $u_1, u_2, u_3, u_4$  are weight parameters assigning significance to each corresponding coefficient which are pre-set as  $u_1 = 0.25, u_2 = 0.25, u_3 = 0.25, u_4 = 0.25$ .

We calculate the precision and recall<sup>[10]</sup> in order to further illustrate our method with 20 test queries and the average precision and recall for all test service requests will listed.

##### Experiment 1: traditional TF-IDF based service discovery.

This method uses none of the semantics and web services documents are modelled as t-dimensional vectors. A weight is assigned for each term in the vector, which means the importance of the term.

##### Experiment 2: semantics and clustering based service discovery.

We re-weight each term in the service document based on WordNet::similarity and re-compute the service similarity matrix which is then used to cluster web services. While matching services, we take the functional similarity of services cluster into account and narrow the range of service discovery.

**Table 5.** 4. Empirical result

	Experiment 1	Experiment 2
Precision	0.25	0.05
Recall	0.79	0.13

Experimental results show that Experiment 2 has a more acceptable result than Experiment 1 and the efficiency of service discovery is improved.

#### 5. Related Work

This section summarizes some previous work of web service discovery based on semantically clustering. [11] proposed an approach to automatically classify web services based on WordNet and suffix tree clustering(STC) algorithm. This paper also proposed a “concept and instance” tree structure and realized the similarity calculation in this structure. But the proposed method only concerned textual description of services and did not explicitly model web services and has no any weight consideration about the different part of description information such as the input / output of services. [12] presented the method SWSC which uses OWL-S and WordNet to improve semantics description and utilizes the clustering algorithm to group services into clusters so as to enhance searching. In this work, after extracting terms form OWL-S, WSDL and service description, each of the extracted terms is expanded using WordNet ontology to enhance its semantics. But it did not set a threshold to limit the expanded term range. [13] presents a web service semantics and clustering based discovery method, and its experiment result is much improved than that of full scan simple search method. [14] presents a service discovery approach based on service clustering and refining service clusters which uses WordNet to compute semantics similarity and refine service clusters with Concept Position Vector. When modeling services, the service description is ignored which contains important service information.

## 6. Conclusion

We proposed an approach for service discovery. The whole process contains web services information extraction, information pre-process, web services modeling based on WordNet terms similarity and services clustering based on the services similarity matrix. The next step work is to conduct compensatory test and do more research about service discovery based on semantics.

## References

- [1] Guo G, Yu F, Chen Z, et al. "A method for semantic web service selection based on QoS ontology", *Journal of Computers*, 2011, 6(2): 377-386.
- [2] Pedersen T, Patwardhan S, Michelizzi J. "WordNet:: Similarity: measuring the relatedness of concepts", *Demonstration Papers at HLT-NAACL 2004. Association for Computational Linguistics*, 2004: 38-41.
- [3] Varelas G, Voutsakis E, Raftopoulou P, et al. "Semantic similarity methods in wordNet and their application to information retrieval on the web", *Proceedings of the 7th annual ACM international workshop on Web information and data management*. ACM, 2005: 10-16.
- [4] Salton G. "Automatic text processing: the transformation, analysis, and retrieval of information by computer", *Addison-Wesley series in computer science*, 1989.
- [5] Agirre E, Alfonseca E, Hall K, et al. "A study on similarity and relatedness using distributional and WordNet-based approaches", *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics*, 2009: 19-27.
- [6] Tsai Y H, Hwang S Y, Tang Y. "A hybrid approach to automatic web services discovery", *Service Sciences (IJCSS), 2011 International Joint Conference on. IEEE*, 2011: 277-281.
- [7] Lin, D. (1998) "An Information-Theoretic Definition of Similarity", *Proc 15th Int'l Conf. on Machine Learning (ICML-98)*, 296-304.
- [8] Marjan Kuchaki R, et al. "A survey of hierarchical clustering algorithms", *The Journal of Mathematics and Computer Science*, 2012, 5(3): 229-240.
- [9] Du Y Y, Zhang Y J, Zhang X L. "A Semantic Approach of Service Clustering and Web Service Discovery", *Information Technology Journal*, 2013, 12(5): 967-974.
- [10] E. Voorhees. "WordNet: An Electronic Lexical Database" *The MIT Press*, Cambridge, MA. 1999, 285–303.
- [11] Deng F. "Web service matching based on semantic classification", *Thesis (Master). Kristianstad University. Enck, W., 2011...* 2012.
- [12] Nayak R, Lee B. "Web service discovery with additional semantics and clustering", *2007 IEEE/WIC/ACM International Conference on Web Intelligence*. 2007.
- [13] Wen T, Sheng G, Li Y, et al. "Research on Web service discovery with semantics and clustering", *Information Technology and Artificial Intelligence Conference (ITAIC), 2011 6th IEEE Joint International*. IEEE, 2011, 1: 62-67.
- [14] Du Y Y, Zhang Y J, Zhang X L. "A Semantic Approach of Service Clustering and Web Service Discovery", *Information Technology Journal*, 2013, 12(5): 967-974.

## Acknowledgments

The research is supported by Universities outstanding young teachers' overseas training project of Jiangsu Provincial Education Department, China. I would like to thank I-ling Yen, Wei Zhu for their guidance and help of my work.

# A Preprocessing Algorithm Before Conflict Detection Based on Hybrid Surveillance

Yu-bo WANG<sup>a</sup>, Hai-jun WANG<sup>a</sup>, Wei-dong JIAO<sup>b,1</sup>, Bo ZHOU<sup>a</sup> and  
Xiao-yun SHEN<sup>b</sup>

<sup>a</sup>Hubei Branch, Air Traffic Management Bureau of Middle & Southern Region, CAAC,  
Wuhan 430302, China

<sup>b</sup>Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China,  
Tianjin 300300, China

**Abstract.** The probability of flight conflict is increased because of the increasing of flight flow and uncertainty of flight route. Considering the problem that redundant operation which caused by a large number of non-related targets. A preprocessing algorithm before conflict detection was proposed. Aircrafts within the surveillance scope of Automatic Dependent Surveillance-Broadcast (ADS-B) IN were divided into 26 different regions, and the correlation between targets was distinguished by special rule of each region. Then the correlated targets were further processed, but uncorrelated ones were blocked. So the calculation of conflict analysis for surveilling surrounding targets was decreased. Above all, the processing principle that security first should be followed, and rather false alarm but never leakage alarm. Finally, the necessity and validity of the proposed algorithm is examined by using Monte Carlo experiment. The simulation results show that not less than 30% of 30 random objects are blocked. The proposed algorithm is effective and accurate for detecting the correlation of them.

**Keywords.** ADS-B, multi-target, correlation, conflict

## 1. Introduction

The probability of flight conflict is increased because of the increasing of flight flow and uncertainty of flight route, and controllers almost can not detect the conflict for free flight in advance. In addition, The aircraft of general aviation often lost contact with other aircrafts and Air Traffic Management System (ATMS), so the aircraft can not get the efficient service of air-traffic control. Hence it is necessary to prevent aircraft incidence in advance through the flight conflict and resolution based on the advanced hybrid surveillance technology.

Although the resolution advisories (RA) logic of TCAS II is the worldwide standard for airborne anti-collision technology, it should be realized that possible variations of evasive maneuvers currently included in the TCAS II system are limited to only the vertical dimension, and its application procedure still be conditional on the limitation of other traditional radar system such as the update rate and cover range of signal [1,2]. The perceptively more advanced TCASIV system will incorporate RA logic for both

---

<sup>1</sup> Corresponding author. Civil aviation engineer, PhD; E-mail: wdjiao@cauc.edu.cn.

vertical and horizontal escape maneuvers. In other hand, it also led to the introduction of the Automatic Dependent Surveillance-Broadcast (ADS-B) technology to overcome the limitation and promote the efficiency of TCAS in the airborne. ADS-B derives the aircraft's precise position in accordance to a GNSS satellite constellation (GPS, Galileo, etc.). This information is combined with several other pieces of data, such as the aircraft type, flight number, barometric altitude, speed, direction and rate of climb or descent; and is converted into a digital message [3–6]. The data pack is then automatically broadcasted via a radio transmitter at a regular interval of nominally 1 second, without the need for any external stimulation, to all other ADS-B capable aircraft, ground stations or satellite communication transceivers that relay the data packets to ATC centers; the data message is processed and displayed to pilots on the Cockpit Display of Traffic Information (CDTI) while the ground air traffic controllers see the ADS-B targets on the regular traffic display screen with the other radar targets. Since in this case the aircraft position data is attained from direct broadcast and not through the processing of some returned target signals, the accuracy of the ADS-B system is not seriously degraded with range, atmospheric condition and altitude.

The space-borne hybrid surveillance technology of next generation based on TCAS and ADS-B will make the horizontal escape maneuvers possible [7–10]. It uses more precise and higher update rate data of ADS-B on TCAS system of aircrafts to compute the data accurately and process all conflicts. So it can provide the save conflict resolution, and it is adaptive to the development of general aviation. However, because the future flight track of current aircraft is depended on flight states, intentions and plans of around aircrafts, most aircrafts is not conflict with the current airborne or their conflict probability is less than the warning threshold, but their flight states have been processed by current airborne and then displayed on CDTI of the current airborne, and this case led to waste of memory resource. So it is necessary to pre-filter those aircrafts that their conflict probability with the current aircraft is less than the warning threshold before the conflict detection and resolution for the large flight flow airspace. The pre-filtering course can make the calculation of conflict analysis for non-conflict objects and memory consuming of Flight Manage Computer (FMC) decreased, at the same time, it can make the display state of CDTI keep concise. So the pilot can operate the current airborne equipment directly based on the clear and concise information on CDTI.

References [8–14] introduce the character that ADS-B datum is displayed on CDTI, and discuss potential application price of ADS-B/CDTI using to TCAS. With the application of ADS-B technology, many countries in Europe and America have started to develop TCASIV. TCASIV system as the next generation of collision avoidance of airborne equipment, will effectively reduce the flight distance and improve air traffic flow. Pilots can directly observe the traffic conditions through CDTI and process many kinds of unexpected situations [15], so filtering algorithm for unrelated targets is necessary to reduce the hardware consumption and improve visual effect. Some researchers propose a series of filtering algorithms for the problem that the display density of aircrafts is too high on CDTI and for conflict aircrafts in references [16,17]. For reducing the unnecessary computing consumption, the targets detected should be less. In the relative coordinate system, the unrelated targets should be eliminated, and the possible conflicted ones need to be remained. What's more, the filtering algorithm should follow the principle that it prefers false alarm to miss alarm. However, the conflict filter-

ing in these methods is considered only for the commercial aviation from two-dimension and their conflict condition is too ideal. Especially there is blind area where the filtering condition is difficult to judging whether it should be remained in the critical region of partition. In the blind region, a situation is that when the state variables obviously changed the judgment result is wrong or unknown. The critical region should meet the continuous gradual transition, which is conducive to accurately tracking the targets in the critical edge region. In this regard, the literature method is obviously too “stiff” so that can not apply to practice.

Considering the shortcoming of these methods, a target pre-filtering algorithm is proposed for the high-density aircrafts airspace under the mixed surveillance of TCAS and ADS-B according to the free flight characteristics of low-altitude reporting airspace and the increasing of low-altitude flight flow. In the algorithm, the surveillance range of ADS-B is divided into 26 different zones according to the flying state of objects. The possibility of approaching of different objects in each section can be judged by the set rules correspondingly. At last, only the potentially conflicting aircrafts are remained to calculate their conflict probabilities further. The pre-filtering course is served for the last computation of conflict probability which was introduced in references [18–22].

The rest of this paper is arranged as follows, Section 2 introduces the filtering approach based on mixed surveillance of TCAS and ADS-B, and judging rules of flight conflict filtering are analyzed. In Section 3, the necessity, validity and accuracy of the presented algorithm are illustrated in two aspects by Monte Carlo experiments. The conclusion is given in the last section.

## 2. Filtering Approach

After transforming the geodetic coordinate to Cartesian coordinate, the state of current aircraft can be expressed as:

$$S_a = (x_1, y_1, z_1, \theta_1, \beta_1, \gamma_1, v_{x1}, v_{y1}, v_{z1})^T$$

in which,  $x_1, y_1, z_1$  represent the component of position of the  $x$ -axis,  $y$ -axis and  $z$ -axis direction respectively, and  $v_{x1}, v_{y1}, v_{z1}$  are the velocity component in different directions correspondingly.  $\theta, \beta$  and  $\gamma$  are heading angle, pitch angle and roll angle respectively. The status information of the object aircraft can be acquired via ADS-B data link communications, and the object can be described also as:

$$S_b = (x_i, y_i, z_i, \theta_i, \beta_i, \gamma_i, v_{xi}, v_{yi}, v_{zi})^T$$

$(i = 1, 2, 3, \dots \text{ and } i > 1)$

So the relative value of position, direction, speed between the current and object aircraft can be expressed as

$$\begin{bmatrix} \Delta x & \Delta v_x & \Delta \theta_i \\ \Delta y & \Delta v_y & \Delta \beta_i \\ \Delta z & \Delta v_z & \Delta \gamma_i \end{bmatrix} = \begin{bmatrix} x_i - x_1 & v_{xi} - v_{x1} & \theta_i - \theta_1 \\ y_i - y_1 & v_{yi} - v_{y1} & \beta_i - \beta_1 \\ z_i - z_1 & v_{zi} - v_{z1} & \gamma_i - \gamma_1 \end{bmatrix}$$

### 2.1. Zone Dividing

According to ATC regulations, the horizontal and vertical radius of the protected area for one aircraft is 5 NM and 1200 ft correspondingly, and other aircrafts are prohibited to break into the area, otherwise there will be a possibility of collision [22–24]. The effective range of TCAS will be greatly expand to 40NM with the help of ADS-B [10,15–17], but the objects need to be more precise handled and especially be surveilled in the area within 40 NM of the current airborne which is the hybrid surveillance area of TACS and ADS-B and also is the last region for avoiding collision. So only that objects outside 40 NM are processed in the proposed filtering algorithm. Namely the effect range of the filtering algorithm is 40 NM to 100 NM from current aircraft. Figure 1 is a sketch map of filtering area, in which the inner small circle part is the effect scope of TCAS, and its radius is denoted with letter  $r$ . Some objects need to be filtered in the area which is out of the effect scope of TCAS but in ADS-B (between the big circle whose radius is denoted with letter  $R$  and the small circle). Of course, considering the aircrafts of general aviation flying in low altitude, the objects which are higher than 3000 meters above local aircraft will be filtered out. However, the area under the local aircraft should be filtered too because of the need for the scene monitoring of the terminal area to prevent the occurring of accidents such as some intruder inroad to the runway.

According to the character of objective aircrafts are far away the current airborne (refer to Section 2.2), the range around the current aircraft is divided into 26 zones which including 8 octant zones, 6 axis orientation zones and 12 quadrant zones. Octant zones are first octant, second octant and so on. And axis orientation zones are positive direction of  $x$  axis zone, negative direction of  $x$  axis zone, positive direction of  $y$

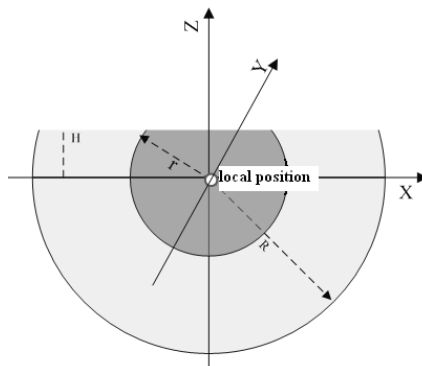
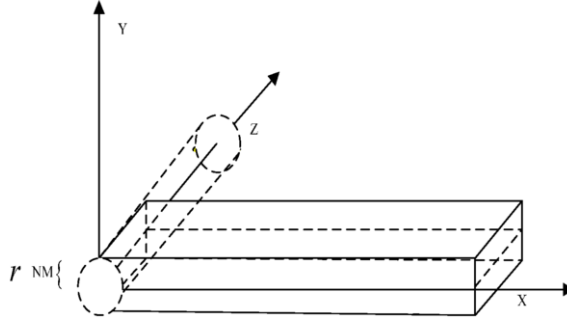


Figure 1. The sketch map of filter range.



**Figure 2.** The axis orientation zone and the quadrant zone.

axis zone, and so on. While quadrant zones are  $+x0 + y$  quadrant,  $+x0 - y$  quadrant,  $-x0 + y$  quadrant etc.

Note that the octant zone is not including the range whose distance from the adjacent axis and quadrant is less than  $r$  NM, and the axis orientation zone is a cylinder around the axis whose radius is  $r$  NM. Similarly, the quadrant zone is a cuboid which is formed by extending  $\pm r$  NM from quadrant. Figure 2 shows an axis orientation zone which is the extension field of the positive direction of  $z$  axis and is an infinite cylinder whose radius is  $r$  NM. Also, the figure shows an quadrant zone which is the extension field of the  $+x0 + z$  quadrant, it is formed by extending  $\pm r$  NM from the  $+x0 + z$  quadrant respectively in positive and negative directions along  $y$  axis, and its final shape is a box whose long is infinite.

## 2.2. Filtering Rules

For octant zones, no matter what the object locates in which octant, if its relative move is away from the current aircraft, it should be filtered out of conflict aircrafts. The case can be classified into four kinds and can be expressed by the component of direction respectively.

- (1) The object aircraft moves away from the current aircraft in all  $x, y, z$  directions.
- (2) The object aircraft moves away from the current aircraft in  $x, y$  directions and closes to the current aircraft in  $z$  direction, but the departure of  $x$  or  $y$  direction is bigger than the closing in  $z$  direction.
- (3) The object aircraft moves away from the current aircraft in  $z, y$  directions and close to current aircraft in  $x$  direction, but the departure of  $z$  or  $y$  direction is bigger than the closing in  $x$  direction.
- (4) The object aircraft moves away from the local aircraft in  $x, z$  directions and close to local aircraft in  $y$  direction, but the departure of  $x$  or  $z$  direction is bigger than the closing in  $y$  direction.

According to these rules, for example, the filtering rule of first octant for which  $\Delta x > r, \Delta y > r, \Delta z > r$  is described as:



$$\begin{aligned} & (\Delta v_x \geq 0 \cap \Delta v_y \geq 0 \cap \Delta v_z \geq 0) \cup \{ \Delta v_x < 0 \cap [(\Delta v_y > 0 \cap \Delta v_y + \Delta v_x > 0) \cup \\ & (\Delta v_z > 0 \cap \Delta v_z + \Delta v_x > 0)] \} \cup \{ \Delta v_y > 0 \cap [(\Delta v_x > 0 \cap \Delta v_x + \Delta v_y > 0) \cup \\ & (\Delta v_z > 0 \cap \Delta v_z + \Delta v_y > 0)] \} \cup \{ \Delta v_z < 0 \cap [(\Delta v_x > 0 \cap \Delta v_x + \Delta v_z > 0) \cup (\Delta v_y > 0 \cap \Delta v_y + \Delta v_z > 0)] \} \end{aligned}$$

Also the filtering rule of second octant for which  $\Delta x < -r, \Delta y > r, \Delta z > r$  is described as:

$$\begin{aligned} & (\Delta v_x \leq 0 \cap \Delta v_y \geq 0 \cap \Delta v_z \geq 0) \cup \{ \Delta v_x > 0 \cap [(\Delta v_y > 0 \cap \Delta v_y - \Delta v_x > 0) \cup \\ & (\Delta v_z > 0 \cap \Delta v_z - \Delta v_x > 0)] \} \cup \{ \Delta v_y < 0 \cap [(\Delta v_x > 0 \cap \Delta v_y - \Delta v_x > 0) \cup \\ & (\Delta v_z > 0 \cap \Delta v_z + \Delta v_y > 0)] \} \cup \{ \Delta v_z < 0 \cap [(\Delta v_x > 0 \cap \Delta v_z - \Delta v_x > 0) \cup (\Delta v_y > 0 \cap \Delta v_y + \Delta v_z > 0)] \} \end{aligned}$$

Other octant filtering rules can be deduced by the same manner.

For the axis orientation zone, regardless of which axis and which direction, the filtering rule is that the object deviates from the current aircraft based on the current departure. For example, the filtering rule of the  $x$  axis orientation zone for which  $\Delta x > r, |\Delta y| < r, |\Delta z| < r$  is described by  $\Delta v_x > 0$ , and the rule of the  $y$  axis orientation zone for which  $|\Delta x| < r, \Delta y > r, |\Delta z| < r$  is  $\Delta v_y > 0$ . And rules of other axis orientation zones can be deduced in the same manner, and are not enumerated here.

For quadrant zones, no matter what the object locates in which zone, the filtering condition meets one of following rules:

- (1) The relative position of the object aircraft in the quadrant region is both away from origin corresponding to two directions respectively.
- (2) The object is close to the current aircraft along one axis of the quadrant, but is away from the current aircraft along others axis, and the departure distance of later is bigger than the closing of former.

For example, the filtering rule of the  $+x0+y$  quadrant zone (namely  $\Delta x > r, \Delta y > r, |\Delta z| < r$ ) is described as:

$$\begin{aligned} & (\Delta v_x \geq 0 \cup \Delta v_y \geq 0) \cup \{ \Delta v_x < 0 \cap [(\Delta v_y > 0 \cap \Delta v_y + \Delta v_z > 0) \cup \\ & (|\Delta v_z| + \Delta v_x > 0)] \} \cup \{ \Delta v_y < 0 \cap [(\Delta v_x > 0 \cap \Delta v_x + \Delta v_y > 0) \cup (|\Delta v_z| + \Delta v_y > 0)] \} \end{aligned}$$

And the filtering rule of the  $+x0-y$  quadrant zone (namely  $\Delta x > r, \Delta y < r, |\Delta z| < r$ ) is described as:

$$\begin{aligned} & (\Delta v_x \geq 0 \cup \Delta v_y \leq 0) \cup \{ \Delta v_x < 0 \cap [(\Delta v_y < 0 \cap \Delta v_x - \Delta v_y > 0) \cup \\ & (|\Delta v_z| + \Delta v_x > 0)] \} \cup \{ \Delta v_y > 0 \cap [(\Delta v_x > 0 \cap \Delta v_x - \Delta v_y > 0) \cup (|\Delta v_z| - \Delta v_y > 0)] \} \end{aligned}$$

By the same manner, rules of other quadrant zones can be deduced, and are not enumerated here.

### 2.3. Filtering Steps

By the zone dividing in Section 2.1 and the filtering rules for non-conflict aircrafts in Section 2.2, the step of the filtering algorithm involves:

Step 1: Deriving the state datum of object aircrafts by analyzing broadcasted datum of object aircrafts (or received datum of the current airborne).

Step 2: Calculating the relative value of position, direction and velocity between the current aircraft and object aircrafts.

Step 3: Detecting whether the object is in the scope of TCAS. If not, the next step 4 will be carried out. Otherwise, the hybrid surveillance procedures of TCAS and ADB-S should be called.

Step 4: Detecting the object locates in which zone of the 26 zones around the current aircraft.

Step 5: Judging whether the object should be filtered according to different filtering rules of different zone.

As far as the calculating time of the proposed algorithm is concerned, the most filtering rules are Boolean calculation which is fast and simple for the digital circuit CMOS, so the consuming time on filtering can be neglected.

## 3. Experiment Results

### 3.1. Verify the Necessity and Validity

In order to test the necessity and validity of the presented algorithm, the experiment of the proposed algorithm is carried out from two aspects. In the first aspect, 30 traced objects are randomly generated and their relative information such as position, course, attitude, velocity are pseudo randomly generated in the filtering airspace around the current aircraft. Then their conflict situation with the current aircraft is traced and surveilled per 0.1 seconds, and some objects which can not conflict with the current aircraft are filtered according to the filtering algorithm. In the second aspect, 30 traced objects whose statuses are set artificially are traced and surveilled, and some objects that can not conflict with the current aircraft are filtered by use the proposed algorithm. At last, both the results of simulation and theoretic analysis are compared to judge whether they are consistent. So the accuracy of the proposed algorithm is tested.

Figure 3 is the traffic situation map of 30 objects generated randomly in 500 seconds in which the word position is the origination tracing point. And the range within 40 NM from the current aircraft is the mixed surveillance range of TCAS and ADS-B, and the range from 40 NM to 100 NM is only the surveillance range of ADS-B. Because there are necessary restrictions for above area of the current aircraft in the low attitude airspace, so the height of the above range is 3000 meters in the paper. But the height of the under range is not limited or requested, and this state can satisfy the need for the scene monitoring of terminal area. Figure 4 is the filtering result of the proposed algorithm in 500 s, in which shown objects include all objects within 40 NM and remained objects after filtering from 40 NM to 100 NM. From Fig. 4 we can obtain that

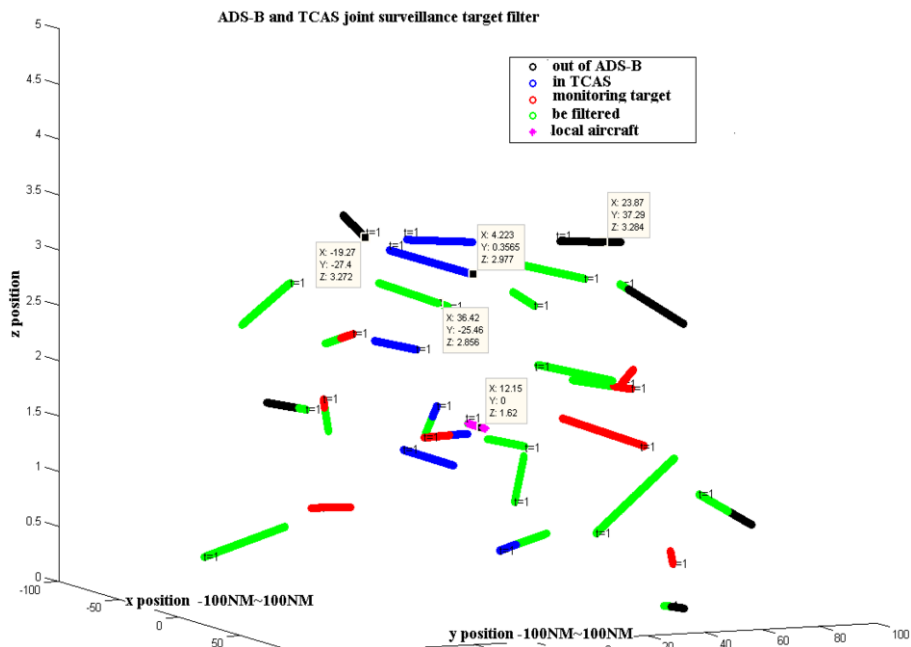


Figure 3. The traffic situation map of 30 objects generated randomly in 500 seconds.

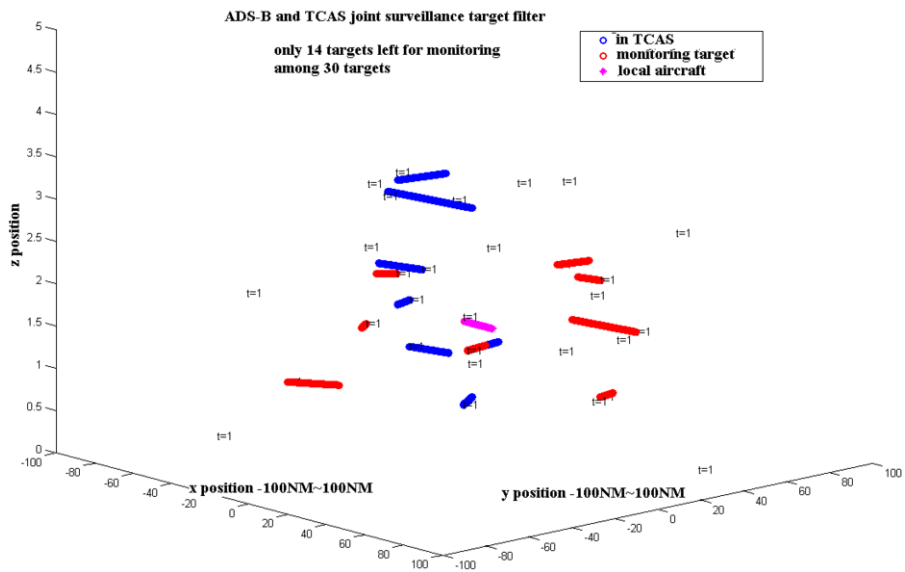
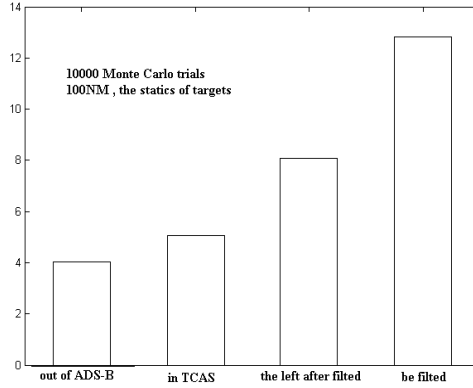
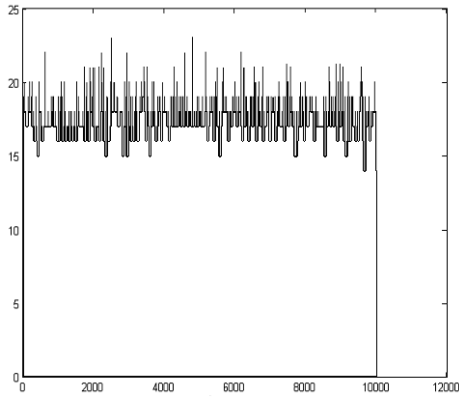


Figure 4. Remained objects after filtering.



**Figure 5.** The static mean of ten thousand times.



**Figure 6.** The filtered object number of ten thousand times.

only 14 objects within 40 NM not be filtered in 500 s, and the total of filtered objects is close to half of all objects. The results of filtering show that the pre-filtering is very necessary.

Monte Carlo experiment analysis method is used here additionally on flight scene for more persuasion of the validity. Figures 5 and 6 are the statistical results of ten thousands experiments of filtering, and in each experiment there are 30 objects randomly generated around current aircraft. Figure 5 shows the static mean of ten thousand times filtering from which we can see that the mean number of filtered aircraft is 13 which is 45% of all 30 aircraft. Figure 6 (in which  $x$ -coordinate denotes the number of experiment and  $y$ -coordinate denotes the number of filtered objects) illustrates the total of remained objects after each filtering. From the figure we can see that the filtered aircraft number is no less than 10 in each filtering, that is to say, the filtering rate is above 33%. These datum show that the necessary and validity of the filtering algorithm from other aspect.

### 3.2. Verify the Accuracy and Reliability

Table 1 shows the status information of 30 objects which is set artificially, in which the unit of position is NM, the unit of velocity is NM/h. The aircraft whose serial number is “0” denotes the current aircraft, and others aircrafts whose serial number is “1” to “30” respectively denote objects within 100 NM of the current aircrafts in Table 1. The position information is represented by component of  $x, y, z$  orientation respectively. Similarly velocity is represented by  $v_x, v_y, v_z$ . We directly calculate heading angle, angle of pitch, angle of bank etc. by Cartesian coordinate system but not WGS84 geographic coordinates system. According to the judging rules of conflict and the theoretic

**Table 1.** The status information set artificially of 30 objects.

Serial number	position			velocity			$\theta$ (heading angle)	$\varphi$ (pitch angle)	$\gamma$ (bank angle)	Conflict	Simulation results
	$x$	$y$	$z$	$v_x$	$v_y$	$v_z$					
1	20	0	3100/1852	160	0	0	$-\pi/2$	0	0	T-A	T-A
2	40	40	2980/1852	100	20	0	-1.3734	0	0	F	F
3	0	30	3000/1852	0	150	0	0	0	0	T-A	T-A
4	0	45	2000/1852	0	-100	5	$-\pi$	0.05	0	R	R
5	-50	0	3000/1852	120	0	0	$-\pi/2$	0	0	F	F
6	-40	-50	3000/1852	-10	-100	0	3.0419	0	0	F	F
7	0	50	2900/1852	140	20	0	-1.4289	0	0	R	R
8	45	60	2899/1852	-20	-100	0	-3.3390	0	0	F	F
9	60	60	3000/1852	-30	100	0	0.2915	0	0	F	F
10	65	30	2000/1852	160	10	5	-1.5084	0.0312	0	F	F
11	50	10	4000/1852	150	10	-2	-1.5042	-0.0133	0	F	F
12	-50	20	3000/1852	100	20	0	-1.3734	0	0	F	F
13	-40	42	3400/1852	90	30	0	-1.2490	0	0	F	F
14	-50	20	3000/1852	170	3	0	-1.5532	0	0	R	R
15	-10	-50	3500/1852	-20	-100	0	2.9442	0	0	F	F
16	-20	-35	4000/1852	-10	60	2	0.1651	0.0329	0	F	F
17	-10	-5	3000/1852	10	80	0	-0.1244	0	0	T-A	T-A
18	-48	-25	5000/1852	-12	98	0	0.1218	0	0	F	F
19	0	30	2000/1852	10	100	1	-0.0997	0.0100	0	T-A	T-A
20	0	-55	4000/1852	-20	80	0	0.2450	0	0	F	F
21	20	-50	3610/1852	-10	-100	0	3.0419	0	0	F	F
22	30	-35	3000/1852	10	-98	0	-3.0399	0	0	F	F
23	10	-60	2000/1852	-5	-120	0	3.100	0	0	F	F
24	58	2	3000/1852	160	10	0	-1.5084	0	0	F	F
25	50	30	4500/1852	100	60	0	-1.0304	0	0	F	F
26	-80	5	2080/1852	120	40	0	-1.2490	0	0	F	F
27	-20	-55	3100/1852	100	-70	0	-2.5309	0	0	F	F
28	-10	9	3500/1852	50	90	0	-0.5071	0	0	T-A	T-A
29	50	-10	3800/1852	120	-60	0	-2.0344	0	0	F	F
30	60	-5	2800/1852	-10	80	0	0.1244	0	0	R	R
0	0	0	3000/1852	150	0	0	$-\pi/2$	0	0	—	—

\* “T-A” denotes that the object aircraft is surveilled by TCAS and ADS-B; “R” denotes that the object aircraft should be remained for surveillance; “F” denotes that the object aircraft can be filtered.

analysis, the aircrafts possessing of serial number “1,3,17,19,28” are the mixed surveillance objects of TCAS and ADS-B, and they must be remained (which is denoted with “T-A” in “Conflict” of Table 1). And the aircrafts possessing of serial number “4,7,14,30” will have the possibility to conflict with the current aircraft, so they should be remained to be traced and surveilled (which is denoted with “R” in “Conflict” of Table 1). Others aircrafts can be filtered (which is denoted with “F” in “Conflict” of Table 1) because they can not happen the conflict with the current aircrafts.

By using the proposed algorithm to the datum of Table 1, aircrafts possessing of serial number “1,3,4,7,14,17,19,28,30” are remained to be traced and surveilled continually (which is denoted with “R” in “Simulation results” of Table 1). Others aircrafts are filtered (which is denoted with “F” in “Simulation results” of Table 1). The experiment results of the proposed algorithm are fully consistent with the theoretical analysis results. So the proposed algorithm is accurate and reliable.

In order to further verify the reliability of the algorithm, 26 interval rules should be validated. The following methods are adopted. For each interval, 30 targets with random position and relevant state parameters including heading, ground speed, attitude, etc are generated. Then the proposed algorithm is used to judge the correction between each target and the local one. It is regarded as the experimental result. At the same time, the change value of distances between them in the following T seconds which is defined as 3 here are calculated artificially. When the distance of the targets is reduced, they are defined as correction, otherwise, they are regarded as unrelated. This is the theoretical real result. The validation experiments of each interval are carried out in accordance with the above process 20 thousand times, then the statistical results show that the accuracy rate of discriminate for each interval is 100%, that is, the experimental results are consistent with the theoretical values.

### 3.3. Time-Consumption Analysis

As the core of the filtering rules involved in the proposed algorithm is logic judgment, so it is fast and simple for the digital circuit chip, and the computation time is short.

The following verification is done on a computer which is configured as 4 core named core i3-2100, basic frequency of CPU 3.30 GHz and internal memory 3 GB. In the scope of the proposed algorithm and around local aircraft, 30 targets are randomly generated, then the algorithm is executed. The 1 million Monte Carlo simulation experiment was carried out, only 11 seconds is consumed, in other words, the average time is about 0.011 milliseconds.

## 4. Conclusion

IN the proposed non-conflict object pre-filtering algorithm for low-altitude airspace, objective aircrafts within the surveillance scope of ADS-B are divided into 26 different zones according to their flight state, and whether objects and the current aircraft are closing is judged by their relative motion under the set rules in each zone. At last the necessity and validity of the proposed algorithm is examined by using Monte Carlo experiment. The simulation results show that, for 30 random objects, 30% objects are filtered, and for 30 given objects, the filtering results of proposed algorithm are fully

consistent with the theoretical analysis results. So the proposed pre-filtering algorithm is necessary, effective and accurate. Additionally, most calculation of the proposed algorithm are Boolean calculation, the calculating time is very fast so that can be neglected for digital circuit CMOS.

## Acknowledgment

This work was supported by Joint Fund of the National Science Foundation of China and The Civil Aviation Administration of China (U1533115).

## Reference

- [1] Brandao R L, Brandao Sr R C, Haissig C M, et al. Traffic alert collision avoidance system (TCAS) devices and methods: European Patent EP 1798572 [P]. 2010-4-14.
- [2] Li L, Fleming E S, Pritchett A R. Influence of demographics on pilot compliance to TCAS resolution advisories [C]//Digital Avionics Systems Conference (DASC), 2012 IEEE/AIAA 31st. IEEE, 2012: 1–25.
- [3] Garcia M, Gilbert T. ADS-B mops updates & impact on surveillance & broadcast services system (SBSS) [C]//Integrated Communications Navigation and Surveillance Conference (ICNS), 2010. IEEE, 2010: L6-1-L6-10.
- [4] Rodrigues C V C, Silva J M R, Bousson K. Advanced air traffic management technologies: the ADS-B impact over ATM concepts. The case for Portugal [J]. International Journal of Aviation Management, 2012, 1(3): 162–180.
- [5] Erzberger H. Transforming the NAS: The next generation air traffic control system [C]//24th International Congress of the Aeronautical Sciences, Yokohama, Japan. 2004.
- [6] Purton L, Abbass H, Alam S. Identification of ADS-B system vulnerabilities and threats [C]//Australian Transport Research Forum, Canberra. 2010.
- [7] da Silva J L R, Brancalion J F B, Fernandes D. Data fusion techniques applied to scenarios including ADS-B and radar sensors for air traffic control [J]. 12th International Conference on Information Fusion Seattle, WA, USA. 2009, 6(9): 1481–1488.
- [8] Haissig C M, Brandao R. Using TCAS surveillance to enable legacy ADS-B transponder use for in-trail procedures [C]//Digital Avionics Systems Conference (DASC), 2012 IEEE/AIAA 31st. IEEE, 2012: 5D5-1-5D5-14.
- [9] Peng L, Lin Y. Study on the model for horizontal escape maneuvers in TCAS [J]. Intelligent Transportation Systems, IEEE Transactions on, 2010, 11(2): 392–398.
- [10] Romli F I, King J D, Li L, et al. Impact of Automatic Dependent Surveillance-Broadcast (ADS-B) on Traffic Alert and Collision Avoidance System (TCAS) Performance [J]. AIAA Guidance, Navigation and Control Conference and Exhibit, Hawaii: AIAA, 2008(8): 18–21.
- [11] Barhydt R, Warren A W. Development of intent information changes to revised minimum aviation system performance standards for automatic dependent surveillance broadcast (RTCA/DO-242A) [M]. National Aeronautics and Space Administration, Langley Research Center, 2002.
- [12] Smith A E, Hulstrom R, Evers C A, et al. Method and apparatus for ADS-B validation, active and passive multilateration and elliptical surveillance: European Patent EP 1906204 [P]. 2009-12-30.
- [13] Zeitlin D, Hammer J, Cieplak J, et al. Achieving early CDTI capability with ADS-B [C]//USA/Europe ATM R&D Seminar. 1998.
- [14] Domino D A, Bateman H, Mundra A, et al. CDTI Enabled Delegated Separation (CEDS) in the vertical domain: Initial feasibility assessment [C]//Integrated Communications, Navigation and Surveillance Conference (ICNS), 2012. IEEE, 2012: L3-1-L3-15.
- [15] Thomas L C, Rantanen E M. Human factors issues in implementation of advanced aviation technologies: A case of false alerts and cockpit displays of traffic information [J]. Theoretical Issues in Ergonomics Science, 2006, 7(5): 501–523.

- [16] Dai C C, Xiao G, Jing Z L. A Selection Algorithm of Conflict Aircrafts Based on ADS-B [J]. *Electronics Optics & Control*, 2011, (10): 11–14.
- [17] Xiao G, Xu Y, Dai C, et al. A selection algorithm for conflict aircrafts and performance analysis based on ADS-B [C]//*Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA 30th. IEEE, 2011: 1D3-1-1D3-6.*
- [18] Paielli A R, Erzberger H. Conflict probability estimation generalized to non-level flight [J]. *Journal of Guidance, control and Dynamics*, 1999, 7(3): 195–222.
- [19] Paielli A R, Erzberger H. Conflict probability estimation for free flight [J]. *Journal of Guidance, Control and Dynamics*, 1997, 20(3): 588–596.
- [20] Erzberger H, Paielli R A, Isaacson D R, et al. Conflict detection and resolution in the presence of prediction error [C]//*1st USA/Europe Air Traffic Management R&D Seminar, Saclay, France. 1997: 17–20.*
- [21] Shen X Y, Zhou B, Cao B, et al. A free flight conflict detection algorithm of low-attitude airspace based on conflict probability [J]. *Electronics Optics & Control*, 2014, 21(6): 47–51.
- [22] Deng W, Zhang J. Collision probability prediction algorithms applicable to the route change [J]. *Beijing: Beijing University of Aeronautics and Astronautics*, 2005, 31(12): 1327–1332.
- [23] Prinzel L J, Shelton K J, Kramer L J, et al. Flight deck interval management and delegated separation for equivalent visual operations [C]//*Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA 30th. IEEE, 2011: 6B2-1-6B2-14.*
- [24] Zhou B, Cao B, Tang P, et al. A new algorithm for track monitoring based on ADS-B [J]. *Electronics Optics & Control*, 2014, 21(7): 45–49.



# Combined Bayesian Network-Based Recognition of Lane Changing Behavior

Yan Shan<sup>1</sup>, Chang Wang<sup>1</sup>, Juan Gao<sup>1</sup>, Dingbo Song<sup>1</sup>  
Aisheng He<sup>1</sup> and Ruibin Zhang<sup>1,2</sup>

<sup>1</sup>*School of Automobile, Chang'an University, Xi'an, China*

<sup>2</sup>*Department of Automobile and Traffic Engineering, Guilin University of Aerospace Technology, Guilin 541004, China*

**Abstract.** Aiming at the lane change behavior recognition requirements of lane change warning system, natural lane change samples were captured by using a test vehicle. Steering angle and distance between vehicle and lane mark were used as characteristic parameters of lane change behavior. Support vector machine (SVM) method was used to establish recognizing model of lane change. According to the high-low identification accuracy of the different time window, 1.2 seconds was selected to be as the window length. The sample data in each time window were filtered by Kalman filter. Then, by using principal component analysis method, the first and second principal components were extracted from all principal components to be as the new variable. Support vector machine-bayesian filter identification model was built to identify general lane change behavior. Final recognition results show that the recognition rate for the real lane change samples can reach more than 95% and the proposed model can also meet the real time and reliability requirements of lane change warning system.

**Keywords.** lane change, behavior recognition, support vector machine, Bayesian filter, ROC curves

## 1. Introduction

Considered as a common driving behavior, lane changing is an important factor related to traffic safety. Unnecessary lane changing can cause traffic accidents. European Union statistics indicate that traffic accidents caused by lane changing account for about 4% to 10% of the total traffic accidents [1]. In particular, traffic accidents related to lane changing caused by human factors account for 75% of the total [2]. Although the traffic fatalities caused by lane changing account for about 0.5% of the total number of deaths, traffic delay accounts for 10% of the total traffic accident delay. Researchers consider lane changing as a complicated driving process, in which drivers do not only focus on vast regions but also operate their steering wheel and pedals frequently. Therefore, drivers are likely to feel nervous and thus easily cause traffic accidents.

At present, a considerable number of both local and international studies have probed into safety driving assistance systems and have obtained significant achievements [3]. However, we cannot establish a precise model that can completely and accurately reflect the lane changing process in an actual traffic setting because of limitations in current research on computer technology and artificial intelligence.

Reliably identifying lane changing behavior is difficult, which results in its low recognition rate and high error rate. Existing recognition methods cannot also effectively distinguish the lane change and movement of vehicles along curves. These factors limit the applicability of the SWA system. Moreover, the lack of actual testing data hinders the development of technology on lane changing behavior.

This study uses actual results from a vehicle test as its primary data and disregards the hidden Markov model because it causes signal distortion as a result of data discretization [4]. This study does not also adopt the neural network model because its robustness and generalization ability are very poor [5], as well as image recognition technology [6] because of its low accuracy and tendency to be easily affected by illumination conditions. Because of these reasons, this study proposed the support vector machine (SVM) [7] combined with Bayesian theory [8] to achieve behavior recognition. SVM does not require discretization of continuous data, and it has a short recognition time and strong generalization ability with the use of small samples. This research will provide strong theoretical support for SWA and will help enhance the effectiveness and instantaneity of the SWA system. The results can also provide an effective reference time for early warning strategy and prediction of danger state and of vehicles that cross the lane line.

## 2. Classification of lane changing behavior

To identify lane changing behavior accurately before a vehicle goes over the lane line, the time at which the vehicle crosses the lane line must be considered as the reference time. The duration of each lane change and the time spent to cross the lane line are different depending on human factors and the external environment. Duration directly determines model validity, but using a single model to identify different periods of lane changing behavior is difficult. This study therefore establishes different recognition models for different types of lane changing behavior. The rate of time consumed is used to categorize lane changing behavior into two types. The rate of time consumed, which is represented by  $\lambda$ , is the ratio of time consumed for the lane change (identified by the system) to the time spent to cross the lane line. Lane changing behavior can be categorized into general lane changing behavior and fast lane changing behavior. The consumed time for lane change identified by the system is marked as  $t_0$  (where  $t_0 = 1.2$  s), and the time spent to cross the lane line is marked as  $t_c$ .

I If the lane changing process meets the following formula, lane change is categorized as general lane change.

$$\lambda = \frac{t_0}{t_c} < 1$$

The system can identify lane change before a vehicles goes over the lane line. The smaller the value of  $\lambda$ , the more reliable is the system. The system can effectively predict lane change conflicts as early as the identification of lane change. This capability can guarantee that the driver has relatively sufficient reaction-operating time and can thus adopt the best lane changing strategy. General lane change accounts for 95% of the total number of lane changes.

II If the lane change process meets the following formula, lane change is categorized as fast lane change.

$$\lambda = \frac{t_0}{t_c} \geq 1$$

Therefore, the system can identify lane change after a vehicle goes over the lane line. Conflicts commonly occur before and after a vehicle goes over the lane line. After this, the high-risk period of lane change conflict is over, and identifying lane change becomes meaningless for SWA.

We can also accurately identify lane change by the lane line, which is marked as 0. The system must accurately identify lane change before a vehicles goes over the lane line. This study establishes different identification models for different lane changes. To analyze different types of lane changing behavior, we construct the support vector machines-Bayesian filtering (SVM-BF) and variance-Bayesian filtering (V-BF) models. These models are shown as figure 1.

The method mentioned above to classify lane changing behavior is suitable for offline identification and not for online identification because the system cannot obtain the time spent across the lane line in advance. This limitation directly affects the feasibility of the classification method mentioned. Out of the 334 samples analyzed, 329 recorded a lane line distance less than 50 cm during the starting point, which accounts for 98.5% of the total number of fast lane changes. Therefore, whether a lane line distance is less than 50 cm is set as the open or not open condition of the model in identifying fast lane change. When the system detects a lane line distance less than 50 cm, the model on fast lane change recognition begins to work. Two models can work separately under online identification.

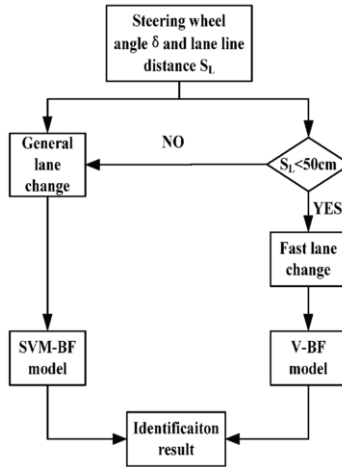


Figure 1. System structure.

### 3. Recognition model based on SVM-BF

A Bayesian network is based on the results of SVM. BF considers the number of lane changes and the traffic environment, and it predicts subsequent driving behavior. The predicted value and the actual output of SVM are comprehensively analyzed to derive the recognition results of SVM-BF. The recognition process is shown in Figure 2.

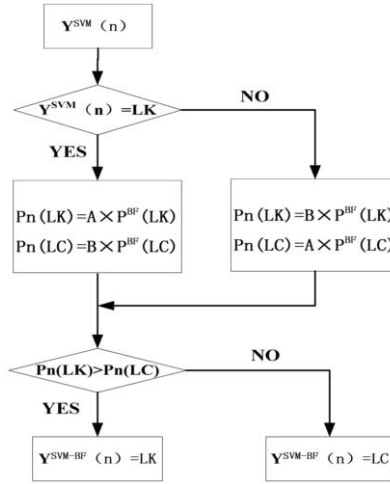


Figure 2. BF process.

where LK represents lane keeping and LC represents the lane change.  $Y^{SVM}(n)$  is the identification result of SVM at  $n$  moment.  $P^{BF}(LK)$  and  $P^{BF}(LC)$  are respectively the probability of lane keeping and lane change, which are predicted by BF.  $P_n(LK)$  and  $P_n(LC)$  are the corrected probability of lane keeping and lane change, respectively.  $Y^{SVM-BF}(n)$  is the identification result of the SVM-BF model. Equations 1, 2, and 3 shows that the model synthesizes the results of SVM and BF prediction and provides the final results of the model.

$$P'_n(LK) = \begin{cases} A * P_n^{BF}(LK) & \text{if } y_n^{SVM} = LK \\ B * P_n^{BF}(LK) & \text{else} \end{cases} \quad (1)$$

$$P'_n(LC) = \begin{cases} B * P_n^{BF}(LC) & \text{if } y_n^{SVM} = LC \\ A * P_n^{BF}(LC) & \text{else} \end{cases} \quad (2)$$

$$y_n^{SVM-BF} = \begin{cases} LK & \text{if } P'_n(LK) > P'_n(LC) \\ LC & \text{else} \end{cases} \quad (3)$$

In these equations, A and B are weight values. To ensure the validity and credibility of the recognition results, we take A and B as variables determined by range (Xmax-Xmin). The maximum lane line distance range of lane keeping is statistically from 0 cm to 30 cm based on the sampling precision of the instruments. The maximum ranges of lane change are 50.97 cm, and the standard deviation is 35.58 cm. Based on three sigma criteria, the maximum lane line distance range in a 1.2 s time window is from 0 cm to 160 cm. Therefore, the abscissa of the weight distribution curve is set between 0 cm and 160 cm. The following figure illustrates the weight distribution curve of A and B.

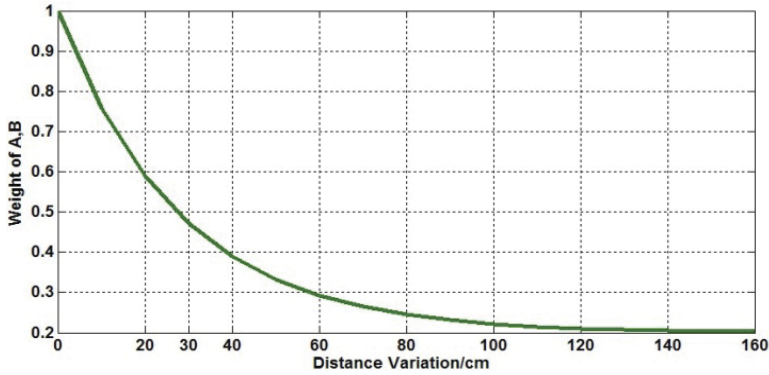


Figure 3. Weight distribution curve of A and B.

The expression of curve in this figure can be achieved by the following equation:

$$y(\Delta s) = p + qe^{r\Delta s} = 0.2 + 0.8e^{-0.036\Delta s} \tag{4}$$

Through verification, if the values of p, q, and r are respectively 0.2, 0.8, and 0.036, the recognition rate of the SVM-BF model is the highest. If the output of SVM is lane keeping, the value of A starts from the curve (as shown above), and B is equal to one minus A. If the output of SVM is lane change, then the value of B starts from the curve (as shown above), and A is equal to one minus B. We include 10 samples to explain the weight distribution process.

Table 1. Correction of the SVM output

SVM recognition result	-1	1	-1	1	-1
BF prediction value	0.37	0.56	0.39	0.57	0.40
	372	8333	3509	9367	4344
SVM-BF output	-1	1	-1	1	-1
Actual result	-1	1	-1	1	-1
SVM recognition result	-1	-1	-1	-1	-1
BF prediction value	0.41	0.61	0.41	0.61	0.41
	0104	0388	4279	2011	3311
SVM-BF output	-1	1	-1	1	-1
Actual result	-1	1	-1	1	-1

The SVM output of the seventh and ninth samples represents lane keeping, and the probability of lane change that BF predicts exceeds 0.5 (Table 1). Therefore, the output of BF and SVM is inconsistent and needs to be comprehensive. The values of A and B are 0.4 and 0.6, respectively.

$$P(LC) = 0.61 \times 0.4 > P(LK) = 0.39 \times 0.6$$

The calculation shows that the probability of lane change is greater than that of lane keeping, so we correct the SVM output. The corrected and the actual results are consistent. Therefore, BF can correct the inaccurate output of SVM to some extent and can improve model recognition rate. In the test, we collect 1,200 samples, of which 900 and 300 samples are the training and test sets, respectively. The recognition results of the model are presented in the following table:

**Table 2.** Recognition results of the SVM-BF model

SVM		SVM-BF	
1% False positive rate (FP)	5%	1%	5%
	(FP)	(FP)	(FP)
73.14%	93.	77.	98.
	91%	21%	66%

In general, the false positive (FP) rate accepted in engineering practice is 5%; therefore, the contrastive analysis on the recognition rates of the SVM and SVM-BF models is performed at 5% FP rate. The SVM-BF recognition rate is almost 5% higher than the SVM recognition rate. Therefore, the generalization ability and adaptability to samples of the SVM-BF model are better than those of the SVM model.

#### 4. Model evaluation

Classifier quality is measured not only by the accuracy of identification of the selected samples. Even if the classifier can accurately identify positive and negative samples, we cannot generalize that the classifier has the same recognition rate. The receiver operator characteristic (ROC) curve [49] [50] can illustrate classifier performance. Therefore, this study uses the ROC curve to evaluate the performance of the SVM-BF model. The ROC curve, also called the receiver operating characteristic curve, is originally used to evaluate radar performance. According to a series of two different classification patterns (boundary value or decision threshold), the ROC curve considers a true positive rate (sensitivity) as the ordinate and a false positive rate (specificity) as the abscissa [51]. In traditional diagnostic test evaluation methods, test results are usually categorized into two groups before they are subjected to statistical analysis. The ROC curve evaluation method is different from the traditional method and is not restricted to dichotomy. Based on the actual situation, the ROC curve allows intermediate states. It can categorize the test results into orderly classifications, such as normal, roughly, suspicious, and abnormal. Statistical analysis can only be conducted after this division. Therefore, the applicable scope of the ROC curve evaluation method is wide, simple, and spontaneous, and the accuracy of the classifier can be observed by the naked eye.

The ROC curve combines sensitivity and specificity through graphical method, and this curve can accurately reflect the relationship between the two factors by analysis. Researchers weigh the influence of false positive and false negatives, and they choose better cut-off point as the judgment reference value. The ROC curve is close to top left corner representatives. It has a large diagnostic value that is significant in comparing indices. The area under the ROC curve (AUC) is generally used as the mark of classifier performance. The AUC value generally ranges between 0.5 and 1.0. The greater is the value of AUC, the higher is the capability of the classifier. Otherwise, the performance is poor. Therefore, this study adopts the ROC curve to compare the three kinds of classification methods, namely, SVM-BF, SVM, and radial basis function (RBF) neural network. The result was shown in Figure 4.

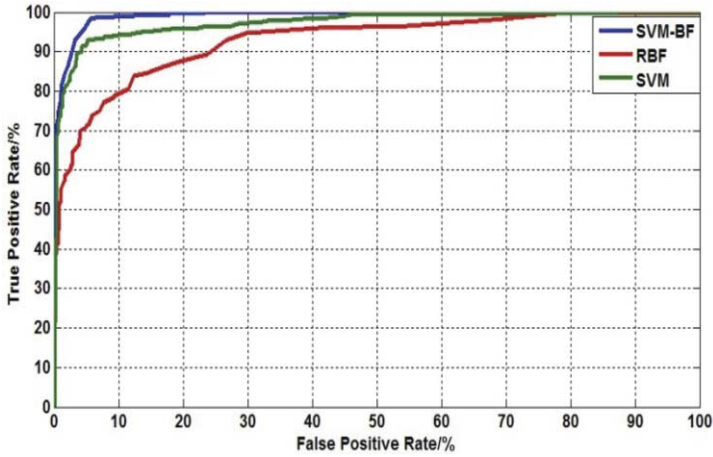


Figure 4. ROC curve.

The area under the SVM-BF model is the largest, followed by that under the SVM model. The greater is the area, the higher is the hit probability of the model. At a low FP rate, the SVM-BF model has the highest hit probability. The SVM-BF recognition rate reaches more than 98% at 5% FP rate. Therefore, 98% of all lane changes can be accurately identified, and 5% of lane change samples are falsely identified as lane keeping.

## 5. Conclusions

This study establishes the SVM-BF model for general lane change. Genetic algorithm is combined with the optimal kernel parameter to establish the best recognition model. The continuous Bayesian network is also used to modify the output of the SVM model. The SVM and SVM-BF models are tested with the same set of samples. The SVM-BF model demonstrates a higher recognition rate than the SVM model. Moreover, the performance of the SVM, SVM-BF, and RBF models is examined via comparison analysis. The ROC curve is used to evaluate the three models, and the results indicate that SVM-BF has the highest true positive rate at 5% FP rate.

## 6. Acknowledgements

This paper was supported National Natural Science Foundation of Shaanxi (2016JQ5096), the Fundamental Research Funds for the Central Universities(310822151028, 310822161009, 310822171118) and Scientific Research and Technology Development Program of Guilin(Grant No.2016010201).

## References

- [1] Holger Berndt, Klaus Dietmayer. Driver Intention Inference with Vehicle Onboard Sensors. 2009:102~:107.

- [2] W. Bouslimi, M. Kassaagi, D. Lourdeaux, P. Fuchs. Augmented Naive Bayesian Network for Driver Behavior Modeling[J]. IEEE Intelligent Vehicles Symposium Conference 2005, Las Vegas, USA, June 7th, 2005:236~242.
- [3] Bing-Fei Wu, Wei-Hsin Chen, Chai-Wei Chang. A New Vehicle Detection with Distance Estimation for Lane Change Warning Systems. Proceedings of the 2007 IEEE Intelligent Vehicles Symposium Istanbul, Turkey, June 13-15, 2007:698~703.
- [4] Wei Liu, XueZhi Wen, Bobo Duan. Rear Vehicle Detection and Tracking for Lane Change Assist. Proceedings of the 2007 IEEE Intelligent Vehicles Symposium Istanbul, Turkey, June 13-15, 2007:252~257.
- [5] Javier Díaz Alonso, Eduardo Ros Vidal, Alexander Rotter, and Martin Mühlenberg. Lane-Change Decision Aid System Based on Motion-Driven Vehicle Tracking. IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, 57(5), SEPTEMBER 2008:2736~2746.
- [6] Sam-Yong Kim, Se-Young Oh&Nam-Gyu Ryu. A New Lane Departure Warning System using a Support Vector Machine Classifier and a Fuzzy System. October 16-19, Jeonbuk, Korea, ICCAS, 2002.
- [7] Joel C. McCall, David P. Wipf, Mohan M. Trivedi, and Bhaskar D. Rao. Lane Change Intent Analysis Using Robust Operators and Sparse Bayesian Learning. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 20078(3): 431~440.
- [8] W. Bouslimi, M. Kassaagi, D. Lourdeaux, P. Fuchs. Augmented Naive Bayesian Network for Driver Behavior Modeling. IEEE Intelligent Vehicles Symposium Conference 2005, Las Vegas, USA, June 7th, 2005:236~242.



# Development of an Aircraft Turnaround Time Estimation Model Based on Discrete Time Simulation

Haiqiang WANG<sup>a,1</sup>, Min WANG<sup>a</sup> and Yang WU<sup>a</sup>

<sup>a</sup> *Department of General Configuration and Aerodynamics, Shanghai Aircraft Design and Research Institute, China*

**Abstract.** According to the actual operation data, the operation cost of a typical medium range single aisle commercial transport aircraft will increase 30 dollars if it stays on ground for one more minute, leading to 1 million dollars increase in its life cycle cost. For the purpose of decreasing cost, airlines require the airplanes can turnaround within a short time period. Boarding and deplaning time accounts for more than half of the total turnaround time, however the current estimation method uses empirical data and cannot evaluate the effects from the aircraft design parameters. An aircraft turnaround time estimation model is developed based on the discrete time simulation, which deems boarding and deplaning as a dynamic process. The interaction between one passenger and other passengers, and the interaction between passengers and cabin are taken into account. The model has been compared with the boarding and deplaning data suggested by aircraft manufacturers. A twin aisle airplane is studied and the sensitivity of cabin design parameters is investigated. The results show that the model is suitable for aircraft designers to determine aircraft design parameters to meet the constraint regarding to aircraft turnaround time.

**Keywords.** Discrete Time Simulation, cabin, boarding, deplaning.

## 1. Introduction

There is a common sense in aerospace industry that airplanes only earn revenue in the air. According to actual airline data, the operation cost of a typical medium range single aisle aircraft will increase 30 dollars if it stays on ground for one more minute. Supposing the design service goal is 25 years for a single aisle airplane with 4 turnarounds between two airports, more than 1 million dollars decrease in the life cycle cost can be achieved if 1 minute can be saved for each turnaround.

Airlines prefer short aircraft turnaround time and avoid the cost due to staying on the ground in order to increase revenue and save operation cost. The success of low cost airline, such as Ryanair and Southwest has already shown that it is feasible and effective to reduce the operation cost per seat by shortening the turnaround time and ground taxiing time. However, the ground taxiing time depends on the effectiveness of air traffic management and the queue of the airplanes waiting for taking off, and it is less likely to be affected by the airplane capability and airline operations.

The turnaround time is defined as the time spent between aircraft arrival in gate and the next departure from the gate. The following factors should be considered when predicting aircraft turnaround time:

---

<sup>1</sup> Corresponding author, Department of General Configuration and Aerodynamics, Shanghai Aircraft Design and Research Institute, China; E-mail: wanghaiqiang@comac.cc.

- The speed of service provided by airlines, such as the cabin cleaning and catering, refueling, water and waste service, cargo unloading and loading.
- Air traffic management. The turnaround will be delayed if traffic volume is not well controlled or the ground equipment is malfunctioned.
- Low visibility or bad weather.
- The aircraft capability. For example, the aisle width could reduce the boarding and deplaning rate. The maximum refueling pressure of the refueling system determines the refueling rate. Number of catering trolleys can determine how much time it will be spent on the trolley exchange.

The scope of this work is specified below: from the perspective of aircraft manufacturers, develop a more accurate aircraft turnaround time estimation model, which is sensitive to cabin design parameters and try to meet the requirement of turnaround time in the early design phase.

## 2. Current estimation method

Three types of service are normally performed during aircraft turnaround: passenger service which includes passenger deplaning, cabin cleaning, galley service and passenger boarding; aircraft service which includes the refueling and water and waste service; cargo loading and unloading. Three types of service could be conducted concurrently. For example, refueling and cargo loading/unloading can be carried out during passenger boarding process. The time spent on these service should be first estimated and a Gantt chart can be drawn according to the sequence of these services.

Using the current estimation method, the time for one single service is calculated as below:

$$t = \frac{N}{V} \quad (1)$$

where N is the passenger number, cargo container number or fuel volume, V is the passenger boarding/deplaning rate, number of cargo containers which are loaded/unloaded per minute or refueling rate.

From the airport characteristics and airplane planning manuals[1-4], it can be found that the critical path for turnaround comprises of passenger boarding, cabin service and passenger deplaning. Figure 1 shows the typical turnaround table for B737-800.

There are two methods to determine the V value: using the statistical data collected from the airplane operation, which can apply for predicting turnaround time for the airplane with similar cabin configuration; using simulation, which can predict the value in the early aircraft design phase without statistical data. Boeing developed PEDS program based discrete event simulation in which the simulation clock is triggered and recorded when an event occurs. Richter et al. adopted discrete time simulation which deems each passenger as an agent and the status of each agent is updated according defined rules [5-7].

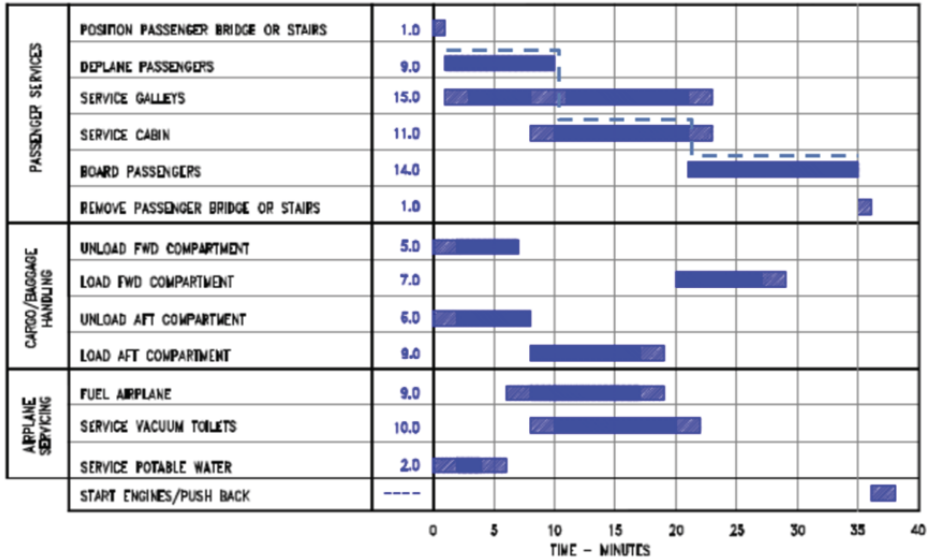


Figure 1. Turnaround table for B737-800[1]

### 3. Aircraft turnaround time estimation model based on discrete time simulation

From the table 1, the boarding and deplaning time accounts for about 50% of the total aircraft turnaround time for twin aisle aircraft and more than 60% for single aisle aircraft. In addition, passenger boarding and deplaning are on the critical path of aircraft turnaround. Therefore, accurate estimation model for boarding and deplaning plays an important role for total aircraft turnaround time prediction. The method proposed in this research employs a discrete time simulation for boarding and deplaning time, and calculates other service time based on the statistical rate collected from the operation of similar airplanes.

Table 1. Proportion of boarding and deplaning time in aircraft turnaround time

B737-800	A320	B777-200LR	A340-500
23/38=61%	28/38=74%	21/45=47%	19/40=48%

Note: data adapted from [1-4]

#### 3.1. Cabin model

The cabin model used in this research is similar with the one described in [8] by Van Landeghen and Beuselinck. Cabin is divided into several rectangular elements, as shown in Figure 2. Each element represents a seat or cabin equipment and element length and width are scaled according to the actual dimensions. The status of each seat is stored in Cell State Matrix, CSM=[ $a_{i,j}$ ] $m \times n$  in which matrix elements representing cabin equipment are always set to 1 and matrix elements representing cabin seats follows the rule below:

$$a_{i,j} = \begin{cases} 0, & \text{passenger not yet seated} \\ 1, & \text{passenger seated} \end{cases} \quad (2)$$

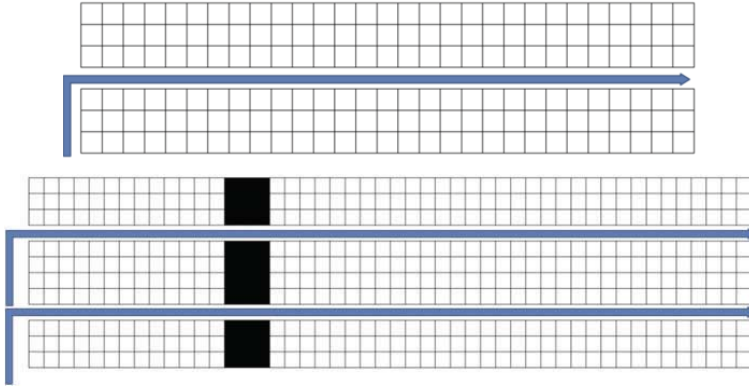


Figure 2. Cabin discretized model

### 3.2. Passenger behavior model

Passenger behavior described in [8] is adapted: passengers wait outside the cabin and follow the "first arrive and first enter" rule. Passengers move in the aisle towards to the row of their seats. Three types of passenger status in the aisle can be observed: moving forward, waiting and being seated. Waiting in the aisle should accounts for two scenarios: passengers in the front put luggage and stop the rear passengers in the aisle moving forward, leading to aisle conflict; to let window passenger in, aisle passengers stand up and block the aisle, leading to seat conflict.

When two conflicts occur, passengers in the rear aisle are blocked and keep waiting. The waiting time due to aisle conflict uses the equation proposed in [9] and seat conflict uses the equation described in [10]:

$$T_{\text{aisle}} = \theta + \frac{\lambda * n_{\text{lug}}}{(C + 1) - (n_{\text{bin}} + n_{\text{lug}})} \quad (3)$$

$$T_{\text{seat}} = \mu(1 + 2n_{\text{seat}}) \quad (4)$$

where  $\theta$  is the basic seated time (unit is s),  $\lambda$  is calibration value from the actual boarding process,  $n_{\text{bin}}$  is the number of luggage already in the overhead compartment,  $n_{\text{lug}}$  is the number of luggage to be stored,  $\mu$  is an coefficient for calibration obtained from the actual boarding process,  $n_{\text{seat}}$  is the number of passengers stand up and let the passenger seated.

### 3.3. Simulation process

Deplaning is similar to boarding but with several differences. For example, passenger can get the luggage during waiting in the aisle; no seat conflict because aisle passengers will enter the aisle first and window passengers last. Due to the similarity, only the boarding simulation is shown in Figure 3.

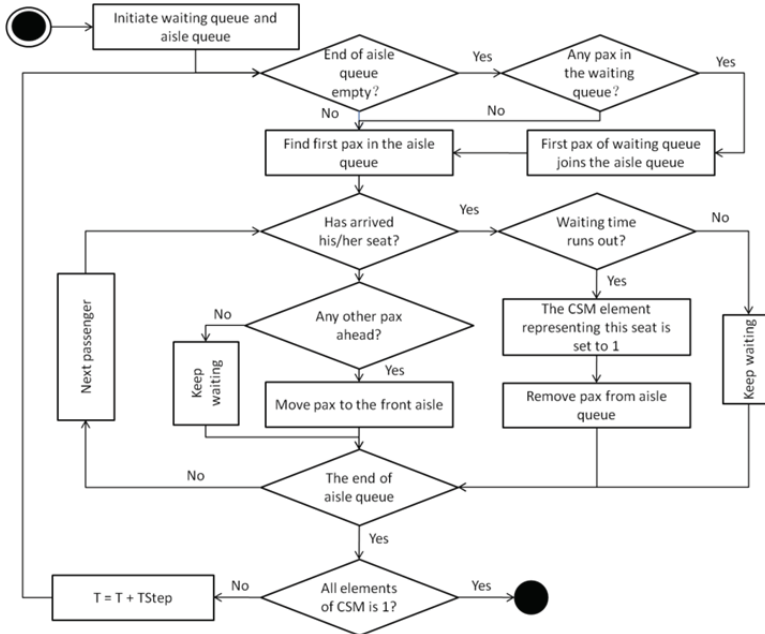


Figure 3. Simulation workflow

### 3.4. Validation for boarding and deplaning simulation

The model has been compared with the boarding and deplaning data suggested by aircraft manufacturers. The validation results are shown in Table 2.

Table 2. Model validation results.

Airplane	Boarding time from ACAP manual* (min)	Boarding time estimated** (min)	Deplaning time from ACAP manual* (min)	Deplaning time estimated** (min)
A320	14.0	13.6	8.4	7.7
B737-800	13.3	13.3	8.9	7.5
B777-200	15.0	16.9	7.5	7.9
A340-500	22.0	21.0	9.0	9.4

\*manual data adapted from [1-4];

\*\*passenger randomly wait outside the cabin and 400 simulation are performed;

## 4. Study case: twin aisle airplane turnaround

A twin aisle airplane with about 300 passenger number is studied here. Several empirical rates adapted from [4] are used:

- Container unloading 1.2 min/container
- Container loading 1.4 min/container.
- Bulk 110kg/min unloading
- Bulk 95kg/min loading

- refueling rate 2857L/min;
- Trolley exchange 1.5 min/trolley
- Portable water 15 min
- Waste removal 20min

The boarding and deplaning time for economy and first and business classes are shown in Figure 4 and figure 5. The turnaround time table is shown in Figure 6.

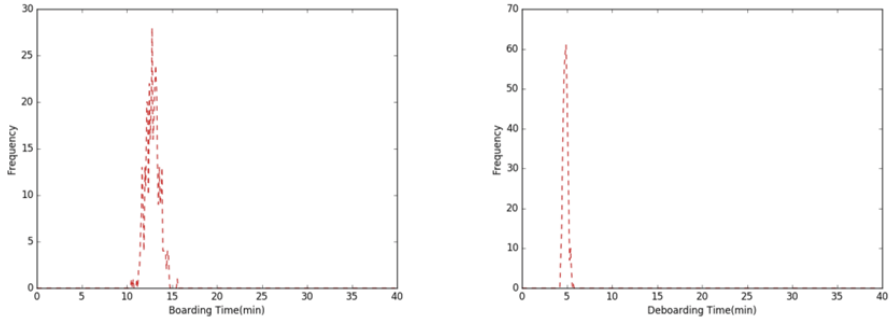


Figure 4. Boarding and deplaning time distribution for economy class (400 simulations)

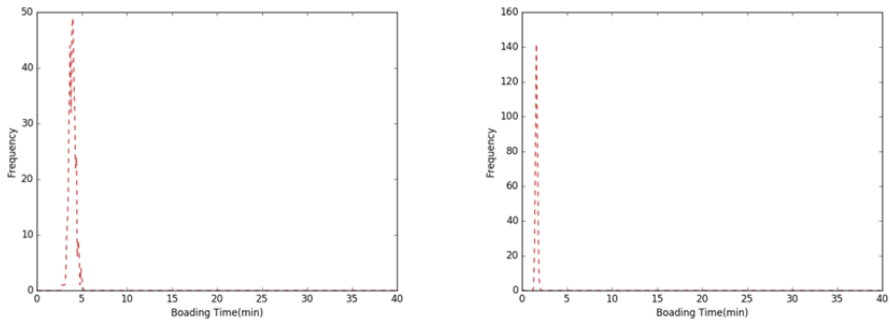


Figure 5. Boarding and deplaning time distribution for first and business class (400 simulations).

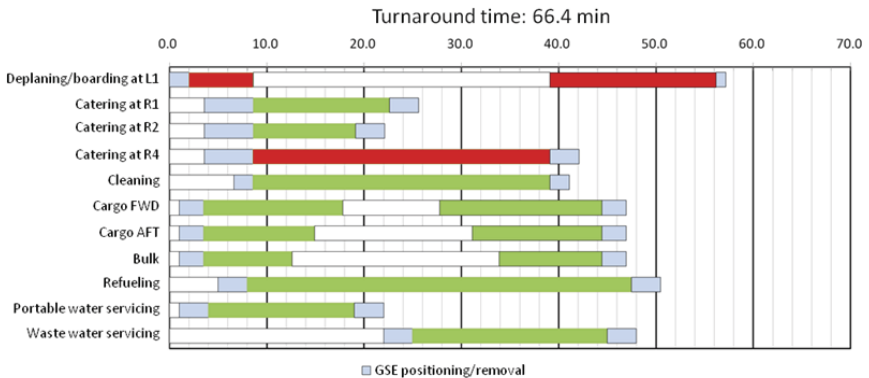


Figure 6. Turnaround time table.

4.1. Effects of cabin design parameters

The effects of cabin design parameters(overhead compartment volume, seat pitch and aisle width) are shown in Figure 7.

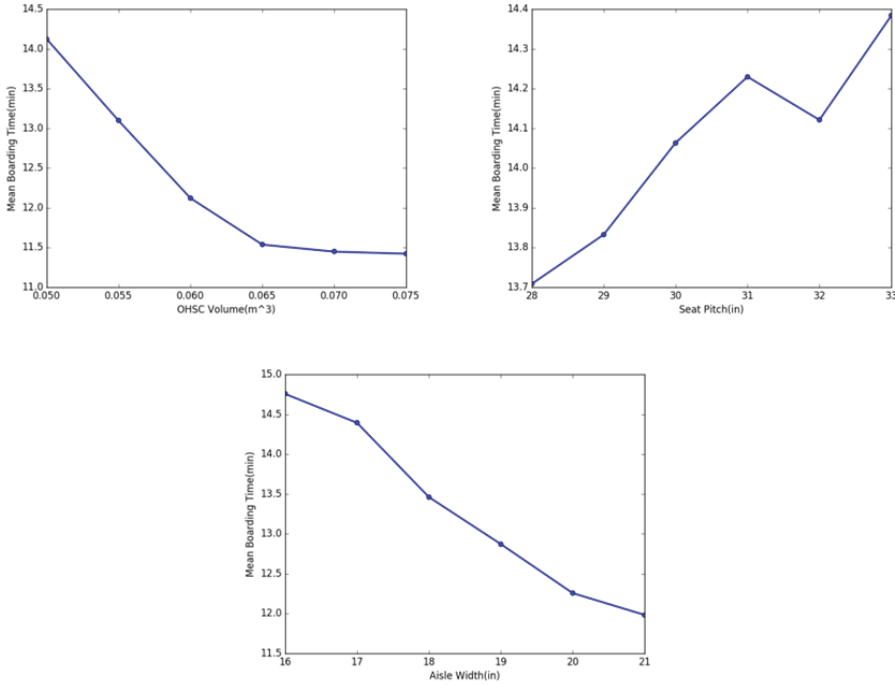


Figure 7. Cabin design parameters on boarding time.

4.2. Effects of doors and boarding strategy

When two doors are available, boarding time and deplaning time are 13.0 min and 4.9 min respectively. When one door is used, boarding time and deplaning time are 20.2 min and 7.1 min respectively.

Three widely adopted boarding strategies have been studied: Random (R) boarding means passengers are randomly waits outside cabin; Outside-in (OI) boarding means window passengers board first and aisle passengers last; Rotating zone (RZ) means the passengers are divided into several groups according to the seat row number and board group by group. Using the proposed simulation model, the boarding time using R, OI and RZ are 13.0 min, 10.7 min and 19.7 min respectively.

4.3. Suggestions on shortening turnaround time

From Figure 6, the critical path for turnaround is Pax deplaning→Galley service→Pax boarding. Under the condition of using two doors, the estimated turnaround time is 57.2 minutes while turnaround time increases to 66.4 when only L1 passenger door is

used. To meet 60 minutes turnaround requirement, several suggestions are given below:

- Speeding up the galley service rate at the fourth service door at right hand(R4);
- Adopting Inside-Outside strategy for boarding, achieving 2.3 min;
- Increasing the overhead compartment volume per passenger from 0.05 m<sup>3</sup> to 0.07 m<sup>3</sup>;
- Increasing aisle width from 17 in to 19 in for 1.4 min saving.

## 5. Conclusions

Boarding and deplaning time account for more than half of the total turnaround time, however the current estimation method uses empirical data and cannot evaluate the effects of aircraft design parameters. An aircraft turnaround time estimation model is developed based on the discrete time simulation, which deems boarding and deplaning as a dynamic process. The model has been validated with the boarding and deplaning data suggested by aircraft manufacturers.

A twin aisle airplane has been studied. Results show the proposed model is able to find the critical path for turnaround, and it is sensitive to the cabin design parameters, such as overhead compartment volume, aisle width and seat pitch. The study case shows overhead compartment volume has considerable effect on boarding and turnaround because this parameter determines the number of aisle conflict.

Further study on the calibrating the passenger behavior model using operation data will be carried out to improve the model accuracy.

## References

- [1] Boeing. 737 Airplane Characteristics for Airport Planning. Boeing Commercial Airplanes. USA: Boeing, 2005.
- [2] Airbus. A320 Airplane Characteristics for Airport Planning. Airbus S.A.S. France: Airbus, 2002.
- [3] Boeing. 777-200LR/-300ER/ Freighter Airplane Characteristics for Airport Planning. Boeing Commercial Airplanes. USA: Boeing, 2007.
- [4] Airbus. A340-500/-600 Airplane Characteristics for Airport Planning. Airbus S.A.S. France: Airbus, 2002.
- [5] Richter T. Simulationsmethodik zur Effizienz- und Komfortbewertung von Menschenflussprozessen in Verkehrsflugzeugen. Munich, Germany: Technical University of Munich, 2007.
- [6] Fuchte J. Enhancement of Aircraft Cabin Design Guidelines with Special Consideration of Aircraft Turnaround and Short Range Operations[D]. Hamburg, Germany: Technical University of Hamburg, 2014.
- [7] Wald A, Harmon M, Klabjan Diego. Structured deplaning via simulation and optimization. Journal of Air Transportation Management, 2014, 36(1): 101-109.
- [8] Van Landeghem H, Beuselinck A. Reducing passenger boarding time in airplanes: A simulation based approach[J]. European Journal of Operational Research, 2002, 142(2): 294 - 308.
- [9] Shang H, Lu H, Peng Y. Aircraft boarding strategy based on cellular automata. Journal of Tsinghua University, 2010, 50(9):1300-1333.
- [10] Ferrari P, Nagel K. Robustness of efficient passenger boarding strategies for airplanes. Transportation Research Board, 2005, 1915(1): 44 - 54.
- [11] Yuan B, Yin J, Wang M. Star: (saving time, adding revenues) boarding/deboarding strategy. Undergraduate Mathematics and Its Applications, 2007, 28(3):371-384.
- [12] Boeing. 757-200/300/ Freighter Airplane Characteristics for Airport Planning. Boeing Commercial Airplanes[R]. USA: Boeing, 2002.



# Research on Wireless Mesh Network's Application of Position Information Sharing Service

Xiao-tao XU, Cai-wang GUO and Shi WANG

National Defense Information Academy, Wuhan, China

E-mails: [superxxt@163.com](mailto:superxxt@163.com), [75239375@qq.com](mailto:75239375@qq.com), [64305264@qq.com](mailto:64305264@qq.com)

**Abstract.** Wireless Mesh Network is a new kind of wireless communication technology widely used in near-field communication, has obvious application advantage in position information sharing service. The paper introduce WMN's technical characteristics used in the fields of Wireless Personal Area Network, analyze WMN's application advantage compared with Ad hoc Network, discuss four kinds of WPAN's traditional application mechanism based on WMN. Based above, describe a new kind of WPAN standard based on WMN called IEEE 802.15.5. Then, put forward a WMN's application case used in vehicle navigation and positioning, has important practical guiding significance of instructing WMN used in the fields of position information sharing service.

**Keywords.** Wireless Mesh Network, information sharing service, technology mechanism

Wireless Mesh Network (WMN) originated in Mobile Ad Hoc Network, as a new broadband Wireless Access Network technology, the Wireless Mesh Network at the same time is a kind of high capacity, high rate of the distributed Network, and has strong robustness, easy extension, and the advantages of low investment costs. In view of the Wireless Mesh Network technology unceasing development, the Wireless Mesh Network technology is applied to the Wireless Personal Area Network (WPAN) field, effectively expand the Wireless Personal Area Network applications, at present, the Wireless Personal Area Network standardization organization will not only Wireless Mesh Network, the key technology applied to traditional Wireless Personal Area Network technology, and specially formulated specifically for Wireless Mesh Network, application of Wireless Personal Area Network technology standards, as to the traditional point to multipoint enhancement and expansion of the Network structure.

## 1. Wireless Mesh Network's Application Advantages Analysis

Wireless Mesh Network structure and the traditional sense of the mobile Ad Hoc Network structure is different, the terminal nodes, the Mesh of the Router nodes and gateway. Mesh routers and Mesh gateway mobility are generally lower, while the Mesh terminal can be either static, also can be any mobile node. Mobile terminals and the

fixed access point can communicate with other nodes directly, and can be used as intermediate routers for data forwarding. Mobile terminal access to the upper Mesh by Mesh router structure of network, the implementation of the network node interconnection and interflow [1]. Wireless Mesh Network has the following distinctive features.

### *1.1. Network Build Convenient*

In the Wireless Mesh Network, the Network structure, Network node configuration is very convenient, you just need to and communication power supply connection, has good performance of adaptation. It is because of this characteristic, which makes the original network can very easily by adding nodes to expand our network coverage, improve the network capacity. Users can easily add new nodes to expand wireless network coverage and network capacity. At the same time, in the Wireless Mesh Network, is not infinite extension Network node, but actively optimize Network structure, increase the efficiency of the application of Network nodes, minimize the number of nodes, so that the Network topology structure is greatly simplified, reduces the cost of Network application and configuration. Wireless Mesh Network, the Network management function is similar to the traditional Wireless LAN, have experience in using WLAN users can quickly grasp the Wireless Mesh Network, the Network application technology.

### *1.2. Network Good Compatibility*

In the traditional single hop wireless communication networks, the equipment must be Shared wireless access nodes. When multiple devices to work at the same time, it is easy to produce channel plugging and results in the decrease of system efficiency and even paralysis. The Wireless Mesh Network, the Network used more jump technology, equipment can find spare backbone Network, access not only ensure the stable operation of equipment, and improve the using efficiency of Network. Good Network compatibility make Wireless Mesh Network can be applied to the corresponding gateway and other Network system (such as the Internet, local area Network (LAN) and other private Network) for connectivity, making Wireless Mesh Network, the end user can be realized with heterogeneous Network communications. In addition, the Wireless Mesh Network also provides many circuitous route, and has the function of information transmission load balance, the Network has good stability, and become the future Internet of things is very important in the process of Network application form.

### *1.3. Multiple Hops Network Function*

Traditional star or a tree network, each terminal must maintain a wireless link and access points directly, so the distance between the terminals and access point will directly affect the data transfer rate, the further away from the access point terminal, data transfer rate is lower. And Wireless Mesh Network with multiple hops Network functions, data communication between adjacent nodes, the data packets forwarded through multiple nodes in the form of more jump, the relay transmission to the destination node. In this way, the wireless channel communication distance between node was effectively shortened, the equipment of the transmission power and transmission performance requirements of the antenna is reduced greatly, but also greatly improve the spectrum

utilization efficiency, improve the network system in complex electromagnetic environment adaptability [2].

#### *1.4. Network Stability Is Good*

Wireless Mesh Network, using distributed Mesh Network structure, the functional nodes have similar function, therefore, don't need to build a large base station or core node, even if some function node because the damage of natural disasters or accidents, other nodes will still be able to independently for appropriate routing, ensure the stable operation of the Network as a whole function.

#### *1.5. Adaptive Balance the Load*

Wireless Mesh Network, the node devices are generally similar routing functions, can the independent forwarding other nodes of information, the user populated areas, when using Wireless Mesh Network, the Network system can adjust the node load, adaptive optimal routing information transmission, the balance of the large capacity information transmission requirements are assigned to each routing node, because node so you can avoid Network congestion caused by the Network paralysis, can adapt to the demand of large capacity information transmission Network.

## **2. Wireless Personal Area Network Expansion Application Based on Wireless Mesh Network**

So far, the Wireless Personal Area Network research group (IEEE 802.15 team) proposed by IEEE 802.15.1 – IEEE 802.15.4 offers special mechanism in essence are not directly support the Mesh Network structure, but only under the way of Point-to-Multipoint Piconet structure, but in the process of practical application by scattering web application forms to constitute the Wireless Mesh Network topology structure, thus effectively extends the traditional Wireless Personal Area Network application scope.

#### *2.1. WMN's Application System Based on IEEE 802.15.1*

The IEEE 802.15.1 is standard bluetooth wireless technology. Bluetooth is a short-range communications system, is used to replace the cable connection portable or fixed electronic equipment, its main feature is the robustness, low power consumption and low cost [3].

In the Bluetooth wireless communication technology, under normal circumstances, a set of devices to share a physical wireless channel, and synchronized to a common clock and FM mode. Synchronous reference (master) of the device is called the main equipment, other equipment is called from the device (slave). In this way a set of equipment constitute a synchronous Piconet (Piconet), it is basic communication mode of Bluetooth technology.

Constitute a Piconet two or more than two devices share the same physical channel, namely they are synchronized to a common clock and FM sequence. A common clock Piconet is the main equipment of the local clock (CLKN). Frequency hopping sequences by the main equipment and the address of master device to determine the local clock.

In a common bluetooth devices application environment, there can be many independent Piconet, each a Piconet has a different physical channel, which is a different kind of main equipment and an independent sequence of Piconet clock and frequency modulation. A device can be based on TDM concurrently to join the no less than two Piconets, because a Piconet is made to the main equipment of the local clock synchronization, constitute, so a device can't be more than two Piconets' master devices, but one device could be used as many independent Piconets' slave device.

When a device belongs to the two above Piconet and Scatternet formed by the device, but the device doesn't have to provide network routing capabilities, core protocols also don't have to provide this functionality, network routing can be done by higher routing protocol.

As a result, more than one Piconet can overlap scatter net, the same bluetooth device can join more than one Piconet, so as to realize the bridging of more than one Piconet. In this way, the bluetooth can flexibly implement multiple hops mesh network structure.

## 2.2. WMN's Application System Based on IEEE 802.15.2

IEEE 802.15.2 mainly solve the Wireless Personal Area Network and Wireless Local Area Network (LAN), these two Wireless networks often work in the same unlicensed spectrum. Web applications based on IEEE 802.15.2 mechanism mainly divided into the coexistence mechanism, cooperative and non-cooperative coexistence mechanism of three [4].

### 1) Coexistence mechanism

There are two kinds of IEEE 802.15.2 coexistence mechanism: cooperation and non-cooperation. Cooperation coexistence mechanism through the exchange of information between two wireless networks to implement; Non-cooperative coexistence mechanism in no exchange of information between two Wireless networks, this mechanism applies only to a Wireless LAN or Wireless Personal Area Network set up after the user data to be sent [5].

### 2) The coexistence of cooperation mechanism

The IEEE 802.15.2 defines the coexistence of the three kinds of cooperation mechanism: Alternating Wireless Medium Access (AWMA), Packet Traffic Arbitration (PTA) and Deterministic Interference Suppression (DIS). AWMA and PTA combination can achieve better coexistence mechanism [6].

The AWMA IEEE 802.11 part of the beacon interval for IEEE 802.15 operation. From the point of view of sequence, the media by the IEEE 802.11 and IEEE 802.15 used interchangeably, prevent interference between two wireless networks.

PTA mechanism is the IEEE 802.11 b site (STA) and the IEEE 802.15.1 node. The IEEE 802.11 b or IEEE 802.15.1 transfer requests should be subject to approval by the PTA, transfer requests of PTA may refuse to lead to conflict.

Determine the interference suppression is mainly used to reduce the IEEE 802.15.1 interference with IEEE 802.11 b, its basic idea is in the receiver to the IEEE IEEE 802.11 b 802.15.1 signal frequency to zero. However, due to the IEEE 802.15.1 each send a new message, skip to a new frequency, therefore, need to know the IEEE IEEE 802.11 b receiver 802.15.1 transmitter frequency hopping patterns in time sequence. This information can be in the IEEE 802.11 b receiver installed IEEE 802.15.1 receiver to receive. Therefore, it is a juxtaposition and collaboration.

### 3) Non-cooperative coexistence mechanism

Non-cooperative coexistence mechanism mainly has the following three forms [7]. First, adaptive interference suppression, mainly in the wireless local area network (LAN) based on signal processing of the physical mechanism; Second, adaptive packet selection and scheduling, IEEE 802.15.1 systems use a variety of configuration packet types (such as packet length and the level of error protection, etc.); Third, adaptive frequency hopping by determining interference channel, so as to change the frequency hopping sequence, avoid channel interference. In summary, the IEEE 802.15.2 is the solution of WPAN and coexistence between WLAN standard, for a variety of technology into Wireless Mesh Network provides support.

### 2.3. WMN's Application System Based on IEEE 802.15.3

Piconet is a self-organizing wireless data communication system, in which many independent data equipment (DEV) communicate with each other. The difference between the Piconet and other types of data network lies in its communications agency is limited to a smaller area around people or objects, usually is static or moving person or object around 10 meters or more [8].

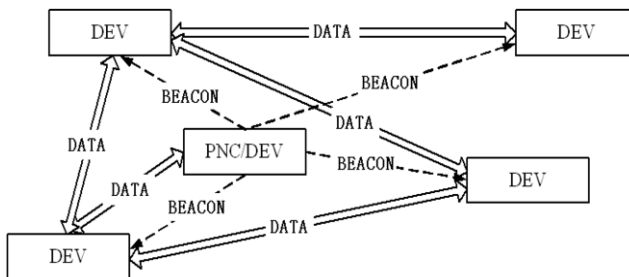
The IEEE 802.15.3 Piconet consists of several parts, as shown in Fig. 1. Basic components is data equipment (DEV), one of the devices are Piconet Coordinator, it USES beacon to provide basic temporal Piconet, management, QoS requirements, energy saving mode, and the access control of Piconet.

IEEE 802.15.3 MAC layer standard goal is: fast connection is established, the self-organizing network, QoS guarantee data transmission, security, members of the dynamic and efficient data transmission.

IEEE 802.15.3 Piconet is PNC for centralized controller structure, poor flexibility, limits the application of UWB technology, so the PNC MBOA is trying to downplay, develop new IEEE 802.15.3 MAC layer protocol to support distributed and centralized topology, which support the mesh network is one of the main goals of the standard.

### 2.4. WMN's Application System Based on IEEE 802.15.4

IEEE 802.15.4 is for LR-WPAN. This kind of simple network structure, lower cost, suitable for the limited power and easy application throughput requirements. Its main goal is easy to install, reliable data transmission, short distance communication, low cost and long battery life, simple and flexible and agreement.



**Figure 1.** The composition of the IEEE 802.15.3 Piconet parts.

According to the application requirements, LR-WPAN mainly run in two kinds of topological mode: star topology and peer entities (peer-to-peer) topology. In a star topology, communication in various devices and central control device, a domain network Coordinator (PAN Coordinator). In the peer entity topology, any device can communicate with a device within the scope of communication, this topology can be used to realize the more complex mesh network structure.

### 3. Wireless Mesh Network Technology Standards Based on WPAN

In order to improve the Wireless Mesh Network technology in Wireless Personal Area Network, the application efficiency of the IEEE organization that made the IEEE 802.15.5 standards, which inherits the IEEE 802.15.1 – 802.15.4 some of the basic ideas, but completely support the reticular structure. Samsung and philips to put forward the preliminary draft of IEEE 802.15.5, divided into a high-speed wireless mesh network and low speed of a domain wireless mesh network two parts. High-speed wireless mesh network, a domain formed by mesh network coordinator self-organizing tree topology structure, and management address pool, address assignment [9]. In this section is proposed based on tree routing policy and centralized routing strategy, including centralized routing policy after the topology server to calculate the optimal path forward data frames between brothers in the tree nodes, achieve the effect of the mesh network, as shown in Fig. 2.

Part a domain at low speed wireless mesh network, and puts forward the adaptive robust tree (ART) and its mesh form – mesh adaptive robust tree (MART) structure, as shown in Fig. 3. Compared with the structure of ART, MART structure be found through a shorter path to grouping the diameter, and avoid the single point of failure.

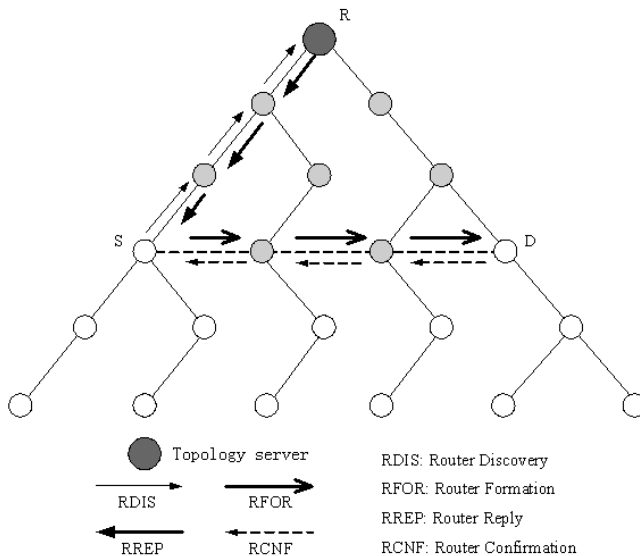


Figure 2. Centralized routing strategy map.

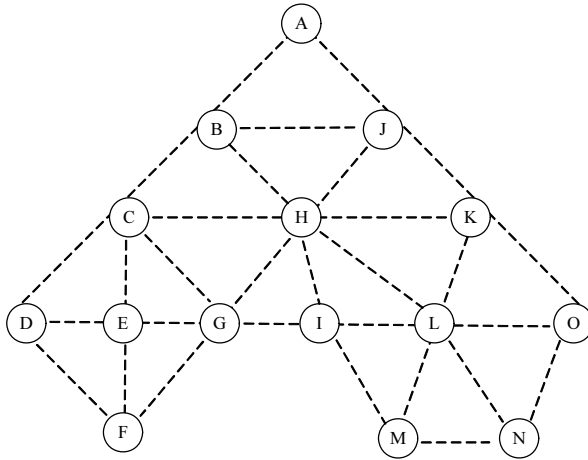


Figure 3. Centralized routing strategy map.

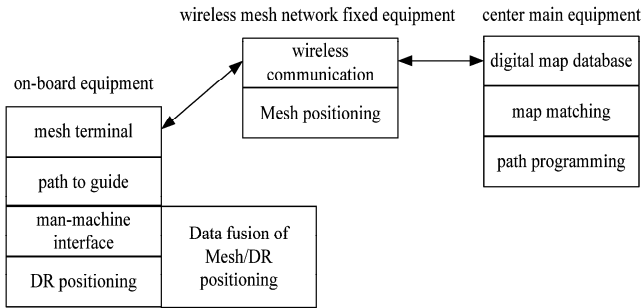


Figure 4. The vehicle navigation system based on wireless Mesh network basic block diagram.

#### 4. Wireless Mesh Network Application in Vehicle Navigation and Positioning System

Vehicle navigation and positioning system is a very important part in intelligent transportation system (ITS) application direction, due to the Wireless Mesh Network with high data rate, large capacity, low communication cost, so the Wireless Mesh Network in the vehicle navigation system obtained the successful application.

Wireless Mesh Network, its application in vehicle navigation system as shown in Fig. 4, the basic framework of electronic map is stored in the control center, not in the on-board equipment, path planning in the control center, complete functions, such as its working process is as follows:

- The fixed equipment for Wireless Mesh Network, the Wireless router and intelligent access point, as a fixed reference point positioning, cooperate with Mesh terminal by measuring the radio of the feedback signal to calculate the location of the terminal;

- Wireless Mesh Network measured terminal location and navigation calculation (DR) technology, the combination of data fusion, can improve the positioning accuracy;
- With the aid of Wireless Mesh Network, the location data back to the center console, coupled with the host on the digital map, map matching, and further improve the positioning accuracy;
- According to the current position of the vehicle and the destination location, path planning on the center console, and then put the planned path information is passed to the terminal, then path guidance, finally with the help of the man-machine interface to realize human-computer interaction.

Compared with the traditional autonomous vehicle navigation system, based on Wireless Mesh Network, the vehicle navigation system has the following advantages:

- Digital map database is stored in the center console, not only save a great deal of terminal equipment storage and computing resources, and when conditions change, such as road reconstruction, road traffic flow changes, such as feature change, only need to center on a host of digital map data update, so the update will be more convenient, fast and efficient, but also reduces the cost;
- Map matching and path planning in the center, further saving the terminal a large amount of storage and computing resources;
- The central decision style design ideas, to achieve effective sharing and distribution of resources, such as path planning, the full source shortest path algorithm will have more practical significance than monophyletic shortest path algorithm.
- With the aid of Wireless Mesh Network positioning method, the control center can be detected in real time traffic information of the whole Network, also can further to dynamic vehicle path planning.
- The vehicle navigation system can be used as an extension of the future mobile communication system and extension, in providing the user with regular phone, Web and multimedia business, such as the vehicle navigation as a value-added service, can increase the penetration of the market.

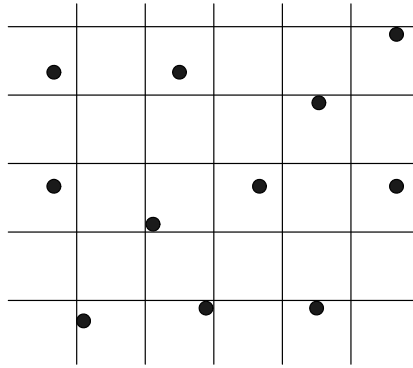
Vehicle positioning methods, based on Wireless Mesh Network, the Network structure of the vehicle navigation system is also different, generally there are two main types of vehicle positioning method. One is to use three or more fixed reference points for vehicle positioning, the network structure as shown in Fig. 5. The second is to use the installed at the intersection of a single fixed reference points for vehicle positioning, the network structure as shown in Fig. 6.

Among them, the line sections, intersection between line segment intersection, solid dot said location fixed reference points. Mobile node to a Mesh of vehicle terminal, fixed node is installed on the roof or a lamppost intelligent wireless router or access point, mobile and fixed nodes communicate in Mesh with each other.

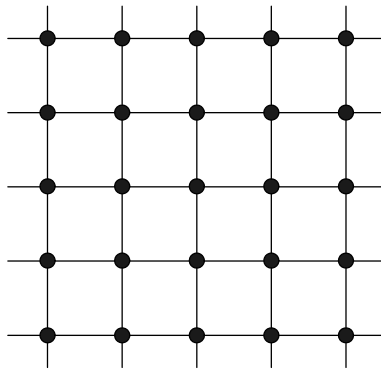
## **5. Conclusion**

Current Wireless Personal Area Network technology has gained more and more extensive application and development, and gradually become one of the key technologies of the application of Internet of things, in order to adapt to the Wireless Personal Area





**Figure 5.** Use the first method of network structure.



**Figure 6.** Use the second method of network structure.

Network in location information service development in the field of application, its traditional tree, linear Network structure already cannot adapt to the development requirements, the introduction of Wireless Mesh Network technology is imminent. Despite the current has been carried out based on Wireless Mesh Network technology of Wireless Personal Area Network, their application and developed specifically for Wireless Mesh Network, the Wireless Personal Area Network application standard, but the standard issue is still in its infancy, still need for Wireless Personal Area Network applications in the field of development to upgrade the standard, make it adapt to the Wireless Personal Area Network technology development.

## References

- [1] Axel Kupper. Location-based Services Fundamentals and Operation. England: John Wiley& Sons Ltd, 2015(2):33–37.
- [2] Uden L. Activity theory for designing mobile learning. International Journal of Mobile Learning and organization, 2007(1):81–102.
- [3] Bellavista P, Corradi A, Foschini L. Enhancing the Scalability of IMS-based Presence Service for LBS Applications. 2009 33rd Annual IEEE International Computer Software and Applications Conference, 2009:29–36.

- [4] Dao D, Rizos Chris, Wang J L. Location-based services. *Technical and business issues*, 2015, 6:169–178.
- [5] Bhuptani M, Moradpour S. *RFID Field Guide: Deploying Radio Frequency Identification Systems*. Prentice Hall PTR, 2015:79–86.
- [6] Medaglia C M, Serbanati A. An Overview of Privacy and Security Issues in the Internet of Things. *The Internet of Things: Proceedings of the 20th Tyrrhenian Workshop on Digital Communications*. Berlin, Germany: Springer-Verlag, 2012:389–394.
- [7] Dao D, Rizos C, Wang J L. Location-based services: technical and business issues, 2012, 6:169–178.
- [8] Kühn P J. Location-Based Services in mobile communication infrastructures. *International Journal of Electronics and Communications*, 2014, 58:159–164.
- [9] Jung E S, Vaidya N H. A power control MAC Protocol for Ad Hoc Networks. *Proceedings of the 8th annual international conference on Mobile computing and networking*, 2012, 36–47.

# Powertrain Parameter Matching and Control Strategy of Electric Drive Chassis for Extended-Range Electric Vehicle

ZHAO Jing-bo and LIU Hai-mei and BEI Shao-yi

*School of Automotive and Traffic Engineering, Jiangsu University of Technology, Changzhou, China*

**Abstract.** In order to evaluate the various performance indexes of extended range electric vehicles, a parameter matching is proposed for a certain extended range electric vehicle. Based on Matlab / Simulink, on the basis of ADVISOR series hybrid electric vehicle model, modify and optimize vehicle dynamics models, motor models and battery models, and propose the control strategy, the dynamic and economic performance was researched in UDDS cycle condition. The results shown that when the motor torque is 0 ~ 200Nm, the engine torque output is 55 ~ 100Nm, the efficiency of the battery is kept above 90% under the CYC\_UDDS cycle condition of the electric vehicle. With the original vehicle dynamics unchanged, the emissions performance and fuel consumption rates have been improved.

**Keywords.** E-REV, parameter matching, ADVISOR, control strategy, power performance

## 1. Introduction

The emergence and development of cars have promoted the development of human civilization, which has brought great convenience for mankind. However, the rapid development of the automotive industry also brought problems such as energy crisis, environmental pollution and other issues at the same time, therefore, energy-saving emission reduction has become the future development target for the car [1,2]. Recent years, new energy vehicles have become the focus of national research. New energy vehicles include pure electric vehicles (BEV), extended range electric vehicles (E-REV), hybrid vehicles (HEV) and fuel cell vehicles (FCEV) and so on.

Compared with pure electric vehicles, the biggest advantage is that mileage is greatly improved, pure electric vehicles driving mileage is still limited due to the battery technology constraints, once the power is exhausted, the car cannot be driven, while E- REV can be refueled at any time. Under the same mileage requirements, the required battery pack is small, without the need for large-capacity battery, and the manufacturing costs significantly reduced, while extending the battery life [3,4].

As the hybrid vehicle uses a complex mechanical power hybrid structure, the engine and the motor compound drive, the battery capacity is very small, it only acts as the auxiliary drive and braking energy recovery [5]. E-REV expands the battery to solve the problem of hybrid battery capacity, the battery is always in a good platform with charge and discharge, extending battery life and reducing maintenance costs.

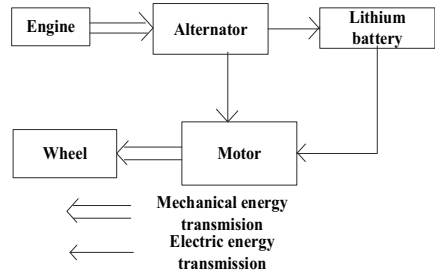
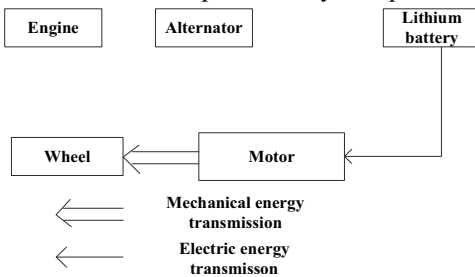
through the control strategy that the engine is always in the best working condition [6]. Besides, extended range of electric vehicles can be an external charge, improving energy efficiency. The battery cost is lower and the technology is more mature [7].

Based on the model of ADVISOR tandem hybrid electric vehicle, the model of vehicle dynamics, generator model and battery model are modified based on Matlab / Simulink platform. This paper has put forward the corresponding control strategy and research the dynamic and economy performance under the UDDS cycle condition, which can provide reference value for the improvement of the control strategy of the extended electric vehicle.

## 2. Parameter Matching of E-REV Powertrain System

### 2.1. Extended-range Mode

Electric car electric chassis is based on the extended range of electric vehicles, it is a power system configuration program. Extended range of electric vehicles can ground charging and one kind of pure electric drive electric vehicles with in-car power supply function. The powertrain system consists of power battery system, power drive system, vehicle control system and auxiliary power units (APU). The car completes the operation control strategy by the vehicle controller. The vehicle operation mode depends on the need work in pure electric mode (Figure 1) or the vehicle power supply mode (HEV), when working in the extended mode (Figure 2), the fuel economy rate with the configuration of the battery capacity increased close to the pure Electric cars infinitely, as a smooth transition model [8]. Because of the low speed high torque, high speed stable operation and the high braking energy recovery efficiency, thus fuel economy, work efficiency and emissions are the under best condition, at the same time the structure is simple and easy to repair.



### 2.2. Drive-motor Parameter Matching

Driving motor is a power source for extended range electric vehicles. Extended range electric vehicles require the drive motor to provide high torque when climbing or in low speed condition, provide high power when accelerating, and large speed range is required, so the motor peak torque must meet the corresponding climbing requirements of the vehicle.

The power of driving motor directly affect the vehicle dynamic performance, the greater the motor power, the backup power is greater when the vehicle is running, the acceleration and climbing performance is outstanding, However, it also increases the

size and mass of the motor itself, thus affect the mass of the vehicle. The rated power of the drive motor is generally determined by the maximum vehicle speed, and the peak power should meet the maximum power requirements corresponding to the maximum speed, acceleration time and climbing performance respectively.

$$P_{m1} = \frac{u_{\max}}{3600\eta_t} \left( mgf + \frac{C_D A u_{\max}^2}{21.15} \right) \quad (1)$$

$$P_{m2} = \frac{u_p}{3600\eta_t} \left( mgf \cos \alpha_{\max} + mg \sin \alpha_{\max} + \frac{C_D A u_{\max}^2}{21.15} \right) \quad (2)$$

$$P_{m3} = \frac{1}{1000\eta_t} \left[ \frac{2}{3} mgf u_f + \frac{1}{5} \rho_a C_D A u_f^3 + \frac{\delta m}{2t_a} (u_f^2 + u_b^2) \right] \quad (3)$$

In three formulas above,  $P_{m1}$  is the motor power required for the maximum speed;  $u_{\max}$  is the maximum speed,  $\eta_t$  is the transmission efficiency of the transmission system;  $m$  is the vehicle mass;  $f$  is the rolling resistance coefficient,  $C_D$  is the drag coefficient,  $A$  is the windward area,  $P_{m2}$  is the motor power required for the maximum slope angle,  $\alpha_{\max}$  is maximum slope angle  $u_p$  is the speed of the vehicle when climbing,  $P_{m3}$  is the motor power required for the acceleration time,  $u_f$  is the final speed after the end of the acceleration,  $u_b$  is the speed corresponding to the rated speed,  $t_a$  is the expected acceleration time, and  $\delta$  is the rotational mass conversion factor. The motor rated power and peak power is shown by

$$P_e \geq P_{m1} \quad (4)$$

$$P_{e\max} \geq [P_{m1} \ P_{m2} \ P_{m3}] \quad (5)$$

$$P_{e\max} = \lambda P_e \quad (6)$$

$\lambda$  is the motor overload factor. The rated torque and peak torque of the motor are

$$T_e = \frac{P_e \times 9550}{n_e} \quad (7)$$

$$T_{e\max} = \frac{P_{e\max} \times 9550}{n_e} \quad (8)$$

$n_e$  is the motor speed.

After the drive motor parameters are determined initially, it is also necessary to verify that the maximum speed and the maximum speed gradient are met, ie.

$$\frac{mg}{T_{e\max}\eta_i}(f\cos\alpha_{\max} + \sin\alpha_{\max} + \frac{C_D A u_p^2}{21.12mg}) \leq \frac{i_0}{r} \leq \frac{0.377n_{\max}}{u_{\max}} \quad (9)$$

By the above formula, the drive motor selects permanent magnet synchronous motor, the peak power of  $P_{e\max}=100\text{kw}$ , overload factor  $\lambda=2$ , then  $P_e=50\text{kw}$ . Rated torque  $T_e=114\text{Nm}$ , the peak torque  $T_{e\max}=178\text{Nm}$ , maximum speed is  $7000\text{r/min}$ , rated speed is  $3000\text{r/min}$ . It meets the maximum speed and maximum climbing requirements.

### 2.3. Battery Parameters Matching

Electric vehicle battery compartments are the main energy sources for the whole vehicle, the battery parameters matching include the battery type selection, the number of single battery cell and the battery sets voltage matching [9]. For extended range electric vehicles, the required voltage level of the battery and the motor voltage level need be consistent, and meet the motor voltage change requirements. Battery energy should meet

$$E_B = \frac{U_m C_E}{1000} \quad (10)$$

$$E_B \geq \frac{mgf + C_D A u_a^2 / 21.15}{3600 \times DOD \eta_i \eta_{mc} \eta_{dis} (1 - \eta_a)} \times S_1 \quad (11)$$

Battery energy should meet the conditions.

$E_B$  is battery energy (kw h),  $U_m$  is battery terminal voltage (V),  $C_E$  is battery capacity (A·h),  $\eta_{mc}$  is the motor efficiency,  $\eta_{dis}$  is the battery discharge efficiency,  $\eta_a$  is the energy consumption ratio of automobile accessories,  $S_1$  is the pure electric mileage (km), DOD is the battery discharge depth.

The maximum discharge power of the battery should be meet

$$P_{bat\_max} \geq \frac{P_{\max}}{\eta_{mc}} + P_A \quad (12)$$

The number of batteries n is calculated as follows

$$n = 1000 \frac{p_e(S/v_c)}{UC} \quad (13)$$

After calculation of the above formula, the battery type is selected the lithium iron phosphate battery, the number of monomer is 90, the rated voltage is 288V, the capacity is 63A.h, SOC is between 30% ~ 100%, the maximum discharge rate is 5C, the maximum charging rate is 3C.

### 2.4. Parameter Matching of Range Extender

The extended range is fixed with the power system of the pure electric vehicle. When the battery SOC value is lower than the set value or the battery fails, the average running speed can be maintained at the same speed. In the extended mode, the engine

provides power source, it requires quite dynamic performance, so matching the engine / generator parameters is necessary [10].

Usually to meet the maximum speed as the standard primary engine power. The engine rated power should be chosen greater than the theoretical value to carry a continuous non-traction load. At the same time, calculate the tank volume according to the engine fuel consumption MAP.

$$P_{RE} = \frac{1}{3600\eta_t} (mgf_{u_{max}} + \frac{C_D A u_{max}^3}{21.15}) \tag{14}$$

$$V = \frac{S_2 f_c}{u_a} \tag{15}$$

$S_2$  is the extended driving mileage.  $f_c$  is the engine fuel consumption.

Engine is the inline four-cylinder gasoline engine, with 43KW power, 1.5L displacement, 15L tank volume; generator is permanent magnet synchronous motor, output power is 32KW, calibration working speed is 4000r / min, rated voltage is 288V.

### 3. Models for E-REV power system main components

#### 3.1. Body Simulation Model

Body simulation model shown in Figure 3, including rolling resistance, slope resistance, wind resistance, acceleration resistance calculation sub-module, and car speed calculation sub-module.

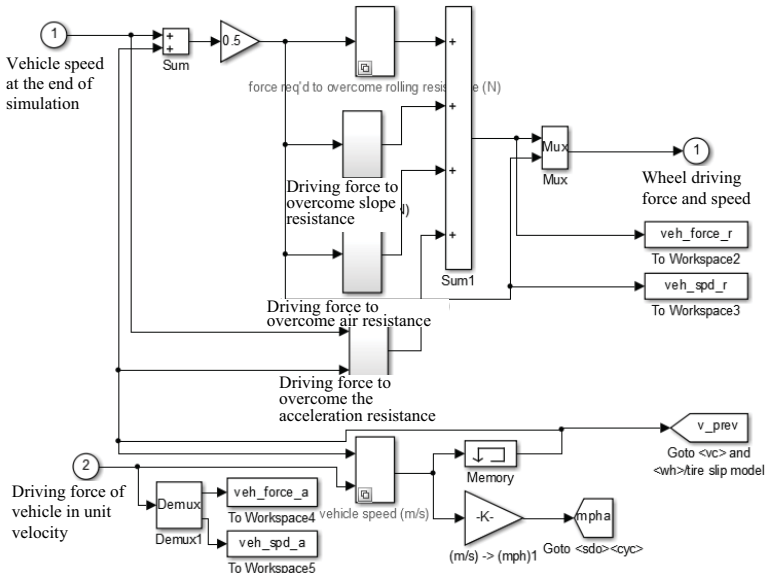


Figure 3. Simulink Model of Electric Vehicle.





#### 4. Dynamic Performance Analysis of E-REV

The constant power has a high efficiency, and the power-follow control mode has a good fuel economy and dynamic, thus in different vehicle operation mode, choose the appropriate power control strategy respectively, switching between the two strategies will make full advantages.

In the pure electric drive mode, the battery is the only power source, during the running course the SOC value is declining, thus the maximum value of SOC is set to 0.9 (to prevent overcharge to the battery damage). When the battery SOC to a minimum, the extension device begins to work, the SOC is not a fixed value, instead be in a range. In order to prevent the battery transition charge and discharge effectively, it should set the maximum and minimum SOC value. When the car's SOC reach to a minimum value, the range extender begins to work, electric vehicles step into the extended mode of operation. In the power-follow model, the engine always work in the lowest fuel consumption area, there are two energy transmission lines currently.

Comparing the dynamic and economy of the two energy transfer routes, calculated that the first route is more economical while the energy loss for the second energy transfer route is higher due to the lower charge and discharge efficiency of the battery. However, from the whole vehicle running conditions, if rely the first energy transfer route, then the corresponding engine running state will be the same as the traditional engine, the energy utilization is lower, therefore, it should be based on different functional conditions and analyzed the two energy transfer routes. The SOC at 0.32-0.75 is the efficient discharge area, define the region positioning as work area.

##### 4.1. Choose of Circulation Conditions

Select the urban road cycle CYC\_UDDS (Urban Dynamometer Driving Schedule) by US Environmental Protection Agency (EPA) in ADVISOR as a cycle for the vehicle performance simulation [11], the design parameters shown in Table 1.

**Table 1.** Design parameters of extended range electric vehicle.

Device	Parameters	Value
Motor	Rated power	50Kw
	Peak power	100Kw
	peak speed	3000r/min
	Peak speed	7000r/min
Power battery pack	voltage	288V
	capacity	63Ah
generator	Rated power	32kw
engine	power	43kw
Actuator main reducer	Transmission ratio / $i_0$	6.058
Whole vehicle	mass / $m$	1430kg
	Windward area / $C_D$	1.97 m <sup>2</sup>
	Tire rolling radius / $r_b$	0.334 m

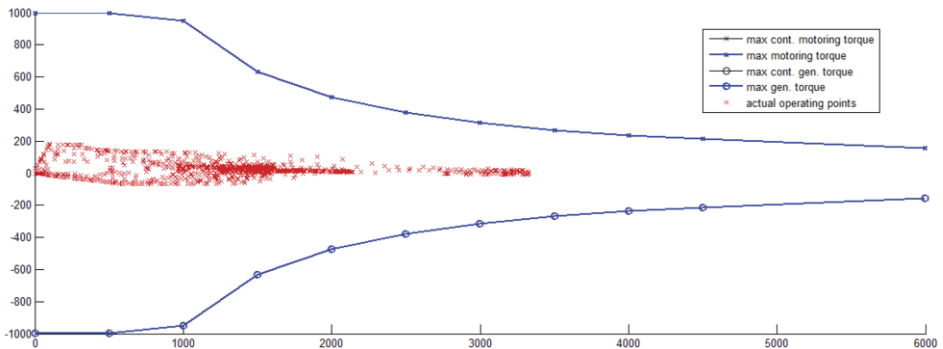
On the base of ADVISOR's original tandem hybrid electric vehicle roof module, modify the corresponding control strategy, whole vehicle and drive motor and other script files. Modify the PRIUS script file, including the motor, power battery pack, generator, engine and vehicle parameters. Load script file in the ADVISOR, select CYC\_UDDS cycle conditions, set the acceleration test and slope test parameters, then have the simulation test. The parameters set shown in Table 2.

**Table 2.** Cycle condition parameter setting.

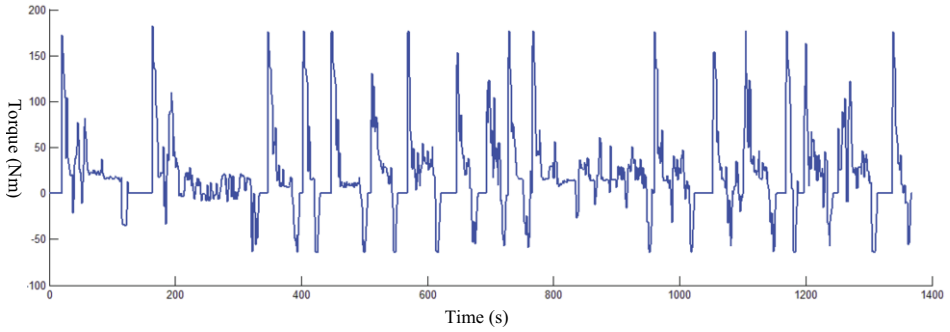
Drive Cycle	CYC_UDDS
Time	1369s
Distance	11.99km
Max Speed	91.25km/h
Average speed	31.25km/h
Average grade ability	5%

#### 4.2. Dynamic Performance Analysis

Automotive dynamic performance is mainly evaluated by the three indicators, the maximum speed, the car acceleration time and the maximum climbing degree [12]. Figure 6 and Figure 7 are the electric motor operating point diagram and the motor torque output diagram in the CYC\_UDDS cycle conditions, the motor torque is concentrated between 0 to 200Nm, and the peak torque is about 200Nm while the peak torque of the selected motor is 178Nm, thus the torque required for normal operation for the motor is less than 178Nm, Obviously the motor can work normally, and meanwhile have a certain overload capacity that can meet the peak requirements for vehicle's dynamic [13]. As can be seen from Table 3, the design has reached the expected development goals for the dynamic performance of electric vehicles.



**Figure 6.** Operating point of motor under UDDS cycle condition.

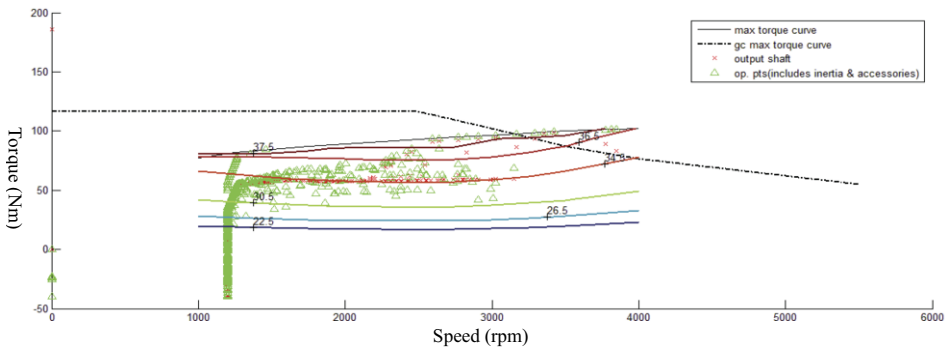


**Figure 7.** Output torque of motor under UDDS cycle.

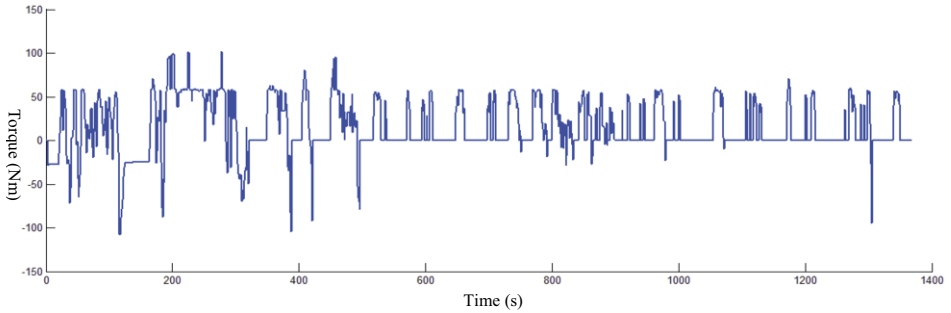
**Table 3.** Velocity, acceleration and emission changes under UDDS cycle.

Parameters	Value
Initial battery SOC	0.7
Simulation finished SOC	0.66
Maximum velocity ( km/h )	142.5 ≥ 120
Maximum acceleration ( $m / s^2$ )	4.8
0—100 km/h acceleration time	9s ≤ 12s
60—100 km/h acceleration time	5.2 ≤ 8s
HC Emissions ( g/km )	0.0664
CO Emissions ( g/km )	0.0771
$NO_x$ Emissions ( g/km )	0.0144
PM Emissions ( g/km )	0

Figure 8 and Figure 9 are the electric car engine working point figure and the engine torque output figure in the CYC\_UDDS cycle, it can be seen from the diagram that the engine torque output within 55 to 100Nm, the peak torque is close to 120Nm, and the selection of engine maximum power is 43kw, maximum torque is 125Nm, so it can work normally and meet the requirements of the vehicle's dynamic in the CYC\_UDDS cycle.



**Figure 8.** The operating point diagram of the engine under the cycle of UDDS.



**Figure 9.** UDDS cycle condition cycle engine torque output diagram.

In summary, through the CYC\_UDDS cycle simulation, the results show that the engine and motor selection are reasonable and can meet the requirements of electric vehicle power.

## 5. Summary

The design and matching of the motorized chassis for the extended electric vehicle are carried out, and calculated the reasonable SOC upper and lower limits, meanwhile proposed the corresponding working mode and strategy. Taking the ADVISOR tandem hybrid electric vehicle model as an example, simulated the CYC\_UDDS cycle performance of the vehicle. The vehicle's dynamic performance was consistent with the expected design goals. It proposes a new idea for the design and parameter matching of the control strategy for the extended electric vehicle, which provided the reference and foundation for the research on the electric vehicle, shortened the research and development process.

## Acknowledgment

Funding from the National Natural Science Foundation of China (Grant No: 61503163), the "333 project" of Jiangsu Province (Grant No: BRA2016440) and the six talent peaks project in Jiangsu Province (Grant No: ZBZZ-024) are gratefully acknowledged.

## References

- [1] Keld Laursen, Toke Reichstein, Ammon Salter, Exploring the effect of geographical proximity and university quality on university-industry collaboration in the United Kingdom, *Regional Studies*, **45** (2011), 507-523.
- [2] Qi X, Wu G, Boriboonsomsin K, Data-Driven Reinforcement Learning - Based Real-Time Energy Management System for Plug-In Hybrid Electric Vehicles, *Transportation Research Record Journal of the Transportation Research Board*, **2572**(2016) 1-8.
- [3] Kim N, Cha S, Peng H, Optimal Control of Hybrid Electric Vehicles Based on Pontryagin's Minimum Principle, *IEEE Transactions on Control Systems Technology*, **19**(2011), 1279-1287.
- [4] Zhu X, Zhang H, Cao D, Robust control of integrated motor-transmission powertrain system over controller area network for automotive applications, *Mechanical Systems & Signal Processing*, **s58-59** (2015), 5-28.

- [5] Rahman K M, Jurkovic S, Stancu C, Design and Performance of Electrical Propulsion System of Extended Range Electric Vehicle (EREV) Chevrolet Volt, *IEEE Transactions on Industry Applications*, **51**(2012), 2479-2488.
- [6] Qiu X, Huang W, Bu F, Torque-Angle-Based Direct Torque Control for Interior Permanent-Magnet Synchronous Motor Drivers in Electric Vehicles, *Journal of Power Electronics*, **13**(2013), 964-974.
- [7] Zhu X, Zhang H, Fang Z, Speed synchronization control for integrated automotive motor-transmission powertrain system with random delays, *Mechanical Systems & Signal Processing*, **s64-65**(2015), 46-57.
- [8] Zhu D, Zhou D, Zhou J, Synchronization Control for a Class of Underactuated Mechanical Systems via Energy Shaping, *Journal of Dynamic Systems Measurement & Control*, **134** (2012), 041007.
- [9] Budde-Meiwes H, Drillkens J, Lunz B, A review of current automotive battery technology and future prospects', *Proceedings of the Institution of Mechanical Engineers Part D Journal of Automobile Engineering*, **227**(2013), 61-76.
- [10] Rahman K, Jurkovic S, Stancu C, Design and performance of electrical propulsion system of extended range electric vehicle (EREV) Chevrolet Volt, *Energy Conversion Congress and Exposition*, **51**(2012), 4152-4159.
- [11] Anwar M, Hayes M, Tata A, Power Dense and Robust Traction Power Inverter for the Second-Generation Chevrolet Volt Extended-Range EV, *Lecture Notes in Computer Science*, **4**(2015), 94-103.
- [12] Sciarretta A, Serrao L, Dewangan P C, A control benchmark on the energy management of a plug-in hybrid electric vehicle, *Control Engineering Practice*, **29**(2014), 287-298 .
- [13] Qian L, Qiu L, Xin F, Energy management and torque coordination control for plug-in 4WD hybrid electric vehicle, *Transactions of the Chinese Society of Agricultural Engineering*, **30**(2014), 55-64.

# Studies on Gravimeter Leveling Algorithm Based on Mixed Sensitivity Robust Control Theory

WANG Hai-miao<sup>1\*</sup>, Qin Yong-yuan<sup>1</sup>, QIAO Xiang-wei<sup>2</sup>, ZHANG Zhao-fei<sup>1</sup>

1. School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

2. Xi'an Aerospace Precision Electromechanical Institute, Xi'an 710100, China

\*E-mail: 139913269@qq.com

**Abstract.** For providing a good condition of measuring, the gravimeter gyro-stabled platform plays an important part in the gravimeter system. By adopting the pure-inertial and the all-digital controlling technology of real-time leveling, the gravimeter gyro-stabled platform will be directed primarily towards isolating the angular motion of the carrier, and keeping the sensitive axis of the gravimeter parallel with the plumb line all the time. In the sea condition, the gravimeter gyro-stabled platform has a low precision in the disturbing of low frequency of sea waves, this paper introduces a leveling algorithm based on Mixed Sensitivity. By adopting a reasonable leveling bandwidth, performance bound function, controller weighting function and bounded uncertainty function, we achieve an optimal control to the leveling loop via the algorithm.

**Keywords.** Sea wave disturbance; leveling; mixed sensitivity robust control.

## 1. Introduction

The gravimeter gyro-stabled platform is an important part of the gravimeter system, by adopting pure-inertial and all-digital controlling technology, it will be directed primarily towards isolating the angular motion of the carrier, and keeping the sensors of gravimeter parallel with the vertical all the time.

In moving base condition, pure inertia leveling error is related to the working environment. When working in the air or on the land, besides accelerometer zero position error, the main error is high frequency disturbance caused by carrier swaying and vibrating. When working on the sea, the main error is low frequency disturbance caused by the wave. For reducing the low frequency disturbance, to design a traditional PID control algorithm is usually by experience when choosing a bandwidth, it is difficult to juggle system responding speed and leveling accuracy. When the winds and currents are strong, the dynamic error of the platform is obvious, and it will cause the measuring precision of the gravity deteriorated.

For solving the problem caused by the bandwidth choosing and the poor dynamic accuracy of the traditional PID control algorithm, this paper introduces a leveling algorithm for the gravimeter gyro-stabled platform based on Mixed Sensitivity Robust Control, and utilizes the real data from the sea trail to simulate, the result shows that the algorithm can reach a high leveling precision in short time.

## 2. Model Building and Bandwidth Choosing

Gravimeter gyro-stabled platform leveling loop is including the accelerometer, ADC, digital filter corrective network, pulse width modulation circuit, gyro torquer and stabilization loop. The concrete structure of the loop is shown in Fig.1.

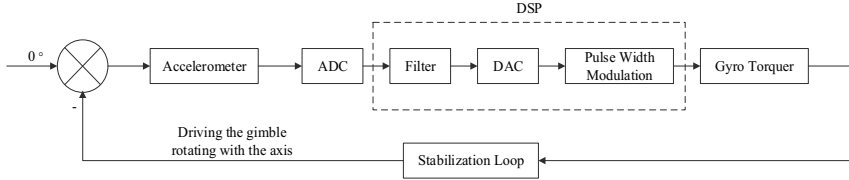


Fig. 1 The leveling process

The platform leveling is using the horizontal gyros and accelerometers, the X gyro-Z accelerometer combination and the Z gyro-X accelerometer combination, they form 2 self-leveling loops to level the horizontal gimbles of platform to parallel with the horizontal plane.

### 2.1 The Principle of the Leveling

The gravimeter gyro-stabled platform leveling is based on gravity vector G, two horizontal accelerometers are taken as sensors, the plane that determined by the output axes of the accelerometers is the initial coordinate plane, and the leveling is to set the inertial coordinate plant coincide with the real horizontal plane. The mathematical model of the platform is shown in Fig.2.

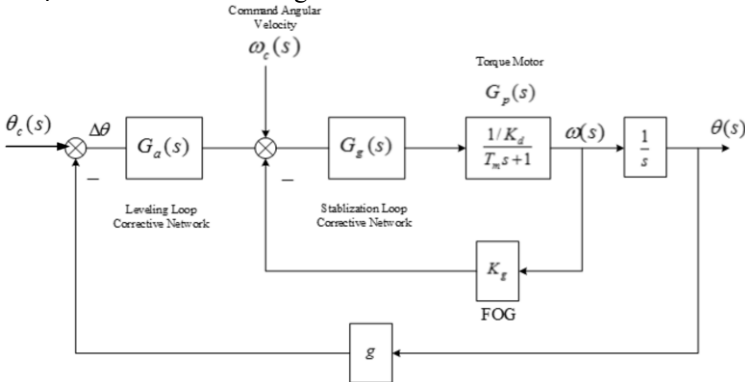


Fig. 2 The Gravimeter Platform Leveling Loop Schematic Diagram

From Fig.2, the platform can be considered as an outer-axis control loop around the stabilization loop, the leveling loop converts the accelerometer output into the axis rotating command angular velocity  $\omega_c(s)$ , it's synthesized with the gyro output, and generates an offset command signal applied to the stabilization loop, then controls the gimbles to rotate, finally the input axis of the two horizontal accelerometers will parallel with the horizontal plane, the leveling is accomplished.

### 2.2 The Leveling Loop Modeling

Taking the X-axis as an example, the digital leveling loop modeling is shown in Fig.3.

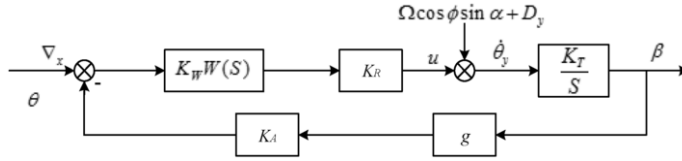


Fig. 3 The model of X-axis leveling loop

From Fig.3, the X-axis leveling open loop transfer function is

$$G(s) = g \frac{K_W K_R K_T K_A W(s)}{s} \quad (1)$$

Where:  $K_A$ —the accelerator ADC scale factor;

$K_W$ —the leveling loop gain;

$K_R$ —the DAC scale factor;

$K_T$ —the stabilization loop moment applied factor

$\nabla_x$ —the X accelerometer zero position

$D_y$ —the Y gyro drift

### 2.3 The Bandwidth Choosing under Low Frequency Disturbance

Under the swaying disturbing acceleration  $a(t) = a_m \sin(\omega_h t)$ , to achieve the goal that leveling precision  $\leq \delta$ , we need to choose a proper loop bandwidth. If the leveling open-loop transfer function is  $G(s) = gG_h(s)$ , the system closed-loop transfer function under the disturbance will be<sup>[1-3]</sup>:

$$\frac{\beta(s)}{a(s)} = \frac{G_h(s)}{1 + gG_h(s)} \quad (2)$$

From (2), when  $s \rightarrow j\omega_h$ , and if  $|G_h(j\omega_h)| \gg 1$ , there is  $\frac{\beta(s)}{a(s)} = \frac{1}{g}$ , and the bandwidth will be  $\omega_c > \omega_h$ , the leveling loop is weak in the disturbance rejection capability, and it is hard to satisfy the demand.

Therefore, when  $s \rightarrow j\omega_h$ , we usually set  $|G_h(j\omega_h)| \ll 1$ , that the bandwidth of the open-loop transfer function will be  $\omega_c < \omega_h$ , and there is

$$\frac{\beta(s)}{a(s)} = \frac{G_h(s)}{1 + gG_h(s)} \approx G_h(s) = \frac{G(s)}{g} \quad (3)$$

And to attenuate the disturbance of  $\left| \frac{G(j\omega_h)}{g} \right| = \frac{\delta}{\alpha_m}$  at the frequency of  $\omega_h$ , the

open-loop function must satisfy the condition that  $G(j\omega_h) \leq \frac{\delta}{\alpha_m} g$ .

From the analysis above, conclusions can be drawn:

a) the bandwidth of the open-loop transfer function must be lower than the disturbing frequency  $\omega_h$ ;





By choosing a proper performance bound function  $W_1(s) = K_1$ , the leveling adjustment time can be guaranteed under the sea condition;

By choosing a proper controller weighted function  $W_3(s) = \frac{T_1s + 1}{s(T_2s + 1)}$ , the amplitude of the signal before  $K_R$  can be limited, and the phase of the closed-loop transfer function will be an integer multiple of  $90^\circ$ .

For cancelling the interference of swaying, the adjustment time must be as short as possible, on the basis of mixed sensitivity robust control theory, there is<sup>[6, 7]</sup>:

$$\begin{bmatrix} W_1(s)S(s) \\ W_2(s)T(s) \\ W_3(s)R(s) \end{bmatrix} \approx 1 \quad (4)$$

Where  $S(s)$  is the sensitivity function of the system,  $T(s)$  is the complementary sensitivity function, and  $R(s)$  is the sensitivity function of the controller, each expression is shown as below:

$$S(s) = (I + K_W W(s)K_R)^{-1} \quad (5)$$

$$T(s) = K_W W(s)K_R (I + K_W W(s)K_R)^{-1} \quad (6)$$

$$R(s) = K_W W(s) (I + K_W W(s)K_R)^{-1} \quad (7)$$

$\begin{bmatrix} W_1(s)S(s) \\ W_2(s)T(s) \\ W_3(s)R(s) \end{bmatrix}$  is the closed-loop transfer function of the leveling loop.

According to the (4)~(7), we obtain the optimum controller of the leveling loop<sup>[8]</sup>:

$$W(s) = \frac{K(Ts + 1)}{s(as^2 + bs + 1)} = \frac{3 \times 10^{-5}(69s + 1)}{s((6.5s)^2 + 9.2s + 1)}$$

According to the optimum controller of leveling loop, we finally get the closed-loop transfer function of the leveling loop.

$$\phi(s) = \frac{B_1s^5 + B_2s^4 + B_3s^3 + B_4s^2 + B_5s + 1}{A_1s^5 + A_2s^4 + A_3s^3 + A_4s^2 + A_5s + 1}$$

The bode and the step response are shown in Fig.5 and 6.

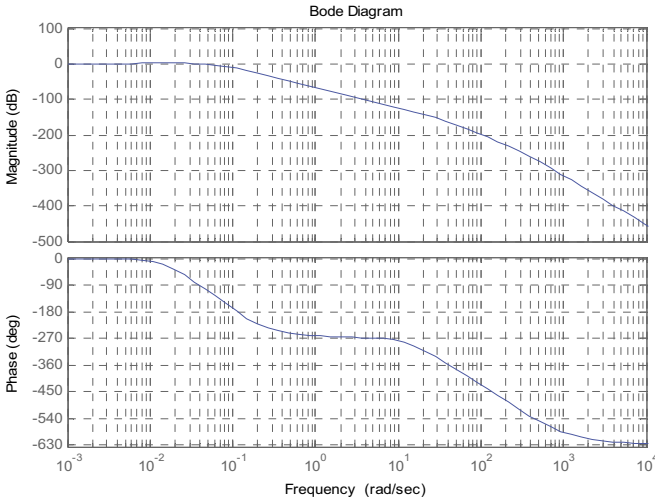


Fig.5 The Closed-loop Bode Diagram of the Gravimeter Leveling

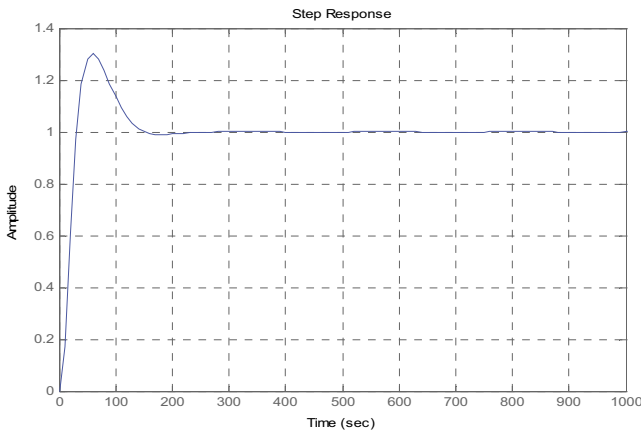


Fig.6 The gravimeter leveling loop closed-loop step response

From Fig.5 and 6, we can see that the system has an attenuation of 56dB at the frequency of 0.1Hz, and the closed-loop Step adjustment time is approximately 200s.

#### 4. Simulation Experiment

For comparison, by using the PID algorithm and the leveling algorithm based on Mixed Sensitivity Robust Control separately, we adopt the real sea wave data acquired by the gravimeter to conduct a simulation of leveling. The leveling precision of X-axis is shown in Fig.7 and 8.

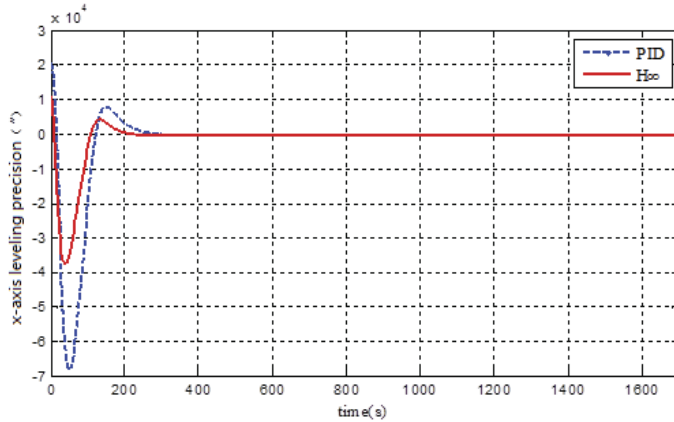


Fig.7 The X-axis Leveling Precision under Two Algorithms

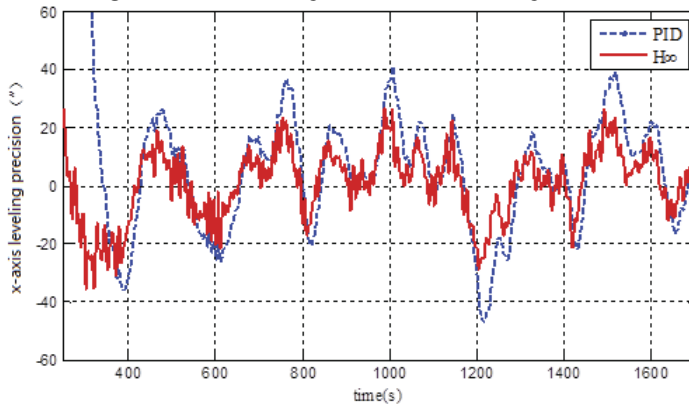


Fig.8 The X-axis Leveling Precision under Two Algorithms (the partial enlargement of Fig.7)

From Fig.7 and 8, in real condition of the sea, it takes about 200 sec to complete a whole procedure of leveling by adopting the leveling algorithm based on Mixed Sensitivity Robust Control, and the maximum error is approximately 28", while it takes about 280 sec to complete the leveling by adopting the traditional PID control algorithm, and the maximum error is over 40". The reason is that the leveling algorithm based on Mixed Sensitivity Robust Control can accommodate the real changes of the sea wave better than the traditional way.

## 5. Conclusion

For the problem that the gravimeter gyro-stabled platform has a low accuracy of leveling affected by the low frequency disturbance from the sea wave, this paper introduces a leveling algorithm based on Mixed Sensitivity Robust Control. By adopting the sea wave data, the behaviors of leveling are simulated through the traditional PID control algorithm and leveling algorithm based on Mixed Sensitivity Robust Control separately. The result shows that the leveling algorithm based on Mixed Sensitivity Robust Control can adapt to the change of the sea wave better than the traditional method, furthermore, it can achieve a better precision of leveling.

## References

- [1] John C. Doyle, Keith Glover, Pramod P. Khargonekar and Bruce A. Francis. State-Space Solutions to Standard  $H_2$  and  $H_\infty$  Control Problems. IEEE Trans. on Automatic Control. 1989, 34(8): 831-847
- [2] A. H. M. A. Rahim, E. P. Nowicki. A Robust Damping Controller for SMES Using Loop-shaping Technique. Electrical Power and Energy System, 2005, 27(4): 465-471
- [3] Kemin Zhou, J. C. Doyle. Robust and Optimal Control. Englewood Cliffs, NJ. Prentice, 1995:41-43
- [4] B. K. Kim, W. K. Chung. Unified Analysis and Design of Robust Disturbance Attenuation Algorithms Using Inherent Structural Equivalence. American Conf., Arlington, VA, 2001: 4046~4051
- [5] B. K. Kim, W. K. Chung. Advanced Design of Disturbance Observer for High Performance Motion Control System. American Control Conf., Anchorage, AK, 2002: 2112-2118
- [6] T. Mita, X. Xin and Brian D. O. Anderson. Extended  $H_\infty$  Control-Solving  $H_\infty$  Servo and Estimation Problem. Proceedings of the 36<sup>th</sup> Conference on Decision & Control, 1997, 42(6): 4653-4658
- [7] Glover K. Pole, Zero cancellations in the general  $H_\infty$ -problem with reference to a two block design. Sys Contro Letter, 1990, 14:295~306
- [8] Tsai M C, Geddes E J M and Postlethwaite I. Pole-zero Cancellations and Closed-loop Properties of an  $H_\infty$  mixed sensitivity design problem, Automatica, 1992, 28(3): 1519-1530

# An Approach of Telecommunication System for Vehicles and Infrastructures

Jiawei Wang<sup>a</sup> Nan Xu<sup>a,1</sup> Zhiyuan Liao<sup>a</sup> Cong Fei<sup>a</sup> Jiaheng Li<sup>a</sup> Wenpeng Zeng<sup>a</sup>

<sup>a</sup>College of Automotive Engineering , Jilin University , Changchun , China

**Abstract.** This paper introduces the technical characteristics of ZigBee and Bluetooth communication protocol, the development trend and some applications in the field of intelligent traffic vehicle communication. This paper discusses the condition that traffic information is sent from the road side to the vehicle and from the vehicle to the road side through ZigBee and Bluetooth communication technology. The model is established to analyze the corresponding transmission frequency and simulate it. Expect to make a positive role in promoting the development of intelligent transportation technology.

**Keywords.** ZigBee, Bluetooth, Android, intelligent traffic, Queuing theory

## 1. Introduction

In the process of intelligent transportation system research, researchers usually pay high attention to choices of wireless communication technique. Techniques such as RFID, GPS, WLAN and LTE wireless network still fail to be popular and accepted by public because of some of the limitations of its own and obstacles in technology. This article discusses two wireless communication techniques : combining ZigBee and Bluetooth to realize transmitting message from roadside device to vehicle-mounted device or transmitting messages from vehicle-mounted roadside device.

ZigBee is a new wireless communication technique. The biggest characteristic is it can constitute a network by its own. What's more, it has simple equipment and low energy consumption. The cost of ZigBee is lower than Bluetooth and WiFi<sup>[1]</sup>.It is really convent to use ZigBee in traffic communication for its simple equipment and low energy consumption and costing. Although, it has low transmission velocity, it can adopt its good point and avoid its shortcomings in traffic communication. It can give full play to its advantages of low energy and consumption costing<sup>[2]</sup>.Bluetooth uses the 2.4 GHz frequency band of radio waves.Its advantage is high communication speed and accurate information transmission. It also has low energy and consumption costing. Standby mode power<sup>[3]</sup>.The comparison of ZigBee and Bluetooth is as table 1.

---

<sup>1</sup>Corresponding Author : Nan Xu , College of Automotive Engineering , Jilin University , Changchun , Renmin Avenue 5988 , Nanling Campus of Jilin University ; E-mail : xn19841231@163.com.

**Table 1.** The comparison of ZigBee and Bluetooth <sup>[4]</sup>

<b>Communication technology</b>	<b>Bluetooth</b>	<b>ZigBee</b>
Operating frequency	2.4GHz	2.4GHz/868/915MHz
Ideal transmission rate /Mb/s	3	0.25
The effective transmission distance	10	>100
Number of nodes	8	255
Transmission power/m W	1~100	1~3

## 2.The intelligence traffic sign based on the technology ZigBee and Bluetooth

### 2.1.Summary of tools and the implementation process

XBee is the module of ZigBee designed by American company DIGI. The feature of this module is its low costing and large Information capacity. BLE LINK Bluetooth 4.0 transmission module is a Bluetooth module designed by DFRobot Company, which uses TI CC2540 chips and can match and transmit data with smart phones Bluetooth.

This device uses two XBee modules and a BLE LINK Bluetooth module, use Arduino IDE software to design the program. This software is based on the C language which can burn the program through wired mode. The road traffic information acquisition is accomplished in three steps. The first step, transmitting the information from the roadside XBee module to on-board XBee module. Roadside XBee module is mounted on Arduino UNO card through XBee shield V1.4.This progress is accomplished by communication protocol of ZigBee. The second step is transmitting the information from on-board XBee module to BLE LINK

Bluetooth module. The on-board XBee module and BLE LINK Bluetooth module are mounted on the same Arduino Mega 2560.This progress is accomplished by Arduino IDE. The third step is transmitting the information from BLE LINK

Bluetooth module to smart phone, and then translating to language, words and pictures. In the same time, reading the information such as speed, fuel consumption through OBD, and write this information to on-board Arduino card which can be read by on-board Xbee module, transmit it to roadside XBee module and cloud. See the design flow chart in Figure 1.

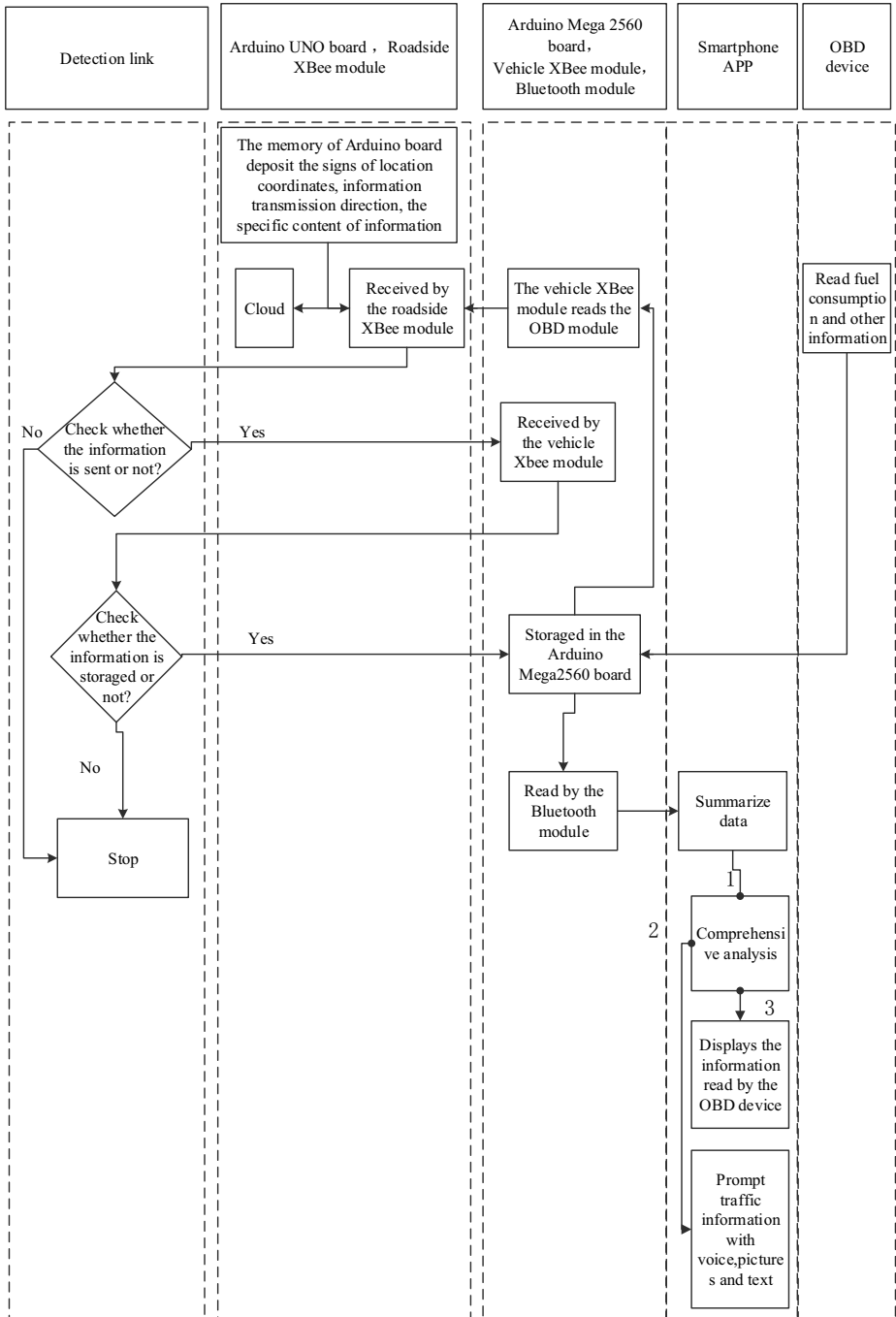


Figure 1. Design flow chart



### 2.2. The coding design of traffic information

On the current national traffic laws and regulations, the existing major road traffic signs can be broadly divided into seven categories. Each type of category has dozens of specific traffic information, after the completion of the collection and classification of these traffic signs, combined with the idea of information transmission, we propose the following traffic information coding scheme.

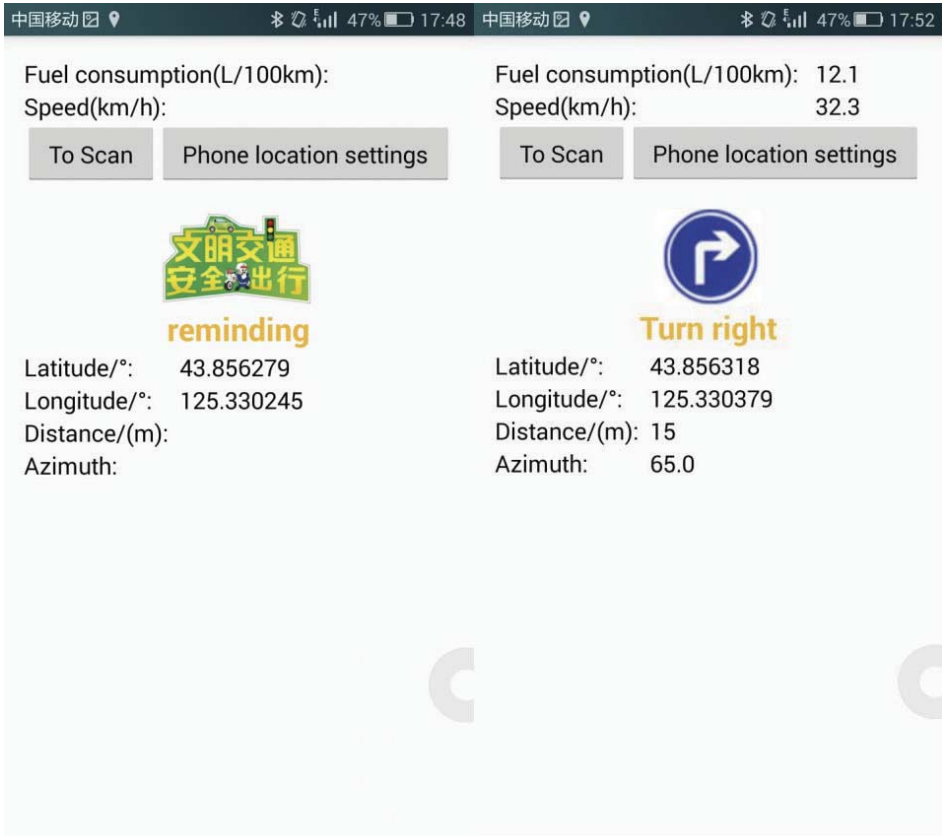
The traffic information is composed of the roadside device coordinate information, the vehicle traveling direction information, and the road sign specific information, Wherein the coordinate information is composed of the longitude and the latitude of the position where the transmitting end is located and the traveling direction information is composed of the angle value formed by the traveling direction of the vehicle and the east direction. The specific information of road signs is divided into warnings, bans, instructions, tourist areas, road construction safety, general road guide, highway six modules. The information under each module consists of a specific code, specifically, a capitalized English alphabet + three Arabic numerals, in which capital letters represent the type of module, G is the warning sign, L indicates ban sign, Z indicates traffic indicator signs, L indicates Tourist area signs, A indicates road construction safety signs, Y indicates general road signs, H indicates highway signs; The two digits represent the ordinal of the information under this type of module, and a check sum is added at the end of the information. As shown in Table 2.

**Table 2.** Information coding rules and examples

The coordinates information of roadside device	Vehicle driving directions	Roadside device coordinates information	Check digit
Including longitude and latitude	The angle between driving direction and east direction	A capital letter + three digits	The number between two "T" indicates the number of Arabic numerals in the message sent.
043.85968777126736N 125.33139811197917E	5.80D	Z003	T41T

### 2.3 Practical Application of Intelligent Traffic Vehicle Communication Device

Assemble this vehicle road communication device according to the above part connection scheme. Two kinds of road traffic information are designed in the test of real vehicle, which are travelling around the island and turning right. The roadside ZigBee device is set near the road signs and is connected the solar panels to the battery. We can see that the Arduino UNO board indicator light is lit and starts flashing, indicating that the device is already working properly. Driving the car forward in the direction of the roadside device, placing the vehicle device in the car and turn on its power, turn on the smartphone application and confirm that the Bluetooth is open and paired with the Bluetooth module on the board. In the course of vehicle travel, we can see the phone application on the screen as shown in Figure 2. And the smartphone is playing the corresponding voice to prompt traffic information.



**Figure 2.** Device connection diagram and the interface when the phone application receives the message

### 3. Exploration on the sending & receiving frequency of vehicle-mounted device and roadside device

#### 3.1 The calculation & simulation of sending frequency from vehicle-mounted device to roadside device

There may be packet loss during the process of sending message of vehicles from vehicle-mounted device to roadside device because of the mutual interference between two cars. In order to ensure the veracity of information transmission, repeat sending is necessary. The key is to find out the times of repetition named  $k$  of feedback from vehicle-mounted device to roadside device per unit time. So it is necessary to import a probabilistic model. The most suitable model is Queuing Theory.

Assumption: the total quantity of vehicles within the scope that roadside device can receive message is  $N$ , the production rate of message of per vehicle is  $\nu_c$ , then  $\nu_c = k \times n$ . The processing speed of roadside device is  $\nu_d$ , the event that vehicle-mounted device sends message to roadside device can be regarded as mutual independence. Therefore Queuing Theory can be used to model. Concrete model is as followed.

There is only one roadside device in this system, that is only one service desk. The service intensity is  $\rho$ , the production rate of vehicle-mounted device is  $v_c$ , that is the reach number of customers per unit time in Queuing System. The event that vehicle-mounted device sends message obeys the Poisson process whose parameter is  $v_c$ . The processing speed of roadside device named  $v_d$  is the quantity of customers who are served per unit time. That is the time that roadside device disposes message obeys negative exponential distribution. The average time to dispose message is  $\frac{1}{v_d}$ .

This model can be regarded as  $M / M / 1 / \infty$  Queuing System., that is a waiting queuing system that input process is Poisson stream, serving time obeys negative exponential distribution, only one service desk in system, the capacity of system is unlimited.

Service intensity

$$\rho = \frac{v_c}{v_d} \tag{1}$$

The possibility that quantity of message to be transferred is  $n$  under steady state in system

$$p_n = (1 - \rho)\rho^n \tag{2}$$

Average quantity of data in system

$$L_s = \sum_{n=0}^{\infty} n \cdot p_n = (1 - \rho) \sum_{n=0}^{\infty} n \cdot \rho^n = \frac{\rho}{1 - \rho} = \frac{v_c}{v_d - v_c} \tag{3}$$

Quantity of data waiting to be sent

$$L_q = \sum_{n=1}^{\infty} (n-1) \cdot p_n = (1 - \rho) \sum_{n=1}^{\infty} (n-1) \rho^n = \frac{\rho^2}{1 - \rho} = \frac{v_c^2}{v_d(v_d - v_c)} \tag{4}$$

Average linger time of data in system

$$W_s = \frac{L_s}{v_c} = \frac{1}{v_d - v_c} \tag{5}$$

Average waiting time of sent data in system

$$W_q = \frac{L_q}{v_c} = \frac{v_c}{v_d(v_d - v_c)} \tag{6}$$

In order to ensure the instantaneity & reliability of data communication, it is necessary to give some constraint conditions to average linger time of data in system & average waiting time of sent data in system. Assume the constraint conditions are as followed.

$$0 \leq W_q \leq t_1 \tag{7}$$

$$0 \leq W_s \leq t_2 \tag{8}$$

The solve to formula above is

$$nk \leq v_d - \frac{1}{t_2} \tag{9}$$

$$nk \leq \frac{v_d^2 t_1}{1 + v_d t_1} \tag{10}$$

Assumption: quantity of lane is  $x$ , the distance of scope that roadside device can receive data from vehicle-mounted device is  $L$ , the velocity of vehicles is  $v$ . The total number of vehicles satisfies the following formula  $n \leq j \frac{xL}{v}$ ,  $j$  is modifying factor, related to weather, road congestion and so forth.

Finally the condition that satisfied for  $k$  is

$$k \leq \min \left\{ v \frac{v_d - 1/t_2}{jxL}, \frac{v_d t_1 v}{(1 + v_d t_1) jxL} \right\} \tag{11}$$

Assume  $j = 0.8$ ,  $t_1 = 1s$ ,  $t_2 = 1s$ ,  $v_d = 50pers$ , the repeat time of sending data  $k(pers)$  & velocity  $v(km/h)$  present linear relationship, its expression is  $k = a \bullet v$ ; whereby the value of  $a$  is related to  $L$  &  $x$ , specific values are as illustrated in table 3.

Table 3. The values of a

x \ a	L/(m)							
	500	600	700	800	900	1000	1100	1200
2	0.0613	0.0511	0.0438	0.0383	0.0340	0.0306	0.0279	0.0255
3	0.0409	0.0340	0.0292	0.0255	0.0227	0.0204	0.0186	0.0170
4	0.0306	0.0255	0.0219	0.0191	0.0170	0.0153	0.0139	0.0128
5	0.0245	0.0204	0.0175	0.0153	0.0136	0.0123	0.0111	0.0102
6	0.0204	0.0170	0.0146	0.0128	0.0113	0.0102	0.0093	0.0085

In order to ensure the instantaneity & reliability of data communication, after the constraint to the parameter that describes time, the repeat time of sending data presents positive correlation with velocity of vehicles. This is corresponded with reality. When the velocity is fairly high, the change of distance between vehicles with roadside device is also fast, the instability during the process of sending data becomes serious. The possibility that data can be transferred successfully at one time is fairly low. [5-6] In order to ensure the instantaneity & reliability of data communication, it is necessary to increase the repeat time of sending data.

The verification of the conclusion above can be simulated by MATLAB. Assume that velocity  $v = 60km/h$ ,  $L = 1000m$ ,  $x = 3$ , so  $k = 1pers$ ; then  $n = 40$ ,  $v_c = nk = 40pers$ ,  $v_d = 50pers$ . Simulation time is 60 seconds, the total number of data is 2400, average linger time of data in system & average waiting time of sent data in system is generated according to negative exponential distribution model. Average linger time of data in system & average waiting time of sent data in system is as shown in figure 3.

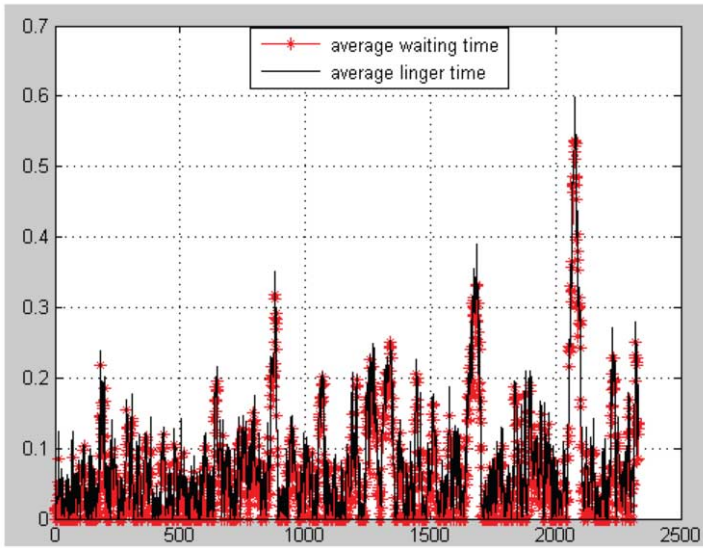


Figure 3. Average linger time & average waiting time in simulation

Out of figure 3, average linger time & average waiting time come within the setting limits. That is to declare this theory is feasible.

3.2. The calculation of sending frequency from roadside device to vehicle-mounted device

The solving of sending frequency from roadside device to vehicle-mounted device can also be done by Queuing Theory. Assume that the total quantity of vehicles within the scope that roadside device can transfer message is  $n$ , the frequency that roadside device transfers data is  $\lambda$ , the frequency that vehicle-mounted receives & disposes message is  $\mu$ , according to the fact that the data transferred by roadside device is often dozens of bits, the message expressed is fairly complex. So the frequency of sending & receiving message shouldn't be too fast. The most suitable frequency is under 2Hz. Queuing Theory can be used to model as followed.

There are  $n$  vehicles waiting to receive data, that is  $n$  service desk; the frequency that roadside device sends message is  $\lambda$ , arrival rate per unit time is  $n\lambda$  in this system. The frequency that vehicle-mounted device receives & disposes message is  $\mu$ , that is the average amount of customers served by service desk per unit time is  $\mu$ .

This queuing model is an  $M/M/n/\infty$  model, which means a multi-service desk waiting queuing system that the arrival of message obeys Poisson distribution, arrival rate per unit time is  $n\lambda$ ; the time that vehicle-mounted device disposes data obeys negative exponential distribution, the average serve time is  $\frac{1}{\mu}$ .

The possibility that there is no message transferred in this system is

$$P_0 = \frac{1}{\sum_{i=0}^{n-1} \frac{(\lambda / \mu)^i}{i!} + \frac{(\lambda / \mu)^n}{n!} \left(\frac{n\mu}{n\mu - \lambda}\right)} \tag{12}$$

The average waiting time of sending message is

$$W_q = \frac{(\lambda / \mu)^n \mu}{(n-1)!(n\mu - \lambda)^2} P_0 \tag{13}$$

It also can be calculated by the method below:

$$W_q = P_{wait} \frac{1}{\mu(n - F)} = P_{wait} \frac{1}{n(\mu - \lambda)} \tag{14}$$

Whereby  $F$  means the arrival load of system,

$$F = \frac{n\lambda}{\mu} \tag{15}$$

$P_{wait}$  means the possibility that customers wait for serve where the arrival load is  $F$  and there are  $n$  servers in service system , also can be regarded as the possibility that data is waiting to be sent in the system.  $P_{wait}$  can be calculated by lingo as  $P_{wait} = @peb(F, n)$  ,whereby  $n$  means the total amount of servers in system.

Solve various amount of vehicles  $n$  corresponded with various frequency  $\lambda$  that roadside device sends message by lingo. Assume that the frequency that vehicle-mounted device receives & disposes message  $\mu = 2$  ,give a constraint to  $W_q$  that  $W_q \leq 0.1s$  ,  $n = 1 \sim 50$  ,  $\lambda(pers)$  is as illustrated in table 4.

**Table 4.** Information coding rules and examples

$n$	1~11	12	13	14	15	16	17	18	19	20~22	>22
$\lambda$	0.1	0.2	0.7	1.0	1.3	1.5	1.6	1.7	1.8	1.9	2.0

In this way, the frequency that roadside device sends message can be confirmed according to the amount of vehicles within the scope that message sent by roadside device can reach.

### 4. Conclusion

This paper have introduced the main feature of Bluetooth and ZigBee. According to their specific advantages, this paper have carried on some exploration and research in their specific application mode at the fields of Intelligent Transportation. ZigBee is a communication medium between roadside and autocars. Bluetooth and smartphones is a medium in one car interaction. This is why this two communication mode can maximize their advantages in each own application area. This paper also aims at how to confirm the frequency of data transmission between roadside device and vehicle-mounted device in ZigBee-Bluetooth Intelligent Transportation System. In order to achieve this goal, this paper have used Queuing Theory to establish models to calculate and simulate, then give a suitable frequency according to the amount and velocity of

autocars. Nevertheless the practical application environment is far more complex than imagine situation. So there should be more research and practice to make ZigBee-Bluetooth Intelligent Transportation System to fact.

## References

- [1] Cai Xiaoyu. Intelligent Smart Home Monitoring System Based On ZigBee and Android, Nanjing: Nanjing Normal University, 2012:6-7.
- [2] Zhao, B. C., & Liu, B. (2008). Research of intelligence traffic control system based on ZigBee technology. *Computer Knowledge & Technology*. 2(10):141-142
- [3] Han Yi, Xu Y J, Hu J. Car Networking Application Research Based On Bluetooth 4.0+HS, *Journal of Highway Communication Technology*, 2013(9):277-279
- [4] Bai Xiaoe. Development and Research on Bluetooth Wireless Networking Technology for Cooperative Vehicle Infrastructure Environments, Harbin: Harbin Institute of Technology:2016:7.
- [5] HAFEEZ K A, ZHAO L, MA B, et al. Performance analysis and enhancement of the DSRC for VANET's safety applications. *IEEE Transactions on Vehicular Technology*, 2013, 62(7): 3069-3083.
- [6] SINGH J P, BAMBOS N, SRINIVASAN B, et al. Wireless LAN performance under varied stress conditions in vehicular traffic scenarios. *Vehicular Technology Conference*, 2002, 2(2): 743-747.

# The Tracker with Online Training Based on the TLD Algorithm

Yongfeng Qi and Peng Zhang

*College Computer Science and Engineering, Northwest Normal University,  
Lanzhou, Gansu, 730070*

**Abstract.** TLD is a real-time long-term tracking system that decomposes the tasks into three components: tracking, learning and detection. The learning estimates detector's errors and updates it to avoid these errors in the future. However, Current implementation of TLD trains only the detector and the tracker stay fixed. As a result, the tracker makes always the same errors. In our paper, we develop a novel training method which combines naive Bayes classifier with the optical flow based on the TLD algorithm to train the tracker. The proposed algorithm mainly consists of two stages: one stage for training and a second stage for tracking. For the training stage, we sample some positive samples near the current target location and negative samples far away from the object center to update the classifier from the current frame. For the tracking stage, Optical flow tracker estimates the object's motion between consecutive frames under the assumption that the frame-to-frame motion is limited and the object is visible. And then we sample a set of image patches, using the classifier to each patch. We determine the target with the maximal classification score. With these definitions, we conduct extensive experiments and comparisons for the proposed method. The comparisons and experiments well demonstrate the effectiveness of our work.

**Keywords:** Tracker, online training, TLD

## 1. Introduction

Object tracking has been extensively studied in computer vision due to its importance in applications such as automated surveillance, video indexing, and traffic monitoring<sup>[1]</sup>, to name a few. To enable the long-term tracking, there are a number of problems which need to be addressed. A successful long-term tracker should handle scale and illumination changes<sup>[2]</sup>, background clutter, partial occlusions<sup>[3]</sup>, and operate in real-time.

Jing Wang et al.<sup>[4]</sup> proposed a particle filter based algorithm for color-guided object tracking to solve problems such as object drifting and lost in complex environment. The proposed algorithm reduces background noise and improves tracking accuracy of objects with changing appearance. However, the proposed algorithm is limited to apply to the scene with slightly changed view angle. In<sup>[5]</sup>, Kaihua Zhang et al present a novel online discriminative feature selection method for object tracking. The tracking task is formulated as a binary classification via a naive Bayes classifier, which is a simple yet effective and efficient tracking system. Zdenek Kalal et al.<sup>[6]</sup> propose a long-term tracking of unknown objects in a video stream. The proposed TLD



tracking algorithm is decomposed into three sub-tasks: tracking<sup>[7]</sup>, learning<sup>[8]</sup> and detection.

However, current implementation of TLD trains only the detector and the tracker stay fixed. As a result, the tracker makes always the same errors.

In this paper, we proposed a novel training method which combines naive bayes classifier with the optical flow based on the TLD algorithm to train the tracker. We divide our training method into two stages: one stage for training and a second stage for tracking. The patches extracted from the frames pass to these two stages and then proceed by iterative bootstrapping. The proposed algorithm can handle the problem that the tracker cannot be trained online in the TLD, which would be able to estimate tracker's errors and updates it to avoid these errors in the future.

The rest of the paper is organized as follows. Section 2 introduces the framework of the tracker with online training and its implementation. Section 3 presents the extensive experiments and comparisons. Section 4 lists the conclusion and future works.

## 2. Proposed Algorithm

In this section, we present our proposed method in details. The problem of the tracking training online in TLD is formulated as a combination of the optical flow with a naive bayes classifier. We divide our training method into two stages: one stage for training and a second stage for tracking.

### 2.1. TLD Tracking Algorithm

Zdenek Kalal et al.<sup>[6]</sup> propose a long-term tracking of unknown objects in a video stream. The proposed TLD tracking algorithm is decomposed into three sub-tasks: tracking, learning and detection. Each sub-task is addressed by a single component and the components operate simultaneously. The tracker follows the object from frame to frame, providing positive example for machine learning. The detector localizes current location and corrects the tracker if necessary. The learning<sup>[8]</sup> estimates detector's errors and updates it to avoid these errors in the future.

The tracking component of TLD is based on optical flow tracker extended with failure detection. The detection is based on the Forward-Backward error, i.e. the tracking is performed forward and backward in time and the discrepancies<sup>[9]</sup> between these two trajectories are measured.

### 2.2. Proposed Algorithm

Our algorithm is shown in figure 1. We assume that the tracking window in the first frame has been determined.

Algorithm 1. Tracker with online training method

Step 1: Training stage

- Sample two sets of image patches,  $D^\alpha = \{z \mid \|I(z) - I_t\| < \alpha\}$  and  $D^{\zeta \cdot \beta} = \{z \mid \zeta < \|I(z) - I_t\| < \beta\}$  with  $\alpha < \zeta < \beta$  where  $I_t$  is the tracking location at the t-th frame and z is each sample.

- Extract the features with these two sets of samples and update the Bayes classifier.

Step 2: Tracking stage

- Input t+1-th video frame.
- Estimate the object's motion by optical flow.
- Sample a set of image patches,  $D^\gamma = \{z \mid \|I(z) - I_{t+1}\| < \gamma\}$ .
- Use Bayes classifier to each patch and find the tracking location  $I_{t+1}$  with the maximal classifier response.

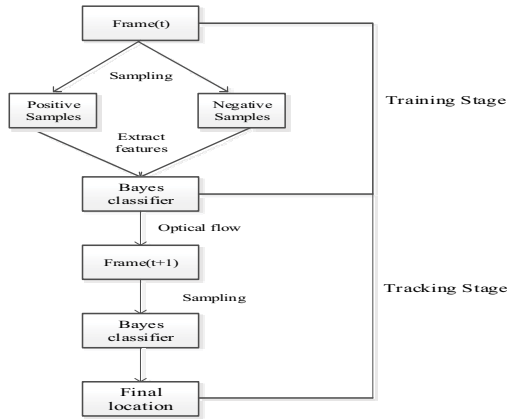


Fig.1. Flow diagram of our proposed algorithm

2.3. Combination of our proposed algorithm and TLD

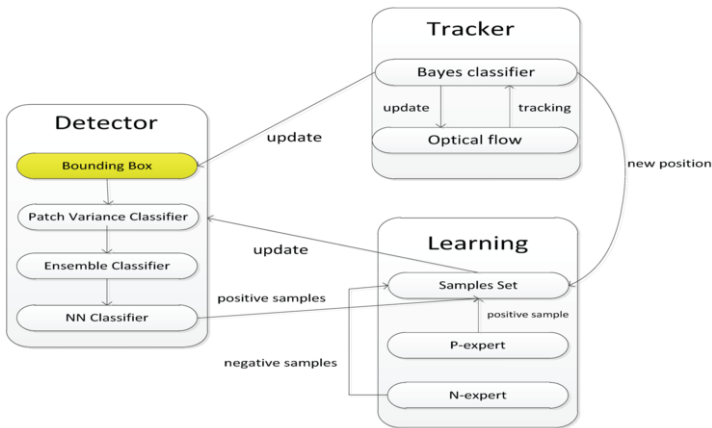


Fig.2. The combination of our proposed algorithm and TLD

As shown by figure 2, tracker estimates the object's motion by the optical flow<sup>[10]</sup> and uses the Bayes classifier<sup>[11]</sup> to determine the target with the maximal classification score. At the end of the tracking task, we sample a set of patches as training data to update the Bayes classifier.

In detector task, the patches scanned from the scanning window are passed through a cascaded classifier which consists of three stages: (1) patch variance, (2) ensemble classifier<sup>[12]</sup>, and (3) nearest neighbor. Each stage either rejects the patch in question or passes it to the next stage.

Learning observes performance of tracker and detector by two types of "experts". P-expert analyzes examples classified as negative, estimates false positive and adds them to training set with positive label. N-expert analyzes examples classified as positive, estimates false positives and adds them with negative label to the training set.

### 3. Experiments

We evaluate our tracking algorithm on 6 challenging sequences where taken it from my own camera, that is: Student, People, Pedestrian, David, Car and Bus. For fair evaluations, we repeat the experiments 10 times on each sequence, and present the averaged results. Table 1 shows the performance of our algorithm. Our tracker is implemented in MATLAB, which runs on a 2.40GHz CPU with 4 GB RAM.

#### 3.1 Experimental Results

We use 6 challenging sequences to evaluate the tracking performance of our algorithm from the aspects of illumination variation, shape deformation, abrupt motion, and occlusion. As is shown in figure 3, the target object in Car sequence undergoes abrupt motion and background clutters. The surrounding background has similar texture. In TLD, the learning can estimate detector's errors and updates it to avoid these errors in the future. However, current implementation of TLD trains only the detector and the tracker stay fixed. As a result, the tracker makes always the same errors. In our paper, we develop a novel training method which combines naive Bayes classifier with the optical flow based on the TLD algorithm to train the tracker. Therefore, the tracker can also analyze its own errors and update it. So even though the texture of the background is similar to the object, our proposed algorithm performs well on this sequence.

**Table 1.** The results of our algorithm



Number	Names	Frames	Not visible frames	Visible frames	Snapshots
1	Student	100	0	100	
2	People	157	0	157	
3	Pedestrian	324	56	268	
4	David	761	0	761	
5	Car	243	30	213	
6	Bus	190	0	190	



Fig. 3 The abrupt motion and background clutters

The blurry images of the pedestrian due to fast motion and be far away from camera make it difficult to track the target object. When the pose of the subject changes rapidly, TLD tracker is easy to lose the target completely for some frames in most of the test sequences. As shown in figure 4, our proposed algorithm performs well on this sequence as we combine naive Bayes classifier with the optical flow based on the TLD algorithm to train the tracker. The tracker determines the target via a naive Bayes classifier with the maximal classification and samples a set of image patches to update the classifier from the current frame, which can not only guarantee the ability of estimating errors, but also improves the stability and robustness of our tracking system. Table 2 shows the dataset of the proposed algorithm.

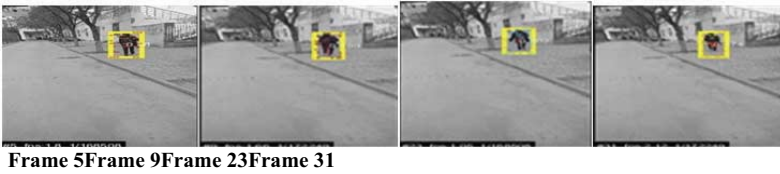


Fig. 4 The blurry images of the pedestrian

Table 2 The dataset of our algorithm

Name	Frames	Mov Camera	Partial occ	Full occ	Pose change	Illum change	Scale change	Similar objects
Student	100	no	yes	no	yes	yes	yes	no
People	157	no	yes	no	yes	yes	yes	yes
Pedestrian	324	no	yes	no	yes	yes	yes	yes
David	761	no	yes	no	yes	yes	no	no
Car	243	no	yes	no	no	yes	no	yes
Bus	190	no	yes	yes	no	yes	no	yes

## 4. Conclusions

In our paper, we develop a novel training method which combines naive Bayes classifier with the optical flow based on the TLD algorithm to train the tracker. We divide our training method into two stages: one stage for training and a second stage for tracking. The proposed algorithm can handle the problem that the tracker cannot be trained online in the TLD, which would be able to estimate tracker's errors and updates it to avoid these errors in the future. Numerous experiments on different challenging

video sequences demonstrated that our tracker achieves favorable performance in terms of accuracy, robustness, and speed.

## References

- [1] Yaochen Li, Yuanqi Su, Yuehu Liu. Fast two-cycle curve evolution with narrow perception of background for object tracking and contour refinement. *Signal Processing: Image Communication*,2016,44:29-43.
- [2] Victor H. Diaz-Ramirez, Kenia Picos, et al. Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. *Optics Communications*,2014,323:32-43.
- [3] Harish Bhaskar, Kartik Dwivedi, Debi Prosad Dogra, et al., Autonomous detection and tracking under illumination changes, occlusions and moving camera, *Signal Processing*,2015,117:343-354.
- [4] Jing Wang, Hong Zhu, Shunyuan Yu, et al., Object tracking using color-feature guided network generalization and tailored feature fusion.2017,238:387-398.
- [5] Kaihua Zhang, Lei Zhang, Ming-Hsuan Yang. Real-time Object Tracking via Online Discriminative Feature Selection, *IEEE Transaction on image processing*, 2013, 22(12):4664-4677.
- [6] Zdenek Kalal, Krystian Mikolajczyk, Jiri Matas. Tracking-Learning-Detection, *IEEE Transactions on Pattern Analysis and Machine in Intelligence*,2012,34(7):1409-1422.
- [7] Zdenek Kalal,Krystian Mikolajczyk, Jiri Matas. Forward-Backward Error: Automatic Detection of Tracking Failures, *International Conference on Pattern Recognition*, 2010:2756-2759.
- [8] Zdenek Kalal, Jiri Matas, Krystian Mikolajczyk. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints, *IEEE*,2010, 238(6):49-56.
- [9] Zhiwu Huang, Ruiping Wang, Shiguang Shan, et al., Face recognition on large-scale video in the wild with hybrid Euclidean-and-Riemannian metric learning, *Pattern Recognition*,2015,48:3113-3124.
- [10] Pengguang Chen, Xingming Zhang, Pong C. Yuen, et al., Combination of spatio-temporal and transform domain for sparse occlusion estimation by optical flow,*Neurocomputing*,2016,214:368-375.
- [11] Sanjana Agarwal, Nirav Jain, Surekha Dholay. Adaptive Testing and Performance Analysis Using Naive Bayes Classifier,2015,45:70-75.
- [12] V. Lepetit, P. Lagger, P. Fua. Randomized trees for real-time keypoint recognition, *Conference on Computer Vision and Pattern Recognition*,2005,2:775-781.

# Research on the Construction of Network Security Attack and Defense Range System in Power Monitoring System

WeiMingxin<sup>a,1</sup>

<sup>a</sup>*Institute of electrical and electronic engineering, North China Electric Power University, China*

**Abstract.** To face the changes of network security situation in industrial control field of power industry and the security needs of important information system and meet the requirements for special network security tools, we carry out the top design work of network attack and defense simulation range in power industry and build a network security test simulation platform which include the power monitoring system, an important information system which similar to the real production environment. In this paper, we list recent developments concerning the construction of network security attack and defense range system in power monitoring system at home and abroad. According to the overall scheme of power monitoring system for security. We found a highly simulated range environment containing all of these parts--sending, changing, matching, usage, scheduling, information, design the network security attack and defense platform architecture and ensure the safe, reliable and stable operation of the power system effectively.

**Keywords.** Network Security, Defense Range, Power Monitor

## 1. Introduction

The definition of traditional national network range[1]: develop a network security test platform which deals with the rapid development of information technology and information weapon equipment, handle the revolutionary changes in the military field, meet the developmental requirements of informational weapons and equipment, and supply the information warfare environment similar to actual warfare.

At this stage, Power monitoring systems and information systems are facing the threat of APT attacks. With traditional safety compliance and consistency risk assessment tools, we are unable to form the effective defense measures against those organized and advanced technical attacks[2].

At present, the power system doesn't have a totally systematical attack and defense environment. Network and information security team is limited to the work of

---

<sup>1</sup>WeiMingxin, Institute of electrical and electronic engineering, North China Electric Power University, Beinong Road, Huilongguan Town, Changping District, Beijing, China; E-mail: weimingxinncepu@163.com

conventional vulnerability detection and single-function point test which restrain the technical capacity and development space by a large margin[3].

Considering the connectivity of the network architecture, we know the security verification of a single link can't reflect the security situation of the entire network system and network security problems of some link may affect other links of the security situation.

To face the changes of network security situation in industrial control field of power industry and the security needs of important information system. To meet the requirements for special network security tools, carry out the top design work of network attack and defense simulation range in power industry, and build a network security test simulation platform which include the power monitoring system, important information system and is similar to the real production environment. To provide basic support for research on network security technology, improvement on attack protection means and verification on full business security technology. And to supply a reliable, controllable and operable experimental environment for evaluation of power monitoring system and safety of important information system.

## **2. Research status**

In 2013, the United States built two test range: the key infrastructure test range of the Idaho National Laboratory and the control system security center of the Sandia National Laboratory. Since 2003, a series of vulnerabilities and risk assessment activities have been carried out, including the assessment of vulnerabilities and pitfalls in the control system, and integrated intrusion detection, protection, and event related capabilities for control system application development[4].

In 2011 the British government built two test ranges: the federated cyber test range and the Breaking point Storm test range. Since the establishment of the range, its main functions and tasks include: simulation of large and complex networks, conduct network testing and evaluation about infrastructure survivability and reliability in a safe and controlled environment.[5]

In 2013, the Canadian National Simulation Laboratory (CASELab) established a corresponding project to carry out the construction of the network test range, its main functions and tasks include: to provide the core research conditions and capabilities of cloud computing, large-scale network security and confidential areas; Providing researchers with systematic analysis and simulation tools, modeling the behavior of real-world, large-scale network systems under fully reproducible experimental conditions.

According to the Internet data analysis, there is no test range for real business simulation test environment in the areas of domestic key infrastructure. The existing test platform mainly for industrial control system single-link application of network structure security verification, there is no industrial control system network test platform which can cover the basic industry. Domestic lack specific detection tool and corresponding standards to device vulnerability, industrial control system, and there is no corresponding means and mechanisms in the industrial control system vulnerability mining, collection, vulnerability risk warning, unified management of loopholes and so on.

### 3. Design

#### 3.1. Design principles

"Unified planning, centralized management". Selecting the existing technology infrastructure and the environmental advantages unit (or relying on the existing national network laboratory simulation environment), distributed to build network security test range of full electricity range, unified planning, centralized management and technical guidance by headquarters.

"combined virtual with reality, simulate deeply". Through the combination of hardware and software, virtual and reality, simulating deeply to business network traffic, services, attacks, etc.[6][7][8][9] To form verification and combat platform of power network security protection technology verification, attack and defense confrontation drill and risk assessment.

"Implement the supervision, combine offensive with defensive". With the help of network test range combat environment, strengthen the power monitoring system and the safe testing of equipment enter the network, achieve the closed-loop management of network and information security technical supervision, to carry out system-level security comprehensive offensive and defensive technology research, enhance the security team offensive and defensive skills level.

"Independent controllable, safe sharing". Strengthen independent controllable of the test range construction process, prevent the loopholes of security and information leakage to third parties, and establish a collaborative sharing mechanism with national security ministries and evaluation agencies.

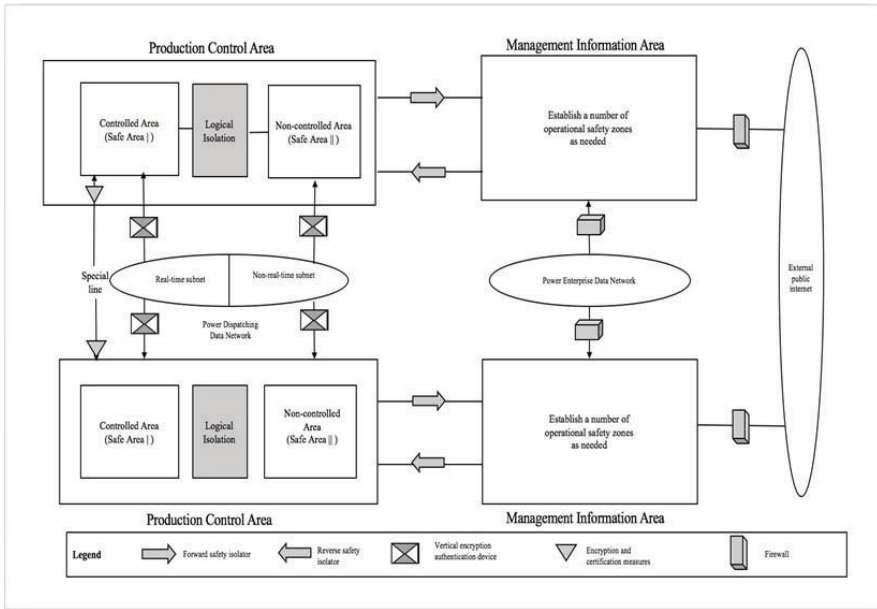
#### 3.2. Safety protection scheme of power monitoring system

According to "Safety protection regulations for power monitoring systems", the framework of the overall safety protection system for power monitoring systems is shown in the figure 1.

##### 3.2.1. Control area (Safety Zone I)

The typical characteristics of the business system or its functional modules (or subsystems) in the control area are as follows. It is an important part of the electricity production and is the focus of security. It can real-time monitoring of the primary power system and vertical use of the power dispatching data network or dedicated channel. The power monitoring system security protection overall framework structure diagram as show in Figure1.





**Figure 1.** Power monitoring system security protection overall framework structure diagram

3.2.2. *Non-controlled area (Safety Zone II)*

The typical characteristics of the business system or its functional modules in the non-control area are as follows. It is the necessary part of electricity production and runs online but does not have the control function and uses the power dispatching data network. It is closely related to the business system which is in the control area or its functional modules.

3.2.3. *Management area of the security zone*

Management information area refers to the collection of business systems which are managed by the electric power enterprise outside of production control area. The traditional typical business systems of management information area include scheduling production management systems, administrative telephone network hanging system, power enterprise data Network and so on. The safe area can be classified by electric power companies according to the specific circumstances, but should no affect the safety of production control area.

3.3. *The simulation range design*

Based on the power protection system security protection program, the network attack and defense simulation range involves an important power monitoring system and important information system, mainly in the "semi-physical simulation" way, to meet

the three major needs of technical research, attack and defense training, and safety verification of all links.[10] Covering the entire process of power system simulation, mainly divided into 7 major areas as power generation, substation, power distribution, electricity, scheduling, information, and supporting technical facilities.

### *3.3.1 Generation side*

Focus on the depth of defense requirements of power grids, and take care of thermal power plants, hydropower plants and new energy power station security protection system features and site protection requirements, we should design DCS and SCADA system architecture simulation environment, which include monitoring host, data communication gateway, control latch, real-time data acquisition device, to control the AGC, AVC power generation, protection, communication, five defense and other business process security control simulation.

### *3.3.2 Substation side*

According to the principle that the intelligent substation applies IEC61850 standard protocol to achieve the principle of interoperability between devices and Typical Structure of Three - tier Two - Substation in Substation, we set up intelligent substation integrated monitoring system simulation environment which has functions of station information collection, equipment monitoring, protection control, traffic analysis and etc. The station control layer equipment includes the monitoring host, data communication gateway machine, data server, integrated application server, operator station, engineer workstation, PMU data concentrator and plan management terminal. The interval layer equipment includes relay protection devices, monitoring and control devices, fault recorder, network recording analyzer and the stability control device. The process layer equipment includes a merging unit, an intelligent terminal, an intelligent component, and the like. They can simulate all kinds of typical network structure, equipment communication, control command and power grid of substation, and provide the whole process data interactive environment for equipment network test and network traffic analysis.

### *3.3.3 Distribution side*

According to the distribution monitoring system operating characteristics and security requirements to build intelligent distribution network system simulation environment. The simulation environment includes power distribution automation master system, front SCADA server, distribution automation terminal equipment (DTU, FTU), column switch, ring network cabinet and other equipment and related system software, to realize real-time data acquisition and monitoring of power distribution system, automatic processing of data information, remote configuration of equipment parameters and status display and operation of sub-gate and it can simulate the typical grid structure, communication mode, fault type, operating condition and network attack and defense scene of the distribution network, and provide environmental support for the implementation of the attack and defense against the distribution system, the verification of the safety risk and the research of the safety technology.

### 3.3.4 Electricity side

According to the data flow collected by the real data collection system, to establish the electricity acquisition system simulation environment. It includes smart meters, concentrators, special change devices, secure access platforms and isolation components. It can realize some crucial functions, such as data acquisition, data transmission, data analysis and display capabilities. Network should support power transmission, wireless, cable and other transmission modes.

### 3.3.5 Scheduling side

According to the schedule I, II and III of partition segregation principle, we build dispatching automatic master-station-system simulation environment. This environment includes the early warning system for real-time monitoring system, schedule-planning system, scheduling management and application system. It also combines the authentication and encryption device, forward and reverse isolation device, data communications gateway machine, time synchronization device, PMU tester, the database server, WEB server, workstations, switches, fortress machine and other hardware devices, with functions of acquiring real-time data of power grid operation, monitoring and controlling. It has the ability to simulate all kinds of typical architecture, network boundary, dispatching command and monitoring operations in the controlling center, providing support for the study of the related penetration test, the isolation system, equipment admittance testing and related safety protection.

### 3.3.6 Information system

The information system puts emphasis on the attack-defense requirements of information management district, in accordance with the "true framework, simulated business" general principles and makes full use of virtualization technology to improve the simulation efficiency. The scope of simulation covers the whole information system of outer network, including websites of outer network towards outer network, 95598 websites, e-commerce platform, power trading, global energy Internet sites, mobile interactive platform and more than 20 other web sites. The simulation also covers the information system of the inner network with three levels, important integration platform system (ERP, collaborative office, financial control, program planning, unified access, etc.) and the centralized purchasing. In addition, it covers widely used operating system, middleware, database and other software, as well as some necessary devices such as routers, switches, firewalls and servers.

### 3.3.7 Supporting facilities

To build a technical support platform for a safe network range, which can meet the demands including attack-defense measures, monitor display, training and competition, the technical support platform should have the following abilities[11]:

The ability of customized configuration for attack and defense scenarios. Technical support platform can simulate nodes like the generic server and the system to combine the simulation nodes with the physical nodes. It also can build a network

environment dynamically according to the practical application scenario, dynamic structures are built based on the firing range simulation environment resources, which is not only applicable to the specific applicate scenario of network environments, but also implement the basic environment support, such as network security defense exercises, training and teaching and products testing for network security.

The ability of supporting security attacking tools. The platform constructs a security attacking toolset consisting of a customizable security attack model, vulnerability scanning and using tools, online customization, analysis and control, providing technical tool support for attack-defense confrontation.

The ability of supervising attack-defense people. Through authority management, security audit and other technical measures, attack-defense and training personnel's security defense tools, devices and systems that they can attack and their behaviors can be restricted by authority management, which is also capable of auditing the operation behavior to prevent illegal operation.

The ability of analyzing attack-defense effectiveness. The platform monitors the host, network equipment, terminal equipment, and the behavior of the safety protection equipment in the simulation environment in real-time. It provides support for defense effect analysis and demonstration. At the same time, it can integrate data both with the platform of network security monitoring for early warning analysis and the vulnerability management platform to verify the effectiveness of security protection strategy, patches installed and so on.

### 3.4 Network Security Attack and Defense Platform Architecture

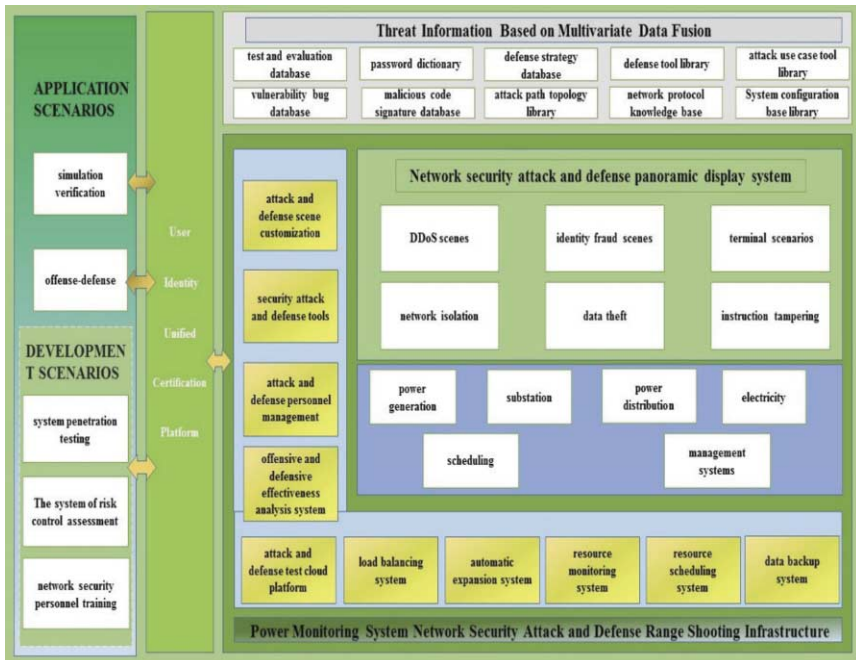


Figure 2. Network security defense platform architecture diagram

As shown in Figure 2, The application scenarios include simulation verification and offense-defense. The development scenarios include system penetration testing, the system of risk control assessment and network security personnel training.

Only through the user identity unified certification, Users could enter the network security attack and defense test system.

The threat information based on multi-data fusion includes test and evaluation database, password dictionary, defense strategy database, defense tool library, attack use case tool library, vulnerability bug database, malicious code signature database, attack path topology library, network protocol knowledge base, System configuration base library and so on. These can be used to attack the system.

Network security attack and defense testing system includes power monitoring system network security attack and defense range shooting infrastructure, business management system, power generation, substation, power distribution, electricity, scheduling, management systems and network security attack and defense panoramic display system.

The business management system includes the functions of attack and defense scene customization, security attack and defense tools, attack and defense personnel management, offensive and defensive effectiveness analysis system, attack and defense test cloud platform, load balancing system, automatic expansion system, resource monitoring system, resource scheduling system and data backup system.

Simulation of network security attack and defense panoramic display system is composed of DDoS scenes, identity fraud scenes, terminal scenarios, network isolation, datum theft, instruction tampering and other components.

#### **4. Conclusion**

Through network attack and defense simulation range environment, we could find security vulnerabilities and threats in power system during the whole process so as to diagnose and carry out health assessments for existing systems, and provides scientific basis and solution for the improvement on application of power system and etc. As a result, we ensure the safe, reliable and stable operation of the power system effectively. Build an attack and defense range demonstration project in the national key infrastructure area and network attack and defense simulation environment covering the power monitoring system and the important information system. Found a highly simulated range environment containing all of these parts--sending, changing, matching, usage, scheduling and information. It was promoted in four aspects and be improved the ability of risk control in power monitoring system, with the level of team building in network, information security and vulnerability detection and the overall level of network security protection.

#### **References**

- [1] Li Q, Hao W, Li C, Xu L, Present Situation and Enlightenment of the Foreign Network Range Technology, *Netinfo Security*, 1671-1122(2014) 09-0063-06.
- [2] Teng Z, Application Research on the Information Security of Power System and Countermeasures to solve security risks, *PUBLIC COMMUNICATION OF SCIENCE & TECHNOLOGY*, 1674-6708(2013)90-0054-02.

- [3] Han Z, Cao Y, Power System Security and Its Prevention, Power System Technology,1000-3673(2004)09-0001-06.
- [4] Ranka J. National Cyber Range[R]. DEFENSE ADVANCED RESEARCH PROJECTS AGENCY ARLINGTON VA STRATEGIC TECHNOLOGY OFFICE (STO), 2011.
- [5] Winter H. System security assessment using a cyber range[C] System Safety, incorporating the Cyber Security Conference 2012, 7th IET International Conference on. IET, 2012: 1-5.
- [6] D Jin, Y Zheng, DM Nicol. A parallel network simulation and virtual time-based network emulation testbed [J]. Journal of Simulation. 2014, 8(8): 206-214.
- [7] Miguel A. Erazo, Jason Liu. Leveraging symbiotic relationship between simulation and emulation for scalable network experimentation [A]. ACM SIGSIM Conference on Principles of Advanced Discrete Simulation[C]. 2013: 79-90.
- [8] ZHANI M F, ELBIAZE H, KAMOUN F. Analysis of prediction performance of training-based models using real network traffic[J]. International Journal of Computer Applications in Technology, 2008, 37(1): 10-19.
- [9] AMIRKHANYAN A, SAPEGIN A, GAWRON M, et al. Simulation user behavior on a security testbed using user behavior states graph[C] Proceedings of the 8th International Conference on Security of Information and Networks. ACM, c2015: 217-223.
- [10] Tang Y, Song X, Liu W, Zhou X, Power System Full Dynamic Simulation, Power System Technology, 1000-3763(2002)09-0007-06.
- [11] Fang B, Jia Y, Li A, Zhang W, Cyber Ranges: state-of-the-art and research challenges, Journal of Cyber Security, 2016, 1(3).

# A Fuel Saving and Safety Driving Guidance System Applied on the Commercial Vehicle Combining GPS and GIS

Liu Hao<sup>a,1</sup> and Hai Yuanquan<sup>b</sup>

<sup>a</sup> *School of Automotive Engineering, Wuhan University of Technology*

<sup>b</sup> *School of materials science and engineering, Jiamusi University*

**Abstract.** EFI engine has the characteristics of DFCO. Sliding on the uphill with gear can save fuel and ensure safe driving. In this paper, a fuel saving and safety driving guidance system applied on the commercial vehicle is proposed. The uphill and downhill road information is predicted based on GPS and GIS. The fuel-efficient and safe speed tips are provided for drivers according to the models with the vehicle's structural parameters and running status parameters. In this paper, the taxiing equations of uphill and downhill are established, the taxiing distance is calculated according to the current vehicle speed and the slope. The vehicle's geographical position is obtained by GPS. The distance between the vehicle and the ramp top is calculated. Energy conservation calculation model and safety evaluation model are established based on energy conservation law. On a road of 8 km, it was found that the braking frequency was reduced by 50%, the energy dissipated by braking was reduced by 2.15% and fuel consumption was reduced by 3.62% after using the system.

**Keywords.** Commercial vehicle, hill, energy saving, safety.

## 1. Introduction

The real-time vehicle position is obtained accurately through the GPS positioning before the uphill[1]. Road information is predicted by calling GIS database[2]. The distance that the vehicle can slide uphill is calculated according to the established model before the vehicle arrives at the slope top. The driver is prompted to relax the throttle when the vehicle reaches the taxiing start point, making the throttle fully open, relying on the vehicle inertia to slide uphill. In general, the driver is used to accelerate uphill[3]. And the driver will step on the brake pedal for safety when the vehicle reaches the top, reducing the initial speed of the downhill[4]. When the vehicle installs the system, the driver will drive in accordance with the prompt speed, changing the part of the acceleration into the sliding, making full use of the commercial vehicle inertia potential energy, which will achieve fuel saving function[5]. At the same time as the

---

<sup>1</sup> Liu Hao, School of automotive engineering, Wuhan University of Technology. No. 122, LuoShi Road, HongShan District, Hubei Province, China.  
E-mail:liu\_hao@whut.edu.cn.

vehicle taxi to reach the slope top, downhill initial speed decreases, the driver does not need to brake before the downhill, so the brake time is relatively reduced, providing security for the subsequent downhill. In the process of downhill, it makes full use of gravity to convert the gravity potential energy into the kinetic energy of the vehicle, so as to realize the fuel saving in the downhill process[6-7].

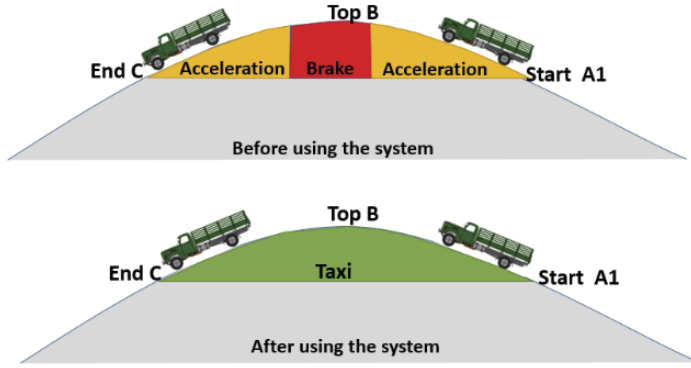


Figure 1. System principle diagram.

## 2. Modeling

### 2.1. Uphill taxiing distance

Through the GIS, the road level is inquired, road speed limit  $v_{limit}$  is determined and the slope  $\alpha_{uphill}$  and length  $L_{uphill}$  of uphill are determined.

$$v_{top} = 0.1v_{limit} \tag{1}$$

$$L_{slide-up} = \int_{v_{start}}^{v_{top}} \frac{Mv_{middle}}{G\sin\alpha_{uphill} + G\cos\alpha_{uphill}f + \frac{C_D A \rho v_{middle}^2}{2} + F_Z(v)} dv \tag{2}$$

$$L_{slide-up} = \text{distance}\{\text{vehicle, ramp top}\} \tag{3}$$

Where:

$v_{start}$ -The speed of the taxiing start point;

$v_{top}$ -The speed of the slope top;

G-Vehicle gravity;

M-The total mass of the vehicle;

f-The rolling resistance coefficient;

$C_D$ -The air resistance coefficient;

A-The windward area of the car;

$\rho$ -The air density;

$v_{middle}$ -The speed of the uphill taxiing process;

$F_Z(v)$ -The engine resistance of the taxiing process;

distance-The distance between vehicle and ramp top;



## 2.2. Downhill taxiing distance

Road speed limit  $v_{\text{limit}}$ , the slope  $\alpha_{\text{downhill}}$  and length  $L_{\text{downhill}}$  are determined through the GIS. Vehicle speed is determined by GPS.

If the vehicle slows down, then

$$L_{v_{\text{top}}-0.3v_{\text{top}}} = \int_{t_{\text{top}}}^{t_{\text{end}}} v_{\text{process}} \times dt \quad (4)$$

Where:

$t_{\text{top}}$ -The moment vehicle reach the top

$t_{\text{end}}$ -The end moment of the downhill slide

$v_{\text{process}}$ - The speed of the downhill taxiing process;

Downhill taxiing distance is as follows:

$$L_{\text{slide-down}} = \min \{L_{\text{downhill}}, L_{v_{\text{top}}-0.3v_{\text{top}}}\} \quad (5)$$

If the vehicle accelerates, then

$$L_{v_{\text{top}}-v_{\text{limit}}} = \int_{t_{\text{top}}}^{t_{\text{end}}} v_{\text{process}} \times dt \quad (6)$$

$$L_{\text{slide-down}} = \min \{L_{\text{downhill}}, L_{v_{\text{top}}-v_{\text{limit}}}\} \quad (7)$$

## 2.3. Uphill energy saving principle

Before using the system: vehicle accelerate to the top with acceleration  $a$ .  $L_{\text{slide-up}}$  is the distance from A1 to B.

The energy consumed by rolling resistance is:

$$W_f = G \times \cos \alpha_{\text{uphill}} \times f \times L_{\text{slide-up}} \quad (8)$$

The energy consumed by air resistance is:

$$W_w = \int_0^{t_{\text{top}}} \frac{C_D A \rho (v_{\text{start}} + at)^2}{2} \times (v_{\text{start}} + at) \times dt \quad (9)$$

The amount of kinetic energy change is:

$$E = \frac{1}{2} M v_{\text{top}}^2 - \frac{1}{2} M v_{\text{start}}^2 \quad (10)$$

The change in the potential energy of the vehicle is:

$$E_p = G \times L_{\text{slide-up}} \times \sin \alpha_{\text{uphill}} \quad (11)$$

The total energy supplied by the engine to the vehicle during this process is

$$Q_1 = W_f + W_w + E + E_p \quad (12)$$

After using the system: the driver relax the throttle completely at the C point. At this point, there is no fuel injection and fuel combustion in the engine cylinder. The engine does not provide energy for the vehicle [8].

$$Q_2 = 0 \quad (13)$$

Compared with the energy consumption before and after using the system during uphill, the engine can save energy:

$$\Delta Q_{\text{uphill}} = Q_1 - Q_2 \quad (14)$$

#### 2.4. Downhill energy saving principle

When the vehicle exceeds the road speed limit, the system prompts the driver to slow down using brakes. Before using the system, the energy provided by the engine from Point B to Point C is  $Q_3$ . Thus, the save energy after using the system is

$$\Delta Q_{\text{downhill}} = Q_3 \quad (15)$$

#### 2.5. Save fuel quality volume

$$V = \frac{\Delta Q_{\text{uphill}} + \Delta Q_{\text{downhill}}}{q\eta\rho} \quad (16)$$

where:

m-The fuel quality of the engine;

q-The calorific value of the fuel;

$\eta$ -The engine fuel efficiency;

$\rho$ -The fuel density;

### 3. Test

#### 3.1. Test vehicle selection

DE8XV5BZ051 is chosen as the target vehicle.



Figure 2. Field test vehicle.

Table 1. The vehicle parameters

Project	Symbol	Parameters	Unit
Vehicle quality	M	1.16	t
The vehicle windward area	A	4.2	m <sup>2</sup>

### 3.2. Test road selection

The test road is located in Suizhou City, Hubei Province, from Qili Tang village to Wan Dian town, total length is 10 km.



Figure 3. Field test road.

## 4. Analysis

Speed, brake, throttle are analyzed to verify that the system is valid. In terms of the whole speed:

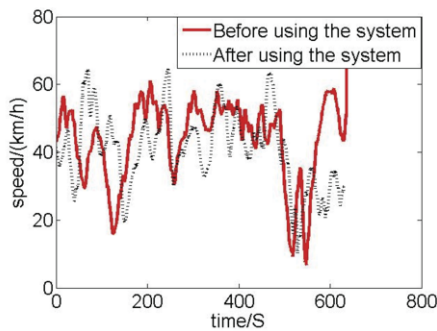


Figure 4. The speed of the entire test section.

The speed trend is basically the same before and after using the system. However, the adjustment time is reduced and the adjustment amplitude is increased after using the system, this is related to the driver's unfamiliarity with the system. The driver just use the system, it is difficult to change the original driving habits in a short time. The driver cannot help controlling the vehicle in accordance with their own driving habits. Driver will change the speed deliberately to close the prompt speed when the system prompts the safety and energy saving speed.

From the brake frequency and brake distribution point of view:

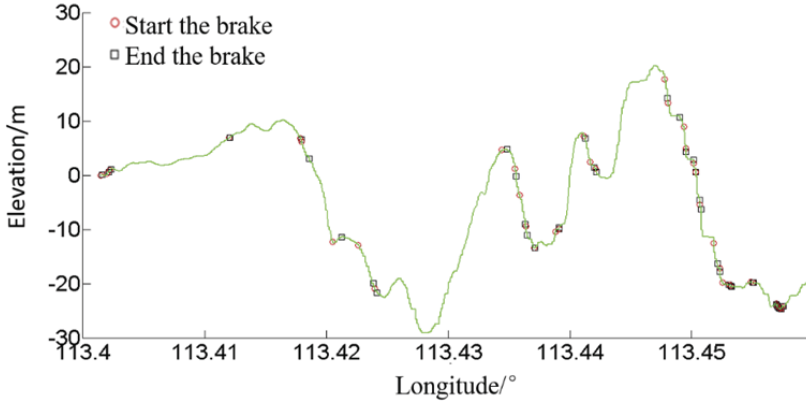


Figure 5. Brake distribution before using the system.

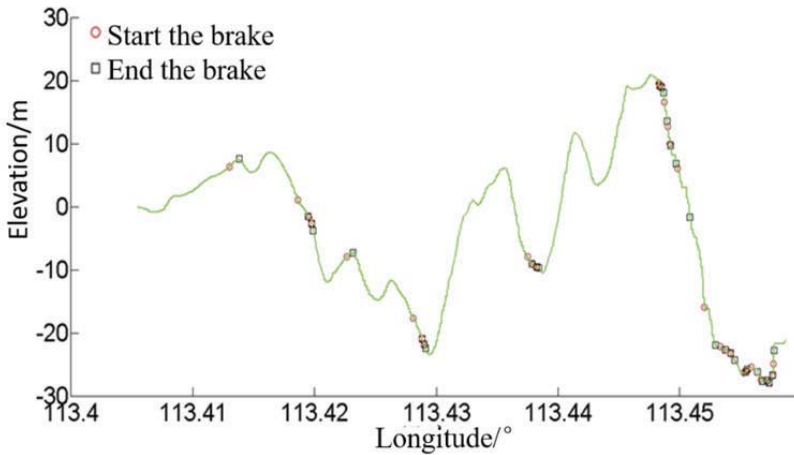


Figure 6. Brake distribution after using the system.

Comparing Figure (5) and Figure (6), it is found that the brake is concentrated in the downhill section. And braking times is reduced in the same downhill section after using the system. From the whole point of braking frequency, the vehicle brake times reduced by twice. The driver brakes 64 times before using the system, but 37 times after using the system. The braking times are reduced by 50%.

The energy dissipation of the brake is studied.

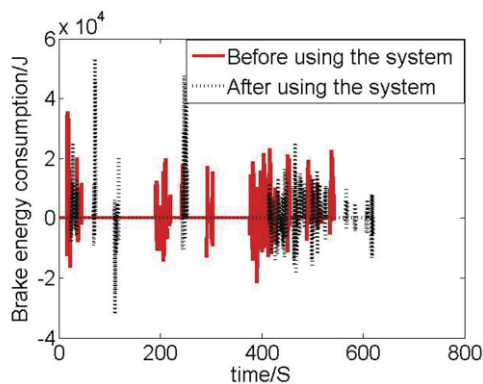


Figure 7. The energy dissipated by the every brake.

It can be seen from the Figure (7), after using the system, brake distribution are more concentrated. Braking is very frequent during 400s-500s, and the vehicle is on the steep slope KL at this time by seeing the corresponding section. After using the system, the start braking time is delayed and braking frequency is reduced, but will appear the emergency brake, which is related to the driver's unfamiliarity with the system. Throughout the experiment, the energy dissipated through the brake is 8567732.487J before using the system, but 8383799.384J after using the system. The energy dissipated by braking is reduced by 183930J, equating to 2.15%.

The throughout consumed energy is studied:

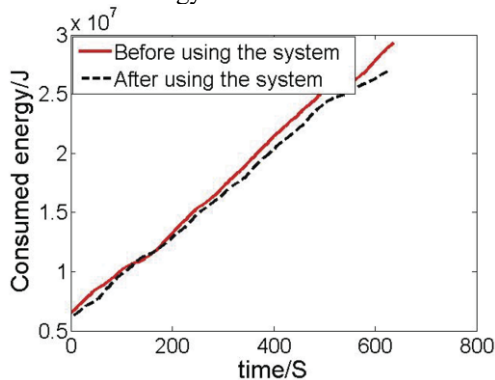


Figure 8. The energy consumption of the whole test section.

Table 2. The energy changes before and after using the system are shown in Table 2

Project	Before	After	Unit
The energy consumed by air resistance	192085.7	155781.41	J
The energy consumed by rolling resistance	22605904.7	20788942.78	J
The energy consumed by brake	8567732.487	8383799.384	J
The energy consumed by engine brake	1185042.42	2043084.13	J
Total energy consumption	32550765	31371607	J
Saved energy		3.62	%

There is a significant difference in the energy provided by engine before and after using the system on the same road. It can be seen from the real vehicle test data that saved energy is 2037200J after using the system, equivalent to 3.62% of the total fuel

consumption before using the system. According to the combustion calorific value of diesel fuel  $3.3 \times 10^7$  J/kg, equivalent to saving 131.2 ml fuel.

**Table 3.** Analysis of energy saving of single ramp before using the system

Uphill mark	Throttle end time(s)	The time vehicle reached the hill top(s)	Taxi duration (s)	Taxi / accelerate to the top hill	Taxi distance (m)	Slope length (m)	Slope (%)
CD	249.76	244.73	0	accelerate	0	20.26	2.63
EF	357.96	356.03	0	accelerate	0	822.5	3.93
FG	373.78	367.84	0	accelerate	0	273.00	2.07
HI	470.23	470.84	0	accelerate	0	845.32	3.80
JK	545.88	548.30	0	accelerate	0	195.38	4.17
LM	816.38	838.24	0	accelerate	0	274.2	5.1

**Table 4.** Analysis of energy saving of single ramp after using the system

Uphill mark	Throttle end time(s)	The time vehicle reached the hill top(s)	Taxi duration (s)	Taxi / accelerate to the top hill	Taxi distance (m)	Slope length (m)	Slope (%)	Fuel saving efficiency (%)
CD	159.67	189.53	6.8	Taxi	20.26	20.26	2.63	100
EF	288.31	290.53	2.21	Taxi	22.83	823.00	3.93	2.77
FG	318.30	311.38	0	accelerate	0	197.12	2.07	0
HI	408.87	410.93	2.06	Taxi	25.87	865.00	3.80	2.99
JK	493.37	494.08	0.71	Taxi	10.51	192.00	4.17	0.55
LM	714.46	729.35	14.8	Taxi	11.5	234.00	5.1	4.91

It can be seen from the table (3) and table (4), after using the system, the system can correctly give the driver tips, making the driver consciously choose to slide up and down under the system guidance which will achieve energy saving in the most uphill. The test section contains six uphill sections. After using the system, the driver drive in accordance with the tip speed, the number of sliding uphill sections increased significantly, indicating that the system can identify the ramp effectively and give the driver a sliding tip when the distance between the vehicle and the ramp top reaches the calculated value according to the model. From the energy-saving effect, for the smaller slope and shorter uphill, the system may prompt the driver to slide up the whole slope, to achieve energy saving 100%. But for most of the long slope, the system will prompt the driver within a short distance from the vehicle to the ramp top. After using the system, the average energy saving effect is 0.55%-5%.

By comparing the data before and after using the system, it is found that the braking times are reduced and the brake energy dissipation is greatly reduced. Energy saving and safety efficiency of the system has been verified.

## 5. Conclusion

1. This paper presents a speed prompt system based on GPS positioning and geographic information prediction, which will make full use of the commercial vehicle inertia to slide downhill and the gravity to slide downhill, realizing the cooperation of energy saving and safety.

2. Fuel saving efficiency of commercial vehicle may reach 100% in a very short and small uphill, but for the common mountain slopes, the average fuel saving efficiency is 0.55%-5% and brake temperature can be decreased by about 3% after using the system.

3. In the future, the system can be developed as a speed planning system for the whole process based on INS and GIS, ensuring that the engine is always in the optimal economic speed conditions and the speed is always below the safe speed.

## References

- [1] Ramadan, H.S., M. Becherif and F. Claude, Energy Management Improvement of Hybrid Electric Vehicles via Combined GPS/Rule-Based Methodology. *IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING*, 2017. 14(2): p. 586-597.
- [2] Inanloo, B. and B. Tansel, Explosion impacts during transport of hazardous cargo: GIS-based characterization of overpressure impacts and delineation of flammable zones for ammonia. *JOURNAL OF ENVIRONMENTAL MANAGEMENT*, 2015. 156: p. 1-9.
- [3] Xu Jin et al. Study on the distribution characteristics of traffic noise in long uphill section of mountain city. *Science and technology and engineering*, 2016.16 (17): 275-282
- [4] Wang Chao, study on the linear design of the downhill section of driver's psychology. *Traffic standardization*, 2011 (3): 84-86.
- [5] Du Muwei et al. Study on the energy saving of engine variable power distribution mechanism. *Energy saving and environmental protection*, 2013 (2): 25-30.
- [6] Scania launched glide (Eco-roll). *The kinetic energy of vehicle technology*, 2013 (6): 6-6.
- [7] Liu Qingbo, Hou Yongping and Yan Conglong. Comparative Analysis of Fuel Economy of Car Coast down in Gear and Neutral. *Journal of Jiamusi University (NATURAL SCIENCE EDITION)*, 2013 (06): 847-849
- [8] Shen Jianjun, Feng Zhongxu, et al. A study on the method of low fuel saving for engine. *Journal of Zhengzhou University (Engineering Science Edition)*, 2012.33 (6): 45-48.

# Study on Passenger Flow Analysis and Prediction Method of the Public Transport Operation Passenger Line of the Adjacent City

Yiming ZHOU<sup>a</sup> and Qingge PANG<sup>b</sup>

<sup>a</sup>*China Academy of Transportation Sciences, Beijing, 100029, China  
30959709@qq.com*

<sup>b</sup>*China Academy of Transportation Sciences, Beijing, 100029, China  
273549893@qq.com*

**Abstract.** With the rapid development of regional integration, the adjacent city mobility is becoming increasingly active, adjacent city masses have new requirements to set the public transport operation passenger line of the adjacent city. To grasp the passenger transport demand of the passenger transportation route in the adjacent city is an important prerequisite for the study of the adjacent city passenger transportation. Based on the analysis of passenger flow characteristics and influencing factors, the paper analyses the models and methods for forecasting passenger demand, according to the characteristics of passenger flow, a method to predict the trend of passenger flow and passenger flow in the public transport operation passenger line of the adjacent city is determined. It has important reference value for the construction of the adjacent city transport network, transport capacity and transportation policy.

**Keywords.** Adjacent city, public transportation, Passenger Flow Forecast Model

## 1. Introduction

### 1.1. Analysis of Passenger Flow Characteristics in Adjacent Cities

The passenger flow is mainly based on the short distance passenger traffic between cities, passenger travel distance is usually less than 50 km. Although it is routes of road passenger transport, in time and space with more significant characteristics of public transport [1–4].

#### *The Distribution in Time*

The main purpose of passenger travel to work, school, shopping, visiting relatives and friends, in case of holidays, there will be peak passenger flow. On weekdays, passenger traffic mainly commuting, Friday afternoon and Sunday afternoon traffic will increase the number of students going home and returning to school. Weekend is mainly near the short distance passenger routes near the residents of leisure and visiting rela-



tives and friends, etc. There will be a noticeable increase in passenger traffic on Friday afternoon and Sunday afternoon.

### *The Distribution in Space*

In the direction of passenger flow distribution, lines are divided into uplink, downlink two directions. According to the difference of two-way passenger flow, the spatial distribution of passenger flow can be divided into two types: two-way and one-way. Two-way line refers to the line, the two directions of the passenger flow is not significant difference, the one-line has a large difference between the upper and lower passenger flow.

### *1.2. Analysis of Influencing Factors of Passenger Flow Forecast*

#### *Macro Factors*

- Social and economic development

The production of passenger traffic mainly comes from production and consumption. The development of economy not only makes the production activities close, also promotes people's consumption of travel, such as shopping malls, the passenger traffic volume shows a rising trend along with the development of economy.

- Population size and structure

The change of population and the change of population structure will cause the change of passenger demand, usually, other conditions remain unchanged, an increase in the number of population will inevitably lead to an increase in passenger traffic, passenger traffic will be reduced accordingly. In addition, changes in population structure, for example, an increase in the number of Non-agricultural population may lead to an increase in passenger traffic.

- Average incomes and consumption level

Per capita income increase, consumer demand is also increase, consumption structure for the basic necessities of life has also changed. Travel, visiting relatives and other aspects of the activities of people's lives has gradually become very common activities, and the corresponding increase in the number of traffic demand.

#### *Micro Factors*

- Ticket Price

In the case of other conditions unchanged, the fare is reduced, will cause an increase in the frequency of passenger travel, the corresponding increase in fares, passengers will reduce the frequency of riding.

The public transport operation passenger line with public welfare and commercial, has the government on public enterprises some subsidies, not lower than before the renovation of the passenger lines fare greatly, ticket price, people travel costs, travel frequency will increase.

- Ride comfort

In the same riding conditions, better ride comfort of passengers more attractive, the passenger seating frequency will increase.

The vehicle condition: before the transformation, the main purpose of the line for profit, the vehicle update slow, average service life is long, the ride comfort is poor; after the transformation, the bus is the purchase of new, ride comfort is better.

Seat condition: before the transformation, each passenger has a seat; after the transformation, allowing standing.

- Travel time cost

In the same riding conditions, passengers travel to the destination to spend the time is shorter, the suction of passengers, the greater the macroeconomic factors, the corresponding frequency of passenger riding will increase. The time spent by passengers mainly includes waiting time and time.

### 1.3. Analysis of The Composition of Passenger Flow

The change of passenger flow is influenced by macro factors and micro factors, we are divided into trend passenger flow and induced traffic. The trend of passenger flow is determined by the factors of society, economy, population and local consumption. This part of the passenger flow according to time, showing a gradual change, so relating to the development of the region. Induced passenger flow is due to the improvement of line service characteristics, that is, changes in the micro aspects of factors, such as lower fares, which makes the increase in passenger transport demand. This part of the passenger flow relating to the service characteristics of factors improvement.

After the transformation, the change of passenger volume is affected not only by the macro factors such as economy and population, but also by the micro factors, such as the reduction of fares and the quality of service. Therefore, the analysis and forecast of the passenger flow, need to increase the passenger flow is divided into two parts of the trend of passenger flow and induced traffic, and then analysis and forecast.

## 2. Methods and Discuss

### Neural Network Method

Neural network has the advantage of high accuracy, but it may not be calculated, appears local minimum.

The steps of using neural network to forecast passenger flow: first of all, it is necessary to initialize the data of passenger flow in order to adapt to the requirement of neural network algorithm; then, the processed data is used as input neural network, that is, the neural network is used to process the data until it is close to the trend of a non-linear curve; finally, the passenger flow forecast is carried out by the correspondence between them.

$$x(k) = f(wx_c(k) + w^2(u(k-1))) \quad (1)$$

$$x_0(k) = x(k-1) \quad (2)$$

$$y(k) = g(w^3x(k)) \quad (3)$$

*Time Prediction Model*

Bus passenger volume is changing with time, so we can forecast the passenger traffic volume by time prediction model. Based on historical data distribution in time, and then use mathematical theory of certain values of related variables predicted. Here are two models: autoregressive moving average model (ARMA) and autoregressive model (AR).

- ARMA model

Autoregressive moving average (ARMA) is a model based on autoregressive average (AR) and moving average technique (MA),

$$Y_t(A(Z)) = B(Z)A_t \tag{4}$$

Where:

$$A(Z) = (1 + \phi_1 Z^{-1} + \phi_2 Z^{-2} + \dots + \phi_p Z^{-p}) \tag{5}$$

$$B(Z) = (1 + \theta_1 Z^{-1} + \theta_2 Z^{-2} + \dots + \theta_q Z^{-q}) \tag{6}$$

- AR model

Autoregressive model (AR) is used to plan and smooth time series. The  $a_t$  in the formula is a linear combination of the observed values of the first P moments and the random perturbations of the present time.

$$Y_t(A(Z)) = a_t \tag{7}$$

$Y_t(A(Z))$  represents the weighted average of the previous moment until the first P observations, where:

$$A(Z) = (1 + \phi_1 Z^{-1} + \phi_2 Z^{-2} + \dots + \phi_p Z^{-p}) \tag{8}$$

Z-P is a delay operator of the first P moments, finding the minimum variance calibration coefficient  $\phi$ S of  $a_t$ .

*Grey Prediction Method*

The grey forecasting method of passenger flow will be regarded as a grey system, then forecast the traffic flow. We divide the system into three categories: White system, grey system and black system. The black system refers to the system needs of the information is unknown, the white system and black system is just the opposite, and the information is known, and the grey system between the black and white system, some known information of some unknown information.

- GM(1,1) model establishment

There are many factors affecting the public transportation system, some of which are known and some are unknown, according to the definition of grey system, the public transportation system can be regarded as a grey system. The establishment of

GM(1,1) grey prediction model, the essence of the grey prediction model is the process that the grey system is gradually whitened to the white system.

$$x^{(0)}(i) + az^{(1)}(i) = u \quad (9)$$

In the formula,  $z^{(1)}$  represents a new sequence obtained by processing the original data.

$x^{(0)}$  is composed of N data:

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\} \quad (10)$$

In the formula,  $x^{(0)}$  is the original time series data,  $a$  is the development of parameters, reflect the trend of development of  $z^{(1)}$  and  $x^{(0)}$ ,  $u$  is the coordination coefficient, reflect the change of the relationship between data,  $x^{(1)}$  can be obtained by  $x^{(0)}$  calculating, the Nth element of  $x^{(1)}$  is:

$$x^{(1)}(i) = \sum_{m=1}^i x^{(0)}(m) = x^{(1)}(i-1) + x^{(0)}(i) \quad (11)$$

The Nth element of the new series  $z^{(1)}$  is:

$$z^{(1)}(i) = 0.5x^{(1)}(i) + 0.5x^{(1)}(i-1) \quad (12)$$

The whitening equation of  $x^{(1)}$  prediction model is established:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (13)$$

Characteristics of grey prediction method:

Less required raw data. Due to all the random variables as grey data, the stochastic process is regarded as a grey process, it does not need the massive historical data, according to some recent data are generated, it can be out of order data to sort out some of the internal rules, and predict.

The method is simple. The method of grey modeling needs a certain mathematical basis, at present widely used computer, some advanced languages have a better way to deal with matrix calculation, simplify the data processing time greatly. The grey system modeling is to analyze the internal law of things with the limited external features, namely the white system. Especially in the use of the grey modeling method of generating function, which was out of order, large fluctuations in the data, described as a smooth curve has certain regularity, and use the grey differential equation can be established, can better reflect the dynamic development of the law of things.

Strong practicability. Gray prediction model can be applied to a variety of situations. Generally speaking, the grey model is closer to the sample data, the predicted value of the model is closer to the actual change curve. When the system changes periodically, it is suggested that the information in the same period should be used as the basis of modeling. In the same period of prediction, a general method for handling available, if the prediction is carried out in different periods, it is necessary to modify

the computer prediction value with the influence of the periodic mutation point on the system, accuracy is still relatively high.

*Gravity Model*

Gravity model is a model which applies Newton’s law of gravity to the analysis and prediction of passenger flow. By the gravity model, set  $Q_{ij}$  for  $i,j$  traffic between the two places, then

$$Q_{ij} = k \frac{(E_i E_j)^\alpha}{R_{ij}^\beta} \tag{14}$$

$k, \alpha, \beta$  are the parameters in the gravity model;  $E_i, E_j$  is the factors of the passenger flow between I, J, called potential energy factor, it can be used the number of population or economic development between the two places;  $R_{ij}$  is transport resistance between the two places, it reflects the cost of passengers from I to j, including the time required, fares and fatigue, etc.

Transportation resistance ( $R_{ij}$ ) is expressed by the following formula:

$$R_{ij} = C_{ij} + T_{ij}W + S_{ij} \tag{15}$$

Where:

- $C_{ij}$  – Transportation costs from i to j (yuan);
- $T_{ij}$  – The time it takes to get from i to j (hour);
- $S_{ij}$  – Fatigue loss from I to j needed (yuan);
- $W$  – Time value (yuan/hour).

Set the current transportation consumption resistance is  $RO_{ij}$ , the future transportation consumption resistance is  $RD_{ij}$ . So:

$$\begin{aligned} \Delta Q_{ij} &= k \frac{(E_i E_j)^\alpha}{R_{Oij}^\beta} - k \frac{(E_i E_j)^\alpha}{R_{Dij}^\beta} \\ &= k \frac{(E_i E_j)^\alpha}{R_{Oij}^\beta} \left[ \left( \frac{R_{Oij}}{R_{Dij}} \right)^\beta - 1 \right] \\ &= Q_{ij} \left[ \left( \frac{R_{Oij}}{R_{Dij}} \right)^\beta - 1 \right] \end{aligned} \tag{16}$$

$Q_{ij}$  is induced passenger flow,  $Q_{ij}$  is the passenger flow of existing passenger line.  $r_A$  is the inducing rate,

$$r_A = \left( \frac{R_{Oij}}{R_{Dij}} \right)^\beta - 1 \tag{17}$$

So,  $\Delta Q_{ij} = Q_{ij} \times r_A$  (18)

### 3. Conclusions

#### *Comparative Analysis of Various Forecasting Methods*

The time series model is mainly used the predictive variables show some regularity with the change of time, this rule is used to predict the amount to be predicted, because it is based on the rule of historical data, it is not suitable for the stochastic system time series model.

The neural network model can be used to analyze the relationship between the influencing factors and the traffic volume through the application of a large number of historical data, then, use this relationship to build a model to predict the traffic. However, the method requires a large amount of data, the general survey of the cost of passenger flow will be too high.

When the historical data are few and the surface regularity between the data is not obvious, the grey model can be used to predict the passenger flow. Need to deal with the data and then get a certain rule of processing data and then to predict the system. Because the grey forecasting model is not accurate for the long term prediction, it can't be used for long-term prediction.

Gravity models are usually used for passenger service characteristics with a greater improvement in line passenger flow forecast, for example, the new line, in terms of fares, security, convenience, etc., the gravity model can reflect the change of these factors. Therefore, the model can be used to predict the passenger flow with these factors.

#### *Determination of Passenger Flow Forecasting Method in Adjacent Cities*

- Determination of trend passenger flow forecasting method

Generally, when the passenger traffic volume of road passenger transport line is less, highway passenger volume of the city will be used to calculate the passenger volume trend, the development trend of highway passenger traffic in this city is obtained, With the growth rate of highway passenger traffic to replace the growth rate of passenger traffic trend of a certain road passenger line, thus, the forecast value of the trend passenger flow is obtained. Therefore, GM(1,1) model can be used to forecast the passenger volume.

- Determination of induced passenger flow forecasting method

Induced passenger flow is caused by the improvement of the service characteristics of the line after the transformation of public transport, therefore, the method of inducing the choice of passenger flow should be able to show the influence of service characteristics on the induced traffic. The gravity model can reflect the influence of the change of the fare and the vehicle condition on the passenger flow after the transformation of the bus, so, the gravity model is used to forecast the passenger flow.

### References

- [1] Richard C.M. Yam, Richard C. Whitfield, Raymond W.F. Chung. Forecasting Traffic Generation in Public Housing Estates. *Journal of Transportation Engineering*, 2014, 7:358–361.

- [2] P. Voshva. Application of cross-nested LOGIT model to mode choice in Tel Aviv, Israel, metropolitan area. *Transportation Research Record*, 2012, 1607:7–15.
- [3] Iwao Okutani. Dynamic prediction of traffic volume through Kalman filtering theory, *Transportation Research Part B: Methodological*, 2010, 18(1):1–11.
- [4] C.O. Tong, S.C. Wong. A predictive dynamic traffic assignment model in congested capacity-constrained road networks. *Transportation Research Part B Methodological*, 2000, 34(8):625–644.

# Research on Frequencies Change Rules of Single-Damaged Simply Supported Beam

*HU Zhaoguang*

Engineering Management Department, China Road and Bridge Corporation, C88  
Andingmenwai Dajie, P. R. China

E-mail: 42048411@qq.com.

**Abstract.** Based on Euler-Bernoulli beam and structure dynamics theory and by means of numerical experimentation, taking a single-damaged simply supported beam as the research object, this paper obtains the frequencies of the beam under the conditions of different damage positions and damage degrees. With the concept of damage factor, rate of change of frequency(RCF), mean value of RCF(MRCF), amplitude of RCF(ARCF), and the ratio of RCF(RRCF), the paper researches frequency characteristics of a simply supported beam. The results show that the frequency and RCF sensitivities are related to damage degree and damage position, and that the RRCF sensitivities are relate to damage position only. The influence of damage degree on the frequency is greater than that of damage position, and that the RCF value variation of the beam is not a monotonic function with the increase of modal order when the damage position and damage degree are constant. The frequency and RCF of the beam at different damage position are oscillation curves, and the oscillation times and amplitude value of RCF curve varies with the change of modal order.

**Keywords.** simply supported beam, single-damaged, RCF, MRCF, ARCF, RRCF

## 1. Introduction

The damage identification method based on dynamic testing is the most promising damage identification technology. It is a kind of structural whole damage detection method, which can comprehensively reflect the structure stiffness, damping and inertia characteristics, and is with small workload on-site, short test time and relatively low economic cost, and more importantly, can be achieved real-time monitoring, so many researchers at home and abroad Many researchers pay attention to this method.

The natural frequency is one of the most readily available indicators in vibration testing, and the test accuracy is high. In the mid-1970s, Cawley and Adams began to study the relationship between natural frequency and damage; by the mid-1980s, a method of determining the damage location and estimating the damage degree was proposed, and the result of damage identification was obtained by comparing the natural frequency of the measured structure with the theoretical natural frequency. The damage identification method based on index change of natural frequency are the frequency difference and the natural frequency change ratio<sup>[1]</sup>. When some structural shapes are more complex and some locations are not suitable for placement of test points, the method of damage based on modal identification is difficult to use, and the



frequency method can play a more significant role. In the process of engineering testing, generally the structure of multi-order natural frequency can be obtained from only one or two test points<sup>[2]</sup>.

Simply supported beam is the most basic and most commonly used form of bridge structure. Based on the deduction of beam rate of change of frequency(RCF) and the ratio of rate of change of frequency(RRCF) for a simply supported beam, this paper focuses on the rules of the beam frequency, the RCF and RRCF caused by the single-damaged of the structure under different damage locations and different damage degrees.

## 2. Theoretical analysis

### 2.1. Relevant definition

- Damage factor  $\alpha$

Reinforced concrete structure has both concrete and steel, whose main damage forms are concrete cracks (width, depth, spacing, etc.), concrete damage, steel corrosion, steel yield and so on. Due to the material complexity of the reinforced concrete structure, this paper will simulate the actual damage state based on the method of reducing the cross-section stiffness, that is, the simulation of all damage models is carried out by reducing the elastic modulus E of the reinforced concrete at the damaged element.

In the structural damage identification, it is considered that the mass matrix is constant, the damage is only caused by the change of stiffness matrix, and the damage degree is achieved by the reduction of stiffness.  $\alpha$  is for the damage factor of the damage element,  $N$  is for the element number, and  $\alpha_N$  is for the damage factor of the Nth element, which represents the damage degree of the Nth element. If the 2nd element damage is 20%, then  $\alpha_2 = 0.8$ .

- Rate of change of frequency (RFC)

$$RCF_i = \frac{f_{ui} - f_{di}}{f_{ui}} \tag{1}$$

Where,  $f_{ui}$  is for  $i$ -order frequency of the beam without damage,  $f_{di}$  is for  $i$ -order frequency of the damaged beam.

- Amplitude of RCF (ARCF)

$$ARCF_i = \max RCF_i - \min RCF_i \tag{2}$$

Where,  $\max RCF_i$  is for the maximum value of the  $i$ th order frequency change rate of damaged beam under a certain damage degree  $\alpha$ ;  $\min RCF_i$  is for the minimum value of the  $i$ th order frequency change rate of damaged beam under a certain damage degree  $\alpha$ .

- Mean value of RCF(MRCF)

$$\overline{RCF_i} = \frac{1}{m} \sum_m RCF_i \tag{3}$$

Where,  $m$  is for the sum of all the damaged elements in the  $i$ -order mode.

- The ratio of RCF(RRCF)

$$RCF_{j,i} = \frac{RCF_j}{RCF_i} \tag{4}$$

### 2.2. Theoretical analysis

According to the basic motion equation of the structure,

$$(\mathbf{K} - \omega^2 \mathbf{M})\boldsymbol{\varphi} = 0 \tag{5}$$

Where,  $\mathbf{K}$  is for the stiffness matrix,  $\mathbf{M}$  is for the mass matrix,  $\omega$  is for the frequency,  $\boldsymbol{\varphi}$  is for the vibration shape, and the following is the same. The vibration form is:

$$[(\mathbf{K} + \Delta\mathbf{K}) - (\omega^2 - \Delta\omega^2)(\mathbf{M} + \Delta\mathbf{M})](\boldsymbol{\varphi} + \Delta\boldsymbol{\varphi}) = 0 \tag{6}$$

$\Delta\mathbf{K}$ 、 $\Delta\mathbf{M}$ 、 $\Delta\omega$  is for the change amount of the stiffness matrix, the mass matrix and the frequency respectively.

The change of the mass is almost negligible, and the structural damage is only considered a change in stiffness, so ( $\Delta\mathbf{M}=0$ ), expand and ignore the second-order items, and simplified:

$$\Delta\omega^2 = (\boldsymbol{\varphi}^T \Delta\mathbf{K} \boldsymbol{\varphi}) / (\boldsymbol{\varphi}^T \mathbf{M} \boldsymbol{\varphi}) \tag{7}$$

The whole stiffness matrix of the structure can be decomposed into the element stiffness matrix, the deformation of the element can be obtained by the vibration mode shapes, which is

$$\boldsymbol{\varepsilon}_m(\boldsymbol{\varphi}) = f(\boldsymbol{\varphi}) \tag{8}$$

For the i-order modal, there are

$$\boldsymbol{\varphi}_i^T \Delta\mathbf{K} \boldsymbol{\varphi}_i = \sum_{m=1}^M \boldsymbol{\varepsilon}_m^T(\boldsymbol{\varphi}_i) \Delta\mathbf{k}_m \boldsymbol{\varepsilon}_m(\boldsymbol{\varphi}_i) \tag{9}$$

Where M is the total number of elements.

Substituting (7) into the formula (9),

$$\Delta\omega_i^2 = \left[ \sum_{j=1}^J \boldsymbol{\varepsilon}_j^T(\boldsymbol{\varphi}_i) \Delta\mathbf{k}_j \boldsymbol{\varepsilon}_j(\boldsymbol{\varphi}_i) \right] / (\boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i) \tag{10}$$

Where, J is the total number of damage elements. For a single damaged element N, equation (10) can be simplified into

$$\Delta\omega_i^2 = [\boldsymbol{\varepsilon}_N^T(\boldsymbol{\varphi}_i) \Delta\mathbf{k}_N \boldsymbol{\varepsilon}_N(\boldsymbol{\varphi}_i)] / (\boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i) \tag{11}$$

Suppose  $\Delta\mathbf{k}_N = \alpha_N \cdot \mathbf{k}_N$ , so<sup>[3]</sup>:

$$\Delta\omega_i^2 = [\alpha_N \boldsymbol{\varepsilon}_N^T(\boldsymbol{\varphi}_i) \mathbf{k}_N \boldsymbol{\varepsilon}_N(\boldsymbol{\varphi}_i)] / (\boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i) \tag{12}$$

Simultaneous equation (5) and (12), so

$$RCF_i = \frac{\Delta\omega_i}{\omega_i} = \sqrt{\frac{\boldsymbol{\varphi}_i^T [\Delta\mathbf{K}] \boldsymbol{\varphi}_i}{\boldsymbol{\varphi}_i^T [\mathbf{K}] \boldsymbol{\varphi}_i}} = \sqrt{\frac{\alpha_N \boldsymbol{\varepsilon}_N^T(\boldsymbol{\varphi}_i) \mathbf{k}_N \boldsymbol{\varepsilon}_N(\boldsymbol{\varphi}_i)}{\boldsymbol{\varphi}_i^T [\mathbf{K}] \boldsymbol{\varphi}_i}} \tag{13}$$

The equation shows that RCF is related to the element damage degree  $\alpha_N$  and the position (element  $N$ ).

When using  $i, j$  two order frequency, the RRCF can be obtained

$$RCF_{j,i} = \frac{\Delta\omega_j/\omega_j}{\Delta\omega_i/\omega_i} = \left( \frac{\frac{\boldsymbol{\varepsilon}_N^T(\boldsymbol{\varphi}_j)\mathbf{k}_N\boldsymbol{\varepsilon}_N(\boldsymbol{\varphi}_j)}{\boldsymbol{\varphi}_j^T[\mathbf{K}]\boldsymbol{\varphi}_j}}{\frac{\boldsymbol{\varepsilon}_N^T(\boldsymbol{\varphi}_i)\mathbf{k}_N\boldsymbol{\varepsilon}_N(\boldsymbol{\varphi}_i)}{\boldsymbol{\varphi}_i^T[\mathbf{K}]\boldsymbol{\varphi}_i}} \right)^{\frac{1}{2}} \tag{14}$$

Obviously, RRCF is only related to the damage position, and the damage degree(damage factor  $\alpha$ ) is not included.

### 3. Numerical experimentation

There are a simply supported beam, whose the calculation span  $l$  is 10.5 meters, section height is 0.6 meter, section width is 0.4 meter, using No.30 concrete, with elastic modulus  $E$  is  $3.0303 \times 10^4$ MPa(see Fig. 1).The finite element model(FEM) unit length is 0.5 meter, with a total of 21 units.

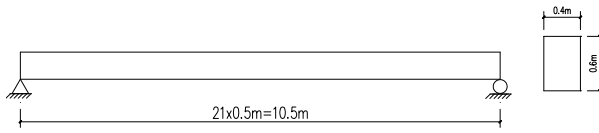


Figure 1. Graphic of the simple supported beam

### 4. Frequency analysis of damaged beam

The numerical simulation of a simply supported beam in this paper is based on the reduction of the elastic modulus of the element to represent the element damage, and the beam is assumed to be a classical beam (Euler-Bernoulli beam). In order to fully reflect the frequency sensitivity to different position and degree of damage, there are 2 to 20 elements (the beam around the bearings should be analyzed according to the Timoshenko beam, this example does not consider the damage near the bearing) with 4 damage degrees, which is  $\alpha=0.8$ ,  $\alpha=0.6$ ,  $\alpha=0.4$ ,  $\alpha=0.2$ , so the beam have 76 kinds of conditions, all of which are single damage.

The damage results in this paper are obtained in an ideal state, that is, the effects of modal incompetence and noise are not considered.

#### 4.1. Frequency variation rule

The single-damage frequency diagram (1 to 4 order frequency only) of the simply supported beam with different damage degree  $\alpha$ , a total of 76 conditions, are obtained by using the beam element number as the abscissa and the single-damage frequency of 2 to 20 elements as the ordinate, see Fig. 2 (a) - (d).

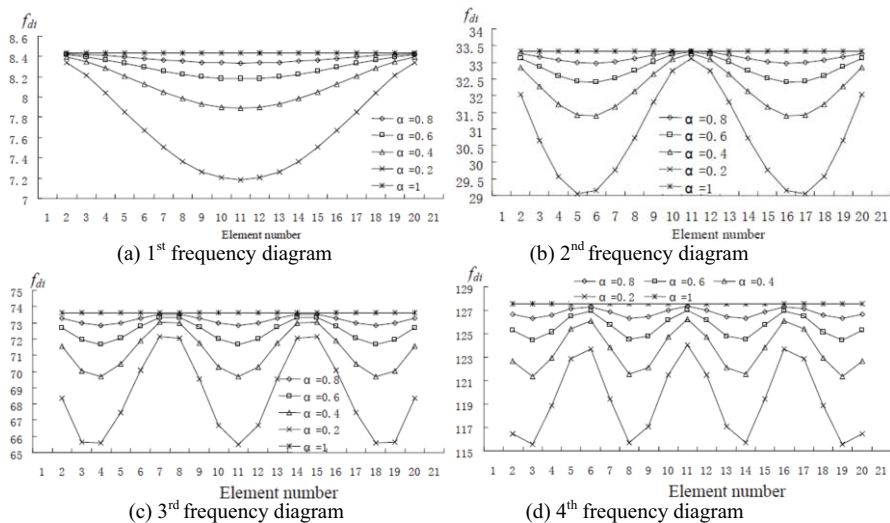


Figure 2. Frequency diagram

It can be seen from Fig.2 (a) - (d) that the frequency of the different damage location  $f_{di}$  of a simply supported beam with single damage is a oscillation curve, the number of oscillation s increases rapidly with the increase of the modal order, and the amplitude of the oscillation varies with the change of the modal order. The sensitivity of beam damage to frequency  $f_{di}$  is related to the beam damage degree, and the greater the damage degree, the lower the frequency  $f_{di}$  of each mode; the sensitivity of beam damage to frequency  $f_{di}$  is related to the beam damage location  $N$  and the modal order  $i$ , and for example, the sensitivity of the beam damage to the first order modal frequency is the largest in the mid-span, and the smaller the sensitivity of the first order modal frequency the closer to the fulcrum. The effect of the damage degree on the natural frequency  $f_{di}$  is greater than that of the damage position  $N$ .

#### 4.2. RCF rule

The RCF diagrams for the simply supported beam with 1 to 10 mode are obtained by taking element number as abscissa, and 2 to 20 element damage RCF as the ordinate, in the case of damage factor  $\alpha=0.8$ (19 conditions), see Fig. 3 (a)- (d).

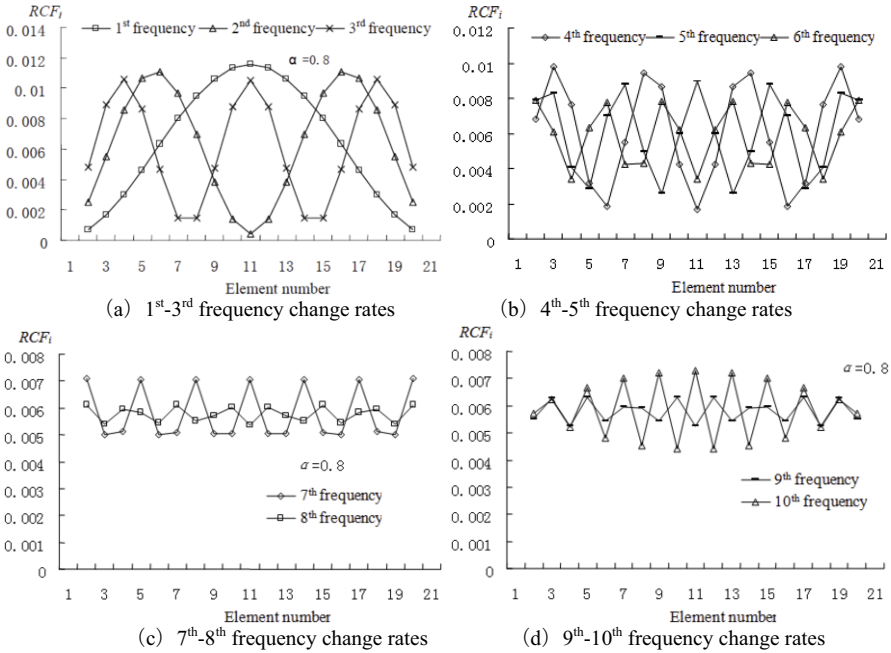


Figure 3. The beam frequency change rates in different damaged positions

It can be seen from Fig.3 (a) to (d) that the RCF of the different damage positions is an oscillation curve when the degree of damage is constant ( $\alpha = 0.8$ ), and the number of oscillations increases rapidly with the increase of modal order, and the oscillation amplitude value varies with the modal order. The sensitivity of the beam damage to RCF is related to the damage position N and the modal order of the frequency  $i$ , for example, the sensitivity of the beam damage to the first order modal RCF<sub>1</sub> is the largest in the mid-span, and the closer to the fulcrum, the lower the sensitivity of RCF<sub>1</sub>.

### 4.3. ARCF variation rule

ARCF variation diagram of single damaged simply supported beam with different damage degree  $\alpha$  from 0.8 to 0.2, a total of 76 conditions, are obtained by taking the modal order as the abscissa, and the amplitude of RCF of the 2-20 elements at the same modal order as the ordinate, see Fig. 4.

It can be seen from Fig. 4 that when the damage factor  $\alpha_N$  is 0.2 or 0.4, ARCF value of the simply supported beam at different damage position decreases with the increase of the modal value for 1 to 8 order modes, and for the 8 to 10 order modes, ARCF value of the curve increases gradually. when the damage factor  $\alpha_N$  is 0.6, ARCF value of the simply supported beam at different damage position decreases with the increase of the modal value for 1 to 7 order modes, and for the 7 to 10 order modes, ARCF value of the curve increases gradually. when the damage factor  $\alpha_N$  is 0.8, ARCF value of the simply supported beam at different damage position decreases with the increase of the modal value for 1 to 6 order modes, and for the 6 to 10 order modes, ARCF value of the curve increases gradually.

From above analysis, we know that

1. ARCF values of the single damaged beam at different damage positions vary with the change of modal order  $i$ .
2. In the case of the same damage factor  $\alpha$ , ARCF values of the single-damaged simply supported beam at different damage positions  $N$  decreases monotonically with the increase of modal order  $i$ , and then increases monotonically.
3.  $ARCF_i$  curve bending point (modal order  $i$ ) of the single damage simply supported beam at different damage positions decreases monotonically with the increase of the damage degree.
4. ARCF value of the single damage simply supported beam at different damage positions, in the non-bending point position, increases with the degree of damage increases.

4.4. MRCF variation rule

MRCF variation diagram of single damaged simply supported beam with different damage degree  $\alpha$  from 0.8 to 0.2, a total of 76 conditions, are obtained by taking the modal order as the abscissa, and MRCF values of the 2-20 elements at the same modal order as the ordinate, see Fig. 5.

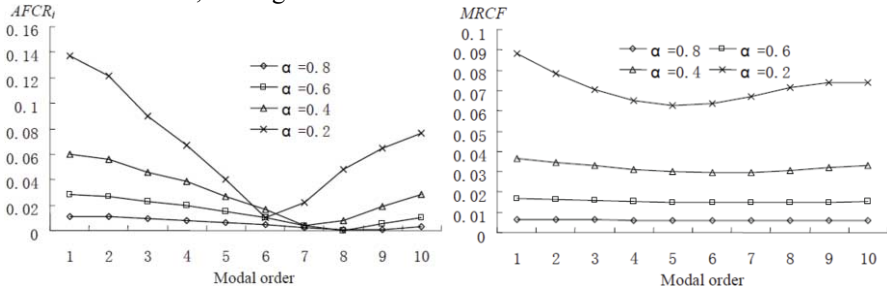


Figure 4. Frequency change rates amplitude variations Figure 5. Frequency change rates average diagram

It can be seen from Figure 6 that MRCF value of the single-damaged beam with different damage is increased with the increase of the damage degree.

4.5. The trend rule of RCF values with a constant damage

The RCF values trend diagram of the simply supported beam with constant damage is obtained by taking the modal order  $i$  as the abscissa with the fixed damage position  $N$  and damage degree  $\alpha_2=0.8$ ,  $\alpha_6=0.8$ ,  $\alpha_{11}=0.8$ (3 conditions), see Fig. 6.

It can be seen from Fig.6 that when the damage position  $N$  and the damage degree  $\alpha$  are fixed, RCF value does not change monotonically with the increase of the modal order  $i$ . RCF values vary from the general trend of increasing to the trend of decreasing with the increase of modal order  $i$  and a constant damage degree  $\alpha$  when the damage position  $N$  of the beam moves from both ends of the beam to the mid-span.

4.6. RRRCF rule

$RCF_{2,1}$ ,  $RCF_{3,1}$  and  $RCF_{3,2}$  variation diagrams of single-damaged simply supported beam with different damage degree  $\alpha$  from 0.8 to 0.2, are obtained by taking element number  $N$  as the abscissa, and  $RCF_{j,i}$  values of the 7-20 elements as the ordinate, see Fig. 7(a)-(c).

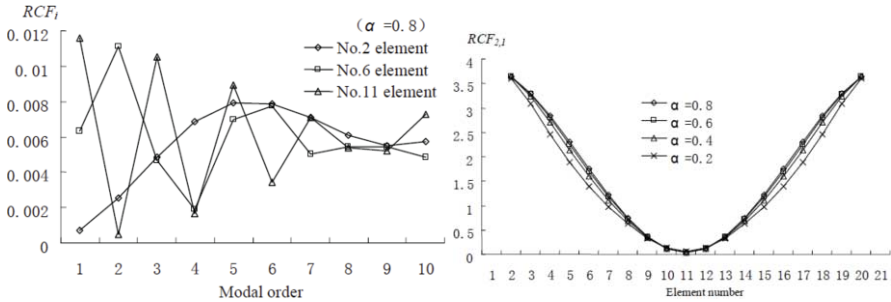
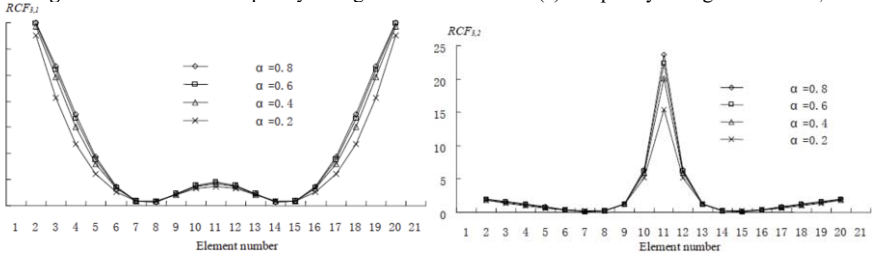


Figure 6. Each order frequency change rates chart

(a) Frequency change ratio  $FCR_{2,1}$



(b) Frequency change ratio  $FCR_{3,1}$

(c) Frequency change ratio  $FCR_{3,2}$

Figure 7. Single damaged beam frequency change ratio

It can be seen from Fig. 7 (a) to (c) that RRCF values of single damaged simply supported beam is independent of the degree of damage, and is related to the damage position  $N$  of the beam, and the graph is symmetrical.

### 5. Conclusion

This paper, based on the numerical experiment, elects 19 damaged elements positions and the 4 kinds of damage degree, a total of 76 kinds of combination conditions, and studies the single-damaged simply supported beam rules of frequency, RCF, and RRCF at different damage positions and degrees, and proves the correctness of theoretical derivation, which can provide a basis for further damage position and degree identification.

1. The frequency and RCF of the single-damaged simply supported beam at different damage position are oscillation curves, and the oscillation times increases rapidly with the increase of the modal order.

2. The oscillation amplitude value of RCF curve of the single-damaged simply supported beam at different damage positions varies with the change of modal order. In the case of the same damage factor, the amplitude of the oscillation curve of RCF decreases firstly and then increases monotonically with the increase of modal order. The bending change point (modal order) the oscillation curve of RCF curve decreases monotonically with the increase of the damage degree. The amplitude of the oscillation curve of RCF at the non-bending point, increases with the increase of the damage degree.

3. The sensitivity of frequency or RCF to beam damage is related to the damage degree, and the greater the damage degree, the more the frequency of each order modal is reduced or the greater the RCF. The sensitivity of frequency or RCF to beam damage

is related to the damage position and modal order of the frequency. The influence of damage degree on natural frequency is greater than that of damage position.

4. When the damage position and damage degree are constant, RCF of the beam does not change monotonically with the increase of the modal order. RRCF of the single-damaged simply supported beam is independent of the damage degree, and is related to the position of the beam, whose graph is symmetrical.

## References

- [1] YAN Ya-xun. Bridge structure damage identification research based on the dynamic testing and the research of bridge comprehensive evaluation theory. Xi'an: Chang'an University, 2008.
- [2] LIU Wen-feng, LIU Chun-tu, YING Huai-qiao. Research on damage orientation by change of eigenfrequency. *Journal of Vibration and Oscillation*, 2004, 23 (2): 28-30.
- [3] KAMINSKI, P C. The approximate location of damage through the analysis of natural frequencies with artificial neural networks. *Journal of Process Mechanical Engineering* 1995(209): 117~123.
- [4] Du Jin-long, GUO Shao-hua. Finite element analysis of dynamic characteristics of damaged reinforced concrete beam. *Journal of Shijiazhuang Railway Institute: Natural Science*, 2007,20(3):18-23.
- [5] JIA Bin, WANG Ru-heng, GUO Wen. Study on damage locating method in beam-like structure by frequency identification. *Journal of Southwest University of Science and Technology*, 2007,22(4): 35-39.
- [6] WANG Zhe, DING Hua. Estimating thresholds of damages based on natural frequency method. *Mechanics in Engineering*, 2009,31(2): 50-54.
- [7] XU Fei-hong, ZHANG Jia-wen. Damage identification of simple supported beam based on the curve intersection of characteristic equation. *Journal of Changsha University of Science and Technology: Natural Science*, 2009,6(3): 11-16.
- [8] ZHAO Ya-ping, WANG Shu-mao, JIAO Qun-ying, et al. Numerical simulation to relation between vibration characteristic and damage condition of reinforced concrete beam. *Chinese Journal of Applied Mechanics*, 2007,24(1): 141-145.
- [9] SHI Zhou, ZHAO Ren-da. Effects of bridge damage on its free-vibration characteristics. *Journal of Earthquake Engineering and Engineering Vibration*, 2007,27(5): 117-123.



# Effect of Vehicle Combinations on the Instability of Heterogeneous Urban Traffic Flow

Cheng Ye<sup>a</sup>, Juan Xu<sup>a</sup>, Shangzhi Xu<sup>a</sup>, Yeqing Qian<sup>a</sup>, Zhipeng Li<sup>a,1</sup>

<sup>a</sup>*Key Laboratory of Road and Traffic Engineering, Ministry of Education & Department of Information and Communication Engineering, Tongji University, Shanghai, 201804 China*

**Abstract.** In this paper, an extended car-following model with consideration of the combination effect in heterogeneous urban traffic flow with slow and fast vehicles is proposed. The combination effect means the order of vehicles in a heterogeneous flow matters, since a fast vehicle would degenerate to a middle state between a fast vehicle and a slow vehicle when it follows a slow vehicle. Specifically, we consider three types of vehicle combinations in the urban traffic flow, namely fast vehicle following fast vehicle (FF), fast vehicle following slow vehicle (FS), and slow vehicle following fast/slow vehicle (SX). Linear stability analysis is conducted to obtain stability criterion, from which we can conclude that the higher penetration of the FS combination can increase the stability of traffic flow.

**Keywords.** Heterogeneous traffic; Combination effect; Car-following model.

## 1. Introduction

Traffic congestion has always been a serious issue to many big cities especially in developing countries, where the speed of construction of infrastructure hardly satisfy the demand of public transportation. Besides, the number of vehicles on the road is larger with the development of economy, which makes the situation worse. In this background, it is urgent to study the complex traffic phenomena and the traffic flow dynamics behind them.

Therefore, many scholars have devoted themselves into the area of traffic flow dynamics, and proposed various models to describe complex traffic phenomenon. Roughly speaking, the existing traffic models in single lane can be classified into macroscopic traffic model [1-5], mesoscopic traffic model [6-8] and microscopic traffic model [9-17] in consideration of the scale of traffic, where macroscopic models describe traffic dynamics at an aggregated level (i.e. evolution of traffic flow and density), while microscopic models replicate traffic dynamics in more detailed level (i.e. the motions of individual vehicles such as acceleration/deceleration, headway, speed etc.), and the mesoscopic models show traffic dynamics at an immediate level (i.e. the dynamics of a group/distribution of vehicles). Among microscope traffic model,

---

<sup>1</sup> Corresponding Author.

one of the most classical single-lane traffic models for homogeneous traffic flow is car-following model which is based on the idea that there exists interactions between neighboring vehicles, and behaviors of one driver would be affected by its front vehicle [18]. Then a general optimal velocity function was introduced to describe influence of the stimulus from the preceding vehicle [9].

The studies mentioned above are restricted to the homogeneous traffic flow in a single lane which ignore the facts that different vehicles would exist in a traffic flow. Therefore, the studies of heterogeneous traffic flow may be closer to real situations. Li et al. [19] explore the stability of heterogeneous traffic flow with different self-stabilizing control vehicles. Li et al. [20] focus on the characteristic of the mix traffic flow consisting of different penetration of fast vehicles and slow vehicles. Besides, the combination effect in heterogeneous traffic flow is also a very interesting phenomenon where driving behaviors would change with the type of preceding vehicle. Specifically, Ngoduy[21] and Yang et al. [22] has investigate the combination effect in the environment of IDM and OVM respectively through the division of mixed car-truck traffic flow into CC, CT, TC, TT, which means the four combinations of car following a car, car following a truck, truck following a car, and truck following a truck. Their researches are the founders to theoretical studies of the combination effect at microscopic level, while it seems that the case is more common in urban traffic environment that the traffic flow only consist of cars distinguished by different maximum velocity, whose combination effect has not been explored.

To this end, this paper focus on the theoretical analysis for the combination effect in urban traffic environment. For the convenience of modelling, it is assumed that there exists only two categories of vehicles in the urban traffic flow as shown in Ref [20], called fast vehicles and slow vehicles. From the empirical observation, a fast vehicle would degenerate to a middle state between a fast vehicle and a slow vehicle when it follows a slow vehicle, while a slow vehicle may not have the middle state since the limitation of its relatively low speed whether it follows a fast vehicle or not. Therefore, there exists only three effective combinations of vehicles, namely FF (a fast vehicle following a fast vehicle), FS (a fast vehicle following a slow vehicle) and SX (a slow vehicle following a fast/slow vehicle).

The remainder of this paper is organized as follows: Section 2 introduces the heterogeneous urban traffic flow with the consideration of combination effect of fast and slow vehicles. Section 3 presents the linear stability analysis for the extended model. Section 4 is the conclusion of the paper.

## 2. Model description

The Bando's optimal velocity model (OVM) [9] is one of most classical microscope traffic model in single lane, which introduces optimal velocity function (OVF) based on the preceding vehicle spacing for each vehicle to describe desired velocity. The framework of the OVM can be written as following:

$$\frac{dv_j(t)}{dt} = a \left[ V(\Delta x_j(t)) - v_j(t) \right] \quad (1)$$

where  $v_j(t)$  and  $x_j(t)$  are the velocity and position of the  $j$  th vehicle at the time  $t$ ,  $a$  is the sensitivity coefficient which depicts sensitivity of a driver,

$\Delta x_j(t) = x_{j-1}(t) - x_j(t)$  is the headway to the preceding vehicle, and  $V(\Delta x_j(t))$  is optimal velocity function which is monotonically increasing with headway and has an upper bound, so hyperbolic tangent function is often chosen.

Then taking into account the difference of maximum velocity among vehicles in a specific traffic flow, the original OVM has been generalized to describe the heterogeneous traffic flow with fast and slow vehicles as shown in Ref [20]. And with driving experience and empirical observation as described above, the combination effect should be taken into consideration, where a fast following a slow vehicle would degenerate into a middle state between a fast vehicle and a slow vehicle with respect to maximum velocity. Whereas, no matter which type of vehicle a slow vehicle follows, the slow may not exist the middle state since the limitation of its velocity. Therefore, there exists three categories of combination in a traffic flow consisting of fast vehicles and slow vehicles, namely FF (a fast vehicle following a fast vehicle), FS (a fast vehicle following a slow vehicle) and SX (a short vehicle following a fast/slow vehicle).

Therefore, in this paper, we generalize the original OVM with the consideration of the combination effect of fast vehicles and slow vehicles, which can be described as follows:

$$\left\{ \begin{array}{l} \frac{dv_j}{dt} = a[V_{FF}(\Delta x_j) - v_j(t)] \\ \text{for vehicle combination of FF with a penetration } p_{FF} \\ \frac{dv_j}{dt} = a[V_{FS}(\Delta x_j) - v_j(t)] \\ \text{for vehicle combination of FS with a penetration } p_{FS} \\ \frac{dv_j}{dt} = a[V_{SX}(\Delta x_j) - v_j(t)] \\ \text{for vehicle combination of SS/SF with a penetration } p_{SX} \end{array} \right. \quad (2)$$

where  $V_{FF}(\Delta x_j)$ ,  $V_{FS}(\Delta x_j)$  and  $V_{SX}(\Delta x_j)$  is the optimal velocity function of the vehicle combination of FF, FS and SS/SF respectively. Besides, these optimal velocity function is the generation of (3) with different parameters inspired by the Ref. [23].

$$V(\Delta x_j(t)) = v_{\max} \frac{\tanh\left(\frac{\Delta x_j(t)}{\Delta s} - \beta\right) + \tanh(\beta)}{1 + \tanh(\beta)} \quad (3)$$

where  $v_{\max}$  is maximum velocity,  $\beta$  denotes form factor and  $\Delta s$  is transition width. Specifically, the maximum velocity for FF, FS and SX are  $120\text{km/h}$ ,  $80\text{km/h}$  and  $60\text{km/h}$ ;  $\beta=1.5$  for all combinations and  $\Delta s = 15\text{m}$ ,  $19\text{m}$  and  $12\text{m}$  respectively.

### 3. Linear stability analysis

In order to test the stability of our proposed model as shown in (2), we conduct linear stability analysis by introducing a small perturbation for the model in this section. Since each single vehicle is described by (1) regardless of different combination effect, we can analyze the stability condition of (1) first. Therefore, we rewrite it as follows:

$$\frac{dv_j}{dt} = E(v_j, \Delta x_j) = a \left[ V(\Delta x_j(t)) - v_j(t) \right] \quad (4)$$

In the equilibrium state of traffic flow, all vehicles share the same velocity  $v_j^0$  and constant headway  $\Delta x_j^0$ , besides their acceleration is extremely close to zero; that is to say all vehicles are in relatively stationary status. Thus, in order to conduct linear stability analysis, when the stability criterion is satisfied,  $\delta v_j$  and  $\delta \Delta x_j$  denoted small deviations of velocity and headway are put into the traffic flow:  $v_j = v_j^0 + \delta v_j$  and  $\Delta x_j = \Delta x_j^0 + \delta \Delta x_j$ . Then, first-order Taylor expansion of (4) is conducted, which leads to:

$$\frac{d\delta v_j(t)}{dt} = \underbrace{E(v_j^0, \Delta x_j^0, \Delta x_{j,his})}_{=0} + E_j^v \delta \Delta v_j + E_j^{\Delta x} \delta \Delta x_j \quad (6)$$

where

$$E_j^{\Delta v} = \left. \frac{\partial E_j}{\partial v_j} \right|_{(v_j^0, \Delta x_j^0)}, \quad E_j^{\Delta x} = \left. \frac{\partial E_j}{\partial \Delta x_j} \right|_{(v_j^0, \Delta x_j^0)} \quad (7)$$

It is assumed that  $\delta v_j(t) = V_j e^{ij\omega + \lambda t}$  and  $\delta \Delta x_j(t) = \Delta X_j e^{ij\omega + \lambda t}$  with constant  $V_j$  and  $\Delta X_j$ , we can simplify (6) to the following terms:

$$V_j \left( \lambda^2 - \lambda E_j^v + E_j^{\Delta x} \right) = V_{j-1} e^{-i\omega} E_j^{\Delta x} \quad (8)$$

Since the analysis is conducted under periodical boundary condition, where last vehicle follows the first one in a platoon, it is obvious to obtain  $V_N = V_0$ . Thus, (8) can be rewritten to (9):

$$\prod_{j=1}^N \left( \lambda^2 - \lambda E_j^v + E_j^{\Delta x} \right) = e^{-iN\omega} \prod_{j=1}^N E_j^{\Delta x} \quad (9)$$

With prior knowledge as shown in Ref [21], the  $\omega = 0$ ,  $\lambda = 0$  is the solution of (9). Thus, we can expand  $\lambda$  as  $i\omega\lambda_1 - \omega^2\lambda_2 + \dots$ , where  $\lambda_1$  and  $\lambda_2$  are both real rational numbers, and replace  $e^{-i\omega N}$  to its second-order Taylor expansion. Then, the substitution can be done in (9), and if we keep the first-order term and second-order term of  $\omega$ , we can obtain the following terms:

$$\left\{ \begin{array}{l} \Theta(\omega) = -i\lambda_1 \sum_{j \in N} E_j^v \prod_{k \neq j, k \in N} E_k^{\Delta x} + iN \prod_{j \in N} E_j^{\Delta x} = 0 \\ \Theta(\omega^2) = -\lambda_1^2 \sum_{j \in N} \prod_{k \neq n, k \in N} E_k^{\Delta x} - \lambda_1^2 \sum_{n < m, (n, m) \in N} E_n^v E_m^v \prod_{k \neq (n, m), k \in N} E_k^{\Delta x} \\ \quad + \lambda_2 \sum_{j \in N} E_j^v \prod_{k \neq j, k \in N} f_k^{\Delta x} + \frac{N^2}{2} \prod_{j \in N} E_j^{\Delta x} = 0 \end{array} \right. \quad (10)$$

By simplifying (10), we can obtain:

$$\left\{ \begin{array}{l} \lambda_1 = \frac{N}{\sum_{j \in N} \frac{E_j^v}{E_j^{\Delta x}}} \\ \frac{\lambda_2}{\lambda_1^2} \sum_{j \in N} \frac{E_j^v}{E_j^{\Delta x}} = \sum_{j \in N} \frac{1}{E_j^{\Delta x}} - \frac{1}{2} \sum_{j \in N} \left( \frac{E_j^v}{E_j^{\Delta x}} \right)^2 \end{array} \right. \quad (11)$$

In order to satisfy the linear stability of the linear stability condition of the model,  $\lambda_1 \leq 0$  and  $\lambda_2 \geq 0$  should be guaranteed. Based on (11),  $\lambda_1 \leq 0$  is always satisfied for the extended car-following model of heterogeneous traffic, whereas  $\lambda_2 \geq 0$  can be reached if:

$$\sum_{j \in N} \left[ \frac{1}{2} \left( \frac{E_j^v}{E_j^{\Delta x}} \right)^2 - \frac{1}{E_j^{\Delta x}} \right] \geq 0 \quad (12)$$

Thus, we obtain the stability condition for homogeneous traffic flow. Then, since our extended model contains three combinations of vehicles (namely, FF, FS and SX), the stability condition should be generalized as follows:

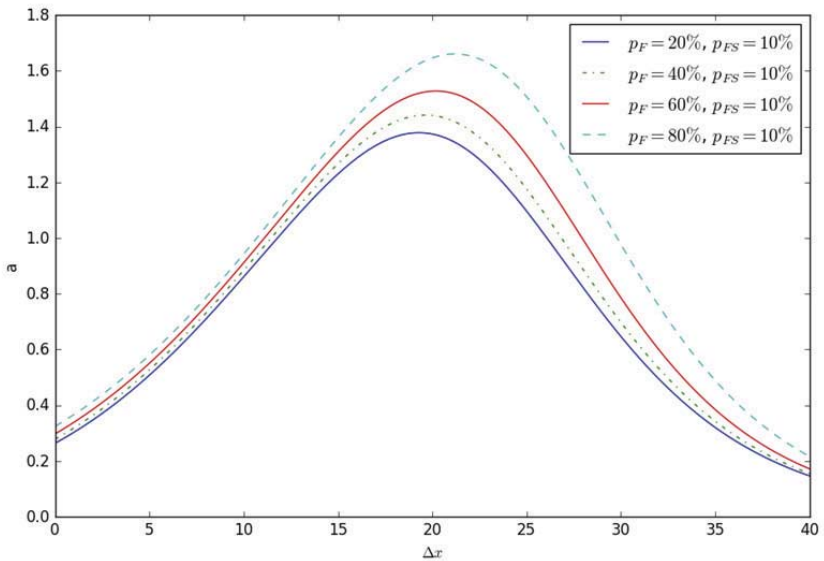
$$p_{FF} N \left[ \frac{1}{2} \left( \frac{E_{j,FF}^v}{E_{j,FF}^{\Delta x}} \right)^2 - \frac{1}{E_{j,FF}^{\Delta x}} \right] + p_{FS} N \left[ \frac{1}{2} \left( \frac{E_{j,FS}^v}{E_{j,FS}^{\Delta x}} \right)^2 - \frac{1}{E_{j,FS}^{\Delta x}} \right] + p_{SX} N \left[ \frac{1}{2} \left( \frac{E_{j,SX}^v}{E_{j,SX}^{\Delta x}} \right)^2 - \frac{1}{E_{j,SX}^{\Delta x}} \right] \geq 0 \quad (13)$$

where  $p_{FF}$ ,  $p_{FS}$  and  $p_{SX}$  denote the penetration of the three combinations respectively, namely a fast vehicle following a fast vehicle, a fast vehicle following a slow vehicle and slow vehicles.

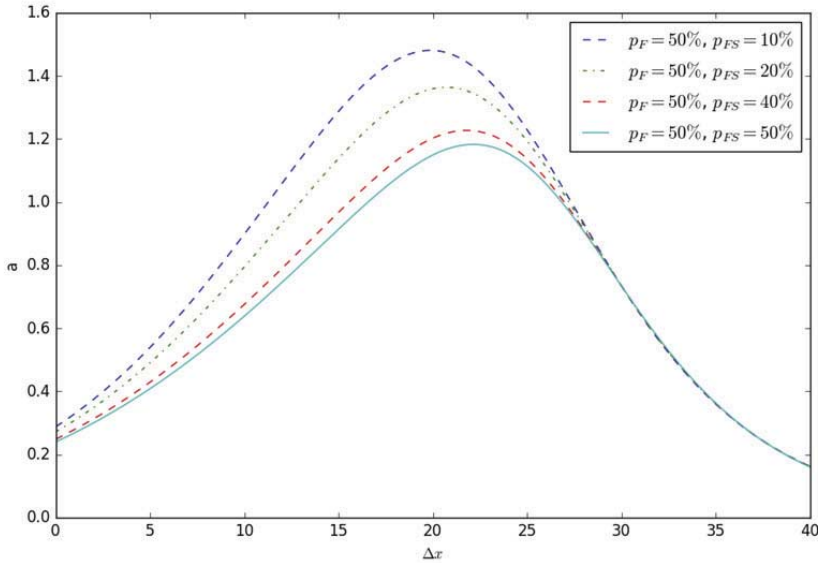
Besides, it is obvious to obtain that  $E_{j,m}^v = -a$ ,  $E_{j,m}^{\Delta x} = aV'_m(\Delta x)$ , where  $a$  is sensitivity coefficient and  $V'_j(\Delta x)$  is the derivative of the optimal velocity function at headway  $\Delta x$  according to the vehicle combination  $m$  (FF, FS or SX). Thus, the stability criterion for heterogeneous traffic flow with vehicle combination can be simplified as:

$$a \geq \frac{2V'_{FF} V'_{FS} V'_{SX} (p_{FF} V'_{FS} V'_{SX} + p_{FS} V'_{FF} V'_{SX} + p_{SX} V'_{FF} V'_{FS})}{p_{FF} (V'_{FS} V'_{SX})^2 + p_{FS} (V'_{FF} V'_{SX})^2 + p_{SX} (V'_{FF} V'_{FS})^2} \quad (14)$$

The stability criterion describes clearly the relationship between the sensitivity of the driver and the percentage of three different combinations of vehicles. Then, Figs. 1 and 2 have been plotted to represent (14), and the area under the each curves in the figure is the unstable region. Specifically, Fig.1 shows the neutral stability curves of different penetration of fast vehicles ( $p_F$ ), under the condition of fixed the penetration of FS combination ( $p_{FS}$ ), where we can tell that the increased penetration of fast vehicles can destabilize the traffic flow with fixed penetration of FS combination. And then we can observe the evolution of traffic stability with different penetration of FS combination under the condition that fast vehicles and slow vehicles share the same quantity (50%) as shown in Fig.2, where we may draw conclusion that the increased penetration of FS can increase the stability of traffic flow.



**Figure 1.** The neutral stability curves of the extended model with different penetration of fast vehicles under the fixed penetration of FS combination.



**Figure 2.** The neutral stability curves of the extended model with different penetration of combination of FS under the fixed penetration of fast vehicles.

#### 4. Conclusions

In this paper, an extended optimal velocity model has been proposed for heterogeneous urban traffic flow with fast vehicles and slow vehicles, which takes the different vehicles combinations into consideration, namely FF, FS and SS. The inspiration of the proposed model based on the facts that the behaviors of a fast vehicle following a slow vehicle are different from following a fast vehicle. And then theoretical linear stability analysis to the proposed model has been conducted to obtain the stability criterion, which we find that the increasing penetration of fast vehicles can destabilize the traffic flow under the condition of fixed penetration of FS combination. Besides, it can be observed that the traffic flow tends to be more stable with the increase of penetration of FS combination when the fast vehicles and slow vehicles share the same quantity.

#### Acknowledgments

This work is supported by the Natural Science Foundation of China under Grant No.51422812 and No.71571107, the Central Universities under Grant No. 0800219308, and the Scientific Foundation of Shenzhen Government of China (GCZX20140508161906699).

## References

- [1] D. Ngoduy, Generalized macroscopic traffic model with time delay, *Nonlinear Dynamics* 77.1-2 (2014), 289-296.
- [2] D. Ngoduy, and R. E. Wilson, Multianticipative nonlocal macroscopic traffic model, *Computer - Aided Civil and Infrastructure Engineering* 29.4 (2014), 248-263.
- [3] D. Ngoduy, and D. Jia, Multi anticipative bidirectional macroscopic traffic model considering cooperative driving strategy, *Transportmetrica B: Transport Dynamics* 5.1 (2017), 100-114.
- [4] N. Davoodi, A.R. Soheili, and S.M. Hashemi, A macro-model for traffic flow with consideration of driver's reaction time and distance, *Nonlinear Dynamics* 83.3 (2016), 1621-1628.
- [5] T. Tang, L. Caccetta, Y. Wu, H. Huang, and X. Yang, A macro model for traffic flow on road networks with varying road conditions, *Journal of Advanced Transportation* 48.4 (2014), 304-317.
- [6] X. Zhou and J. Taylor, DTALite: A queue-based mesoscopic traffic simulator for fast model evaluation and calibration, *Cogent Engineering* 1.1 (2014), 961345.
- [7] S.L. Paveri-Fontana, On Boltzmann-like treatments for traffic flow: a critical review of the basic model and an alternative proposal for dilute traffic analysis, *Transportation research* 9.4 (1975), 225-235.
- [8] D. Chowdhury, L. Santen, and A. Schadschneider, Statistical physics of vehicular traffic and some related systems, *Physics Reports* 329.4 (2000), 199-329.
- [9] M. Bando, K. Hasebe, A. Nakayama, and A. Shibata, Dynamical model of traffic congestion and numerical simulation, *Physical review E* 51.2 (1995), 1035.
- [10] H. Ge, S. Dai, L. Dong, and Y. Xue, Stabilization effect of traffic flow in an extended car-following model based on an intelligent transportation system application, *Physical Review E* 70.6 (2004), 066134.
- [11] F. Liu, R. Cheng, H. Ge and C. Yu, A new car-following model with consideration of the velocity difference between the current speed and the historical speed of the leading car, *Physica A: Statistical Mechanics and its Applications* 464 (2016), 267-277.
- [12] G. Peng, and D. Sun, A dynamical model of car-following with the consideration of the multiple information of preceding cars, *Physics Letters A* 374.15 (2010), 1694-1698.
- [13] G. Peng, W. Lu, H. He and Z. Gu, Nonlinear analysis of a new car-following model accounting for the optimal velocity changes with memory, *Communications in Nonlinear Science and Numerical Simulation* 40 (2016), 197-205.
- [14] T. Nagatani, Stabilization and enhancement of traffic flow by the next-nearest-neighbor interaction, *Physical Review E* 60.6 (1999), 6395.
- [15] S. Yu, and Z. Shi, An improved car-following model considering relative velocity fluctuation, *Communications in Nonlinear Science and Numerical Simulation* 36 (2016), 319-326.
- [16] T. Tang, J. Zhang, L. Chen, and H. Shang, Analysis of vehicle's safety envelope under car-following model, *Physica A: Statistical Mechanics and its Applications* 474 (2017), 127-133.
- [17] H. Kuang, Z. Xu, X. Li and S. Lo, An extended car-following model accounting for the honk effect and numerical tests, *Nonlinear Dynamics* 87.1 (2017), 149-157.
- [18] G.F. Newell, Nonlinear effects in the dynamics of car following, *Operations research* 9.2 (1961), 209-229.
- [19] Z. Li, W. Li, S. Xu, Y. Qian, and J. Sun, Traffic behavior of mixed traffic flow with two kinds of different self-stabilizing control vehicles, *Physica A: Statistical Mechanics and its Applications* 436 (2015), 729-738.
- [20] Z. Li, X. Xu, S. Xu, Y. Qian, and J. Sun, Analytical studies on an extended car following model for mixed traffic flow with slow and fast vehicles, *International Journal of Modern Physics C* 27.01 (2016), 1650004.
- [21] D. Ngoduy, Effect of the car-following combinations on the instability of heterogeneous traffic flow, *Transportmetrica B: transport dynamics* 3.1 (2015), 44-58.
- [22] D. Yang, P. Jin, Y. Pu and B. Ran, Stability analysis of the mixed traffic flow of cars and trucks using heterogeneous optimal velocity car-following model, *Physica A: Statistical Mechanics and its Applications* 395 (2014), 371-383.
- [23] M. Treiber, and A. Kesting, *Traffic Flow Dynamics: Data, Models and Simulation*, Springer-Verlag, Berlin Heidelberg, 2013.



# Design and Implementation of an APP-Based Intelligent Service System

Jia Bao, Jun Tao\*, Chao Wen, and Jingwen Zhang  
*Wuhan University of Technology, Wuhan, PR China*

\*e-mail: [hubei.taojun0411@163.com](mailto:hubei.taojun0411@163.com)

**Abstract.** The article designs an APP-based intelligent service system based on the current university library seat utilization. As a development tool that connects the users with libraries, this system offers online reservation and other people-oriented services (e.g. information push and star-level services). In addition, the system also provides fully functional services with the combination of the QR code scanning and unlocking technology, raising the seat utilization rate to the most possible extent.

**Keywords.** online reservation; APP management; QR code scanning

## Introduction

University library is an ideal place for college students to study. However, seat occupying is a prevalent problem in university library, which has brought negative impacts to students' learning and the utilization of library resources. To solve this problem, many colleges and universities have taken various measures. One of them is the intelligent seat selection system, which enables users to pick seats online. Though this system can precisely control the seats, it fails to verify that the users are sitting on the reserved seats.

Another solution is the C/S mode based LAN (Local Area Network) seat management system, with which the seat is allocated the moment a user swipes his or her entrance card. Though a linkage can be established between the entrance card system and seat management system, it inevitably provides seats for those users who only borrow books in libraries, resulting in a waste of seat resources[1]. These two mentioned systems all require the presence of the user in the library, which does not reduce the queue length and waiting time.

Therefore, here a APP-based intelligent service system is introduced to solve this dilemma. While enabling users to reserve seats at any time any place, it also establishes a one-to-one connection between the user and the seat reserved, using the QR (Quick Response) code unlocking technology. As a result, the system has obvious advantages over the previous instances such as achieving the remote control of the seat to reduce the waiting time, decreasing the dependence of the physical ID card and matching the reserved seat with the seat that is actually taken. These contribute to the increasing utilization of the library seats.

## System Overall Design

The library seat management system consists of two parts: APP client and QR code scanning and unlocking. The APP client consists of the functional module and the QR code generation module; the latter is composed of the QR code scanner, Bluetooth transceiver module, MCU (Microcontroller Unit) control module and the motor-driven unlocking module. Users will receive a QR code after reserving a seat in the APP client. When unlocking the seats, the scanner scans the QR code provided by the user, and the received information is compared with that in the database. If the received information matches the reservation information, the unlocking signal is then sent to MCU control module via Bluetooth to realize the motor-driven unlocking. The flow chart is shown in Figure 1.

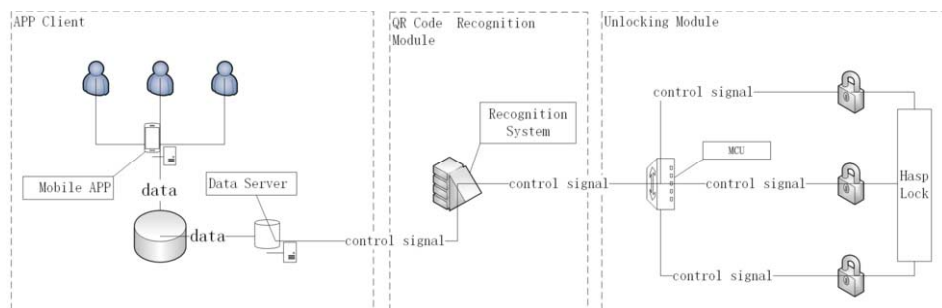


Figure 1. System workflow.

## APP Client Development and Design

### System framework

The APP-based library seat management system, involving the client and server, is independently developed with the C/S framework. The Client is for the user interaction interface and the Server for the operation.

The SSH framework is selected for the server. The SSH framework includes the presentation layer, business logic layer, data persistence layer and domain model layer. The separation of the MVC in the basic framework of the whole system is achieved by Struts, of which the control of the task transition is a part. Hibernate framework supports the persistence layer and Spring manages the Struts and Hibernate. These multiple separated layers have several advantages such as the quick building of the app, clear structure, good replicability, and convenient maintenance.

In the business process, the presentation layer achieves the interaction using the APP interface, whose tasks include sending the request, receiving and displaying the data. Regulated by the configuration file, Struts sends the request of the client APP to the corresponding Servlet (interface), which executed some action logic. Then Hibernate extracts the data for the business logic from the database and saves it into corresponding DAO. Spring assigns the application of these data to different Servlets, and returns the processed results to the Client [2]. The detailed frame is shown in Figure 2.

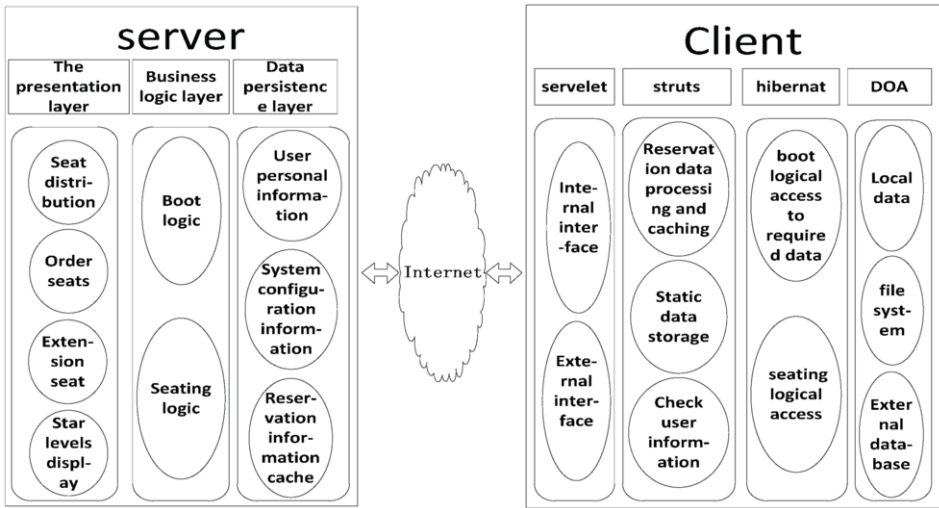


Figure 2. APP system framework.

Functional design

Based on the “user-friendly” principle, the APP is designed to offer seat reservation service ranging from a series of personalized situation, providing users with the best experience. As shown in Figure 3, the App consists of eight functional modules: seat type selection module, seat selection module, cancellation module, navigating and positioning module, query module, duration selection module and information display module. In addition, a two-way circulating service pattern is embedded into the APP, which recommends and customizes the functions based on the users’ behaviors to improve the convenience and interactivity of the APP.

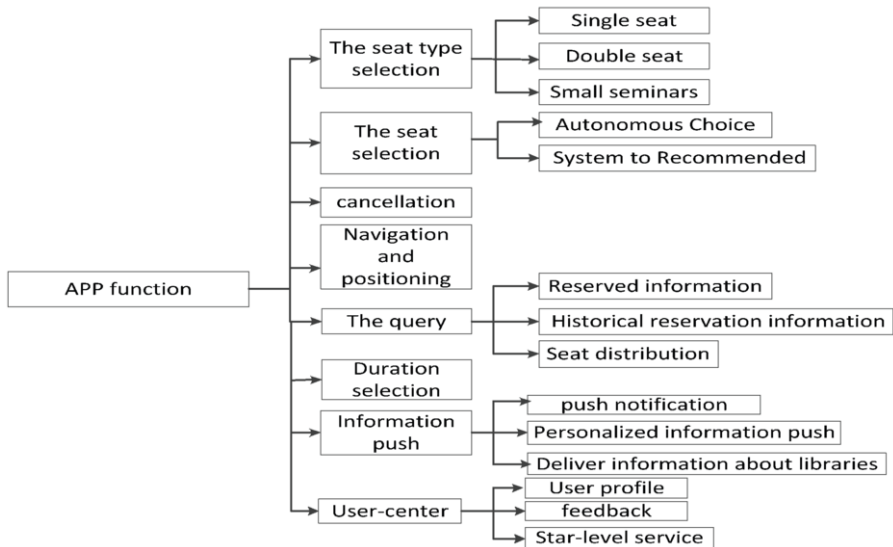


Figure 3. APP functional modules.

The seat type selection module offers multiple options for the users such as single seat, double seats, seats plan for small seminars and other commonly-used seats plan. Users can also submit their customized seat plan to the system for the system to allocate the seats.

Seat selection module provides the available seat option for the users. The seats are displayed with the floor to offer a more intuitive and convenient selection. The system will also recommend seats according to their previous reservation habits.

Both of the above functions are achieved based on the bidirectional cyclic service mode. Two methods are adopted in this mode. One is the customized plan. System allocates the seat by the demand of the user. The other is automatic selection. System selects the most suitable option for the users based on their previous behavior and preference<sup>[3]</sup>.

Cancellation module serves to cancel the existing reservation, including passive and active cancellation. The active cancellation is defined as the cancellation request sent by the users. Passive cancellation is defined as the cancellation due to the unavoidable situations such as the users' being late for more than 15 minutes or absent for more than 45 minutes. After the situation meets the requirement of the passive cancellation for more than 15 minutes, the system will return the seats to the selection pool. The maximum active cancellation time is set to be 3 times a day to avoid the waste brought by the malicious cancellation.

Navigating and positioning module helps users find their seats quickly if the access to the direction sensor is granted. The position data of the sensor will be transmitted to the monitor for the system to determine the relative direction and distance between the users and seats. Combining that with the navigation data in GEO (Geographical Position Coordinates), the module can display the real time location of the user in the floor plan to provide navigation.

Query module checks the seat reservation in a real-time basis and provides users with the current reservation information and history, which assists the management of their own reservation records.

Duration selection module enables users to set the reservation time. The maximum duration is determined by users' credit levels (introduced later).

Information display module includes three parts. The first one has push notification function. Push notifications can be delivered to the client via SDK (Software Development Kit). For instance, the APP will inform users "Time is coming up" 15 minutes prior to the due time. Similarly, when the time is approaching 15 minutes before the end of SDK, the APP will warn users and provide them with two options, "Continue" or "Leave". If the "Leave" option is selected, the locking signal will be sent to the QR code scanning and unlocking module. If "Continue" option is selected, the duration information will be updated and the time will be recounted in the database server. The second part involves users' personalized information push. The system will promptly notify the current seat availability information such as number and the location of the seats that are open to reserve according to users' previous preference. The third part releases information about libraries such as emergency (e.g., power failure) and the opening hours in holidays. The module service process is shown in Figure 4.

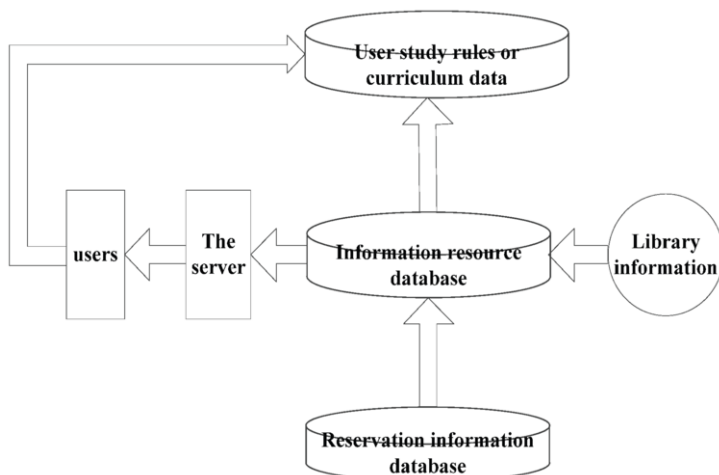


Figure 4. Information push service process.

User-center module is composed of three submodules. They are user profile, feedback and star-level ranking. User profile shows users' basic information. Feedback provides the users with a feedback platform, i.e. users can report others' malicious behaviors during the usage and librarians will regulate the usage of seats under the actual condition. Another function of this module is the QA (Question and Answer) session. A forum, either established by the developers or by third-party, will serve as a platform for the users to communicate either with other users or with the support team to solve the issues encountered during the routine use.

Star-level service module is designed to encourage users' good behavior. It will adjust increase or decrease star levels in their behavior data tables according to the access of users' behavior. Bad behavior refers to the violation of the library rule and good behavior includes not using cellphone for a certain time interval. Default Star levels is set to 4 and one star can be used as 2 hours' study time. When the listener function does not detect the cellphone usage for 45 minutes, a star will be rewarded for being focus. A star will be reduced when the violation adds to three. data flow diagram is shown in Figure 5.

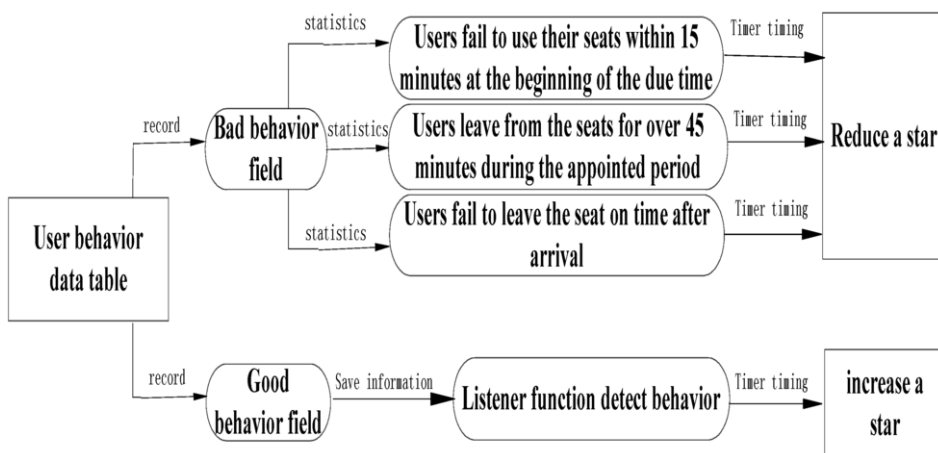


Figure 5. Star-level module data flow chart.

QR code and its generation technology

Major progress has been made in the fields of Internet and big data, resulting in the wide use of QR code. As the parameter distribution shown in Figure 6, the QR code carries large quantities of information within a relatively small area, which makes it becomes an ideal carrier of various types of information. Information is stored as black and white patterns coded according to certain rules horizontally and vertically. By mimicking the basic algorithm of computer, i.e. bit stream of “0” and “1”, the QR Code uses several binary geometric shapes to represent the numerical information, which is read and processed by image input devices such as photoelectric scanners<sup>[4]</sup> to achieve the automatic processing of the information.<sup>[4]</sup>

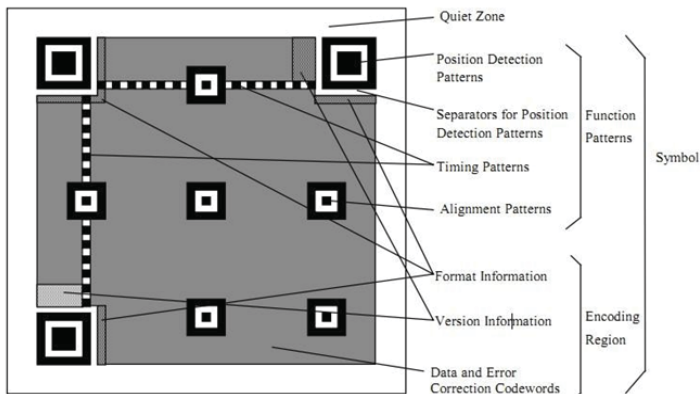


Figure 6. QR code parameter distribution diagram

The common coding format of the QR code are data matrix, QR code, etc. In this system the QR code is adopted to achieve the recording of the reservation information.

QR Code is generated based on the QR code Common in the Two Dimension Code. Various parameters of QR Code graphics including contents, size, type, etc. are transferred to the QR Code Common. Objects are created via QRCode.jar package after the completion of parameter transferration to set the size and contents. The QR code image is ultimately formed using the fill Rect method in Graphics 2D, as shown in Figure 7.

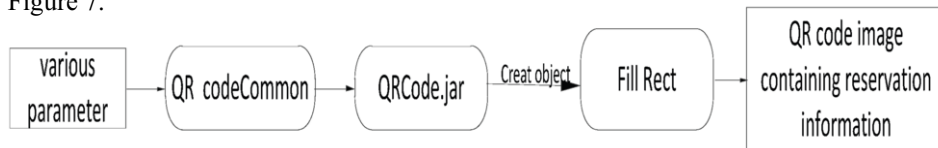


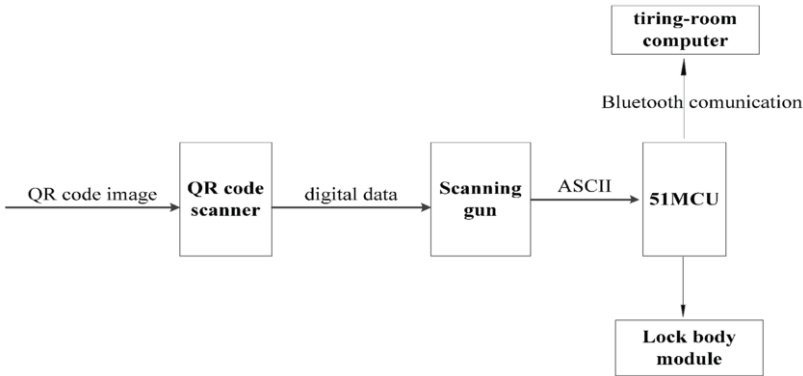
Figure 7. QR code image generation data flow diagram

QR code scanning and unlocking module

The system uses QR code to recognize and check the reservation information. The data storage of QR code has advantages such as low cost, large capacity. It can reduce the developing cost and meet the needs of various users. Unlike Radio Frequency Identification (RFID) which is commonly used in traditional seat management systems, this method reduces the dependence of the physical to avoid situations in which users

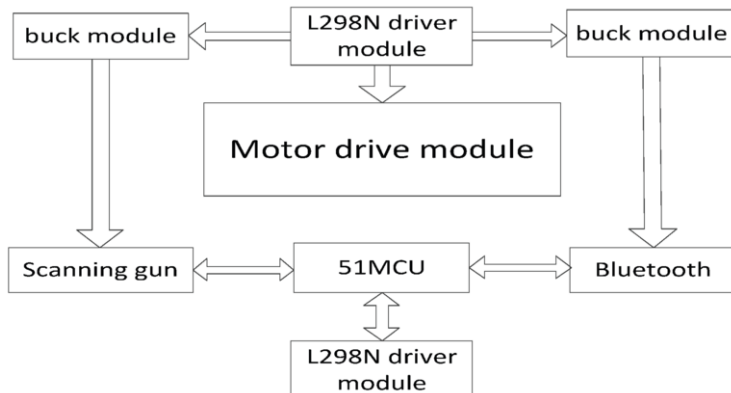
cannot enter the library without the ID card.

The QR code scanning and unlocking module uses an embedded 51 MCU as the main processor. It also integrates QR code scanner, image processor and the lock module. The communication between MCU and computer is established via bluetooth. The image acquired by the scanner is converted from digitized information to ASCII and sent to the MCU by the image processor<sup>[5]</sup>. The lock module receives the MCU commands and the motor drives the lock to unlock. The overall structure diagram is as shown in Figure 8.



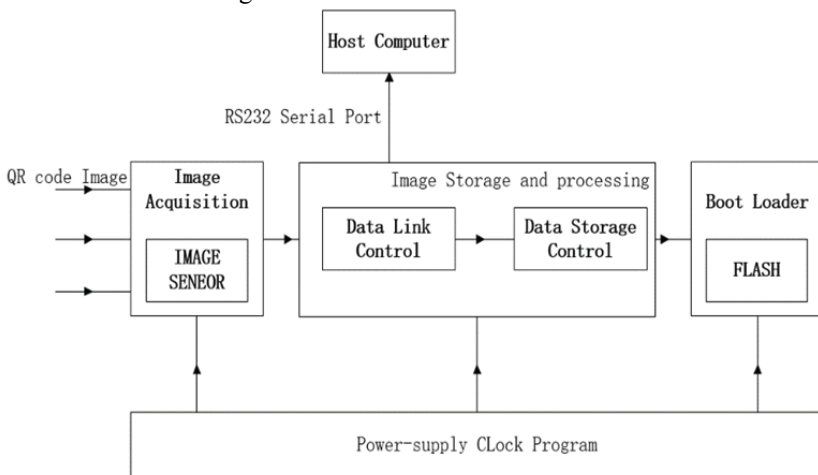
**Figure 8.** Overall structural diagram of QR code scanning and unlocking module. The functions of components are listed as follows:

- 51 MCU minimum system: the core module of the computer bluetooth data storage, instruction judgment, image processor data reception, data comparison and the final motor-driven unlock.
- Image processor: the main function is to convert digitized information into ASCII and send it to MCU serial port.
- Lock module: with a 12V DC (Direct Current) motor, it determines whether the locking operation is continued according to L298N signals; two L298N modules are added as the power-driven module of bluetooth, image processor and 51 MCU; the voltage of the pin is changed through MCU, so as to control the input order from the bluetooth and scanning gun to MCU serial port, and the voltage-driven module is used to control the opening and closing of the lock.



**Figure 9.** Lock body module structure.

- QR code scanner module aims to digitalize the QR code image. Its working areas can be divide to four components, image acquisition, image storage and processing, bootstrap loader and host computer communication. Image acquisition is composed of an image sensor and its peripheral circuit, and used to collect the decoding QR code images. As the core module to decode the QR code, the image storage and processing area consists of Field—Programmable Gate Array (FPGA), First Input First Output (FIFO), Digital Signal Processing (DSP) and Synchronous Dynamic Random Access Memory (SDRAM) and their peripheral circuits. Its main functions include the transmission control, storage, processing and decoding of the images acquired. Bootstrap loader is composed of FLASH and its peripheral circuit to implement the functions of storage system and guarantee the automatic functioning of the system after it is separated from the PC. The PC communication is mainly composed of DSP, FPGA and RS232 chip. According to RS232 communication protocol, a serial port communication control module is written in FPGA to realize the communication between the system board and PC. The schematic diagram of QR code scanner is as shown in Figure 10.



**Figure 10.** QR code scanner schematic diagram.

## Conclusion

This research designed an APP and QR Code based intelligent library manage system based on the precedent library manage system. The high versatility and personalization brought by the APP is successfully combined with the convenience provided by the locking system to offer a better user experience, increase the efficiency of the utilization of the library resource and reduce the malicious behaviors.



## **References**

- [1] Xie Hong, Wang Bingjiang. Development and Application of Library Seat Management System Based on VB. *Library Tribune*, 2010, 30 (5): 58-60.
- [2] Zhang Wanlong. Research on Interaction Design of Networking Products of Cinema Ticketing System. Beijing: North China University of Technology, 2015.
- [3] Cao Zhimei. Research on User Service Mode of Digital Library. *Journal of University libraries*, 2005, (02): 16-19.
- [4] Zhu Wenjing, Xia Cuijuan. Application of QR Code in Library Mobile Service - A Case Study of Shanghai Library. *New Technology of Library and Information Service*, 2012, (Z1): 115-120.
- [5] Chen Qinghua, Qi Yudong, Hong Bei. A Smart Lock Design Based on Fingerprint and QR Code. *Aeronautical Computing Technique*, 2016, (06): 93-95.

# Traffic Incident Severity Detection Model Based on Catastrophe Theory

Hongwei LI<sup>a,1</sup>, Sulan LI<sup>b</sup>, Hongwei ZHU<sup>c</sup>, Xing ZHAO<sup>a</sup>, Xiaoli ZHANG<sup>a</sup>

<sup>a</sup> College of Civil and Transportation Engineering, Hohai University, China;

<sup>b</sup> School of transportation, Wuhan university of technology, China;

<sup>c</sup> Wuhan Transportation Science Research Institute, China

**Abstract.** An expressway traffic incident severity detection model is established in this paper, which can effectively detect the traffic incidents and their severity. By analyzing the relationship between the severity of freeway traffic incidents and the number of lanes, flow, speed and duration of traffic events in the event lane, the following conclusions are drawn: (1) The traffic flow of traffic incident lane and non-traffic incident lane has multi-mode, sudden jump and other abrupt theoretical features When a traffic incident occurs; (2) The severity of traffic accidents controls the degree of sudden and sudden jump of traffic incident lane and non-traffic incident lane. Based on catastrophe theory, this paper proposes a discriminant model of traffic accident severity based on traffic flow velocity and flow rate as independent variables. The measured data of expressway traffic events verifies the effectiveness of the proposed model.

**Keywords.** expressway, catastrophe theory, traffic incident, severity

## 1. Introduction

As an important part of urban traffic, the function of expressway is to reduce the pressure on the city traffic, carry most of the long distance transportation and perfect the urban road network. Such as the road area of Shanghai expressway accounted for 5% of all road downtown area, but the expressway takes 20% of the traffic flow. After outer and inner ring expressway opening, the traffic flow demanded by 10% - 15% in the city center, and the transit traffic demanded by 70<sup>[1]</sup>. Congestion caused by traffic incidents has accounted for 50% to 75% of total traffic of Shanghai expressway<sup>[2]</sup>. How to reduce traffic congestion and deal with traffic accidents are the urgent traffic problem. Traffic Incident Automatic Detection technology (Automatic Incident Detection, AID) can reduce the congestion caused by traffic accident time and reduce casualties and property losses.

---

<sup>1</sup> Hongwei LI, College of Civil and Transportation Engineering, Xikang Road 1#, Nanjing 210098, China; E-mail: lihongwei-2008@163.com.

## 2. Literature review

Wang R proposes an efficient multiple model particle filter (EMMPF) to solve the problems of traffic state estimation and incident detection on a highway in California, which requires significantly less computation time compared to existing multiple model nonlinear filters. The results show the EMMPF is capable of estimating the traffic state and detecting incidents and requires an order of magnitude less computation time compared to existing algorithms, especially when the hybrid system has a large number of rare models<sup>[3]</sup>. Ahmad T. H. use historical data to establish recurrent speed profiles and identifies non-recurrent congestion based on their negative impacts on speeds for expressway. Buffer time is employed to measure TTR. Extra buffer time is defined as the extra delay caused by traffic incidents. A Tobit model is used to identify and quantify factors that affect EBTI using a selected freeway segment in the Southeast Queensland, Australia network<sup>[4]</sup>. Ahmed F. presents a traffic control system that can work standalone to handle various boundary conditions of the recurrent, non-recurrent congestion, transit signal priority and downstream blockage conditions to improve the overall traffic network vehicular productivity and efficiency using field detectors' data. The comparative performance of this control logic is quite satisfactory for some of the most frequently used phase settings in the network with a high number of junctions under highly congested conditions<sup>[5]</sup>. Xinyong ZH presents a freeway network traffic incident situation evaluation model to comprehend and evaluate the severity of traffic incident situations by introducing the invulnerability concept in the complex network theory and perceiving the topological structure of a road network with the help of traffic incident situations<sup>[6]</sup>.

Lun z. presents a naive Bayesian classifier-based algorithm for freeway non-recurrent traffic incident detection to enhance the accuracy and learning ability of intelligent traffic incident detection algorithm using traffic wave theory. the algorithm in freeway traffic incident detection system is of high accuracy and strong robustness even if the traffic volumes increase<sup>[7]</sup>. Shifeng N. improve a traffic incident detection algorithm in urban expressway basing on proposing the definition of lengthways time series of traffic parameters and defining abnormal traffic statuses using gain amplification principle<sup>[8]</sup>. Dewang CH. makes an automatic incident detection (AID) algorithm based on abnormal variations of traffic flows for two-dimensional space and one-dimensional time<sup>[9]</sup>. Yongsheng Q. established a cellular automaton model of highway based on the NaSh model. The paper found that the accident has a great influence on the highway traffic flow and considerably on traffic flow density within a certain range. And the influence of the accident blocking point on traffic flow on lane 1 is less than on lane 2<sup>[10-11]</sup>.

It can be found that existing research has the following characteristics:

- Most of AID technologies detect the incident on the highway.
- Most of AID technologies only detect whether the traffic incidents happen, not can detect the severity of the incident.

Recently, the presence of AID technologies can't satisfy traffic management needs. A traffic incident severity detection technology is urgently needed to make a targeted emergency measure.

### 3. Data collection and process

- Location of survey

The location of survey is the expressway in front of Hohai University. Expressway characteristics are shown in table 1.

**Table 1. Expressway characteristics of survey**

width(m)	number of lanes	traffic facilities
3.5	5 lanes in single direction	central isolation strip no bus station, deceleration zone and signal light

- Types of survey traffic data

Because the microcosmic software-VISSIM has advantage in simulating expressway and urban road network, the study chooses VISSIM to simulate. The calibration of the software is based on the field of investigating statistics. Types of survey traffic data include speed of the vehicle, traffic volume and the proportion of vehicle types.

The speed of the vehicle and the traffic volume is shown in table 2.

**Table 2. The traffic volume and the traffic volume of survey**

vehicle types	light-duty vehicle	oversize vehicle
expectation speed of vehicle (Km/h)	(45,70)	(40,60)
volume(veh/h/lane)	1912	46

- Survey design

Traffic incidents are simulated by the method of signal light. When the traffic incident happens, the signal light is red light, whereas green light. Set up the incident lasts for 15 min. Simulation schemes are shown in table 3.

**Table 3. Simulation schemes**

lanes No.	1	2	3	4	5
1	T	T	T	T	T
2	F	T	T	T	T
3	F	F	T	T	T
4	F	F	F	T	T
5	F	F	F	F	T

### 4. Traffic flow data analysis

It is can be found that the speed and volume of jam lanes and adjacent lanes have great changes when the incident happened. And the speed of jam lanes is different from one lane, two lanes to three lanes. Specific data is shown in table 4.

**Table 4. The Lowest Average Speed of Jam Lanes**

jam Lanes	one lane	two lanes	three lanes
Speed (km/h)	41	32	23

Considering the difference of the speed between the block lane and the adjacent un-block lane, the traffic incident severity detection model is established.

- One lane

Speed has small changes in a short time, and the impact on the traffic capacity is lesser. The impact of changing lane caused by jam is relatively small. The jam has a mainly influence on the speed of adjacent lanes.

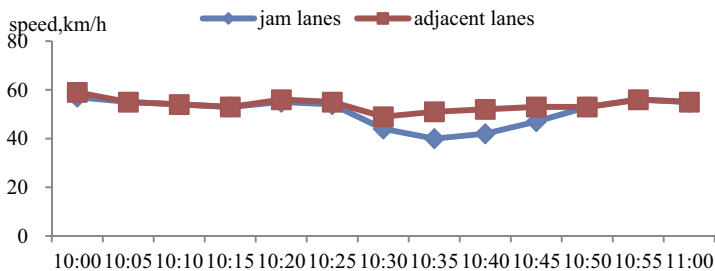


Figure 1. The speed of one lane jam.

Similar to changes in the rate of lane clogging congestion occurs, the cross section of traffic has certain change, but congestion to a lesser degree, has less effect on the whole as well as the adjacent lane. For the comparison of five lanes wide lane, only one lane blockage, slight accidents, effect on the whole is limited, the remaining four lanes will still be able to continue to bear the whole road traffic. For the adjacent lane, lane traffic diversion to the lane is likely to be some influence on the driveway, but only 1 lane of traffic, and share the pressure is not big.

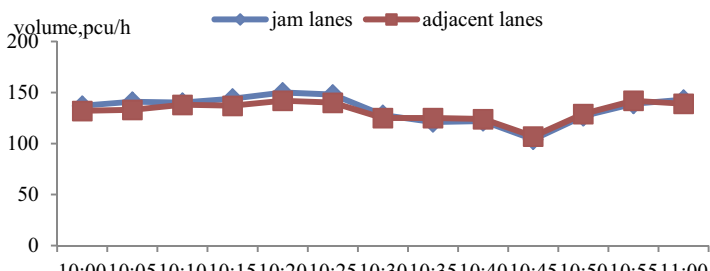


Figure 2. The volume of one lane jam.

- two lane

In the case of two lanes jams, the difference of speed has some changes. There is a rapid change on jam lanes and adjacent lanes when incident happened and speed secondary decline on jam lanes.

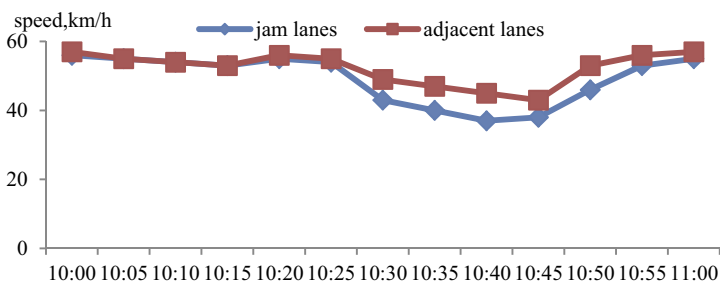


Figure 3. The speed of two lanes jam.

During two-lane jams, the number of vehicles has a certain change, and the impact on adjacent lanes. Jam two-lane, directly affected has two lanes, the two lanes vehicles will try to through the blocked area, sometimes blocking the driveway vehicle will happen many times change, so the blocked section under adjacent lanes of traffic capacity affected degree increases.

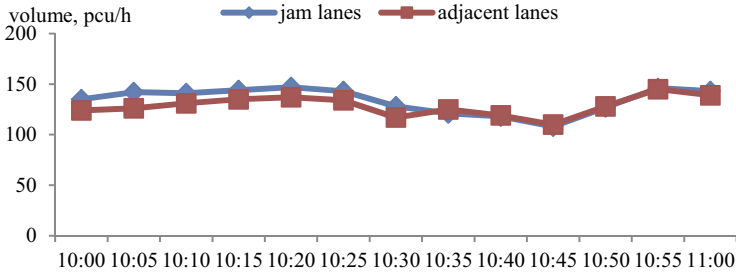


Figure 4. The volume of two lanes jam.

- three lanes

In the case of three lanes were blocked, is linked to a single lane was jam is roughly same, because the three road jam is usually a big accident, so the driver can make a faster response, all is by blocking road cars will immediately slow down.

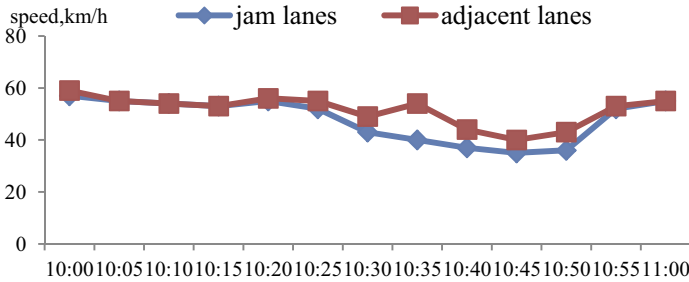


Figure 5. The speed of three lanes jam.

More than three lanes jam, obviously serious blockage accident, vehicle hard by, cross section with very little traffic.

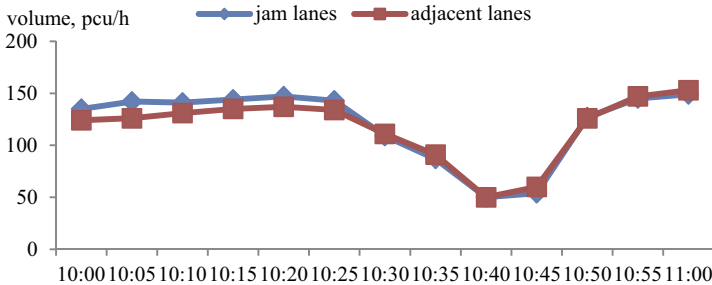


Figure 6. The volume of three lanes jam.

## 5. Traffic Incident Severity Detection Model

### 5.1. Catastrophe theory

Catastrophe theory, which originated with the work of the French mathematician René Thom in the 1960s, and became very popular due to the efforts of Christopher Zeeman in the 1970s, considers the special case where the long-run stable equilibrium can be identified with the minimum of a smooth, well-defined potential function. In mathematics, catastrophe theory is a branch of bifurcation theory in the study of dynamical systems; it is also a particular special case of more general singularity theory in geometry. Small changes in certain parameters of a nonlinear system can cause equilibrium to appear or disappear, or to change from attracting to repel and vice versa, leading to large and sudden changes in the behaviors of the system. However, examined in a larger parameter space, catastrophe theory reveals that such bifurcation points tend to occur as part of well-defined qualitative geometrical structures.

From the preceding analysis, we could conclude that traffic incidents were correlated highly with abnormal traffic status, and we could detect the existence of traffic incidents by identifying the abnormal traffic status. Catastrophe theory provides a means for mathematically modeling the dynamic processes that underlie traffic accidents. The changing trends of traffic parameters when an incident occurs are shown in Figure 7. From 15:05-15:50, the volume decreased sharply and speed also reduced. On the contrast, occupancy increases. Degrees of traffic parameters for sensitivity are volume, speed, and occupancy from high to low when an accident occurs on the urban accident. The results are different with freeway. And the results test that the model could be better described by using the integrated traffic parameters.

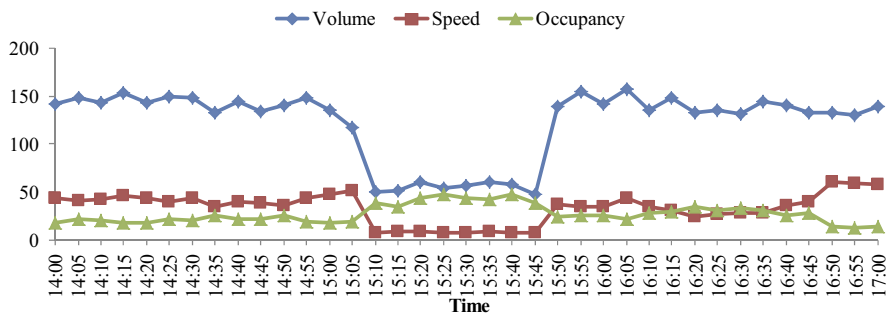


Figure 7. Traffic parameters chance the trend.

Based on the definition of abnormal traffic status, the long-term trend of traffic parameters data represented the normal status of traffic flow and its short-term processing along represented the abnormal status of traffic flow. So long as the actual traffic status on certain moment was compared with the normal traffic status on the same moment, we could determine whether the traffic flow showed the short-term processing along, namely the abnormal status.

### 5.2. Traffic incident severity detection model

In the number of different lane blockage, the speed drop degree is different, also can have very different traffic flow quantity. After multiple comparison results change model is shown in equation (1) and (2).

$$n = k_1 v_0 / v + k_2 q_0 / q \tag{1}$$

$$m = \begin{cases} 1, n \geq 0.7 \\ 2, 0.4 < n < 0.7 \\ 3, n \leq 0.4 \end{cases} \tag{2}$$

Where:  $v_0$  is the minimum speed under incident;  $v$  is the speed of normal traffic flow;  $q_0$  is the volume under incident;  $q$  is the volume of normal traffic flow;  $k_1=k_2=0.5$ ;  $n$  is the amount of jam lanes;  $m$  is the traffic incident severity, “1” means one jam lane and minor accidents, “2” means two jam lanes and ordinary accidents; “3” means three jam lanes and serious accidents.

## 6. Case study

Shanghai expressway is chosen to verify the feasibility and validity of traffic incident severity detection model.

The data is still being tested every five minutes. The speed and volume of one single lane under traffic incident take into the formula (1) and (2) to test the accuracy of the model. Test results are shown in table 5.  $v_0$  is the minimum speed under incident;  $v$  is the speed of normal traffic flow;  $q_0$  is the volume under incident;  $q$  is the volume of normal traffic flow;  $n$  is the amount of jam lanes;  $m$  is the traffic incident severity, “1” means one jam lane and minor accidents, “2” means two jam lanes and ordinary accidents; “3” means three jam lanes and serious accidents.

Table 5. Test Results

Severity of incident	$v$	$v_0$	$q$	$q_0$	$n$	$m$
One lane	52	40	145	117	0.79	1
Two lanes	52	32	145	73	0.56	2
Three lanes and above	52	18	145	52	0.35	3

This result suggests that traffic incident severity detection model has certain reliability and accuracy

## 7. Conclusions

The conclusions are summarized as follows.

- (1) The speed and volume of jam lanes and adjacent lanes have great changes when the incident happened.
- (2) The speed of jam lanes is different from one lane, two lanes to three lanes.



(3) For one lane jams, speed has small changes in a short time, and the impact on the traffic capacity is lesser. The impact of changing lane caused by jam is relatively small. The jam has a mainly influence on the speed of adjacent lanes.

(4) For two lanes jams, the difference of speed has some changes. There is a rapid change on jam lanes and adjacent lanes when incident happened and speed secondary decline on jam lanes.

(5) Three lanes were blocked, is linked to a single lane was jam is roughly same, because the three road jam is usually a big accident, so the driver can make a faster response, all is by blocking road cars will immediately slow down.

(6) The expressway traffic incident severity detection model can effectively detect the traffic incidents and their severity.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 71501061, 51408190 and 51608171), Natural Science Foundation of Jiangsu province (No. BK20150821), and Science and Technology Plan of Hubei Provincial Transport Department (No.2016-13-1-3).

## References

- [1] Yi Y.(2003). "The urban expressway development of our country were reviewed." *Urban Traffic*, 8 : 52-54.
- [2] Beibei J.Y., Xiaoning Z, Lijun S.(2008). "Incident duration prediction grounded on Bayesian decision method based tree algorithm". *Journal of Tongji University(Natural Science)*, 36, 319-324
- [3] Wang R, Shimao F , Daniel B.(2016). "Efficient multiple model particle filtering for joint traffic state estimation and incident detection". *Transportation Research Part C*, 71,521–537
- [4] Ahmad T. H., Luis F, Simon W, Phil C, Ameneh S.(2016). "Reprint of: modelling the impact of traffic incidents on travel time reliability". *Transportation Research Part C*, 70, 86–97
- [5] Ahmed F., Hawas Y.E.(2015). "An integrated real-time traffic signal system for transit signal priority, incident detection and congestion management". *Transportation Research Part C*, 60, 52–76
- [6] Xinyong Z, Shi Na , Haozhe C.(2013). "Traffic incident situation evaluation based on road network reliability of invulnerability". *Journal of Transportation Systems Engineering and Information Technology*, 13, 79–85.
- [7] Lun z., Wenchen Y., Tuo L., Yipin S.(2014). "A naive Bayesian classifier-based algorithm for freeway Traffic incident detection". *Journal of Tongji University(Natural Science)*, 42, 558-563.
- [8] Shifeng N., Guiyan J., Hongwei L., Hui J.(2011). "Automated detection algorithm for traffic incident in urban expressway based on lengthways time series of traffic parameters". *Journal of Harbin Institute of Technology*,43, 144-147.
- [9] Dewang CH.,Yong Y.,Taomei ZH.(2011). "Automatic Incident Detection Algorithm for Urban Expressway Based on 3-D Integration". *China Journal of Highway and Transport*, 24, 94-99.
- [10] Yongsheng Q., Junwei Z., Jiawei D., Yuwen L., Min W., Jun W.(2011). "Cellular automaton traffic flow model considering influence of accedents". *Acta Physica Sinica*, 60, 060505.
- [11] Dongfang M, Dianhai W \*, Yiming B, Sheng J. Recognition of Bottlenecks on Urban Roads Using Queue Detector Data. *KSCE Journal of Civil Engineering*, 2016, Vol.20, No.7, pp. 2955–2964.

# A Method of Constructing Scenario Information Model Based on Power Operation Scene

Zhansheng Hou<sup>a,b,1</sup>, Lin Peng<sup>a,b</sup>, Haiyun Han<sup>a,b</sup>, Min Xu<sup>a,b</sup>,  
Gang Wang<sup>a,b</sup>, He Wang<sup>a,b</sup> and Qiang Zhou<sup>c</sup>

<sup>a</sup> State Grid Smart Grid Research Institute

<sup>b</sup> State Grid Key Laboratory of Information & Network Security

<sup>c</sup> Maintenance Division of State Grid Jiangsu Electric Power Company

**Abstract.** This paper deals with a method and system of scenario information model based on electric power scene. The construction method includes the following steps: (1)Collecting scene information, (2)Scenario information classification, (3)Information model construction, (4) The formation of power operations situational awareness system and scene applications. Based on the scenario information model of electric power operation scene, this paper describes the entity from the six dimensions of human, auxiliary equipment, environment, time, task and service, and finds out the relationship between entity, instance and entity, and constructs information model, To achieve the power of the task of service discovery, interpretation, query and unified scheduling, to provide the necessary services.

**Keywords.** Situational Awareness Information Space Inference Engine Rule Base Case Base

## 1. Introduction

Contextual Perception Computing was first proposed by Schilit in 1994. It is an important research direction of pervasive computing. The goal is to enable the system to adaptively provide the user with the services and information related to the current task of the user according to the change of the situation information. Early research focused on the application of the user's location to detect the user's location, which is the most representative of the two engineers Parc Tab and Active Badge. Scenario Awareness Computing System The well-known research programs in foreign countries are: MIT's CSALL Labs' AIRE program is based on the Agent-based software platform, the prototype system is Intelligent Room; Stanford University's Interactive Workspace program is dedicated to researching how users in rich technical resources Workspace to interact with the prototype iRoom system; Microsoft's Easyliving program focuses on user experience in a smart environment that is full of interactive devices. In the domestic, Smart Classroom, an intelligent classroom built by the Institute of Human-Computer Interaction and Media Integration, Tsinghua University, from the perspective of human-computer interaction and information space, virtual environment modeling and interaction, multi-functional perception and adaptive network information access

---

<sup>1</sup> Corresponding Author State Grid Smart Grid Research Institute; Nanjing China; 210003; E-mail: houzhansheng2003@163.com.

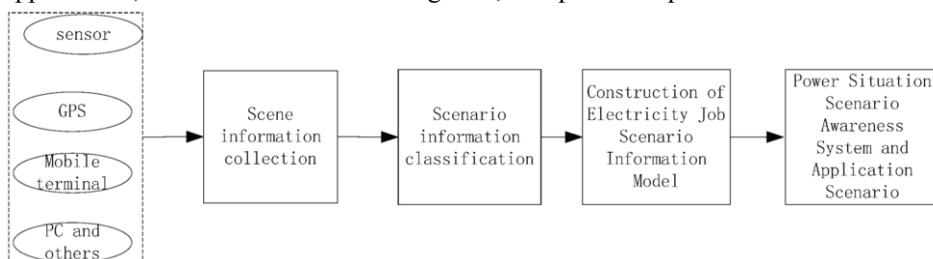
The theory and practice of perceptual computing are discussed in depth and have achieved good results.

In the case of power operation, the types of power equipment are complex and diverse, and the state of the equipment is constantly changing, and the situation of the equipment is very important. There are a wide variety of information needed for power operations, numerous types of field operations and complex relationships, and operational personnel will face massive amounts of data, which presents more new challenges to the information model.

In order to solve the shortcomings of the above-mentioned prior art, this paper provides a method and system of scenario information model based on power operation scene. By constructing information model from six dimensions: human, auxiliary equipment, environment, time, task and service, The physical space and information space logical correspondence, the construction of power scene scenario information model, based on the model to provide users with the necessary resource information and professional services.

## 2. Scenario information model construction method

The scenario information construction method mainly includes scenario information collection, scenario information classification, scenario information model construction, power operation scene perception system and scene application. Through the various sensors, mobile terminals, GPS, PC and other equipment to collect the scene information, the scene Information classification, to build the scene information model, and the formation of power operations situational awareness system and scene applications, the flow chart shown in Figure 1, the specific steps are as follows:



**Figure 1.** Power scene information model construction flow chart

### 2.1. Scenario collection

The purpose of the scene information collection is to obtain the original data as the value of the scene feature, including: through various sensors (infrared sensors, temperature sensors, humidity sensors, RFID sensors, ultrasonic sensors, laser ranging sensors, electromagnetic sensors, gyroscope sensors), GPS, mobile terminals and other access to the environment scenarios, user information and equipment information; through software agents such as PC to collect situational information; user scene input scene information.

## 2.2. Scenario Information Classification

The collected scene information is cleaned by data, the data of the association relation is removed, the filtering and filtering are used to form the effective data layer information, and the power scene information is divided according to the source and the category from six dimensions.

(1) User entity, including the user itself and the user-related information, including the job, the organization (post), skills.

(2) Time entity, according to the situation, time scenario information description of the dimension is different, can be time period or time point, time strategy.

(3) Auxiliary Equipment entity, mainly refers to intelligent interactive terminals (such as smart helmet, smart glasses, smart bracelet), instrumentation, industrial equipment, spare parts and so on.

(4) Environment entity, including physical environment (such as weather, humidity, temperature, light, pollution, noise level, etc.) and geographical environment (such as geographical map, substation 3D map).

(5) Task entity, task time, location, status, etc., including job tasks, task steps, power outage plan, scheduling commands, multi-task management, installation requirements.

(6) Service entity, including the rule base, reasoning engine, case library and service interface, etc .

## 2.3. Information model construction

Based on the ontology method, the context information model COIS (Context Ontology for Information Space) is established to describe the various entities involved in the power scene environment and express it in W3C international standard OWL language. In the COIS model, select User, Task, Equipment, Environment, Time, Service to build the information model.

### (1) User entity

The user-related context information is User Profile, Preference, Role, etc., as shown in Figure 2. User Profile records the user's own basic information, such as name, gender, age, contact information, work department, home address, Email and so on. Preference records the user's personal preferences and work skills, such as good skills, work habits and so on. Role is the user identity information, for example, in the power scene information space, the user role is divided into maintenance personnel, inspectors, experts, other personnel. Different roles of users with different access rights, maintenance personnel can be arranged according to job tasks, the use of auxiliary equipment for electrical equipment maintenance. Inspectors can only view the equipment, experts on the maintenance and inspection work for remote video guidance.

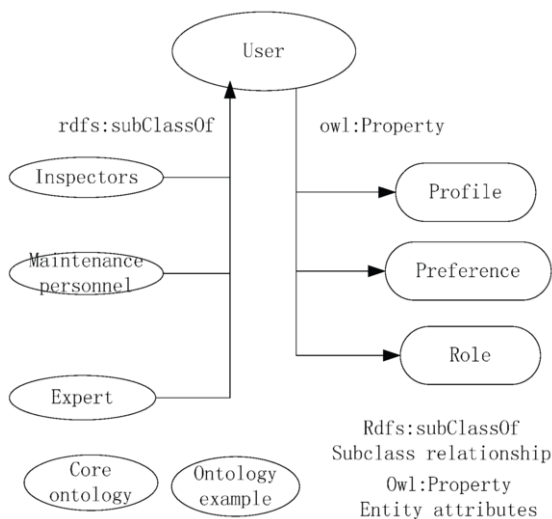


Figure 2. User entity diagram

(2) Task entity

The job task information is the main line in the power job scenario space. The task entity is used to describe the status, location, and time of the job task, as shown in Figure 3. Work task entities can be divided into two main categories: Plan Task and Reasoning Task.

Among them, the scheduled task is generally scheduled to arrange the inspection and maintenance tasks, the task attributes (such as operators, places, start and stop time, etc.) can be pre-set. Another kind of reasoning task is derived from the reasoning of situational information, with some uncertainty.

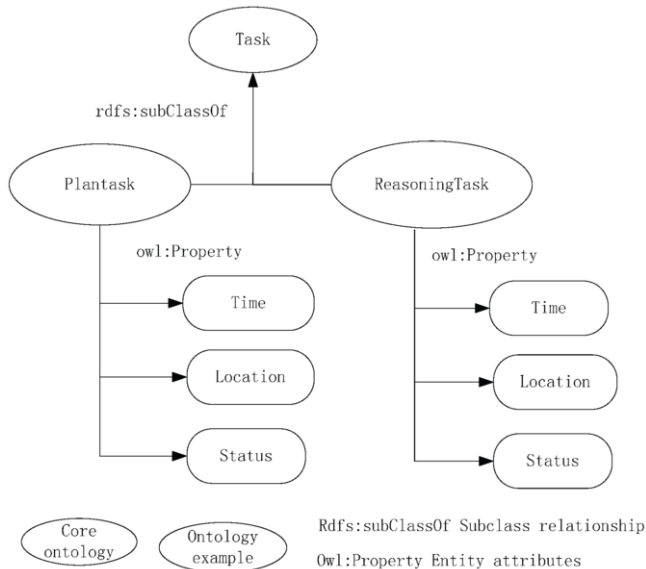


Figure 3. Task entity diagram

(3) Environmental entity

Electricity workers use equipment to interact with the surrounding environment or other users around them, and the surrounding environment can provide useful information to assist users in making decisions. From the point of view of the operator as the center, the environmental information related to the activity or mobility of the operator is considered to be the available context information.

We divide the environment entity into two subclasses: Location, Physical Environment, as shown in Figure 4. In the electric power scene, the geographic location is divided into the geographical map and the 3D map of the substation, which represent the outdoor space position and the outdoor space position respectively. The Physical Environment includes information such as temperature, humidity, light, noise, contamination, and other external environment.

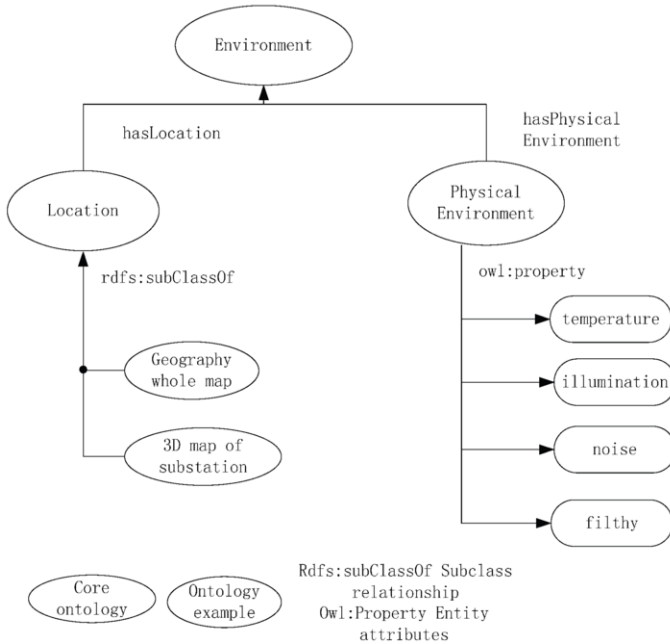


Figure 4.Environmental Entity Chart

(5) Auxiliary Equipment entity

Operators use auxiliary equipment to complete the task, auxiliary equipment examples include intelligent helmets, smart glasses, instrumentation, industrial equipment, spare parts, etc., as shown in Figure 5.

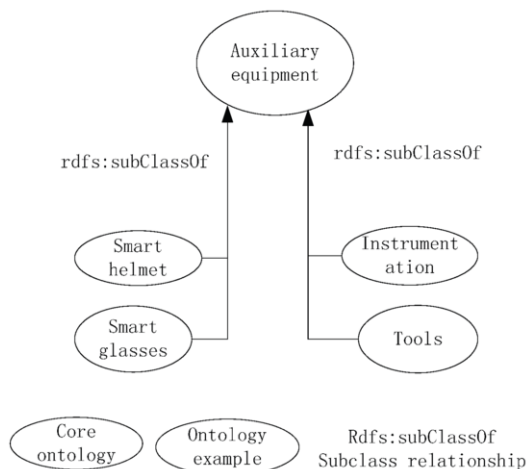


Figure 5. Auxiliary device entity diagram

(6) Service entity

The service entity provides services for the power job scenario model, which includes the rule base, the reasoning engine, the case library, and so on, as shown in Figure 6. When the task is executed, the case is obtained from the case library according to the rules of the rule base. If the case is not the case, it is implemented by the inference engine. Service interface through the event monitoring interface, service query interface, scenario service interpreter, to achieve service discovery, interpretation, query and unified scheduling, external services required.

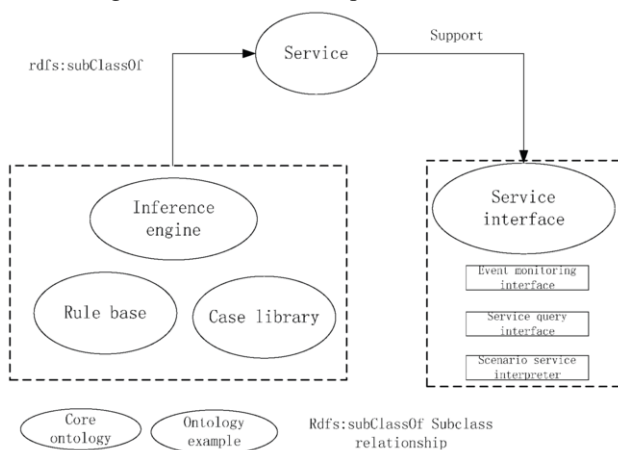


Figure 6. Service Entity Chart

In the scenario scene of the power scene, the relationship between the core entities is shown in Figure 7. After the user enters the power job scenario space, it exists in the information space environment. At a time, the user participates in the current job assignment task and uses Information space in a variety of services, users and auxiliary equipment for information exchange, complete the power equipment inspection and maintenance tasks.

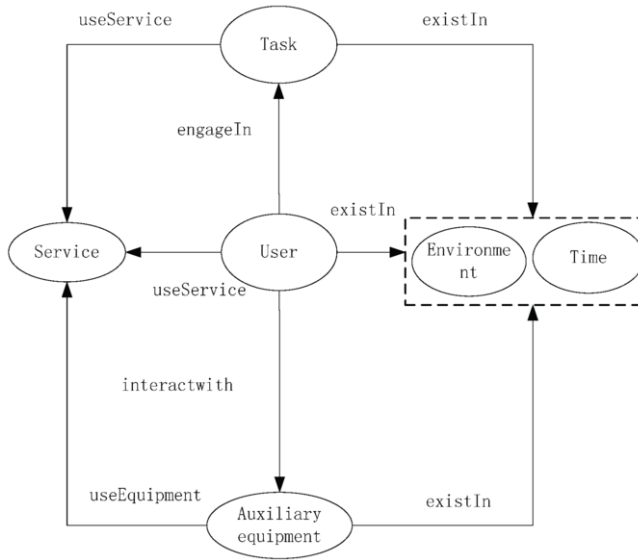


Figure 7. User entity diagram

**3. Electricity situational awareness system prototype and application**

Scenario-aware system diagram is shown in Figure. 8. Through the above steps, it can be seen that the scenario based on the scenario information model, the scenario rule base, the reasoning engine, the service case library and the information service module, Mission services to find, explain, query and unified scheduling, to provide the necessary services. The scene-based perception system based on the electric power operation scene comprises a collection scene information module, a scene information model, a reasoning and decision module and an information service module which are connected in turn.

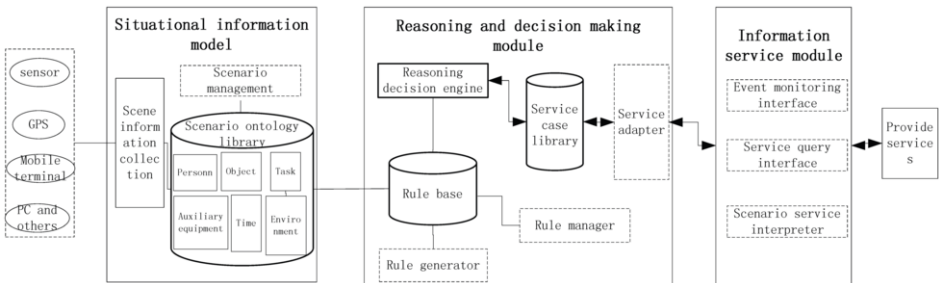


Figure 8 . scenario-aware system diagram

In the application scenario of electric power operation, users in the substation, transmission line, distribution station, information communication room through the smart phone, smart glasses, smart helmet, according to the GPS, GIS map to achieve outdoor navigation, indoor according to the indoor map and operational tasks , Automatic display of nearby equipment running status, automatic display of operating



information, according to set the operating route, automatic navigation, for off-line, automatic alarm, can achieve the path planning and remote assistance.

#### 4. Conclusion

Based on the scenario information model of power operation scene, this paper constructs the information model from six dimensions: human, auxiliary equipment, environment, time, task and service, realizes the physical space and information space logical correspondence, constructs the power scene information model, In the model based on the user to provide the necessary resource information and professional services.

#### References

- [1] Pu Hai tao. Research on Key Technologies of Intelligent Interaction Based on Context Awareness in Internet of Things, Shandong University of Science and Technology, 2011.
- [2] Qin Wei-jun. Ontology - based intelligent spatial scene information model research, Tsinghua University, 2005.
- [3] Xu Bing zhao. A scenario-driven context-aware computing framework, Computer Science, 2012.39 (3): 216-221.
- [4] Qin Huai feng, Context Sensitive Computing for Sensitive Networks, Xi'an, Shaanxi: Northwestern Poly technical Univrsity, 2006.
- [5] Hong, J.I., etal. : An Infrastructure Approach to Context -Aware Computing. HCI Journal,2001,Vol.16.
- [6] Kindberg T. Fox A. System software for ubiquitous computing. IEEE PervasiveComputing,2002 1(1):70-81.
- [7] Erickson T. Some problems with the notation of context-aware computing. Communications of the ACM, 2004.102-104.
- [8] Erickson T. Some problems with the notation of context-aware computing. Communications of the ACM, 2004. 102-104.

# Research on Location Planning of Urban Rail Transit Emergency Facility Considering Vulnerability Node

Lei Zhu and Ning Zhang

**Abstract** A reasonable location planning of emergency facility is an important work in the emergency management of urban rail transit. Based on the complex network theory, we propose a reliable emergency facility location model that aiming to improve the coverage level for the nodes with high vulnerability. This is mainly due to that the nodes with the different vulnerabilities in the network have different requirements for coverage level. We research on the node vulnerability of urban rail transit using complex network theory, and identifies the key nodes in the network before location planning. The vulnerable nodes in the network will be given higher weight to improve their coverage level in the location model, so that the equilibrium and efficiency of emergency are enhanced. To efficiently solve the model, we develop a heuristic based on Lagrangian relaxation. Finally, the model and heuristic are applied to the emergency facility location in urban rail transit (URT) network in Nanjing, China. The proposed method is effective and robust, and contributes greatly to the performance of emergency facility location planning in emergency management of URT.

**Keywords** urban rail transit · complex network · location · vulnerability

## 1. Introduction

Urban rail transit (URT) plays a critical role in the continuous development of modern cities with its multiple advantages of large capacity, high speed, punctuality, low energy consumption. As of December 31, 2016, URT lines amounting to a total length of 4152.8 km were opened to traffic in total in Mainland China [1]. Together with the gradual formation of URT network operation pattern and dense population of the cities, many Chinese URT lines have reached the designed long-term maximum passenger flow volume [2]. Due to the heavily overcrowded population and the situation of underground space, there exist a lot of potential risks during the operation of the metro station. Rescue team in URT companies generally seek the emergency supplies from a nearby facility so as to reduce transportation time and/or cost. Therefore, it will be essential to study location planning of URT emergency rescue station as an important decision-support measure for emergency management.

Many previous studies have considered different factors of location planning, however, which neglect the topological structure and working state of demand points in

---

Lei Zhu  
Intelligent Transportation System Research Center  
Southeast University, Nanjing, Jiangsu  
e-mail: cumtzhu@163.com

Ning Zhang (✉)  
Intelligent Transportation System Research Center  
Southeast University, Nanjing, Jiangsu  
e-mail: seuurt@126.com

the URT network [3][4][5]. To bridge these gaps, we propose a reliable location planning framework for URT emergency rescue station that integrates vulnerability of the node. In this paper, we consider that the vulnerable nodes in the network are given higher weight to improve their coverage level in the location model, so that the equilibrium and efficiency of emergency are enhanced.

The remainder of the paper is organized as follows: in next section we present the development of the model. In the third section we describe the model established in this paper. In Section 4 approximate algorithmic approaches are discussed. Then, a real application to the location of emergency rescue stations in URT is presented. Last section is devoted to the conclusion.

## 2. Development of the model

### 2.1. Topological property indicators of complex networks

Let  $G=(V,E,A)$  represent an undirected graph where  $V=\{1,2,\dots,N\}$  is the set of the nodes and  $E=\{e_{ij}\}$  is the set of the edges.  $A=(a_{ij})_{N\times N}$  represents the adjacent matrix of the graph with  $a_{ij}$  equal to 1 if there is an edge joining node  $i$  to node  $j$  and to 0 otherwise[6][7]. The degree of a node is the number of edges the node is connected to, and is defined as  $d_i = \sum_{i \neq j} a_{ij}$ . The betweenness of the node  $i$  is defined as the ratio of the number of all shortest paths passing the node  $i$  to the number of all the shortest paths, which is defined as  $b_i = \sum_{i,j \in V(j \neq k)} \frac{n_{jk}(i)}{n_{jk}}$ . The compactness of node  $i$  describes the

difficulty of nodes to reach other nodes in the network, which is defined as  $c_i = \frac{N-1}{\sum_{j=1}^N d_{ij}}$ .

Where the entry  $d_{ij}$  is the length of the geodesic from node  $i$  to node  $j$ . The eigenvector of node  $i$  is used to analyze the indirect influence of the node which is obtained by connecting to a height value node. Assuming  $\lambda$  is the principal eigenvalue of adjacency matrix  $A$ ,  $e = (e_1, e_2, \dots, e_N)$  is the eigenvector of matrix  $A$  corresponding to  $\lambda$ , eigenvector is denoted as  $e_i = \lambda^{-1} \sum_{j=1}^N a_{ij} e_j$ .

### 2.2. Vulnerability analysis

We propose a concept that the node vulnerability is a descent characteristic that reflects the performance of the URT network after some stations are attacked [8][9]. Besides, this section focus on vulnerability analysis of URT systems under node attack strategies with different interface design criteria [10][11]. First, an indicator the so-called global efficiency is selected to represent the network efficiency under different attack strategies, which is defined as:

$$E_{glob}(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \tag{1}$$

Where  $0 < E_{glob}(G) < 1$ , the large  $E_{glob}(G)$  is, the higher the connectivity reliability of network.

We test the impact of node failures on the whole URT system, and the responses of the system under different disturbance vectors are considered. The first vector is random failure, in which nodes are removed by random selection, with an equal failure probability for each node. The second vector is deliberate attacks. Here, we consider four types of deliberate attack: degree-based attack, betweenness-based attack, compactness-based attack and eigenvector-based attack. For degree-based attack, the nodes are removed in descending order of degree in the network. If some nodes happen to have the same degree, we randomly choose one of them. Note that we periodically recalculate the degree after each removal, and select the nodes with highest degree next. The degree of a node is a good indicator of its topological importance. This strategy represents a deliberate and intelligent attack in which the attacker chooses to disable nodes with a large number of neighboring components. For betweenness-based attack, disruption scenarios deliberately target the largest betweenness nodes. This vector is used to approximate an attack on high-betweenness nodes, and it has been reported to result in disproportionately large failures. Meanwhile, after each selection the betweenness will be recalculated and the nodes with the highest betweenness will be deleted next. For compactness-based attack and feature vector-based attack, there is the same attack strategy as above.

### 3. Model formulation

#### 3.1. Assumption and notation

In order to simplify the actual factors, we propose the following assumptions to the construction of the model. First, we assume that all the nodes in the network are candidates to the location of facilities, as well as nodes containing demand. Second, rescue stations can only be located at subway station, and only one rescue station can be established at the alternative point. Finally, the model is established for location planning without considering the existing facilities.

Let

$X$  : be the set of nodes;

$V$  : be the set of candidate nodes;

$x_i$  : be the demand node  $i$ ,  $x_i \in X$ ;

$v_j$  : be the candidate node  $j$ ,  $v_j \in V$ ;

$\omega_i$  : be the weight associated with the demand node  $i$ ;

$a_{ij}$  : be the coverage level from emergency facility  $j$  to demand node  $i$ ;

$$x_j = \begin{cases} 1 & \text{if a emergency facility is located at } j \\ 0 & \text{otherwise} \end{cases};$$

$$y_{ij} = \begin{cases} 1 & \text{if node } i \text{ is covered by candidate node } j \\ 0 & \text{otherwise} \end{cases}$$

### 3.2. Formulation

The maximal cover location problem (MCLP) has proved to be one of the most useful facility location models from both theoretical, and practical points of view [12][13]. The objective of MCLP is to establish a set of  $m$  facilities so as to maximize the total weight of “covered” nodes. One of the key assumptions of the MCLP is that coverage is binary, that is, a certain demand point location is either fully covered (if there is a facility within distance  $r$  from demand point’s location), or not covered at all. However, in many applications—particularly the ones connected to the location of facilities—such binary coverage assumption may be unrealistic.

In this paper, we assume that a demand point at node  $i$  can be covered at different levels of coverage, depending on the distance from  $i$  to the closest facility location. Moreover, we assume that the coverage level decreases as a “step function” of the distance to the closest facility. Thus, for each node  $i \in N$  we define  $k$  coverage radii  $r_i^0 = 0 < r_i^1 < \dots < r_i^k = \infty$  with associate coverage levels  $a_i^1 = 1 > a_i^2 > \dots > a_i^k \geq 0$ . Otherwise, we introduces a positive weight  $\omega_i = w_1d_i + w_2b_i + w_3c_i + w_4e_i$  ( $w_1, w_2, w_3, w_4$  are the weights of each indicator.) which can represent the vulnerability of the node in the network. We refer to this problem as the generalized maximal cover location problem (GMCLP). The GMCLP can now be written as:

Problem P1:

$$Max Z = \sum_{i \in I} \sum_{j \in J} \omega_i a_{ij} y_{ij} \tag{2}$$

S. t.

$$y_{ij} - x_j \leq 0 \tag{3}$$

$$\sum_{j \in J} x_j = N \tag{4}$$

$$\sum_{j \in J} y_{ij} = 1 (i \in I) \tag{5}$$

$$x_j \in \{0,1\} (j \in J); y_{ij} \in \{0,1\} (i \in I, j \in J) \tag{6}$$

The objective  $Z$  maximizes the total weight of the demand points covered. Constraint (2) indicates that the demand point  $i$  may be covered by the candidate point  $j$  only when the candidate point  $j$  is selected. Constraint (3) specifies the number of facilities to be located. Constraints (4) defines that demand area  $i$  may be assigned to only one potential facility node  $j$ . Constraint (5) defines that both  $x_j$  and  $y_{ij}$  are binary variables.

In this paper, the coverage level is represented by response time, and is represented by a segmentation function, which extends the binary coverage to multiple coverage.

$$a_{ij} = \begin{cases} 1 & t \leq t_0 \\ 1 - \frac{t-t_0}{t_n-t} & t_0 < t < t_n \\ 0 & t \geq t_n \end{cases} \tag{6}$$

Where  $t$  is the travel time from the rescue station to the demand point,  $t_0$  is the lower limit of the response time,  $t_n$  is the upper limit of the response time.

### 4. Solution approach

The heuristic algorithm proposed in this section is based on the Lagrangian relaxation with using the Lagrange multipliers to relax the original problem's constraint which is difficult to deal with [13][15]. Let the Lagrange multiplier be  $\lambda_i$ , for the problem P1, the Lagrangian problem is:

$$\text{Problem P2: } \text{Max } Z_1 = \sum_{i \in I} \sum_{j \in J} \omega_i a_{ij} y_{ij} + \sum_{i \in I} \lambda_i \left( 1 - \sum_{j \in J} y_{ij} \right) = \sum_{i \in I} \sum_{j \in J} (\omega_i a_{ij} - \lambda_i) y_{ij} + \sum_{i \in I} \lambda_i$$

$$\text{s. t. } y_{ij} - x_j \leq 0 \tag{7}$$

$$\sum_{j \in J} x_j = N \tag{8}$$

$$\lambda_i \geq 0 \tag{9}$$

$$x_j \in \{0, 1\} (j \in J); y_{ij} \in \{0, 1\} (i \in I, j \in J) \tag{10}$$

Obviously, in order to maximize the objection function of Problem P2, If and only if  $\omega_i a_{ij} - \lambda_i > 0$ , Let  $y_{ij} = 1$ , otherwise  $y_{ij} = 0$  Considering constraints (2),  $y_{ij}$  can be further defined as:

$$y_{ij} = \begin{cases} 1 & \text{If } x_j = 1 \text{ and } \omega_i a_{ij} - \lambda_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Let  $Q_j = \sum_{i \in I} \max(0, \omega_i a_{ij} - \lambda_i)$ , the Problem P2 is simplified into the Problem P3:

$$\text{Problem P3: } \text{Max } Z_3 = \sum_{j \in J} Q_j x_j$$

$$\text{s. t. } \sum_{j \in J} x_j = N \tag{12}$$

$$x_j \in \{0, 1\} (j \in J); y_{ij} \in \{0, 1\} (i \in I, j \in J) \tag{13}$$

We can get the optimal value of  $Z_3$  by finding the  $P$  largest values of  $Q_j$ , and let  $x_j = 1$ , the rest is 0, the optimal solution of Problem P3 can be directly calculated. The value of  $y_{ij}$  can be calculated by formula (11), thus the optimal solution and the optimal value of Problem P2 are obtained. For any Lagrange multiplier  $\lambda_i \geq 0$ , the optimal value of the problem P2 is the upper bound of the problem P1, denoted as  $Z_U$ . In addition, we define  $J^*$  as the subscript set of  $x_j = 1$  in the optimal solution of problem P2, and  $J^* = \{j | x_j = 1\}$ .

Since the optimal solution of the Problem P2 is obtained without considering constraint (4), so it is not the optimal solution of the Problem P1, even may not be a feasible solution of Problem P1. For any demand point  $i$ , it is possible to find a facility with the highest coverage level among the optimal site location determined by Problem P2. Hence, we define:

$$\varphi_i = \max_{j \in J^*} \{ \omega_i a_{ij} \}, \forall i \in I \tag{14}$$

The lower bound of the optimal value to Problem P1 is the sum of all  $\varphi_i$ , denoted as  $Z_L$ ,  $Z_L = \sum_{i \in I} \varphi_i$ . Let the value of  $y_{ij}$  which satisfies formula (18) be 1 and the rest is 0, thus we can get the feasible solution of the original Problem P1 with a known  $\lambda_i$ .

According to the above analysis, we can get  $Z_U$  and  $Z_L$  of the optimal value to the Problem P1 with a known  $\lambda_i$ . In order to approach the optimal solution exactly, we can continuously adjust the value of  $Z_U$  and  $Z_L$ . The approximate optimal solution can be obtained using sub-gradient optimization. Let the superscript  $k$  represent the number of iterations, the upper and lower bounds of the optimal solutions for the  $k$ th iteration are denoted as  $Z_U^k$  and  $Z_L^k$  respectively. Correspondingly, the Lagrange multiplier is  $\lambda_i^k$  and the optimal solution of problem P2 is  $x_j^k$  and  $y_{ij}^k$  for the  $k$ th iteration. Let  $\alpha^k$  represent the step size of the  $k$ th iteration, and  $LB$  is the largest lower bound of the  $k$ th iteration,  $UB$  is the minimum upper bound of the  $k$ th iteration.

Then, the Lagrange multiplier in the  $k$ th iteration can be obtained by:

$$\lambda_i^{k+1} = \max \left\{ 0, \lambda_i^k - \left( 1 - \sum_{j \in J} y_{ij}^k \right) \alpha^k \right\}, \forall i \in I \tag{15}$$

A step-by-step description of the heuristic is as follows:

1. Initialize parameters.  $k = 1, UB = +\infty, LB = -\infty, \lambda_i^1, \alpha^k = \frac{Z_U^k - Z_L^k}{\sum_{i \in I} (1 - \sum_{j \in J} y_{ij}^k)^2}$ .
2. Calculate  $\omega_i a_{ij}$  ( $\forall i \in I, j \in J$ ) according to the time matrix  $T = (t_{ij})_{m \times n}$ .
3. Solve the Lagrangian relaxation problem.
4. Find the feasible solution of Problem P1 by formula (14) and calculate  $Z_L^k$ .
5. Update the upper bound and lower bound of the optimal value.  $UB = \min(UB, Z_U^k), LB = \max(LB, Z_L^k)$ .
6. Update step size  $\alpha^k$ .
7. Update Lagrange multiplier  $\lambda_i^k$ .
8. (1)  $UB - LB \leq 0.3$  (2)  $\sum_{j \in J} y_{ij}^k = 1$  (3)  $\alpha^k \leq 0.0001$  (4)  $k = 400$ . If either of the four conditions is satisfied, stop, take the result to the user; otherwise, transferred to the next step.
9. Update iterations.  $k = k + 1$ , go to Step3 and restart the procedure.

### 5. Model implementation

We applied the proposed model and solution method to a real-world case study. Let the  $G = (V, E, A)$  represent an annotated and undirected URT network graph. Till 2016, Nanjing has 6 subway lines, with a total length more than 225 kilometers of track transport network. The URT network representation consists of 113 nodes (including 7 transfer stations) and 115 directed links. The analysis procedures are coded in MATLAB R2016a and run on a notebook computer with 2.53 GHz CPU and 4.00 GB memory. Table 1 shows calculation results after normalization of each indicator to a subset of nodes.

**Table 1.** Calculation Results of Property Indicators

Node	Degree	Betweenness	Compactness	Eigenvector
V1	0.2509	0.1491	0.7780	0.0023
V2	0.7511	0.2196	0.7854	0.0047

Node	Degree	Betweenness	Compactness	Eigenvector
V3	0.2509	0.2148	0.7920	0.0045
V4	0.2509	0.2267	0.7994	0.0065
V5	0.5010	0.3135	0.8078	0.0119
V6	0.2509	0.2154	0.8089	0.0130
V7	0.2509	0.2172	0.8103	0.0208
V8	0.2509	0.2193	0.8119	0.0393
V9	0.7511	0.2977	0.8138	0.0777
V10	0.2509	0.1038	0.8077	0.0023

The normalized property indicators shown in 0 reveals that the identification results of different indicators to the vulnerability nodes are different. The effectiveness of each index is different in the process of vulnerability nodes identification. We analyze the effective degree of each index by comparing the relationship between the number of failed nodes and the global efficiency of the network under different attack strategies. 0 illustrates global efficiency of the URT network in dependence of the fraction of removed nodes under five types of attack strategies. The abscissa indicates the number of failure nodes, and the ordinate indicates the global efficiency of Nanjing URT network. Different interface design criteria based on features such as degree, betweenness, compactness, eigenvector are considered.

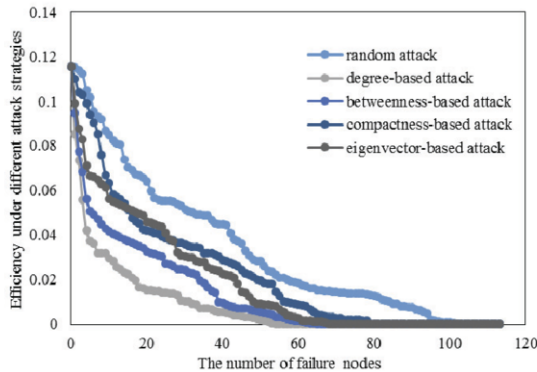


Figure1. The global efficiency under different attack strategies

The global network efficiency shown in Figure 1 reveals that the URT network suffers large global efficiency declines as the fraction of nodes removal increases. Meanwhile random removal of edges will cause less damages to the network than other attack strategies. This is mainly because when these nodes with high topological property are attacked, other elements will likely to be assigned more loads, which may exceed their maximum capacity and cause more performance losses. For different strategies, efficiency variation for the high degree-based attack is significantly faster than other four kinds of attack strategies. Comparing with the efficiency variation of compactness-based attack and eigenvector-based attack have some differences, the compactness indicator is difficult to identify the nodes with lower vulnerability than the eigenvector indicator, but the eigenvector indicator index has a better ability to identify the less vulnerable nodes. In conclusion, degree indicator is most effective, followed by betweenness, compactness and eigenvector. Therefore, according to the analytic hierarchy process the comprehensive vulnerability index is calculated as:

$$w_i = 0.3165d_i + 0.2891b_i + 0.1741c_i + 0.2203e_i \tag{16}$$

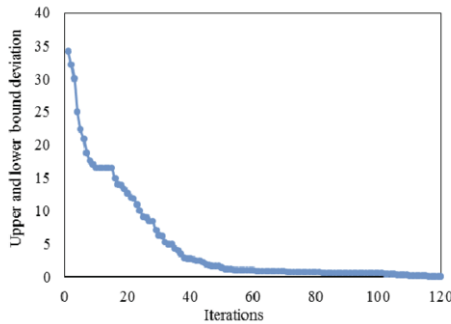


Hence, the comprehensive vulnerability indicator of each node is obtained by formula (16), and it is classified as grade I (severe-vulnerability), grade II (moderate-vulnerability), grade III (mild-vulnerability) and gradeIV(non-vulnerability), as shown in Table 2. With this method we are able to identify the critical nodes of the network that are crucial to the functioning of the URT system, these nodes need to be considered as the first concern for the protection purpose in selecting locations for rescue station.

**Table 2.** Classification of node vulnerability

Vulnerability level	Vulnerability	Node
Grade I	0.45-0.60	V9,V14, V26,V44, V73
Grade II	0.30-0.45	V2,V5, V93,V94, V95,V96, V97, V98,V99, V100
GradeIII	0.15-0.30	Others
GradeIV	0-0.15	V16,V17,V40,V55, V64,V65, V72, V88,V89, V113

To test the efficiency of the GMCLP (presented above), and the heuristic, we conduct a series of computational experiments in the Nanjing URT network. In our experiments, we generate a large number of instances of GMCLP by varying the key problem parameter such as the number of facilities. The number of facilities to be located is between 4 and 20, the heuristics are programmed in MATLAB, and we can conclude that the computational performance of the heuristic was quite satisfying for the model established in this paper, since all problems obtained the optimal solution. Upper bound and lower bound deviation of only 2 out of 17 problems have exceeded 3%. In addition, the time used by the heuristic is almost insignificant. All problems are solved in less than 1 second of CPU time.



**Figure 2.** Astringency of the Lagrange heuristic (m=113, p=20)

Figure 2 illustrates the astringency of the heuristic when the number of facilities is 20, and the heuristic has a quick convergence when iterations is less than 10. When the iteration times are greater than 50, upper bound and lower bound deviation of the problem is less than 3%, which shows that the Lagrange heuristic has a good astringency in solving the model.

If six rescue stations can be selected in the network, the facilities are located in Yuantong Station, Nanjing railway station, Daxinggong station, Nanjing south railway station, Hedingqiao, Taifenglu. Additionally, we set another scenario that the vulnerability of the nodes are not considered. Thus, the vulnerability weights of each nodes in this scenario are a same value which is less than 1, and the coverage level of each node are represented by the reciprocal of the distance from the demand node to the rescue station. With the same method, the facilities are located in NanJing South

Railway Station, Xiangyulunan, Lvboyuan, Gulou, Mingguogong, Dachang, and the facility locations in both scenarios are shown in Figure 3.

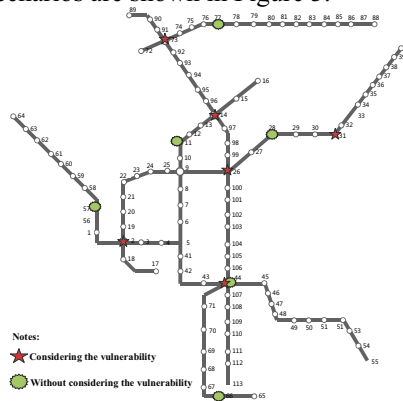


Figure 3. The rescue station locations in both cases

The following conclusions can be drawn from comparing our experimental results:

Without considering the vulnerability of nodes in the network we can get the six most critical nodes for establishing rescue stations are displayed in the Figure 3. We find that the rescue facilities are inclined to be evenly distributed throughout the network and cover all other nodes in the network. This is mainly due to the functional characteristic of the GMCLP which aim to ensure all nodes in the network are covered. However, the vulnerable nodes in the network can not be the targets to protect for normal operation, these nodes without adequate coverage level from the rescue facilities and timely rescue may lead to massive loss in the face of serious emergencies.

There is a completely difference for considering the vulnerability node in the network: location results are almost different, there is just 1 result in this case to be coincident. These nodes are the targets to protect for normal operation. Although the method illustrated above just considers a global efficiency decline, these nodes can help to better protect the network. To ensure the reliability for the network, these nodes were conceived with high coverage level. In this analysis the result of location, we can find that location facilities were all located on vulnerable nodes, this is due to the break of the vulnerable nodes is easy to impact the entire network.

## 6. Conclusions and future research

This paper studied the location planning of URT emergency rescue station where the theory and results of complex network research were introduced. Based on the topological property of the network, degree, betweenness, compactness and eigenvector are selected to measure the importance of nodes. We analyzed the effective degree of each indicators under different attack strategies, and calculate the comprehensive vulnerability index of each node. The results reveal that the method could find the critical node, these nodes need to be considered as the first concern for the protection purpose. Then, we developed a reliable location model that endowed each node in the network with the suitable weight to embody difference protection for different vulnerability nodes. The objective of the model is to maximize the total weight of the covered nodes. Finally, the model was applied to the location of URT emergency

rescue stations in the city of Nanjing, China. Our computational experiments showed that the GMCLP and the Lagrangian relaxation heuristic presented in this paper are quite useful in a number of facility location applications.

Future research can be conducted in a few directions. Several assumptions in this paper may not always be realistic. It will be interesting to investigate possible ways to relax these assumptions. We calculate the node vulnerability mainly using static topological property indicators, but we do recognize that static indicators and dynamic indicators such as personal factors, management factors and environmental factor may be present in reality. It will be worthwhile in the future to explore ways to measure dynamic indicators with each nodes.

## References

- [1] Statistical Analysis Report on Urban Rail Transit in 2016. Message of China Association of Metros. Beijing: China Association of Metros.
- [2] C. Shi, M. Zhong, X. Nong, L. He, J. Shi and G. Feng(2012) "Modeling and safety strategy of passenger evacuation in a metro station in China, *Safety Science*, pp. 1319-1332.
- [3] A.M. Caunhye, X.F. Nie and S. Pokharel(2012) Optimization models in emergency logistics: A literature review, *Socio-Economic Planning Sciences*, vol. 46, no. 1, pp. 4-13.
- [4] H.Z. Jia , O. Fernando and D. Maged(2007) A modeling framework for facility location of medical services for large-scale emergencies, *IIE Transactions*, vol. 39, no. 1, pp. 41-55.
- [5] Y.Y. Zhao, Y.R. An and Q. Ai(2014) Research on size and location of distributed generation with vulnerable node identification in the active distribution network. *IET Generation, Transmission & Distribution*, vol. 8, no. 11, pp. 1801-1809.
- [6] S.L. Wang, L. Hong and M.O. Yang(2013) Vulnerability analysis of interdependent infrastructure systems under edge attack strategies, *Safety Science*, vol. 51, no. 1, pp. 328-337.
- [7] S. Boccaletti, V. Latora and Y. Moreno(2006) Complex networks: Structure and dynamics, *Physics reports*, vol. 424, no. 4-5, pp. 175-308.
- [8] J. Johansson and H. Hassel(2010) An approach for modelling interdependent infrastructures in the context of vulnerability analysis, *Reliability Engineering & System Safety*, vol. 95, no. 12, pp. 1335-1344.
- [9] E. Jenelius(2009) Network structure and travel patterns: explaining the geographical disparities of road network vulnerability, *Journal of Transport Geography*, vol. 17, no. 3, pp. 234-244.
- [10] M.A.P. Taylor and S. Susilawati(2012) Remoteness and accessibility in the vulnerability analysis of regional road networks, *Transportation Research Part A: Policy and Practice*, vol. 46, no. 5, pp. 761-771.
- [11] A. Brenkert and E. Malone(2005) Modeling vulnerability and resilience to climate change: a case study of India and Indian States, *Climatic Change*, vol. 72, no. 1, pp. 57-102.
- [12] V. Marianov and C. ReVelle(1995) Siting emergency services. In: Drezner Z, editor. *Facility location*, Berlin: Springer, pp. 199-223.
- [13] O. Berman, Z. Drezner and D. Krass(2010) Generalized coverage: New developments in covering location models, *Computers & Operations Research*, vol. 37, no. 10, pp. 1675-1687.
- [14] S. Lim and M. Kuby(2010) Heuristic algorithms for siting alternative-fuel stations using the Flow-Refueling Location Model, *European Journal of Operational Research*, vol. 204, no. 1, pp. 51-61.
- [15] M.S. Daskin(1995) *Network and Discrete Location: Models, Algorithms, and Applications*, John Wiley & Sons: New York.

# Channel Estimation in OFDM System Based on LS and FEC Iteration<sup>1</sup>

Wenyan Chen<sup>a,1</sup>, Ying Huang<sup>a</sup>, Xiang Zou<sup>b,1</sup>, Wenbo Fan<sup>a</sup>, and Jianxin Guo<sup>a</sup>

<sup>a</sup>*School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China*

<sup>b</sup>*School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China*

**Abstract.** In the Orthogonal Frequency Division Multiplexing (OFDM) system, a channel estimation algorithm based on least squares (LS) and forward error correction (FEC) iteration was proposed for the shortcoming of traditional LS algorithm. In this paper, LS algorithm and FEC algorithm were introduced, and the improved algorithm was explained in detail. The improved algorithm used the FEC correction performance to correct channel estimation results. Simulation results showed that the improved algorithm can effectively improve the performance of channel estimation.

**Keywords.** Orthogonal Frequency Division Multiplexing, channel coding, Forward error correction, channel estimation, least squares

## 1. Introduction

OFDM is a special multi-carrier technology, which is one of the key technologies in the Power Line Carrier (PLC) communication system<sup>[1]</sup>. OFDM communication technology divides the data transmission channel into several orthogonal sub-channels, converts the high speed data signal into parallel low speed sub-data stream, modulates and transmits on each sub-channel, has strong anti-multipath fading and interference between symbols. Since the spectra of the sub-carriers overlap each other, the OFDM has a high spectral efficiency<sup>[2]</sup>. However, in order to maintain the high frequency band utilization rate and to achieve high speed data transmission, the receiver must know the accurate channel state information, and the channel estimation algorithm is one of the key technologies of OFDM<sup>[3]</sup>. In this paper, a channel estimation algorithm based on LS and FEC iteration is proposed. The improved algorithm reduces the bit error rate and improves the performance of the communication system by using the error correction performance of FEC.

---

<sup>1</sup> Wenyan Chen, Associate Proessor, School of electrical and control engineering, Xi'an University of Science and Technology, Yanta Road No.58, Xi'an 710054, Shannxi Province, China. E-mail: ch\_w\_y@163.com

## 2. OFDM channel model

In this paper, the channel estimation of OFDM communication system was studied, the system utilizes the technology of modulation and reuse. The spectrum utilization of the whole system is significantly improved due to the orthogonality of sub-carriers.

The source data stream is decomposed into multiple sub-streams after channel coding, interleaving and symbol modulation. Each sub-data stream is subjected to Inverse Fast Fourier Transform (IFFT) transform and a pilot signal for channel estimation is inserted into the data signal. The receiver estimates the channel response by the pilot signal. The data is converted to time domain data  $x(n)$  via IFFT:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \exp\left(\frac{j2\pi nk}{N}\right) \quad (1)$$

In formula (1),  $0 \leq n \leq N - 1$ ,  $X(k)$  is the modulated signal on the  $k$  sub-carrier,  $N$  is the total number of sub-carriers on an OFDM symbol, usually an exponential power of  $2^{[4]}$ . Before sending a signal, the cyclic prefix (CP) is inserted into the front of the time domain signal. The time domain model of the multipath channel impulse response (CIR) can be summarized as follow:

$$h(n) = \sum_m \alpha_m \delta(n - m) \quad (2)$$

In formula (2),  $\alpha_m$  is the gain of the  $m$  multipath.  $m$  is the time delay normalized to the sampling period. The time domain signal received by the OFDM receiver can be expressed as:

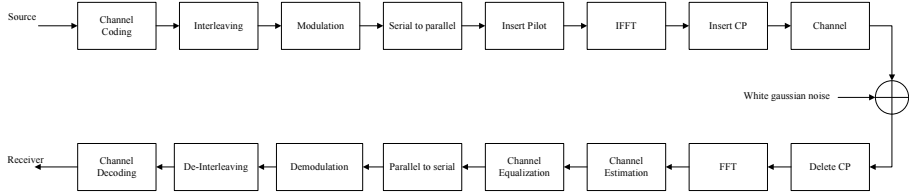
$$y(n) = x(n) \otimes h(n) + w(n) \quad (3)$$

In formula (3),  $0 \leq n \leq N - 1$ ,  $\otimes$  is the cyclic convolution of  $N$  points,  $w(n)$  is time domain additive Gaussian white noise. The frequency domain expression of equation (3) is:

$$Y(k) = H(k)X(k) + W(k) \quad (4)$$

In formula (4),  $0 \leq k \leq N - 1$ ,  $W(k)$  is the frequency domain additive Gaussian white noise,  $H(k)$  is the frequency domain response of the channel.

It is a typical block diagram for OFDM system<sup>[5]</sup> as shown in figure 1.



**Figure 1.** The block diagram of OFDM communication system.

### 3. LS algorithm

The channel transmission characteristics of OFDM communication system are very complex, the receiver is difficult to restore the perfect transmission signal, so the channel estimation in the actual communication system is very important.

The channel estimation is the process of estimating the channel frequency response of the data transmission channel. The channel estimator estimates the transmission characteristics of the channel through the pilot signal inserted by the transmitting. The equalizer to estimate the value of the receiver to receive the received signal to improve the transmission performance of the communication system<sup>[6]</sup>. Assume that the OFDM system model is represented by equation(5):

$$Y_p = X_p H + W_p \tag{5}$$

In formula (5),  $H$  is the channel response,  $X_p$  is a known pilot transmission signal,  $Y_p$  is the received pilot signal.  $W_p$  is the Gaussian white noise superimposed on the pilot sub-channel. The frequency domain characteristics of the channel response can be obtained by transmitting the signal  $X_p$  through the known pilot. Using the estimated value of the channel response to correct other information without prior agreement<sup>[7]</sup>.

The LS algorithm goal is to obtain the minimum square error of the sender and the receiver, so that the sum of the error between the estimated result and the observation point is the smallest, that is:

$$\hat{H}_{ls} = \arg \min[(Y_p - \hat{H}_p X_p) * (Y_p - \hat{H}_p X_p)] \tag{6}$$

In formula (6),  $X_p$  is the pilot signal inserted at the sender,  $Y_p$  is the signal at the received pilot position,  $\hat{H}_p$  is the channel response estimate at the pilot location.

The channel estimation at the pilot point of the LS algorithm can be obtained by transforming as following:

$$\hat{H}_{ls} = \frac{Y_p}{X_p} = \frac{H_p X_p + W_p}{X_p} = H_p + \frac{W_p}{X_p} \quad (7)$$

In formula (7),  $W_p$  is the frequency domain noise at the pilot point. The estimated value of the channel frequency response at the pilot based on the LS algorithm can be obtained by the above method. The channel frequency response at the data point can be obtained by the data interpolation method at the pilot point. In this paper, linear interpolation was selected for simulation.

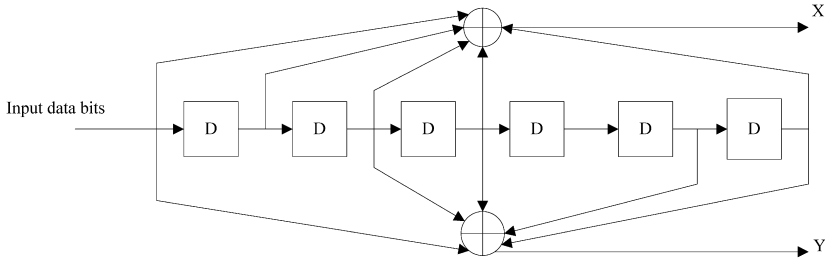
The advantage of LS algorithm is its low computational complexity and easy implementation. However, the algorithm has low estimation accuracy, especially when the SNR is lower.

#### 4. The coding principle of FEC

FEC is one of necessary links to achieve reliable communication in OFDM communication system. It is a way to improve correctness of data transmission. In OFDM system, because the noise and the presence of fading in the channel, the information received is wrong. FEC has the ability of error detection and correction, can detect and correct digital signals, enhance the digital signal to resist various channels in the interference of the ability to increase data communication reliability, improve system performance<sup>[8]</sup>.

FEC uses the redundant part of the transmitted information to achieve error detection and correction. The basic idea is that the sender uses the forward error correction code to carry out redundant coding of the transmitted data, through the channel to the receiver, the receiver through the decoder to receive redundant data to check and correct the information stream in the error bit, improve the reliability of the data. FEC code is often divided into group code, convolutional code and Turbo.

The improved method of this paper used convolutional code, so the emphasis on convolutional code. The core idea of the convolutional code is: the encoder converts the  $m$  bits data information into  $n$  bits data information and transmits it. The representation is  $(n,k,m)$ ,  $n$  is the bit length of the output information,  $k$  is the information bit length contained in the codeword,  $m$  is the coding memory degree. The output  $n$  bits information is related not only to the  $k$  bits information of the current segment but also to the preceding  $(m-1)$  bits information.



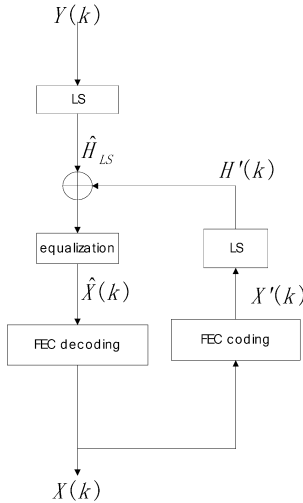
**Figure 2.** The structure diagram of convolutional code encoder.

Figure 2 shows the (2,1,7) convolutional code of the encoder structure, the generating polynomial is (171,133).

### 5. Based on LS and FEC Iteration Algorithm

The traditional LS channel estimation algorithm has low computational complexity and easy implementation, but the estimation accuracy is lower and more obvious in low SNR. In order to improve the performance of the traditional LS channel estimation algorithm, this paper proposed a channel estimation method based on LS and FEC iteration.

The improved channel estimation algorithm was proposed in this paper mainly used the error detection and error correction performance of the forward error correction code to improve the accuracy of channel estimation. The algorithm used the FEC to correct the channel response value obtained by the traditional LS algorithm and improve the accuracy of the channel estimation after the LS algorithm obtains the channel response.



**Figure 3.** The flow chart of the improved algorithm.



The process flow diagram of the improved algorithm is shown in figure 3. The concrete implementation steps are as follows:

- a. The receiver extracts the pilot signal from the received signal and performs LS channel estimation to obtain an estimated value of the channel response  $\hat{H}_{ls}$ .
- b. The receiver extracts the data signal from the received signal, the data signal and the channel estimate  $\hat{H}_{ls}$  are sent to the equalizer to obtain the equalized signal  $\hat{X}(k)$ .
- c.  $\hat{X}(k)$  does Viterbi decoding to get the original signal from the sender  $X(k)$ .
- d.  $X(k)$  is encoded in the same way as the sender, the encoded data  $X'(k)$  will be get.
- e. The receiver extracts the data signal and  $X'(k)$  do division operations so that new channel estimates  $H'(k)$  can be obtained.
- f. The estimate of the response by two channels is revised, and the corrected formula can be expressed as:

$$H(k) = (1 - \alpha)\hat{H}_{ls} + \alpha H'(k) \quad (8)$$

- g. Finally, use the modified channel to estimate  $H(k)$  for equilibrium, channel decoding, and restore the original signal from the sender.

This paper simulated the channel estimation for tradition algorithm and improved algorithm, the performance of two algorithms were detected by the error rate.

The specific parameters of the OFDM simulation system are shown in table 1:

**Table 1.** Parameters of OFDM simulation system

Simulation parameter	Numerical values
Carrier number	64
Symbol number	50
Modulation mode	16QAM
Pilot interval	5
FEC Coding	(33,23) Convolutional coder
FEC Decoding	Viterbi Decoding

The simulation channel adopted the additive Gaussian white channel model, the cyclic prefix took 1/8 of the total carrier, the pilot interval was 5, and each OFDM symbol contained 11 pilot signals.

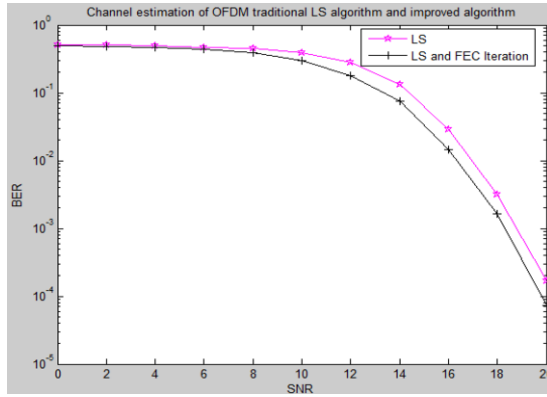


Figure 4. The simulation figure of improve algorithm.

From the simulation results in figure 4, it can be seen that the bit error rate (BER) of two channel estimation algorithms decreases with increasing SNR. The performance of the traditional LS channel algorithm is poor, and the performance of the improved algorithm is superior to the traditional algorithm.

In order to confirm the value of the correction factor  $\alpha$  in the improved algorithm, the simulation was performed about the effect of  $\alpha$  on BER under different SNR in the same simulation environment as above. The simulation results in a low SNR with 1 dB and 2 dB, in high SNR with 10 dB and 11dB were shown in Fig.5(a) and Fig.5(b).

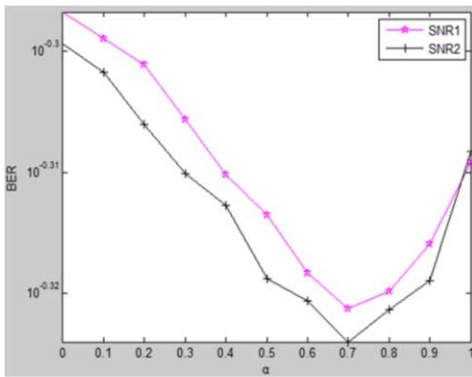


Figure 5(a). The simulation figure in low SNR.

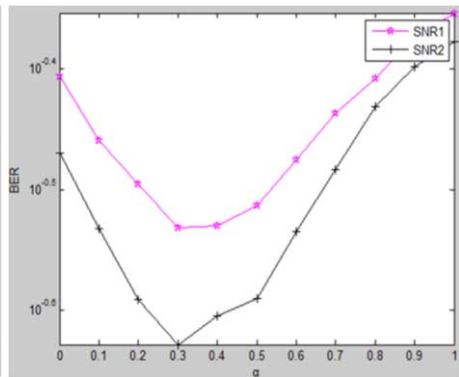


Figure 5(b). The simulation figure in high SNR.

It can be seen from Fig.5(a) and Fig.5(b) that in the case of low SNR, when the value of  $\alpha$  was about 0.7, the BER was low. In the case of high SNR, when the value of  $\alpha$  was about 0.3, the BER was low.

Although the channel estimation algorithm based on LS and FEC iteration requires two channel decoding and channel coding, the computational complexity and memory space of the communication algorithm are increased to a certain extent. However, the method has good performance and can reduce the BER of system. For systems with less stringent storage and computing complexity, the improved channel estimation algorithm has advantages over traditional LS channel estimation algorithm.

## 6. Conclusion

In this paper, the LS algorithm of channel estimation and the convolution coding of FEC in OFDM system were introduced in detail. The improved channel estimation algorithm based on LS and FEC iteration was proposed, and compared with the traditional LS algorithm. According to the simulation results analysis, the improved algorithm is more accurate for channel estimation, which can effectively reduce the BER and improve the performance of OFDM channel estimation.

## References

- [1] Zhang Pingze, Based on low voltage power line carrier communication method comparison, Journal of electronic design engineering, **18** (2010):26-28
- [2] Li Zhenna, OFDM baseband signal processing technology research. Wuhan University of technology, 2007
- [3] Tomasoni A, Gatti D, Bellini S, et al. Efficient OFDM Channel Estimation via an Information Criterion, IEEE Transactions on Wireless C-ommunications, **12** (2013):1352-1362
- [4] Xi Yaming, The power line communication system based on OFDM standard research and performance analysis of the contrast. Chong qing: Chongqing University, 2012
- [5] Hou Jinfeng, Liu Jian, A Novel Channel Estimation Algorithm for 3GPP LTE Down System Using Joint Time-frequency Two dimensional Iterative Wiener Filter[C]. Nanjing: Communication Technology (ICCT), 2010 12<sup>th</sup> IEEE International Conference on, (2010):289-292
- [6] Zhang Qing, Power line carrier channel estimation algorithm based on OFDM research. Chongqing: Chongqing University, 2013
- [7] Lu Lizhi, Feng Ye, Yu Renqing, Hu Xiaowen, Integration of radar and communication based on OFDM waveform research, Journal of information technology, **2** (2014):144-147
- [8] Sun Hengyang, Huang Guoce, Wu Yong. The application of adaptive HEC in satellite mobile communication research. Journal of network and communication, **14** (2003):7-10

# Container Terminal Operation System in Emission Control Area Using Modelling Based on Multi Agents

Jue HOU<sup>1</sup>

*China Waterborne Transport Research Institute, Ministry of Transport of the People's Republic of China, Haidian District, Beijing, China*

**Abstract.** Improving operation efficiency of container terminal, is important for terminal operators. Meanwhile, the air pollution problem is increasingly serious in container terminal, with the establishment of the Emission Control Areas (ECAs) of China in 2015, environmental protection pressure is increasing to terminal operators. In order to comply with the ECA regulation, optimizing the operation efficiency, reducing air pollution emissions, which, is an urgent problem to terminal management in China. This paper based on the characteristic of Container Terminal Operation System (CTOS), including berth allocation problem, truck dispatching problem, and yard allocation problem, establishes the architecture of CTOS based on the Multi Agents System. A model and simulation on a port container terminal is built on the JADE platform to support the decision-making of container terminal, which can reduce air pollution emissions in port, and maintain service levels. CTOS allows the container terminal operator to be more flexible in their decision to meet the ECA regulation.

**Keywords.** Container terminal, operation system, emission control area, simulation, multi agents

## 1. Introduction

In the last more than a decade years, with China's domestic and foreign trade continues to expand, container transport is increasingly valued. With construction of container terminal improving, and ascension of operation and management level of container terminal. Meanwhile, ship exhaust emissions like nitrogen oxide (NO<sub>x</sub>), sulphur oxide (SO<sub>x</sub>) and particulate matters (PM), have contributed to the worsening of the atmospheric environment in port areas, even threatening the health of people living in port cities and coastal communities. China has established the three Emission Control Areas (ECAs) in 2015, Three ECAs include the Pearl River Delta, the Yangtze River Delta, and in the Bohai Sea. They have been created to reduce the levels of ship-generated air pollution and mainly focus on the sulphur content of fuels. To comply with the regulation, ships are to use fuel oil with a sulphur content of no more than 0.5% m/m, the marine gas oil (MGO), or other equivalent measures to reduce emissions including exhaust gas scrubbing, alternative clean fuels and shore power (cold ironing).

---

<sup>1</sup> Corresponding author. Jue Hou, China Waterborne Transport Research Institute, Ministry of Transport of the People's Republic of China, Haidian District, Beijing, China; E-mail: mars8510@163.com.

Port operators and ship owners both face massive environmental protection challenges resulting from ECA regulation in China. In order to comply with the ECA regulation, optimizing the operation efficiency, reducing air pollution emissions, which, is an urgent problem to terminal management in China. Container Terminal Operation System (CTOS) could solve the conflicts between operation efficiency and emission constriction in reasonable strategy of container terminal. Multi-Agents System (MAS) is the system of the completion of some complex task and target. There are research results applying the MAS technology to CTOS.

This paper based on the characteristic of CTOS, including berth allocation problem (BAP), yard allocation problem (YAP) and truck dispatching problem (TDP), then built the simulation model of CTOS with MAS based on the JADE platform. Different strategies are designed, which can reduce air pollution emissions in container terminal, and maintain service levels, and allows the container terminal operator to be more flexible in their decision to meet the ECA regulation.

## 2. Model Description of CTOS

### 2.1. Container Terminal Operation System

Container Terminal Operation System (CTOS) is complex system which reflect a series of operation process in container terminal. CTOS is mainly for three problems in this study: when container ships arrive at terminal through anchorage, which berth should be allocated to the mooring ship, called berth allocation problem (BAP). Then, ships are handled by quay cranes, and containers of ships are transhipped by trucks between quay and yard, called the truck allocation problem (TDP). In container terminal yard stack, yard crane and truck need cooperate with each other, to transport and storage container. called the yard allocation problem (YAP), as shown in Fig. 1. Above problems will be discussed in next section.

With the ECA regulation, container terminal operator seeks to reduce ship emissions for mooring period, without lowering the service level, by reducing ship waiting time and depart time as much as possible in the berth plan. Considering the berth which equipped shore power, as shown in circle of Fig. 1, grey rectangle represents the ship with on-board shore power installation. The SPS represents the shore power system, each SPS could supply power to one berth/ship. When ship is supplied by SPS, which shutdown the auxiliary machine, and no air pollution emit.

The subsystems structure of CTOS in this study as shown in Fig. 2, including location subsystem and equipment subsystem. Each subsystem represents the transshipment of location and handle equipment of container flow in CTOS.

### 2.2. Problems of CTOS

- *Berth Allocation Problem*

Container terminal operator gives the order to assign a berth for each arriving ship, which to make the ship handle and depart as sooner as possible, in order to reduce the cost and time for both of terminal and ship owner. Because of BAP is decision problem in container terminal allocation, which, directly impact efficiency of YAP and TDP. This paper uses the following strategies: the first come first serve (FCFS) [1,2], the

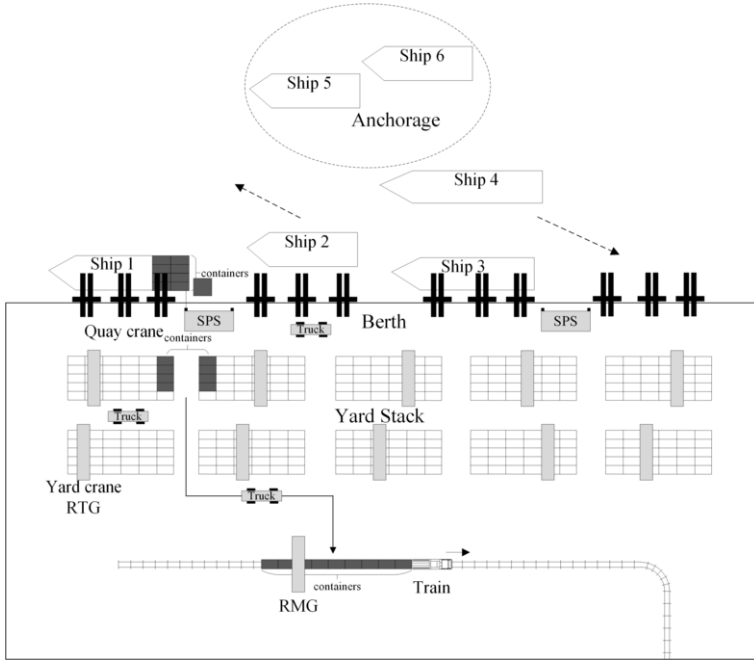


Figure 1. A typical container terminal and its equipment.

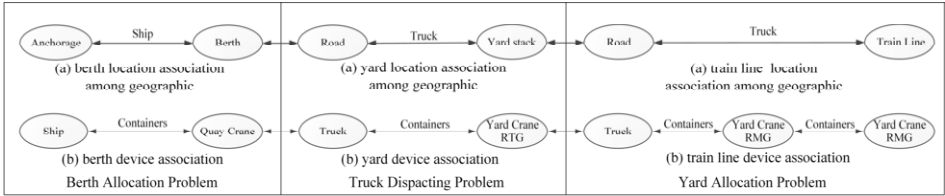


Figure 2. Flowchart of container terminal operation system.

dynamic berth allocation problem (DBAP) [3–5], which includes discrete and continuous berth (DBAPD and DBAPC), and following the ECA regulation of China, reducing air pollution emissions in port, maintaining service levels. The dynamic berth allocation problem with emission consideration (EDBAP) can be described as follows:

$$\min f_1 = \sum_{i \in B} \sum_{j \in V'} \sum_{k \in U} \left\{ (T - k + 1) C_{ij} + S_i - A_i \right\} x_{ijk} \alpha_j \omega_1 + \sum_{i \in B} \sum_{j \in V'} \sum_{k \in U} (T - k + 1) y_{ijk} \alpha_j \omega_1 + \sum_{i \in B} \sum_{j \in V'} \sum_{k \in U} \left\{ (T - k + 1) C_{ij} - A_i \right\} e_{ijk} s_{ijk} \omega_2 \quad (1)$$

$$\alpha_j = \frac{W_j - W_{\min}}{W_{\max} - W_{\min}} \quad (2)$$

s.t.

$$\sum_{i \in V'} \sum_{m \in P_k} (C_{il} x_{ilm} + y_{ilm}) + y_{ijk} - (A_j - S_i) x_{ijk} \geq 0 \quad \forall i \in B, j \in W, k \in U \quad (3)$$

$$\sum_{i \in B} \sum_{k \in U} x_{ijk} = 1, \quad \forall j \in V \tag{4}$$

$$\sum_{j \in V} x_{ijk}^k \leq 1, \quad \forall i \in B, k \in U \tag{5}$$

$$x_{ijk} \in \{0,1\} \quad y_{ijk} \geq 0 \quad \forall i \in B, j \in W, k \in U \tag{6}$$

The objective function (1) minimize the total service time (waiting and delayed departure time) and emissions for all vessels in discrete quay shoreline. Equation (2) is volume weight of ship  $j$ . Equation (3) assure that ships are serviced after their arrival. Equations (4) and (5) the constraint of  $x$ .  $T$  the arrival time of ships,  $T \in \mathbb{N}$ ;  $k$  the set of service orders;  $C_{ij}$  the handling time spent by ship at berth  $i$ ;  $S_i$  the time when berth location  $i$  becomes idle for the planning horizon,  $S \in \mathbb{N}$ ;  $A_i$  the arrival time of ship  $j$ ;  $x_{ijk}$  ship  $j$  is serviced as the  $k$ th ship at berth  $i$ ;  $y_{ijk}$  the idle time of berth  $i$  between the departure of the  $(k - 1)$ th ship and the arrival of the  $k$ th ship when ship  $j$  is serviced as the  $k$ th ship;  $i$  is set of berth;  $V$ , the set of ships,  $V \in \mathbb{N}$ ;  $B$  the set of berths;  $e_{ijk}$  the emission of berth location  $i$  with each ship  $j$ ,  $k$  the set of service orders,  $\omega_1$  the weight of total service time,  $\omega_2$  the weight of total emission,  $\omega_1 + \omega_2 = 1$ .

- *Truck Dispatching Problem*

Container truck is important transport equipment which greatly affects the loading and unloading efficiency in container terminal, including utilization of yard crane, ship waiting time and so on. Truck allocation problem can be described as follows [6,7]:

$$\min f_3 = \omega_1 \frac{\sum_{i \in V} \sum_{j \in V} x_{ij}^k d_{ij}}{\nu} + \omega_2 \sum_{i \in V} \sum_{j \in V} x_{ij}^k d_{ij} c \tag{7}$$

s.t.

$$\sum_{k \in N} \sum_{j \in V} (x_{ij}^k + x_{ji}^k) \geq 1 \tag{8}$$

$$\sum_{k \in N} \sum_{i \in V} x_{ij}^k \leq 1 \quad \forall j \in V \tag{9}$$

$$\sum_{k \in N} \sum_{i \in V} x_{ij}^k \leq 1 \quad \forall j \in V \tag{10}$$

$$x_{ij}^k \in \{0,1\} \quad \forall i, j \in V, k = 1, 2, \dots, N \tag{11}$$

The objective Eq. (7) is to the minimum truck operation time and emission,  $\omega_1$  the weight of total operation time,  $\omega_2$  the weight of total emission,  $\omega_1 + \omega_2 = 1$ . Equation (8) the truck  $k$  is dispatched from point  $i$  to  $j$  once at least, Eqs (9) and (10) the task by truck  $k$  from point  $i$  to  $j$  is finished only once, and no re-dispatched, Eq. (11) the  $x_{ij}$  is 0–1 integer variable,  $d_{ij}$  the distance form point  $i$  to  $j$  in upload/load container operation.  $d_{ij}$  the distance form point  $i$  to  $j$  in upload/load container operation.

- *Yard Allocation Problem*

In container terminal yard stack, yard crane and truck need cooperate with each other, to transport and storage container. In order to reduce the waiting time between yard crane and truck, terminal operator takes yard allocation to improve yard stack operation efficiency, the YAP problem can be described as follows [8,9]:

$$\min f_2 = \sum_{i=1}^n \sum_{j=1}^n d_{ij} x_{ij} + \sum_{i=1}^n P_i(s_i) \quad (12)$$

s.t.

$$P_i(s_i) = \begin{cases} a_i(ET_i - s_i) & s_i < ET_i \\ 0 & ET_i \leq s_i \leq LT_i \\ b_i(s_i - LT_i) & s_i > LT_i \end{cases} \quad (13)$$

$$\sum_{i=1}^n x_{ij} = 1, \quad \sum_{j=1}^n x_{ij} = 1 \quad (14)$$

$$f_i = T_i + \max \{ f_{i-1} + t_{i-1,i}, R_i \} \quad (15)$$

$$d_{ij} = |B_{ki} - B_{kj}| \cdot l \quad (16)$$

$$s_i = f_{i-1} + t_{i-1,i} \quad (17)$$

The objective function (12) minimize the move distance of yard crane and waiting time of truck. Equation (13) is time window, to limitation the waiting time of truck for yard crane. Equation (14) each container of uploading or loading can only be operated once. Equation (15) the time of yard crane and truck. Equation (16) the move distance of yard crane. Equation (17) the operation time limitation of yard crane.  $d_{ij}$  the distance of yard crane form task  $i$  to task  $j$ ,  $x_{ij} = 1$  if yard crane form task  $i$  to task  $j$ ,  $P_i(s_i)$  the function of time window limitation,  $a_i$  and  $b_i$  the penalty factor,  $ET_i$  the start time of operation,  $LT_i$  the end time of operation,  $T_i$  the time of yard crane operation,  $R_i$  the arrive time of truck  $i$ ,  $B_{ki}$  and  $B_{kj}$  the bay  $k$  of task  $i$  and  $j$ ,  $t_{i-1,i}$  the move time of task  $i - 1$  to  $i$ ,  $l$  the width of bay.

### 3. CTOS Modelling Based on MAS

#### 3.1. Multi Agents System of CTOS

Operational research, such as Eqs (1), (7) and (12) might provide the exact solution for the given problems: BAP, TDP and YAP. However, these are NP-hard problems by modelling using mixed integer non-linear programming (MINLP), size and complexity of MINLP exponential growth with the size of variables, leading too long time to find the optimal solution (e.g. branch and bound algorithm). Furthermore, the CTOS comprise many interrelated complex subsystems of whole system is often not feasible. Although separate models of respective subsystems can be defined, this means that important interrelations of subsystems will be neglected.



Simulation is not so strongly limited by the size and complexity of the CTOS and can usually cover all required subsystems in a single model, respecting their mutual relations as well [10–12]. Multi-Agents System (MAS) is a complex system that consists of agents which are the characteristics of autonomy, cooperation, and sociality [13,14], and it is able to reflect container terminal operating process. MAS is typical discrete, distribute and dynamic system, could meet functional requirements of the CTOS. This paper study the CTOS based on MAS, using the below characteristics: the autonomy, when the CTOS sends request, the agent do not accept request until the task is finished. The cooperation: when the container transport or operation task could not be completed by single agent, the task by coordination and negotiation with other agents. The sociality: the agents bring about communication, collaboration, coordination and negotiation with the others. This study has a sequence of discrete time steps (1 sec), at each time step, the agent receives a state from the environment. At the next time step, as a consequence of its action, the agent receives a new task and moves to a new state.

### 3.2. Subsystems Based on MAS of CTOS

This study with the CTOS consists of two subsystems, the location subsystem and equipment subsystem. Location subsystem is composed of anchorage agents, berth agents, stack yard agents, road agents and train loading and unloading line agents. They are responsible for handling locations of the CTOS. The interaction sequence between these location agents as shown in Fig. 3. Equipment subsystem is composed of ship agents, quay crane agents, truck agents, yard crane agents and train agents. They are responsible for handling operation of the CTOS. The flow of containers is bi-direction, the import and export, the interaction sequence of import container between these equipment agents as shown in Fig. 4.

Based on the above subsystems structures, there is much communication among agents in the CTOS in order to complete the cooperation with each other by exchanging information. Therefore, the communication is based on the cooperation among the agents. Furthermore, there is interactive rule of information to guarantee efficient operation when the cooperation is appeared in the agents. As shown in Figs 3 and 4, there is

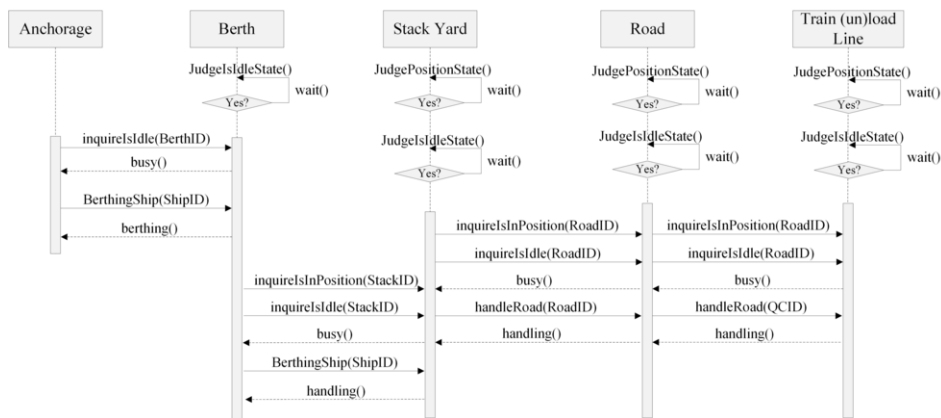


Figure 3. Location subsystem interaction sequence diagram of container terminal operation system.

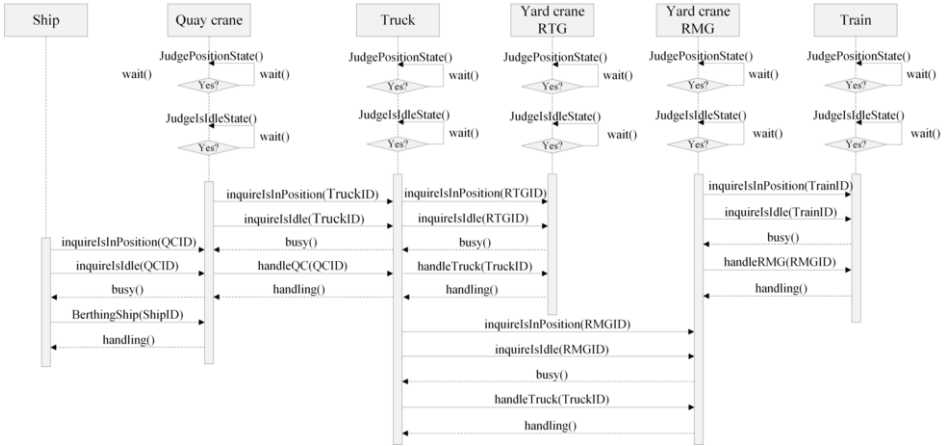


Figure 4. Equipment subsystem interaction sequence diagram of container terminal operation system.

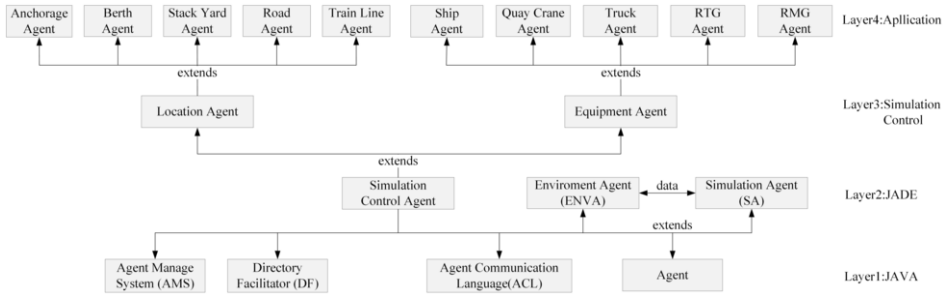


Figure 5. Frame diagram of container terminal operation system based on JADE platform.

a complex collaboration among agents, which is the whole handling and transportation process of the CTOS.

### 4. Simulation Platform

In this paper, all the container terminal agents are developed over the Java Agent Development framework (JADE) platform, which is a programming and simulation environment based on Java that matches with Unified Modelling Language (UML), and conforms to the Foundation for Intelligent Physical Agents (FIPA) specifications. MAS based on JADE platform allows the user to combine different techniques and approaches such as differential equations and discrete events with container terminal operation [15]. This paper build a CTOS simulation model based on JADE platform, the framework of which is divided into 4 layers, as shown in Fig. 5.

The first layer is Java platform, which ensure the flexibility and portability of the MAS system. The second layer is JADE platform, located in the upper Java platform, JADE platform is the API based on Java, completely, which has good compatibility with the Java platform, including Agent Manage System (AMS), Directory Facilitator (DF) and Agent Communication Language (ACL) services. The third layer is the simu-

lation control system that provided AMS, DF and ACL services, mainly includes three important Agent: Simulation Agent (SA), which is the design and implementation of a simulation for all common Agent class. Simulation Controller Agent (SCA) the manager of SA, including start, termination and synchronization. Environment Agent (ENVA) is the main work of storage simulation using data. The fourth layer is application layer, which is extended through the simulation control layer, to control location and equipment agent.

### 5. Empirical Simulation Study

This study performs a simulation research on Yangshan port container terminal in Shanghai of China, which has a throughput capacity of 4,000,000 TEU per year. In order to evaluate possible impacts both on environmental and service level, from the ECA regulations. Meanwhile, takes 4 shore power systems (planned) into consideration, there are 2 shore power systems installation forms: the discrete form (berth 1, 3, 5, 7), and continuous form (100 m, 900 m, 1600 m, 2100 m, begin in the west of quay shoreline), as shown in Fig. 6.

The Yangshan port container terminal design parameters as in Table 1, including location subsystem and equipment subsystem.

Due to the transport distance of container is variable, which is calculated by distance divide to truck move speed, as shown in Table 2. From Table 1, RTG and RMG in operation for stack yard and train line, respectively. The coordinates of the southwest corner of the stacks are used for positioning the stack in the yard, and which, are used by the truck for determining distances to the quay and stack.

In order to decrease the scale of the simulation model, this study perform a computational study on arrival time data of 61 container ships in two weeks with dynamic strategy, as shown in Table 3, and assumes 28 trains, each one has different volume and arrive time in 8:00 and 16:00 in one day.

In this study, due to the absence of measurement equipment of emission factors installed in container ships, this study adopts the emissions factors in Table 4, which are widely used in reference studies [16–18], then NOx, SOx and PM emissions will be calculated by fuel consumed multiply their emissions factors. According to the ECA regulation, when the ship without on-board shore power system installing in berthing period, they need to take the 0.5% sulphur MGO fuel, and 0.1% sulphur diesel fuel for truck.

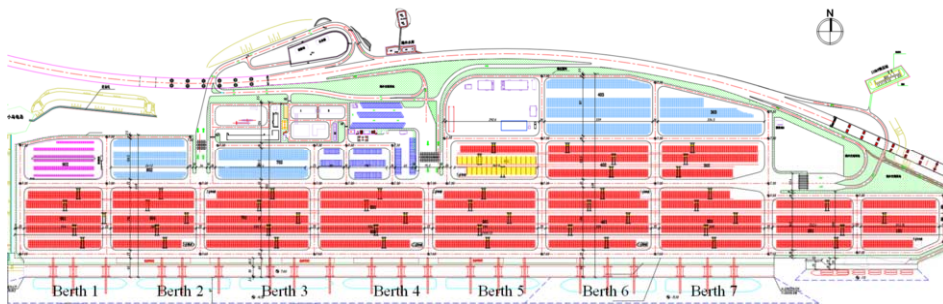


Figure 6. The layout of Yangshan port container terminal (Shanghai, China).

**Table 1.** The Yangshan port container terminal design parameters.

Agent	Item	Parameter
Quay	shoreline	2350 meter
	berth	7
	shore power system	4 (planned)
Stack Yard	yard area	98.6 kilo square meter
	stacks	11
	each stack yard length	150 meter
	each stack yard width	50 meter
	each stack yard capacity	180 TEU
	each length	200~268 meter
	each width	156.5~178.82 meter
	stack yard total capacity	28052 TEU
Road	north-south road	12
	east-west road	4
	road width	25 meter
Equipment	quay crane	16
	yard crane	88
	truck	80
	empty container stacker	6
	train	2
Train Handling Line	rail loading line	2

**Table 2.** The equipment data in container terminal.

Agent	Power (kW)	Size L×W×H (meter)	Operation Time (sec/TEU)	Move Speed (meter/sec)	Number	Energy
Quay crane	900	65×30×10	60	0.40	16	electric
Yard crane (RTG)	450	30×31×19	30	0.30	88	electric
Yard crane (RMG)	480	30×31×19	35	0.25	8	electric
Truck	100	12.5×2×3	variable	1.10	80	diesel
Train (locomotive)	2500	21×3×4	variable	variable	2	diesel

**Table 3.** The arriving ship data in container terminal.

Dead weight tonnage (DWT) (t)	Handle Container (TEU)	Length (m)	Auxiliary engine (kW×Numbers)	Number of ships	Number of on-board SPS
10000~20000	701~1050	141~183	430×3~700×3	8	2
20001~50000	1051~3500	183~293	700×3~1960×3	4	1
50001~70000	3501~5650	293~399	1960×3~2320×4	12	7
70001~90000	5651~6630	300~399	2320×4~2820×4	21	17
90001~110000	6631~9500	346~399	2820×4~3020×4	8	8
110001~150000	11001~15500	367	3020×4~3820×4	5	3
>150000	>15500	>367	3850×4	3	3

**Table 4.** Emission factors.

Sulphur	Emission	Emission factor (g/kg-fuel)
0.5%	NO <sub>x</sub>	64.8
	SO <sub>x</sub>	15.9
	PM	4.5
0.1%	NO <sub>x</sub>	64
	SO <sub>x</sub>	2.0
	PM	2.6

**Table 5.** Empirical simulation result of waiting time.

Strategy	Waiting time (min)						Total (min)
	Berth		Yard stack		Train line		
	Ship	QC	Truck	RTG	RMG	Train	
FCFS	12736.88	1826.63	2964.60	3350.71	1350.09	698.61	22927.52
DBAPD	9830.10	1350.71	2646.85	3031.06	950.73	626.40	18435.85
DBAPC	9394.73	1165.27	2301.03	2834.42	1150.16	663.97	17509.58
EDBAPD	11054.94	1485.45	3126.21	3173.71	1383.49	641.52	20865.32
EDBAPC	9866.20	1361.11	2990.94	3103.24	1271.61	639.85	19232.95

**Table 6.** Empirical simulation result of air pollution emission.

Strategy	Berth emission (kg)			Yard stack emission (kg)			Train line emission (kg)			Total/kg
	NO <sub>x</sub>	SO <sub>x</sub>	PM	NO <sub>x</sub>	SO <sub>x</sub>	PM	NO <sub>x</sub>	SO <sub>x</sub>	PM	
FCFS	22398.52	5495.93	1555.45	3247.36	101.48	131.92	388.47	12.14	15.78	33347.05
DBAPD	19576.71	4803.54	1359.49	3010.09	94.07	122.28	276.88	8.65	11.25	29262.96
DBAPC	18597.87	4563.37	1291.52	2931.92	91.62	119.11	230.67	7.21	9.37	27842.66
EDBAPD	14095.23	3458.55	978.84	3288.62	102.77	133.60	332.09	10.38	13.49	22413.57
EDBAPC	12967.61	3181.87	900.53	3105.35	97.04	126.15	300.20	9.38	12.20	20700.33

The CTOS based on MAS simulation model is formulated by JADE 4.40, and run on a personal computer with Intel Core i7 GHz CPU and 16 GB RAM. As in the Table 5, take  $\omega_1 = 0.5$ ,  $\omega_2 = 0.5$  in Eq. (1) with EDBAP, the empirical test result of different strategies shows the waiting time of different problems.

Table 6 shows empirical simulation result of emissions with different strategies of container terminal.

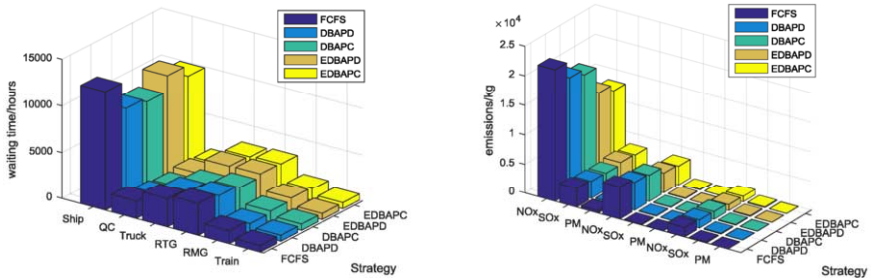
Table 7 shows the utilization ratio of SPS, which means the ship with on-board shore power installation is allocated to the berth, which has supply shore power to the ship.

Figure 7 shows all these strategies empirical test results by using MAS simulation model of the CTOS.

The total waiting time of DBAPD is 11.64% lower than EDBAPD, but its total emissions are 30.56% higher than EDBAPD. The total waiting time of DBAPC is 8.96% lower than EDBAPC, but its total emissions are 34.50% higher than EDBAPC. The EDBAPD and EDBAPC both can reduce air pollution emissions in container in-

**Table 7.** Empirical simulation result of utilization ratio of SPS.

Strategy	Utilization ratio of SPS/%
FCFS	65.09%
DBAPD	63.41%
DBAPC	60.98%
EDBAPD	73.17%
EDBAPC	85.37%



**Figure 7.** Empirical simulation result of all strategies.

terminal transport, maintain service levels, and increase utilization ratio of SPS, which allows the terminal operator to be more flexible in their decision to meet the ECAs regulation. Adjusting the weight of Eq. (1), the container terminal managers can according to different adjust the weights and evaluate the implementation effect, which allows the terminal operator to be more flexible in their decision to meet the ECA regulation.

### 6. Conclusions

In order to comply with ECA regulations, container terminal operator need to decrease the emissions and keep high efficiency as much as possible, to minimize the total service waiting time that includes berth, yard stack and train loading and unloading line, and the emissions during operation and waiting period. This paper study the characteristic of the CTOS, then build a simulation model of the CTOS with MAS, based on the JADE platform. Finally, an empirical simulation research based on different strategies is designed, which can reduce air pollution emissions in container terminal, and maintain service levels, allows the container terminal operator to be more flexible in their decision to meet the ECA regulation.

There are also some limitations in this study. First, this study does not consider the avoidance mechanism of truck agent in operation period. Second, stacks for hazardous and refrigerated containers are not considered, meanwhile container gate agent is omitted.

Verification and validation of the CTOS simulation model will be tested. Overcoming these limitations and improving the model could be the topics for future research.

## Acknowledgements

This research was partly supported carried out with financial support from the Conserve Energy and Reduce Emissions Grant (CEREG) 2014-JNJP-008-037 and 2015-JNJP-010-056, from the Ministry of Transport of the People's Republic of China.

## References

- [1] Imai A, Nishimura E, Papadimitriou S. (2001). The dynamic berth allocation problem for a container port. *Transportation Research Part B Methodological*, 35(4):401–417.
- [2] Imai A, Nishimura E, Papadimitriou S. (2003). Berth allocation with service priority. *Transportation Research Part B Methodological*, 37(5):437–457.
- [3] Imai A, Sun X, Nishimura E, Papadimitriou S. (2005). Berth allocation in a container port: using a continuous location space approach. *Transportation Research Part B Methodological*, 39(3):199–221.
- [4] Golias M M, Saharidis G K, Boile M, Theofanis S, Ierapetritou M G. 2009. The berth allocation problem: optimizing vessel arrival time. *Marit. Econom. Logist.* 11(4):358–377.
- [5] Du Y, Chen Q, Quan X, Long L, Fung R Y K. (2011). Berth allocation considering fuel consumption and ship emissions. *Transportation Research Part E Logistics & Transportation Review*, 47(6): 1021–1037.
- [6] Yan S, Lai W, Chen M. (2008). Production scheduling and truck dispatching of ready mixed concrete [J]. *Transportation Research Part E*, 44(1):164–179.
- [7] Liao C H, Tseng P H, Lu C S. (2009). Comparing carbon dioxide emissions of trucking and intermodal container transport in Taiwan [J]. *Transportation Research Part D*, 14(7):493–496.
- [8] Canrong Zhang, Zhihai Zhang, Li Zheng, et al. (2011). A decision support system for the allocation of yard cranes and blocks in container terminals [J]. *Asia Pacific Journal of Operational Research*, 28(6):803–829.
- [9] Ya X, Qiushuang C, Xiongwen Q. (2010). A robust integrated approach to yard space allocation and crane scheduling in container terminals [C]//International Conference on Service Systems and Service Management. IEEE Xplore, 1–6.
- [10] Sun Z, Lee L H, Chew E P, et al. (2012). MicroPort: A general simulation platform for seaport container terminals[J]. *Advanced Engineering Informatics*, 26(1):80–89.
- [11] Chhetri P, Jayatilleke G B, Gekara V O, et al. (2016). Container terminal operations simulator (CTOS) – Simulating the impact of extreme weather events on port operation [J]. *European Journal of Transport & Infrastructure Research*.
- [12] Kocifaj M, Adamko N. (2014). Modelling of container terminals using two-layer agent architecture [C]//IEEE, International Symposium on Applied Machine Intelligence and Informatics. IEEE, 251–256.
- [13] Ye Z, Li W, Li B. (2010). Modeling of Container Terminal Logistics Operation System based on multi-agents [C]//International Conference on Computer Supported Cooperative Work in Design. IEEE, 324–328.
- [14] Henesey L, Davidsson P, Persson J. (2006). Evaluating Container Terminal Transshipment Operational Policies: An Agent-Based Simulation Approach [J]. *Travail Et Emploi*, 10(115):81–93.
- [15] Wooldridge M. *Developing Multi-Agent Systems with JADE* [M]. John Wiley, 2001.
- [16] Chang Y T, Roh Y, Park H. (2014). Assessing noxious gases of ship operations in a potential emission control area. *Transportation Research Part D Transport & Environment*, 27(2):91–97.
- [17] Corbett J J, Wang H, Winebrake J J. (2009). The effectiveness and costs of speed reductions on emissions from international shipping [J]. *Transportation Research Part D Transport & Environment*, 14(8):593–598.
- [18] Jiang L, Kronbak J, Christensen L. (2014). The costs and benefits of sulphur reduction measures: sulphur scrubbers versus marine gas oil. *Transportat. Res Part D: Transp. Environ.* 28:19–27.

# Flight Delay Propagation Analysis Based on the Mechanism of the Susceptible-Infected-Susceptible Model

Weiwei WU<sup>a,1</sup>, Haoyu ZHANG<sup>a</sup>, Jinfu ZHU<sup>a</sup> and Frank WITLOX<sup>b,a</sup>

<sup>a</sup>*College of Civil Aviation, Nanjing University of Aeronautics and Astronautics,  
210016 Nanjing, Jiangsu, China*

<sup>b</sup>*Department of Geography, Ghent University, Krijgslaan 281/S8,  
B9000 Gent, Belgium*

**Abstract.** This paper investigates flight delay propagation in the air transport networks. An integrate flight-based susceptible-infected-susceptible (FSIS) model is generated using the mechanism of epidemic spreading. Furthermore, the propagation probability in the FSIS model is analyzed through the regression model and later applied to China Easter Airline. The results show that propagation probability varies from different routes, which related to the flight frequency of airports, route distances, scheduled buffer times, and propagated delay times, and the FSIS model can efficiently reveal the process of flight delay propagation, and evaluate the number of delayed flights.

**Keywords.** Air transport network, flight delay propagation, susceptible-infected-recovered model, delay propagation probability

**CLC number.** [U8]      **Document code.** A

## 1. Introduction

In recent years, China has experienced unprecedented growth in the demand of aviation industry, the total flight movements has double increased from past 10 years. Meanwhile, the problem of flight delay has caused more public attention. Generally, propagated delay occurs because of connected resources such as aircraft, crew, passengers and airport resources that involved in an initially delayed flight and flights downstream (Kafle and Zou, 2016). Similarly, a congestion airport may propagate the delay to its connected airports through the affected delay flights, and eventually involve a significant part of the network Fleurquin et al., 2013). On the other hand, with the different route and airport attributions such as route distance and airport flight frequency, as well as the flight scheduling of airlines such as scheduled buffer time for each flight, the behavior of delay propagation is more complicated, and consequently its characteristics are more diverse.

---

<sup>1</sup> Corresponding author. Funding: National Natural Science Foundation of China (71201081); Key scientific research projects of NUAA (NZ2016109).



At present, it is found that delay of the upstream flight is the main reason for the delay of downstream flight, and several researchers focused on queuing network modeling for the propagation of delays in the air transport network. Baspinar et al. (2016) constructed an airport based queuing network model for simulating delay propagation and analyzing the total network delay. Airport-specific critical capacity values are observed and airports that are operating below these critical value, can be significant jump in the value of total delay. Pyrgiotis et al. (2013) proposed an approximate network delays model based on queuing engine and delay propagation algorithm to study the complex phenomenon of delay propagation between airports, and is applied to the network composed of 34 busy airports in the United States. The results show that, in some major airports, especially in hub airports, the propagated delay tends to put off traffic demand. Meanwhile, studies also found that delay may magnified under the coupling of multi-airport system, and consequently lead to more delay flights and longer delay time (Kondo A., 2009). Shervin et al. (2010) analyzed slack between sequential flights in the planned scheduled when delay occurs, and showed how delay propagation can be reduced by redistributing the existing slack in the planning process. The econometric analysis from Kafle and Zou (2016) quantified how much propagated delay will be generated out of newly formed delays that occur to each sequence of flights, and reveal the effects of various influencing factors on the initiation and progression of propagated delays.

The literature above investigated the flight delay propagation mainly focus on propagated delays between successive flights, and still lack of researches on the process of delay propagation in the view of whole network. Moreover, due to the factors on different routes and airport attributions, the characteristics of delay propagation in complex air transport network became more complicated, especially considering both the flight scheduling of airlines and the actual operation in airports. Thus, the SIS model is utilized to understand the characteristics of delay propagation in air transportation network, by assuming the process of disease spreading and delay propagation are similar (Baspinar and Koyuncu, 2016).

In the epidemic model literature, the basic compartment model was first introduced by Bernoulli (1760) in the 18th century (Bacaër et al., 2011). The epidemic model has long been used to study the behavior of spreading process such as disease or rumor (Lin et al., 2014; Wang et al., 2014). There are also several researches on epidemic spreading base on different network (Sun et al., 2016; Zhu et al., 2013). From Yang et al. (2012), a bus transportation network is abstracted into complex network by using P approach, and the Susceptible-Infective-Susceptible (SIS) model is applied to simulate the epidemic spreading behavior and obtained the theoretical spreading threshold, which enabled the disease disappeared. Infection delay and propagation are considered by Xia et al. (2012) when analyzed epidemic spreading through individuals traveling.

Our paper aims to provide regression analysis of propagated delays between different routes as well as propose the flight-based SIS (FSIS) model to analyze the process of flight delays. As a result, the characteristics of delay propagation in the whole network can be concluded.

The whole paper is divided as the following three issues: Firstly, the FSIS model is improved from the SIS model. Under the mechanisms of epidemic spreading in the SIS model, flight delay spreading rules and the effective delay propagation probability

is redefined to properly describe flight delay propagation in the air network; Secondly, a regression model is proposed to analyze the delay propagation probability on different routes, which is related to airport flight frequency, route distance, scheduled buffer times, and propagated delay times. And the estimation results show that the delay propagation probability varies with different routes; and, finally, the process of flight delay propagation is studied based on the FSIS model, and factors that impact delay propagation are analyzed from the actual delay profile.

The rest of the paper is organized as follows. Section 2 explains the basics of epidemic spreading process, and describes flight delay propagation in air transportation network by redefining the spreading rules and effective delay propagation probability. Consequently, the regression model for delay propagation probability is presented. In Section 3, the delay propagation probability in specific airports are focused on, and the process of flight delay propagation is analyzed by using the FSIS model. And Section 4 is the conclusion.

## 2. Delay Propagation Model

### 2.1. Delay Propagation Process

The classical epidemic model is effective to observe the spreading behavior of some kind of disease between individuals (Dietz and Ja, 2002). One of the basic models is the SIS model (Bacaër et al., 2011), where  $S$  represents healthy individuals but susceptible to be infected, and  $I$  represents infected individuals and able to recover. With the dynamics of the population (or network), the healthy and infected individuals get into contact, thus, the infected individuals can transfer from  $I$  to  $S$  with recovery rate  $\delta$ , and healthy individuals may transfer into  $I$  with infection rate  $\beta$ . The cyclic relationship is depicted in Fig. 1.

The flight delay spreading in air network can be realized in a similar vein. A delayed flight may propagate the delays to its succession flight through some shared resources, such as aircraft, crew, and passengers (Asfe et al., 2014). It may also lead to new added delay to other flights through shared airport resources during its ground phase (Zámková and Prokop, 2015). This delayed flight causes an infection, and the infection may spread to the whole network by more infected delay flights.

In a methodological manner, epidemic spreading and delay propagation resemble each other. Let  $N$  to be the total number of flights in the network, and  $N_i(t)$  is the number of delayed flights at time  $t$ . Due to the sufficient scheduled buffer time, a flight with departure delay may have an early arrival; conversely, a flight succeeding to its early/on time arrival inbound flight may still depart late for the tight schedule time or lack of

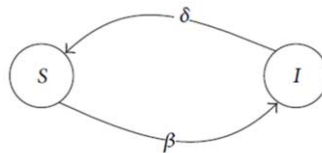


Figure 1. Epidemic processes with  $S$  susceptible and  $I$  infected.

airport resources. Thus, all flights affect and are affected in probability, that is, the infection rates  $\beta_{ij}$  and recovery rates  $\delta_i$  varies between each sequence flights  $i$  and  $j$ , which can be related to route traffic, block time, scheduled buffer time, and departure delay times. Suppose there are only two kinds of flights in the network at time  $t$ , one is normal flights, and the other is delayed flights. The dynamics of SIS model can be written as (Nowzari et al., 2016):

$$\dot{I}_i(t) = \delta'_i I_i(t) S_j(t) - \mu'_i I_i(t) \tag{1}$$

$$\dot{S}_j(t) = -\delta'_j I_i(t) S_j(t) + \mu'_j I_i(t) \tag{2}$$

where  $I(t) = Ni(t)/N$  and  $S(t) = (N - Ni(t))/N$  are the fraction of delayed and normal flights.

Generally,  $\lambda'_{ij} = \delta'_{ij}/\mu'_i$  is the effective infected rate, and without loss of generality, let  $\mu'_i = 1$ . We will focus on the value of  $\lambda'_{ij}$  in the next section. As  $S(t) = 1 - I(t)$ , the fraction of delayed flights at time  $t$  can be translated into the following form:

$$\dot{I}_i(t) = \sum_{j=1}^N \lambda_{ij} I_j(t) (1 - I_i(t)) - I_i(t) \tag{3}$$

The first item in the right of the Eq. (3) means normal flight  $i$  can be infected with probability  $\lambda'_{ij}$  through all its connected delayed flight  $j$ .

For an air transportation network with elements such as: airports, flight routes and so on, it can be abstracted as a directed graph with nodes and edges similar to epidemics network model. In our paper, as we focus on delay propagation between connected flights, it appears to be a flight-based network with individual flight as the nodes, and the connectivity between sequence flights as the edges. With the mechanism of epidemic spreading, flight delay propagation over air traffic network can be observed through a FSIS model. In this model, an infected/delayed flight is defined as to be delayed greater than 15 minutes. The following data-driven algorithm is used to construct FSIS model, which allows us to estimate the fraction of infected flights  $I_i(t + 1)$  through data-driven statistical analysis, with the input of  $I_i(t)$  at each time  $t$ . To better understand the behavior of flight propagation, we utilized historical flight data and verified by certain flight delays in the following section.

The information from historical flight data should consists of two parts: the schedule flight dataset  $Fs$ , including the schedule departure/arrival time; and the actual flown dataset  $Fa$  (flight-track data), which include the actual arrival/departure time. In addition, route traffic is also needed. From dataset  $Fs$  and  $Fa$ , the state of individual flight can be calculated, and consequently obtain the fractions of  $I_i(t)$  at time  $t$ . Then, fractions of  $I_i(t + 1)$  are estimated by  $I_i(t)$  and  $\lambda'_{ij}$  base on the flight delay propagation process of Eq. (3), where the effective infected rate  $\lambda'_{ij}$  is constructed from a Regression-Model in Section 3.

Data-driven algorithm based on FSIS model:

**Input:**  $I_i(t)$ ,  $Fs$ ,  $Fa$ ;

**output:**  $I_i(t + 1)$ ;

```

for  $i$  do
  for  $j$  do
    Generate  $\lambda_{ij}^t$  from Regression-Model;
    Calculate  $I_i(t + 1)$  through Eq. (3);
  
```

### 2.2. Propagated Delay

To better understand the amplifying or mitigating factors impact on the propagated delay between flight sequences, we focus on delay propagated from upstream flight, and consequent arrival delay of its succeed flight, together with other route attributes such as flight distance, scheduled buffer time, and frequency, which influence both the schedule block time, and the overall network delay performance.

Since our aim in this section is to find out the different degree on how these factors affect the delay propagation along flight legs, we firstly confirm the propagated delay by using the definition (Lan, 2006): delay that occurs when the aircraft to be used for a flight leg is delayed on its prior flight leg. This delay is a function of an aircraft’s routing. Moreover, to further explain how propagated delay occurs, Fig. 2 illustrates the relationships between departures, arrivals, and delays with the schedule flight data set and the actual flown data set.

The solid arrows represent the original schedule for two sequence flight  $i$  and  $j$ . The dotted arrows represent the actual departures and arrivals of these flight legs. Denote explanations are as follows:

- PDT**: the scheduled departure time
- ADT**: the actual departure time
- PAT**: the planned arrival time
- AAT**: the actual arrival time
- PCT**: the scheduled turn-around time
- MCT**: the minimum turn-around time

Then the slack of flight  $j$  (which specified the ground buffer time) is the difference between **PCT** and **MCT**:

$$PCT_j = PDT_j - PAT_j \tag{4}$$

$$Slack_j = PCT_j - MCT_j \tag{5}$$

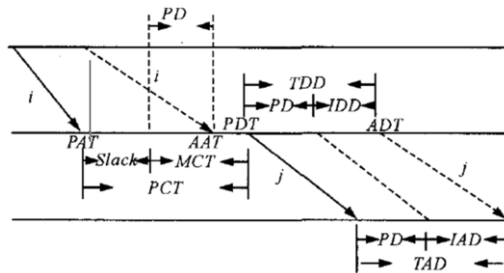


Figure 2. Generate of propagated delay.

Consequently,  $PD_{ij}$ , the propagated delay between sequence flight  $i$  and  $j$  can be determined:

$$\begin{aligned} TAD_j &= AAT_j - PAT_j \\ &= PD_{ij} + IAD_j - Buffer_j \end{aligned} \quad (6)$$

$$PD_{ij} = \text{Max}(TAD_j - IAD_j - Buffer_j, 0) \quad (7)$$

Where  $TAD_j$  refers to the total arrival delay,  $IAD_j$  is the independent arrival delay, and  $Buffer_j$  comprises two parts, including ground  $Slack_j$  and adjusted time during block phase.

### 2.3. Delay Propagation Probability

Scheduled turn-around and block-to-block times play an important role in absorbing and reducing primary and subsequent propagated delays. In flight schedules, buffer times for flight legs are often set in both airport ground and block time phase. The method proposed by Jetzki (2009) is used to determine the buffer time here (Jetzki, 2009), which correspond to the scheduled flight buffer time and ground buffer time respectively, using flight dataset  $F_s$  and  $F_a$ .

The traditional notation  $V'_{ij}$  is used to represent the propagated delay utility between upstream flight  $i$  and its succeed flight  $j$  at time period  $t$ , where  $V'_{ij}$  is a function of flight and route attributes:

$$V'_{ij} = f(PD'_{ij}, Buffer'_j, Traffic'_{D_j}, Distance_{ij}, K'_{ij}) \quad (8)$$

Where

$V'_{ij}$  is an utility function of propagated delay of flight  $j$ , which related to the propagated delay from flight  $i$ ;

$PD'_{ij}$  is the propagated delay between sequence flight  $i$  and  $j$ ;

$Buffer'_j$  is total scheduled buffer time of flight  $j$ ;

$Traffic'_{D_j}$  is the traffic of airport  $D_j$  at time  $t$ , where  $D_j$  is the departure airport of flight  $j$ , which is also the arrival airport of flight  $i$ ;

$Distance_{ij}$  is the route distance between where flight  $i$  departure and flight  $j$  arrived;

$K'_{ij}$  is an route-specific factor including characteristics of its individual stochastic and efficiency at time  $t$ .

Then, the utility function of Eq. (8) is specified as follow:

$$V'_{ij} = \alpha \ln(Traffic'_{ij}) + \beta \ln(Distance'_{ij}) + \gamma \left( \frac{Buffer'_j}{PD'_{ij}} \right) + K'_{ij} \quad (9)$$

The whole process of delay propagation for those resource shared flights, always start at receiving propagated delay from its upstream flight at the departure airport, until arriving at its arrival airport (and may propagate the delay to its own succeed

flight if there's any arrival delay). During the whole process, including both ground and block phase of the infected flights, the propagated delay may be magnified or absorbed. Propagated delay from upstream, congestion in the departure airport, etc., may even magnify the propagated delay that will propagate to the succeed flight; on the other hand, the propagated delay may not experience as greatly because the effects are dampened out through the buffer time airlines place in their schedule to allow for uncertainty factors (Wang, 2003). Thus, with the scheduled buffer time and the adjusted block time during fly time, an infected flight is considered in the process of recovery. If the departure delay is totally absorbed, and there's no arrival delay, the "infected" flight is no more infectious, which may also regard as the ending of the propagated delay with no longer propagation.

A specific example: flight  $j$  is an infected flight with propagated delay from its upstream flight  $i$ , however, due to the scheduled buffer time and adjusted block time, there's no arrival delay when it arrived. That is, the propagated delay of flight  $i$  isn't propagated again by flight  $j$ . Thus, at the time of arrival, the recovery rate of flight  $j$  in Eq. (1) is  $\mu_i = 1$ .

As delays will eventually diminished and disappeared with infinite time, what we focus on is how the propagated delay that flight  $j$  received will impact its own arrival time and will further propagate to its succeed flight. Thus, delay propagated and absorbed interaction is considered in both ground and block phase, and arrival delay of an infected flight is utilized to exhibit the impact of delay propagated from its upstream flight. Moreover, recovery rate is set to be 1, not only for model simplified, but also to observe the effective infected rate, namely  $\lambda'_{ij} = \delta'_{ij}/\mu'_i$ , that enables us to identify the infectious relationship between sequence flight  $i$  and  $j$ .

A multiplier is provided here to describe the probability of delay propagation (Kondo, 2009).

$$p'_{ij} = TAD_j / TDD_j \tag{10}$$

Where  $TDD_j = ADT_j - PDT_j$ , is total departure delay time. Multiplier  $p'_{ij}$  is defined to see "how much" percentage of delay time is reserved during the whole process, where the value of  $p'_{ij}$  can be either greater than 1 or less than 1. Since a delayed (infected) flight affects the flights mostly on their actual arrival slots, the probability  $\lambda'_{ij}$  that an infected flight  $i$  may propagate the delay to its succeed flight  $j$ , is the exponential function of its utility divided by the total arrival delay of all other flights in the arrival airport of flight  $j$  during time period  $t$ :

$$\lambda'_{ij} = \frac{e^{V'_{ij}}}{\sum_{m \in F'_{A_j}} e^{V'_{im}}} = \frac{P'_{ij}}{\sum_{m \in F'_{A_j}} P'_{km}} \tag{11}$$

Where  $A_j$  is the arrival airport of flight  $j$ ;  $F'_{A_j}$  is the set of all the arrival flights in  $A_j$  at time  $t$ ;  $F_m$  is the set of all upstream flights for the flights in  $F'_{A_j}$ .

Based on the utility function proposed in Eq. (9) and the definition of infected rate in Eq. (11), the infected rate  $\lambda'_{ij}$  between sequence flights at time  $t$  can be expressed as follows:

$$\lambda_{ij}^t = \frac{\exp(\alpha \ln(Traffic_{ij}^t) + \beta \ln(Distance_{ij}^t) + \gamma (\frac{Buffer_j^t}{PD_{ij}^t}))}{\sum_{m \in F_{A_j}^t}^{k \in F_m} \exp(\alpha \ln(Traffic_{km}^t) + \beta \ln(Distance_{km}^t) + \gamma (\frac{Buffer_j^t}{PD_{ij}^t}))}$$

$$= \frac{(Traffic_{ij}^t)^\alpha + (Distance_{ij}^t)^\beta + (\exp(\frac{Buffer_j^t}{PD_{ij}^t}))^\gamma}{\sum_{m \in F_{A_j}^t}^{k \in F_m} (Traffic_{km}^t)^\alpha + (Distance_{km}^t)^\beta + (\exp(Distance_{km}^t))^\gamma} \tag{12}$$

Thus, the quantified delay propagation probability  $\lambda_{ij}^t$  enables us to identify which factors during flight delay propagating should be pay more attention and/or adjusted.

### 3. Delay Propagation Based on the FSIS Model

#### 3.1. Regression Analysis of Delay Propagation Probability

The data for each variables in the regression model above are either directly available or can be obtained by using simple calculations from the AIMMS database product. The data of China Eastern Airline (MU Airline) are available from September to November in year 2016. We select 33 routes, mainly focus on 3 departure airports with other 11 arrival airports. The 3 departure airports include PEK, PVG, and SHA, and the other 11 airports are HGH, HRB, KHN, KMN, LHW, NGB, NKG, TAO, TYN, WUH, and XIY.

The statistical results from least square regression are shown in Table 1. The t statistics for all estimated coefficients are significant.

Form table above, the signs of estimated coefficients for the variables of frequency, distance, and the ratio of buffer time and propagated delay time are expected, and all are significant at the 5% level. The corrected  $R^2$  is 0.847. The traffic coefficient, however, is the only positive number which is 0.862. Based on the estimated coefficients, we find that, with the same distance and buffer/PD, if the traffic on route  $i-j$  is increased by 10%, the probability for a flight on route  $i-j$  to propagate its delay may increase by 8.563%. On the other hand, when the distance is longer, or the value of buffer/PD is greater by 10%, the probability may decrease by 2.801% and 11.971% respectively. The estimation results indicate that delay propagation probability  $\lambda_{ij}^t$  will increase with higher route traffic; while decrease with longer flight distance, longer buffer time, and/or shorter propagated delay.

**Table 1.** Statistical Estimation Results for delay propagation probability.

	Coefficients	t Statistic	P-value	Multiple R	R <sup>2</sup>
$K_{ij}^t$	-6.195	-4.109	3.32E-4		
$Traffic_{ij}^t$	0.862	10.156	1.01E-10		
$Distance_{ij}^t$	-0.298	-1.792	8.43E-2	0.920	0.847
$(Buffer/PD)_{ij}^t$	-1.275	-6.328	8.96E-7		

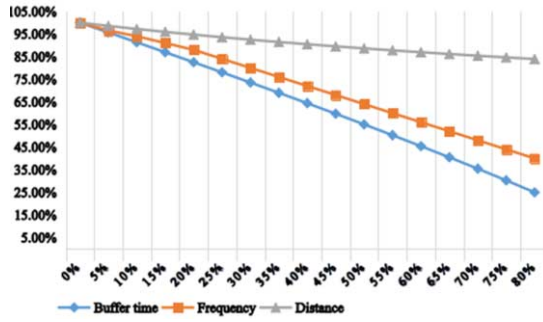


Figure 3. Changes of  $\lambda_{ij}$  based on frequency, distance and buffer time changing.

We also find that the coefficient for buffer/PD is  $-1.275$ , which means that, the flight with longer buffer time or/and shorter propagated delay will have an even lower probability for delay propagating. Moreover, the absolute value of buffer/PD coefficient is much higher than those of distance and traffic coefficients. Therefore we expect that flights can obtain lower  $\lambda'_{ij}$  from increasing buffer time than from decreasing frequency or longer distance. Figure 3 shows how  $\lambda'_{ij}$  changes in three circumstances, where all variables are remain the same with the base data on the 33 routes above, except the changing variable in each condition. (1) decreasing traffic from 10% to 70%; (2) increasing distance from 10% to 70%; (3) increasing buffer time from 10% to 70%. The chart shows clearly that  $\lambda'_{ij}$  decreases much faster with the increase of buffer time than from decreased route traffic or longer distance.

It can be seen, curves of  $\lambda'_{ij}$  are linear decreased in all conditions, where delay propagation probability is more sensitive to frequency and buffer time, and far less sensitive to distance. These results indicate that traffic in departure airport and the scheduled buffer time play an important role in impacting flight delay. It is easy to understand: more flight movements may lead the airport to congestion, which leave the aircraft queuing on the runway, and straightly increases the flight ground waiting time. The scheduled buffer time can also be a great help in decreasing  $\lambda'_{ij}$ . However, consider the utilization of aircraft rotation, it is often less other than sufficient buffer time in flight scheduling. Lastly, there is usually little to do in the block phase, except to speed adjustment, and it is easier to shorten the block time with longer distance.

From Table 1, we first calculate the delay propagation probability between 21 selected airports, which is shown in Fig. 4. The increasing value of  $\lambda'_{ij}$  is marked from light to dark, indicating that the delay propagation probability are various with different routes.

### 3.2. Flight Delay Analysis of Airline

In this section, the evaluated delayed flight rates are presented based on the FSIS model to analyze the process of delay propagation, and the efficiency of delay propagation probability is also verified by comparing with the actual delay profile of MU Airline.

As flight delay in the network often involves time delays due to its actual flight time, the distribution of flight times of China Easter Airline is analyzed and it is observed that 79.48% of flights have flight times between 1 hour to 3 hours, and the av-



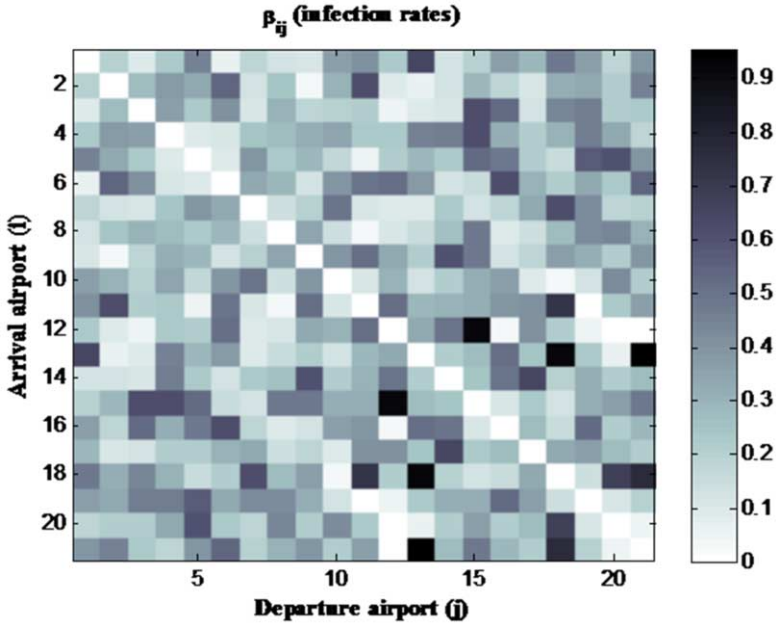


Figure 4. Delay propagation probability  $\lambda'_{ij}$  between 21 selected airports based on data from MU Airline.

erage flight time is 122 minutes. Thus, a 2-hour time window is chosen to be the delay spreading periods that enable us to see the complete transition on the dynamics of the network. It means that  $I_i(t + 2)$  is the fraction of delay flights which evaluated from  $I_i(t)$  2 hours later, and can be approximated through the updated delay propagation probability  $\lambda'_{ij}$  by solving the data-driven algorithm in Section 2.

The real and evaluated rates of delayed flights of MU Airline are shown in Fig. 5. It is seen that the model provides a reliable approximation on the delay spreading. However, two noticeable difference can be concluded. In the solid line, there are two obvious peak at time around 8:30 and 10:30 in the morning (where they are just at a 2-hour interval, we also call it a “flight bank”), while the increasing of delay flight rate in the evaluated line is flatter. On the other hand, the rate of the two curves are both fluctuate smoothly but remain at a high level rate in the afternoon, except that the evaluated curve doesn’t decreased as sharp as the real one did. That is because the propagated delays absorbed in the afternoon is less than it absorbed in the morning which lead to high proportion of delayed flights. However, the smooth fluctuation reveals the impact of all propagated delay is lower in the afternoon due to the lower number of rotational frequencies, and the impact of primary propagated delay from morning flights plays a less important role on afternoon flight delays.

Jetzki (2009) pointed out that, airlines appear to be focus more on schedule adherence, that they may schedule more buffer time to ensure flights punctuality; while flight connectivity is more cared about in the afternoon. It can also be confirmed from  $\lambda'_{ij}$  in the FSIS model.

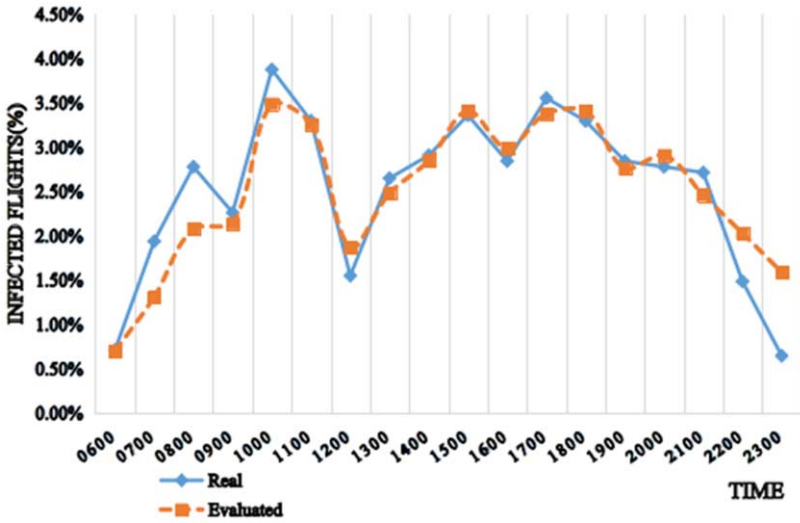


Figure 5. Comparison between real and evaluated fractions of delayed flights through the FSIS model.

#### 4. Conclusion

As more frequent flight delay nowadays, the objective of the study is to better understand the processes and characteristics of flight delay propagation in the air transportation network, as well as to identify factors which amplify or mitigate the delay propagation.

When an aircraft arrives late at its destination, the delayed inbound flight may not only propagate the delay to its next flight leg, but may also affect other flights within the network through their sharing resources. The FSIS model is presented based on the mechanisms of epidemic spreading in SIS model, by improving the spreading rules and the effective delay propagation probability in order to describe flight delay propagation in the air network properly. Then, a regression model is also constructed under the definition of delay propagation probability, to see how the infectious delayed flight is impacted by different airports and routes on a detailed microscopic level. Finally, with the actual flight-by-flight data of airlines, the simulation results show a reliable approximation on delay propagation, and provide some useful conclusion on flight delay controlling:

- (1) delay propagation probability of flight delay are various from different routes, which related to the flight frequency of airports, route distances, scheduled buffer times, and propagated delay times; And the reasonable scheduled buffer times is effective in absorbing and reducing primary and subsequent propagated delays.
- (2) The effect of airport flight frequency and scheduled buffer time on flight delays is contrary. Furthermore, the smaller airport is more sensitive to buffer time; while, flight frequency play a more important role on larger airport, which indicates that larger traffic during peak hours and shorter scheduled

buffer time for aircraft utilizing improvement are allowed in small and large airports respectively.

- (3) In the whole network, morning flights of airlines are more sensitive to primary propagated delays than afternoon flights which depends on the connectivity of flights, thus, schedule longer buffer time on the first sequential flight leg is an effective way to restrain delay propagation in the whole network.

The findings can help to improve airline and airport planning in order to achieve a higher level of resilience towards primary and propagated delays. Furthermore, the FSIS model aims to bridge the gap in the existing modeling tools between micro-simulations and macroscopic models, by providing more detailed insights on delay propagation, which can be useful for macroscopic analyses and simulations. For micro analysis often focus on delays between sequential flight legs, which require extensive resources and computational effort, but regardless of the delay propagation in the whole network; while macroscopic models are simple to use but typically lack the detailed impact of airport and route characteristics on delay propagation, and credible regression model for flight delay propagation probability on different routes.

## References

- S. AhmadBeygi, A. Cohn, M. Lapp, Decreasing airline delay propagation by re-allocating scheduled slack, *Iie Transactions*, **42** (2010), 478–489.
- M.K. Asfe, M.J. Zehi, M.N.S. Tash, N.M. Yaghoubi, Ranking different factors influencing flight delay. *Management Science Letters*, **4** (2014), 1397–1400.
- N. Bacaër, *A Short History of Mathematical Population Dynamics*, Springer, London, Dordrecht, Heidelberg, New York, 2011, pp. 21–30.
- B. Baspinar, E. Koyuncu, A Data-Driven Air Transportation Delay Propagation Model Using Epidemic Process Models, *International Journal of Aerospace Engineering*, **7** (2016), 1–11.
- B. Baspinar, N.K. Ure, E. Koyuncu, G. Inalhan, Analysis of Delay Characteristics of European Air Traffic through a Data-Driven Airport-Centric Queuing Network Model, *IFAC PapersOnLine*, **49** (2016), 359–364.
- D. Brockmann, D. Helbing, The hidden geometry of complex, network-driven contagion phenomena, *Science*, **342** (2013), 1337–1342.
- K. Dietz, H. Ja, Daniel Bernoulli’s epidemiological model revisited, *Mathematical Biosciences*, **180** (2002), 1–21.
- P. Fleurquin, J.J. Ramasco, V.M. Eguiluz, Systemic delay propagation in the US airport network, *Scientific Reports*, **3** (2013), 142–154.
- M. Jetzki, The propagation of air transport delays in Europe, *Department of Airport and Air Transportation Research of RWTH Aachen University*, 2009.
- N. Kafle, B. Zou, Modeling flight delay propagation: A new analytical-econometric approach, *Transportation Research Part B – Methodological*, **93** (2016), 520–542.
- A. Kondo, Delay Propagation and Multiplier. *Transportation Research Board Annual Meeting, Washington, US*, 2009.
- S. Lan, J.P. Clarke, C. Barnhart, Planning for Robust Airline Operations: Optimizing Aircraft Routings and Flight Departure Times to Minimize Passenger Disruptions, *Transportation Science*, **40** (2006), 15–28.
- Y. Lin, D. Jiang, S. Wang, Stationary distribution of a stochastic SIS epidemic model with vaccination, *Physica A Statistical Mechanics & Its Applications*, **394** (2014), 187–197.
- C. Nowzari, V.M. Preciado, G.J. Pappas, Analysis and Control of Epidemics: A Survey of Spreading Processes on Complex Networks, *Control Systems IEEE*, **36** (2016), 26–46.
- N. Pyrgiotis, K.M. Malone, A. Odoni, Modelling delay propagation within an airport Network, *Transportation Research Part C: Emerging Technologies*, **27** (2013), 60–75.

- M. Sun, H. Zhang, H. Kang, G. Zhu, X. Fu, Epidemic spreading on adaptively weighted scale-free networks, *Journal of Mathematical Biology*, **74** (2016), 1263–1298.
- J. Wang, L. Zhao, R. Huang, SIRaRu rumor spreading model in complex networks, *Physica A Statistical Mechanics & Its Applications*, **398** (2014), 43–55.
- P.T.R. Wang, L.A. Schaefer, L.A. Wojcik, Flight connections and their impacts on delay propagation, *Digital Avionics Systems Conference*, **9** (2003), 51–61.
- X.H. Yang, B. Wang, S.Y. Chen, W.L. Wang, Epidemic dynamics behavior in some bus transport networks, *Physica A Statistical Mechanics & Its Applications*, **391**(2012), 917–924.
- M. Zámková, M. Prokop, The Evaluation of Factors Influencing Flights Delay at Czech International Airports, *Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis*, **63** (2015), 2187–2196.
- G. Zhu, G. Chen, X.J. Xu, X. Fu, Epidemic spreading on contact networks with adaptive weights, *Journal of Theoretical Biology*, **317** (2013), 133–139.

# Research on Constant Power Charging and Discharging of Battery Based on LCL Filter

HUANG Mengtao<sup>a,1</sup>, ZHAO Jiamei<sup>a,2</sup>

<sup>a</sup>*School of Electric and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China*

**Abstract.** In allusion to the application of LCL filter in three-phase voltage rectifier circuit, this paper will apply LCL filter to battery energy storage system, compared with traditional single LC filter, we can use smaller inductance value to achieve better filtering effect, eliminating the problems caused by conventional filters, such as large volume, high cost, poor dynamic response of system. According to the requirements of the design of the whole system, by analyzing the circuit principle and control strategy of the system, using the three-phase voltage type PWM rectifier (VSR) and bidirectional DC/DC converter, through a simple design method of the parameters of the LCL filter, ultimately, the correctness of the theoretical analysis is verified by simulation and experiment and realize the constant power charging and discharging of the battery.

**Keywords.** LCL filter, bidirectional DC/DC converter, battery, VSR

## 1. Introduction

In the traditional power network, because of the defect of its own power supply mode, cannot meet the demand of power supply quality in current society. With the continuous development of new energy technology, micro-grid has become the main development direction of electric power industry in twenty-first Century. However, due to the power device in the micro network vulnerable to the impact of the external environment, so that the output power of the grid is intermittent, instability and unpredictability, which has a certain influence on the operating characteristics of the micro-grid. In order to ensure the balance and continuity of power supply, energy storage device has become an indispensable part of micro-grid. Storage battery energy storage system as one of the most mature energy storage technology, which can not only play the role of energy buffer, but also to improve the quality of micro-grid, meanwhile, it plays an important role to enhance the economic benefits of micro-network<sup>[1-4]</sup>.

In the battery energy storage system, the three-phase VSR as a connecting bridge between with power grid and energy storage system, it exist some advantages, such as,

---

<sup>1</sup> Huang Mengtao(1965-), female, Ph. D, Professor, master tutor, mainly engaged in computer control, intelligent systems and image based measurement and recognition of research .E-mail: 656228336@qq.com.

<sup>2</sup> ZHAO Jiamei (1992-), female, master's degree, research direction: Electrical theory and new technology .E-mail:493563551@qq.com.

bidirectional energy flow, controllable DC voltage and unity power factor operation and so on. The traditional three-phase VSR uses a single LC filter circuit, which can filter out the network side of the current harmonics in a certain range. However, in high-power applications, the switching frequency of three-phase VSR is relatively low, in order to limit the harmonic current within allowable range, we need for larger inductance to meet the design requirements of power system, which will lead to the stability of the system and the dynamic response performance is reduced. In order to solve this problem, this article will use LCL filter instead of the traditional single LC filter and employ it in battery energy storage system. Through the design of LCL filter parameters, by means of voltage and current double loop control and current control strategy of monocyclic, adopts SVPWM modulation method, we can achieve the better filtering effect with smaller inductance value by the comparison of experiment and simulation, solves the problems caused by the traditional filter, so as to realize the constant power charging and discharging of the battery.

### 2. Circuit Principle

The main circuit of the constant power charging and discharging of storage battery is shown in Figure 1. The circuit is made up of  $L_g$ CL filter circuit, three-phase voltage type PWM rectifier circuit and bidirectional DC/DC converter circuit which can realize the function of constant power charge and discharge of battery.

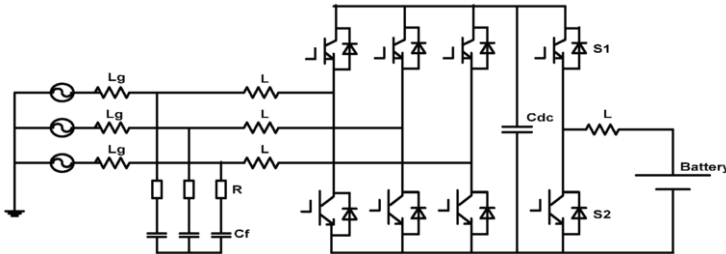
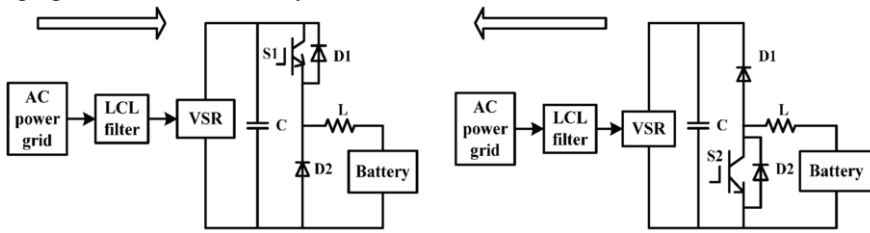


Figure 1. The main circuit of constant power charging and discharging of battery.

Figure 2 shows the equivalent circuit diagram in the charge and discharge of the battery. Figure 2 a) is the equivalent circuit when the battery is charged. The overall energy in the circuit flows from the AC power grid side to the battery, the battery absorbs energy from the grid, the three-phase voltage type PWM is in the rectified state, voltage and current of the grid is the same phase, at the same time, bidirectional Buck/Boost converter working in Buck mode, the switch S1 is on, S2 is off, realize charging function of the battery.



a) The equivalent circuit of battery charging      b) The equivalent circuit of battery discharging

Figure 2. The equivalent circuit diagram in the charge and discharge of the battery.

Figure 2 b) is the equivalent circuit when the battery is discharged. The overall energy in the circuit flows from the battery to the AC network side, the battery release energy to the power grid, the three-phase voltage type PWM is in the inverter state, voltage and current of the grid is the opposite phase, meanwhile, bidirectional Buck/Boost converter working in Boost mode, the switch S1 is off, S2 is on, achieve the function of battery discharge.

### 3. The Control Strategy of the System

#### 3.1. Double closed-loop control of three-phase voltage source PWM rectifier

To realize constant power of storage battery charging and discharging, it is necessary to the three-phase voltage-type PWM converter working in rectifier and inverter mode so as to realize the bidirectional power flow, thus, we adopt the voltage and current double-loop control strategy. Among them, the outer loop of the voltage is used to realize the constant voltage of the DC bus voltage, and the current inner loop is used to control the current of the net side as sine wave to realize the unit power factor operation. This control strategy is based on the instantaneous reactive power theory to implement, therefore, it is necessary to transform coordinate in the design, we should convert the three-phase static coordinates abc to two-phase rotation coordinates dq to achieve the control of the current loop, followed by two-phase rotation coordinates dq to the two-phase static coordinate system  $\alpha\beta$  to implement the control of three-phase voltage converter [6-7].

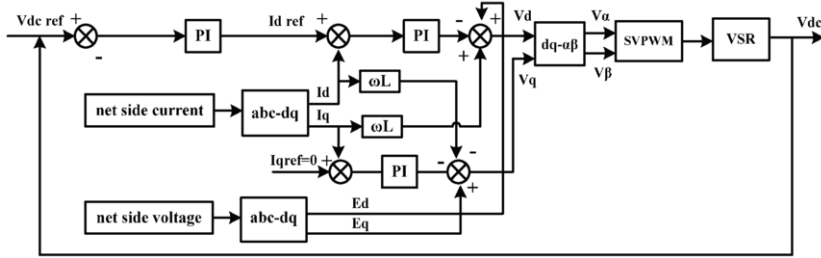


Figure 3. Block diagram of the control strategy of VSR.

The block diagram of the control strategy of three-phase voltage type PWM converter is shown in Figure 3. By comparing the direct current output voltage  $V_{dc}$  with the given direct current reference voltage  $V_{dcref}$  obtain a value pass the outer loop PI control to acquire the net side current active component reference  $I_{dref}$ . According to the reference value of the given active and reactive power from the network side compared with the current value after the coordinate transformation respectively and then sent to the PI regulator. In accordance with the nonlinear decoupling equation, under the two-phase rotating coordinate, the voltage value of three-phase AC side is as following formula:

$$\begin{cases} V_d = E_d + \omega L I_q - \left( K_{ip} + \frac{K_{ii}}{S} \right) (I_{dref} - I_d) \\ V_q = E_q - \omega L I_d - \left( K_{ip} + \frac{K_{ii}}{S} \right) (I_{qref} - I_q) \end{cases} \quad (1)$$

After the coordinate transformation, we obtain the voltage component of the three-phase network side voltage in the two-phase stationary coordinate system, which as the input control signal of voltage space vector. Eventually, we acquire the pulse signal of the three-phase PWM converter so as to achieve control of the converter.

### 3.2. The control strategy of bidirectional DC/DC converter

The bidirectional DC / DC converter acts as a bridge between the energy storage device and the grid side converter, and has the function of transmitting energy. The bidirectional DC / DC converter needs to be controlled in order to achieve the function of constant power charging and discharging of the battery. Due to the terminal voltage of the battery changes slowly in the process of charging and discharging, so we can through control the current of charging and discharging to control its power. The control of DC/DC converters is achieved by using current single ring when the voltage of direct current bus is constant. The block diagram of the control strategy is shown in figure 4. The current reference value  $I_{ref}$  is gained by using a given constant power divided by the voltage on both ends of the battery, which compared with current actual value of circuit  $I_b$ , get through PI regulating we obtain the control signal of the PWM so as to achieve constant power charging and discharging of battery.

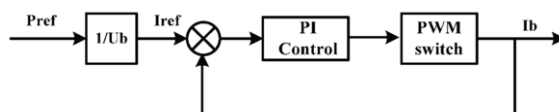


Figure 4. Control strategy of bidirectional DC/DC converter.

## 4. Parameter Design of LCL Filter

### 4.1. The choice of filter inductance $L_g$ on the network side<sup>[8]</sup>

When we choose filter inductance of the network side, we usually use the peak value of the current harmonic  $i_{rpmax}$  as a reference standard. In practical applications, the peak of the current harmonic  $i_{rpmax}$  usually use the maximum fundamental current peak of 15% can meet control requirements. The parameters selecting of filter inductance on the side of the net is as follows:

$$L_g = \frac{U_s}{2\sqrt{6}i_{rpmax}f_{sw}} \tag{2}$$

### 4.2. The choice of filter equivalent total inductance

The design of the total filter inductance  $L_T$  ( $L_T=L_g+L$ ) should not exceed 0.1pu, if the total inductance value is too large, we require a higher direct current voltage to achieve good current control performance, which will increase the loss of power switch<sup>[9]</sup>.Considering the three-phase voltage type PWM operating under the unit power factor, the total inductance value  $L_T$  of the LCL filter should satisfy the following formula:



$$L_T \leq \frac{\sqrt{(MU_{dc})^2 - U_{sm}^2}}{\omega_b I_{1m}} \quad (3)$$

Among them: M is the maximum utilization of the PWM phase voltage (related to the PWM control mode). If using SPWM control, M=0.5; if using SVPWM control, M=0.577.  $U_{dc}$  is the voltage value of the direct current side;  $U_{sm}$  is the phase voltage amplitude of the grid;  $I_{1m}$  is the phase current peak of the fundamental wave of the grid;  $\omega_b$  is the base wave angular frequency of the grid.

In order to reach certain harmonic filtering requirements, the inductance value cannot be too small, the smaller the inductance value, the worse the ability to filter the harmonic. Thus, the maximum of ripple current<sup>[10]</sup>:

$$i_{rp\max} = \frac{U_{dc}}{8L_T f_{sw}} \quad (4)$$

Among them:  $f_{sw}$  is the switching frequency for three-phase PWM rectifier.

Therefore, the total inductance of the LCL filter should satisfy the following equation:

$$L_T \geq \frac{U_{dc}}{8i_{rp\max} f_{sw}} \quad (5)$$

In conclusion, the range of the total inductance LCL filter  $L_T$  is as follows:

$$\frac{U_{dc}}{8i_{rp\max} f_{sw}} \leq L_T \leq \frac{\sqrt{(MU_{dc})^2 - U_{sm}^2}}{\omega_b I_{1m}} \quad (6)$$

#### 4.3. The choice of filter capacitor

In order to avoid the filter capacitor is too large so that the system produces larger reactive power, resulting in the power factor of the system is decreased. Therefore, in the design of the capacitance value, the reactive power absorbed by the capacitor should not exceed 5% of the rated active power of the system<sup>[11]</sup>: which satisfied the formula:

$$3U_c^2 \omega_b C_f \leq 5\% P_N \quad (7)$$

Among them:  $U_c$  is the voltage value of the filter capacitor;  $P_N$  is the rated power of the system.

If the voltage drop on the network side inductance is relatively small, it is considered that the filter capacitor voltage  $U_c$  is approximately equal to the grid voltage  $U_s$ , which satisfied the formula:

$$C_f \leq 5\% \frac{P_N}{3\omega_b U_s^2} \quad (8)$$

#### 4.4. The calculation of resonant frequency<sup>[9]</sup>

In order to prevent the LCL filter from occurring resonance in the low frequency domain near the fundamental wave and in the high frequency region near the switching

frequency harmonics, the resonant frequency of the LCL filter should satisfy the following formula:

$$10f_b \leq f_{res} = \sqrt{\frac{L_g + L}{L_g LC_f}} \leq 0.5f_{sw} \tag{9}$$

Among them:  $f_b$  is the fundamental frequency of the grid;  $f_{sw}$  is the switching frequency of the three-phase PWM rectifier.

#### 4.5. The calculation of damping resistor $R_d$

Capacitor branch series damping resistance is to prevent the LCL filter from resonance occurs, but resistance cannot too big, so as to bring too much loss so that lead to the stability of the system performance become poor. Therefore, the value of  $R_d$  is set to the resonance point about 1/3 of capacitive reactance<sup>[12]</sup>, as shown in equation (10):

$$R_d = \frac{1}{3\omega_{res} C_f} \tag{10}$$

Among them:  $\omega_{res}$  is resonant angular frequency.

### 5. System Modeling and Simulation Analysis

#### 5.1. System modeling

According to the above analysis and design, we use Matlab/simulink to build simulation model. The simulation model of the main circuit is shown in Figure 5, of which, the subsystem module is three-phase voltage source PWM control model; Subsystem1 module is bidirectional DC/DC converter control model. The known parameters set in the simulation are as follows: The rated power  $P_N = 50kW$ ; The effective value of the grid phase voltage is 220V; The fundamental frequency of the grid  $f_b = 50HZ$ ; The switching frequency of the three-phase voltage type PWM  $f_{sw} = 5kHz$ ; The output voltage of the direct current side  $V_{dc} = 700V$ ; The switching frequency of bidirectional DC / DC converter  $f_{PWM} = 10kHz$ ; Battery voltage is 400V, power is 10kW; Simulation time is 2s.

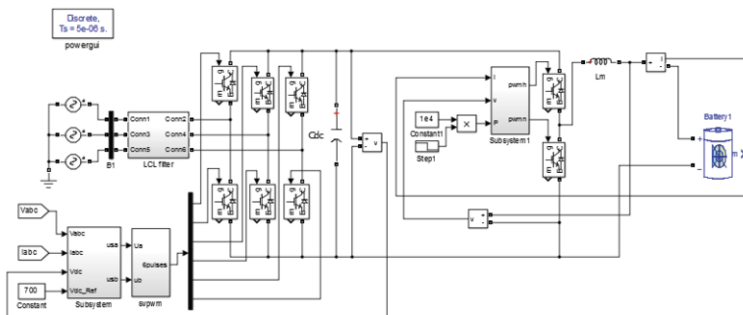


Figure 5. The simulation model of the main circuit

The control circuit diagram of three-phase voltage type PWM as shown in Figure 6, because of MATLAB / Simulink software identify the normalized value, so the voltage and current should be standardized, the voltage value through the PLL to obtain

the value of sine and cosine that synchronization with the grid voltage, so as to know the angle of transformation. We obtain the value of voltage and current of the network side in the two-phase rotation coordinates through the coordinate transformation. Eventually, we realize the control of the three-phase PWM converter by adopting the control strategy of voltage outer loop and current inner loop and SVPWM control algorithm.

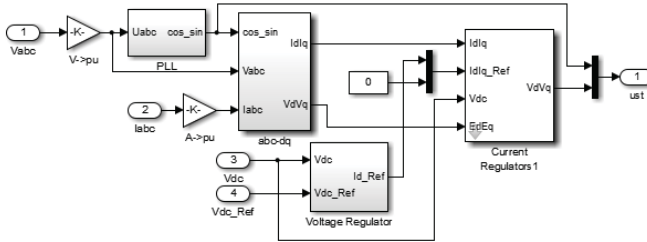


Figure 6. The control circuit diagram of three-phase voltage type PWM

5.2. Simulation and analysis of the system

By given the known parameters, using the formula (2) - (10) to calculate all parameters of the LCL filter: Filter inductance of net side  $L_g = 0.557$  mH; The range of filter total inductance values is:  $1.09\text{mH} \leq L_T \leq 7.65\text{mH}$ ; Filter capacitance  $C_f = 54.8$  uF; The scope of resonance frequency is:  $500\text{HZ} \leq f_{res} \leq 2500\text{HZ}$ . Other simulation parameters are set as shown in Table 1.

Table 1. Simulation parameters of LCL filter

Total inductance of LCL filter $L_T$	Inductance of alternating current side $L$	Damping resistance $R_d$
1.2mH	0.623mH	0.8Ω
3mH	2.443mH	0.96Ω
6mH	5.443mH	3Ω

By debugging and calculation set PI parameters of the current inner ring is:  $K_{pi} = 1$ ,  $K_{ii} = 5$ ; PI parameters of the voltage outer loop is:  $K_{pv} = 0.005$ ,  $K_{iv} = 0.02$ . The filter capacitor on the direct-current side:  $C = 6000\mu\text{F}$ . PI parameter of the bidirectional DC/DC current loop is: the  $K_{pi} = 10$ ,  $K_{ii} = 20$ , inductance of direct-current side  $L_m = 5$  mH.

(a) When the total inductance of LCL filter  $L_T = 6$  mH , we carry out simulation for the system, the network side of the voltage and current simulation waveform is shown in Figure 7; Figure 8 is the voltage waveform of the DC side; Power waveforms at both ends of the battery are shown in Figure 9.

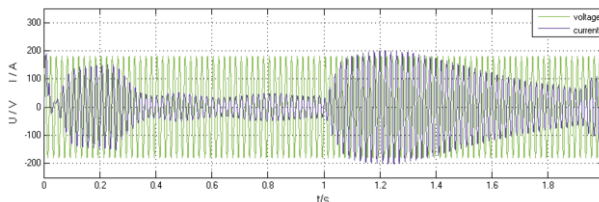


Figure 7. The network side of the voltage and current simulation waveform

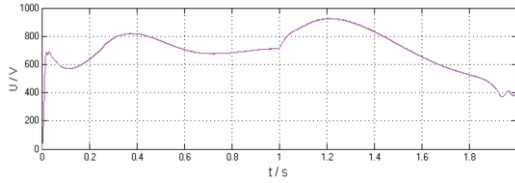


Figure 8. The voltage waveform of the DC side

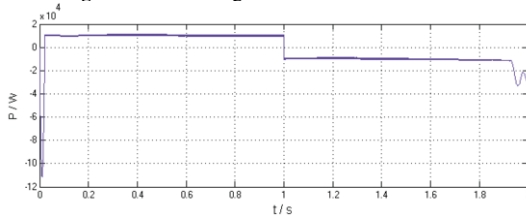


Figure 9. Power waveforms at both ends of the battery

(b) When the total inductance of LCL filter  $L_T = 3 \text{ mH}$ , we carry out simulation for the system, the network side of the voltage and current simulation waveform is shown in Figure 10; Figure 11 is the voltage waveform of the DC side; Power waveforms at both ends of the battery are shown in Figure 12.

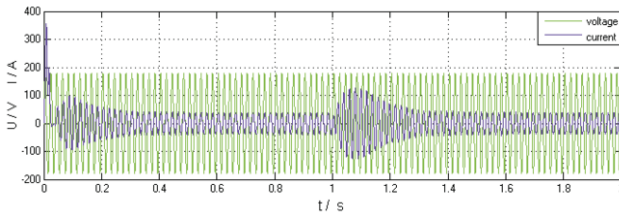


Figure 10. The network side of the voltage and current simulation waveform

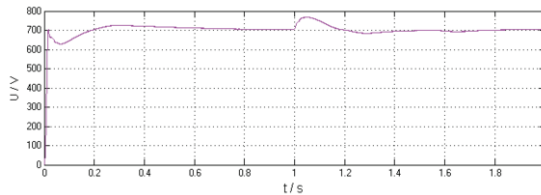


Figure 11. The voltage waveform of the DC side

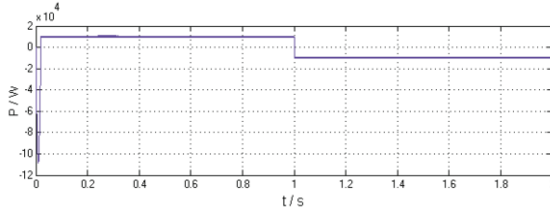


Figure 12. Power waveforms at both ends of the battery

(c) When the total inductance of LCL filter  $L_T = 1.2 \text{ mH}$ , we carry out simulation for the system, the network side of the voltage and current simulation waveform and its simulation waveform enlarged drawing is shown in Figure 13 and Figure 14, Figure 15 is the voltage waveform of the DC side; Power waveforms at both ends of the battery are shown in Figure 16.

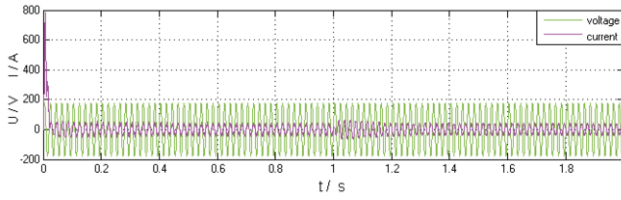


Figure 13. The network side of the voltage and current simulation waveform

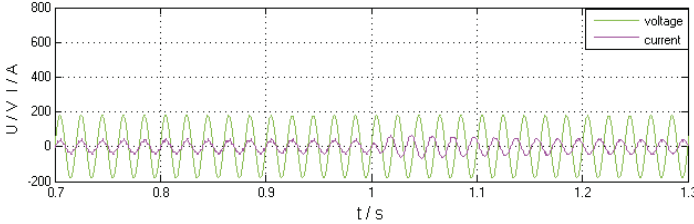


Figure 14. The network side of the voltage and current simulation waveform enlarged drawing

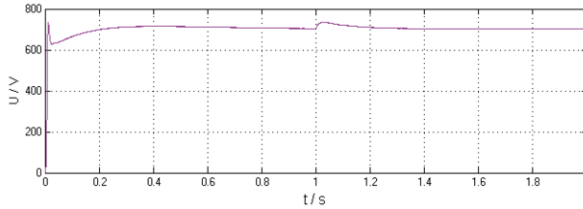


Figure 15. The voltage waveform of the DC side

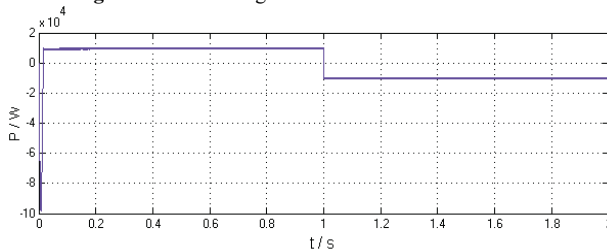


Figure 16. Power waveforms at both ends of the battery

(d) Analysis the simulation results.

From figure 7 to figure 15 can be seen: when the inductance value of the AC side is bigger, the oscillation amplitude of network side current is larger in the initial phase, The DC side voltage of the system is unstable so that the filtering effect of the system is poor and the charging and discharging power of the battery is also unstable. With the decrease of the inductance value of the AC side, we can see that current fluctuation of the grid side has been reduced so as to make the system achieve stable state. Based on the above analysis, when the network side of the inductance is 0.623mH, the performance of the system is the best, we achieve through a smaller inductance value to reach a better filtering effect. As we can see from figure 14 and figure 16, the power of the battery is 1kW before the 1s, the grid is charging for battery use constant power meanwhile, voltage and current of the grid is the same phase; After the 1s, the power of the battery is -1kw, and the battery is discharging by constant power to the grid, voltage and current of the grid is the opposite phase.

## 6. Conclusion

This paper proposes a simple design method for LCL filter, elaborates its design principle in detail. The simulation and experimental results show that the designed LCL filter can use the smaller inductance to filter out the current harmonics on the network side, which can not only effectively reduce the harmonic pollution of the device to the power grid but also come true the operation of the system in the unit power factor. Eventually, we realize the constant power to the battery charging and discharging and verify the correctness of the design parameters.

## References

- [1] Kong linyi, Liao liying, Zhang haiwu, etc. Application of battery storage system in power system [J]. Electrical switch, 2008 (5) : 61-64.
- [2] Jia hongxin, Zhang yu, Wang yufei, etc. The application of energy storage technology in wind power system [J]. Renewable energy, 2009, 6:10-15.
- [3] Zhou Lin, Huang yong, Guo ke, etc. A survey of energy storage technology for micro-grid [J]. Power system protection and control, 2011, 4 (39) : 147-152.
- [4] Hua Guanghui, He Weiguo, Zhao Dawei. Role of Energy Storage Technologies in the Construction of Strong Smart Grid[J].Power supply,2010,8(27) : 22-29.
- [5] Chen Yao, Jin Xinmin, Tong yibin, etc. LCL filter for three-phase voltage-type PWM rectifier [J]. Journal of electrical engineering, 2007,9,22 (8): 124-129.
- [6] Lin Fei, Du Xin. MATLAB Simulation of Power Electronics Application Technology [M]. Beijing: China Electric Power Press, 2009.
- [7] Liang Tao. Study on Simulation of Battery Energy Storage System Based on Different Models [D]. Chongqing University, 2015.
- [8] Zhang Zhengzhan. Study on control strategy of PWM rectifier based on LCL filtering [D]. Henan University of Science and Technology, 2014.
- [9] Guo Xizheng, You Xiaojie, Zhang Liwei,etc. Design method of LCL filter for three-phase voltage source PWM rectifier [J]. Applications device, 2010 (3):33-38.
- [10] Dahono P A. A control method for DC-DC converter that has an LCL output filter based on new virtual capacitor and resistor concepts[C]//2004 IEEE 35<sup>th</sup> Annual Power Electronics Specialists Conference. Aachen(Germany), 2004(1) : 36-42.
- [11] Liserre M, Teodorescu R, Blaabjerg F. Stability of photovoltaic and wind turbine grid-connected inverters for a large set of grid impedance values[J]. IEEE Transactions on Power Electronics, 2006, 21(1) : 263-272.
- [12] Liserre M, Blaabjerg F, Hansen S. Design and control of an LCL-filter-based three-phase active rectifier [J]. IEEE Transactions on Industry Applications, 2005, 41(5): 1281-1291.

# An Improved Quantum Evolutionary Algorithm for High Dimension Knapsack Problem

Li Hao , Ma Lei, Qin Na

*School of Electrical Engineering, Southwest Jiaotong university, Chengdu, China*

**Abstract.** Knapsack problem is a typical combinatorial optimization problem, which has attracted wide attention because of its simple description, difficult solution and wide application background. For the knapsack problem, we proposed an improved quantum evolutionary algorithm. In this algorithm, the qubits are initialized according to the value densities of their corresponding items, the angle of Q-gate is determined through direct comparison between the quantum chromosome and the optimal solution, random crossover method is taken to share information among chromosomes, and He gate is used to prevent from premature convergence. Finally, experiment demonstrates the effectiveness of the algorithm.

**Keywords.** knapsack problem, quantum evolutionary algorithm, qubit, Q-gate, He gate

## 1. Introduction

Knapsack problem is a typical combinatorial optimization problem, which has a wide range of application, such as resource allocation, investment decisions, material cutting, budget control, etc.<sup>[1]</sup>. The research on knapsack problem is of great significance both in theory and in practice. However, this problem belongs to a class of NP complete problems with  $O(2^n)$  complexity, the difficulty to solve it will increase exponentially as the problem size increases linearly. It is often difficult to solve in polynomial time through the traditional optimization method. Therefore, to use the stochastic optimization method to get its satisfactory solution has become the main solution to the current<sup>[2-5]</sup>.

Quantum Evolutionary Algorithm (QEA) is an intelligent optimization algorithm based on quantum computation. Compared to other intelligent optimization algorithm, it has many advantages such as small population, strong searching ability, ability to treat the balance between exploration and exploitation. In 2002, Han et al proposed a quantum-inspired evolutionary algorithm(QEA) based on the concept of quantum bit and superposition of quantum states<sup>[6-8]</sup>. This algorithm introduced the quantum state expression into the genetic code, taken the quantum rotation gate to evolve the chromosome, and using it to solve 0-1 knapsack problem, the experiment shows that this algorithm has better performance than conventional genetic algorithm. After then, QEA gained wide attention and was used in many fields<sup>[9-11]</sup>, but until now, the solution of knapsack problem is still one of its typical applications<sup>[12-14]</sup>.

In this work, we propose an improved Quantum Evolutionary Algorithm(IQEA) for knapsack problem. In this algorithm, the qubits are initialized according to the value densities of their corresponding items, the angle of Q-gate is determined through direct comparison between the quantum chromosome and the optimal solution, random crossover method is taken to realize information sharing among chromosomes,  $H\epsilon$  gate is used to prevent from premature convergence. Using it to solve the knapsack problem, the simulation results show that the algorithm has better optimization performance than common genetic algorithm(CGA) and QEA.

This paper is organized as follows. Section 2 describes the knapsack problem. Section 3 introduce QEA. Section 4 presents the improved QEA. And the experimental results are shown in section 5. Finally, concluding remarks follow in section 6.

## 2. Knapsack problem

The knapsack problem can be described as: given  $n$  items and a knapsack, select a subset of the items to put in the knapsack so as to maximize the profit, that is

$$\max f(X) = \sum_{i=1}^n p_i x_i \quad (1)$$

$$\text{Subject to } \sum_{i=1}^n w_i x_i \leq C$$

where  $w_i$  and  $p_i$ ,  $1 \leq i \leq n$  is the weight and the profit of the  $i$ -th item respectively.  $C$  is the capacity of the knapsack.  $x_i \in \{0,1\}$ ,  $x_i = 1$  if the  $i$ -th item is selected, otherwise  $x_i = 0$ .

## 3.QEA

### 3.1 QEA

QEA is a kind of probability optimization algorithm based on quantum computation concept and theory. In this algorithm, the smallest unit of information is called qubit. A qubit may be not only in the states 0 and 1, but also in a linear superposition of the two states, the state of a qubit can be represented as

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (2)$$

where  $\alpha$  and  $\beta$  are the probability amplitudes of the states  $|0\rangle$  and  $|1\rangle$  respectively,  $|\alpha|^2$  and  $|\beta|^2$  give the probability that the qubit will be found in the state  $|0\rangle$  and  $|1\rangle$ . Normalization of the state to unity guarantee

$$|\alpha|^2 + |\beta|^2 = 1 \quad (3)$$

Further, a chromosome can be represented as a string of  $n$ -qubits, that is

$$\left( \begin{array}{c|c|c} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \beta_1 & \beta_2 & & \beta_n \end{array} \right) \quad (4)$$



Observing the quantum chromosome, we can get a binary string. And using Q-gate, the chromosome can be updated. The  $i$ -th qubit is updated as follows:

$$\begin{pmatrix} \alpha'_i \\ \beta'_i \end{pmatrix} = \begin{pmatrix} \cos \Delta\theta_i & -\sin \Delta\theta_i \\ \sin \Delta\theta_i & \cos \Delta\theta_i \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \quad (5)$$

where  $\Delta\theta_i$  is the rotating angle. QEA keeps running until the population converge, and the optimization to the problem is implemented eventually.

The procedure of QEA is described as follows:

Procedure QEA

Begin

Initialize  $Q(t)$  at  $t=0$

Make  $P(t)$  by observing the states of  $Q(t)$

Evaluate  $P(t)$

Store the best solutions among  $P(t)$  into  $B(t)$

While (not termination condition) do

$t = t + 1$

Make  $P(t)$  by observing the states of  $Q(t-1)$

Evaluate  $P(t)$

Update  $Q(t)$  using Q-gates

Store the best solutions among  $B(t)$ ,  $B(t-1)$  and  $P(t)$  into  $B(t)$

Store the best solutions  $b$  among

If (migration condition)

Migrate  $b$  or  $b'_j$  to  $B(t)$  globally or locally, respectively

End

End

where  $Q(t) = \{q_1^t, q_2^t, \dots, q_m^t\}$ ,  $q_j^t = \begin{bmatrix} \alpha_{j1}^t & \alpha_{j2}^t & \dots & \alpha_{jn}^t \\ \beta_{j1}^t & \beta_{j2}^t & \dots & \beta_{jn}^t \end{bmatrix}$ ,  $P(t) = \{X_1^t, X_2^t, \dots, X_m^t\}$ ,

$B(t) \in X_j$ ,  $X_j^t = \{x_{j1}^t, x_{j2}^t, \dots, x_{jn}^t\}$ ,  $B^t \in X_j$ ,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, m$ ,  $m$  is the size of the population.

In the step of "Initialize  $Q(t)$  at  $t=0$ ",  $\alpha$  and  $\beta$  of all  $q_j^0$  in  $Q(0)$  are initialized with  $1/\sqrt{2}$ . It means that in each chromosome,  $q_j^0$  represents the linear superposition of all possible states with the same probably.

To obtain the binary solution, the step of "Make  $P(t)$  by observing  $Q(t)$ " is implemented. When  $Q(t)$  is observed,  $x_{ji}^t = 0$  or 1 of  $P(t)$  will determine according to the probability  $|\alpha|^2$  or  $|\beta|^2$ .

In the "update" operation, Q-gate is used to update the chromosome, it is as follows

$$U(\Delta\theta_{ji}) = \begin{pmatrix} \cos \Delta\theta_{ji} & -\sin \Delta\theta_{ji} \\ \sin \Delta\theta_{ji} & \cos \Delta\theta_{ji} \end{pmatrix}$$

where  $\Delta\theta_{ji}$  is the rotating angle of the qubit. It is determined from a lookup table according to the  $i$ -th bit of the best solution  $b'_j$  and the binary solution  $x_j^t$ .

In the “migration” operation, If the migration condition is met, a global migration or a local migration is taken to implement information sharing among chromosomes.

### 3.2 QEA and EDA

Essentially, QEA is a kind of estimation of distribution algorithm(EDA). Its each quantum chromosome can be seen as a probability model based on the solution space of the problem, the chromosome collapsing into a binary string is the sampling process based on the probability model, the chromosome evolution based on Q-gate is the learning process of the probability model. But differing from other EDA, QEA can have many probabilistic models. For most of EDA, the probability model is obtained through learning the best solution or part of better solutions, which will make the best solution and its adjacent area become the search focus in order to achieve good optimization effect. But on the other hand, the model is often just a simple unimodal function, its peak is corresponding to the best solution. Other solutions which are similar to the best solution in genes form will have a larger probability to be obtained, and the solutions far from the best solution, although which are possible to be obtained according to the model, the obtained probability is relatively small. Due to the complexity of the practical problems, the objective function is often a multi-peaks function. If the global best solution is far away from the peak of the model, it is difficult to be found so that the algorithm will be trapped into the local optimum. The single model feature of EDA limits its performance. In QEA, each model also have one peak like EDA, but many models can have different search focus, besides the focus of each model itself, the overlap area of more than one models will also have greater opportunity to be searched, which improved the optimal performance of QEA.

## 4. IQEA

### 4.1 Qubit with angle expression

In quantum computation theory, qubit can also be expressed with angle  $\theta$ . That is,  $\alpha$  and  $\beta$  can be expressed as  $\alpha = \cos \theta$ ,  $\beta = \sin \theta$ , correspondingly,  $|\psi\rangle = \cos \theta |0\rangle + \sin \theta |1\rangle$ , It also meet  $\cos^2 \theta + \sin^2 \theta = 1$ . When qubit is measured, 0 or 1 will be gotten with the probability of  $\cos^2 \theta$  or  $\sin^2 \theta$ , respectively.

Further, a quantum chromosome can be expressed as  $q = (\theta_1 | \theta_2 | \dots | \theta_n)$ , The Q-gate operation can be expressed as the sum of the angle  $\theta$  and Q-gate angle  $\Delta \theta$ , that is,  $\theta' = \Delta \theta + \theta$ . Obviously, the expression and evolution of the chromosome can be simplified sharply if the qubit is expressed with angle.

### 4.2 Determination of Q-gate angle

In QEA, how to determine the Q-gate angle is the key to affect the algorithm performance. The lookup table method used in the algorithm has many disadvantages, such as complicated, poor flexibility, difficult to determine the parameters, and not easy to modify once determine. Considering a binary string, its every bit is "0" or "1" can corresponding to a special case that the qubit angle is 0 "or  $\pi/2$ ". Thus, we can directly

compare the qubit and the optimal solution to determine the rotating angle. For a binary number  $b$ , we define its angle.

$$\theta_b = \begin{cases} 0 & b = 0 \\ \pi / 2 & b = 1 \end{cases} \tag{6}$$

Further, we define the calculation formula of the rotating angle:

$$\Delta\theta_i = (\theta_{b_i} - \theta_i) \cdot r \tag{7}$$

where  $\theta_i$  is the  $i$ th qubit of a quantum chromosome,  $\theta_{b_i}$  is the angle of  $i$ th bit of historical best solution of this chromosome, and  $r \in [0,1]$  is an adjusting factor.

### 4.3 Population initialization

Aim at the knapsack problem, we define the value density of the  $i$ th items

$$d_i = \frac{P_i}{w_i} \quad i = 1, \dots, n \tag{8}$$

Obviously, in the case of the knapsack capacity is used as much as possible, the greater the value density of items in the knapsack, the better.

In initialization of QEA, all qubits are set to  $\pi / 4$  so that all solutions are obtained with the same probability. But obviously, if the probability that the items with high value density be selected is greater, and the probability that the items with low value density be selected is smaller, the better effect will be gotten. Therefore, we consider to initialize qubit differently according to the value density of their items.

Firstly, all items are sorted by value density. Next, put the items into the knapsack according to the order until the requirements of capacity is not met. Let the number of items in the knapsack is  $m$ , we set the qubits corresponding to the first  $m/2$  items to be  $0.4\pi$ , the last  $(n-m)/2$  qubits to be  $0.1\pi$ , and the rest to be  $\pi / 4$ .

### 4.4 Information sharing

In QEA, it is very important to share the information about optimal solution among the chromosomes. [15] has proved that in QEA, stochastic information sharing is better to deal with complex large scale optimization problem than information sharing based on the fixed neighborhood structure. Therefore, we consider to take this way to replace immigration operation to achieve information sharing among chromosomes.

### 4.5 He gate

To avoid premature convergence, the concept of He gate is introduced, it can be defined

$$\theta_i'' = \begin{cases} \varepsilon & \theta_i' \leq \varepsilon \\ \theta_i' & \varepsilon \leq \theta_i' \leq \pi / 2 - \varepsilon \\ \pi / 2 - \varepsilon & \theta_i' \geq \pi / 2 - \varepsilon \end{cases} \tag{9}$$

where  $0 < \varepsilon \leq \pi / 2$ ,  $\theta_i'$  is the qubit angle after rotating. Can be seen, He gate makes the angle always in the area  $[\varepsilon, \pi / 2 - \varepsilon]$ . When  $\varepsilon \rightarrow 0$ , He gate will become a ordinary quantum rotation gate, and when  $\varepsilon \rightarrow \pi / 4$ , the quantum gate will be out of work.

The procedure of IQEA is described as follows:

### Procedure IQEA

Begin

Sort items by value density

Initialize  $Q(t)$  at  $t=0$

Make  $P(t)$  by observing the states of  $Q(t)$

Evaluate  $P(t)$

Store the best solutions among  $P(t)$  into  $B(t)$

While (not termination condition) do

$t=t+1$

Make  $P(t)$  by observing the states of  $Q(t-1)$

Evaluate  $P(t)$

Update  $Q(t)$  using Q-gates

Store the best solutions among  $B(t-1)$  and  $P(t)$  into  $B(t)$

Store the best solutions  $b$  among  $B(t)$

Randomly select part of individual exchange evolutionary goals

End

End

#### 4.6. Algorithm analysis

According to (7), in every step, qubit angle can be determined in accordance with the formula

$$\theta_i' = \theta_i + \Delta\theta_i = \theta_i + (\theta_{bi} - \theta_i) \times r \quad (10)$$

Suppose before one iteration, the angle  $\theta_i$  is in the first quadrant, that is  $0 < \theta_i < \pi/2$ , because  $r \in (0,1)$ , when the binary bit of the optimal solution is 1,  $\Delta\theta_i = (\pi/2 - \theta_i) \times r > 0$ , and  $\theta_i' = \theta_i + (\pi/2 - \theta_i) \times r = \pi/2 - (\pi/2 - \theta_i) \times (1-r) < \pi/2$ . That means the angle will rotate counterclockwise toward the direction of  $\pi/2$ , which makes the probability that the qubit be found in 1 increase, but after rotating, the angle still less than  $\pi/2$ , and is in the first quadrant. When the bit of the best binary solution is 0,  $\Delta\theta_i = -\theta_i \times r < 0$  and  $\theta_i' = \theta_i - \theta_i \times r = \theta_i \times (1-r) > 0$ , That means the angle will rotate clockwise toward the 0 direction, which makes the probability of getting 0 increase, and after rotation the angle is greater than 0, still in the first quadrant. Since in the beginning of the algorithm, all the angles of qubits in the individual are initialized in the first quadrant, then as the algorithm evolving, all the angles of the qubits in the individual will always be limited to the first quadrant and the evolution direction will be toward the best solution. Eventually, the angles of qubits will tend to the angle 0 or  $\pi/2$  of corresponding bit in the best solution. While all qubit angle reach the angle corresponding the best solution, the algorithm has converged, and the entire population converge to the best solution.

In the algorithm, the rotation angle is proportional to the difference of the angle of the qubit and the angle of corresponding bit of the optimal solution. The larger the difference is, the rotating angle will larger, which make the qubit approximate to the optimal solution at a fast speed, the maximum rotation amplitude of qubit is  $|\Delta\theta_i| \approx \pi/2 \times r$ ; conversely, the smaller the difference is, the rotating angle is smaller. The minimum rotation amplitude will tends to 0 as the qubit approximate to the optimal

solution.

### 5. Experimental results

To verify the performance of IQEA, we use it for the knapsack problems, and compared with CGA and QEA. In this experiments, we set the number of the items to be 100, 500, 1000 respectively. The weight and value of the items is randomly generated. The weights are taken in [5,20] and the values in [50,100]. The capacity of the knapsack are set to 985, 3676, 10410 respectively. In the algorithm,  $f(X)$  is used as the fitness function. Every binary string in P(t) is repaired with the strategy below

```

Procedure repair  $p(t)$ 
  Begin
    All items in the knapsack are sorted in the order
    Knapsack-overfilled = false
    If  $\sum_{i=1}^n w_i x_i > C$  Knapsack-overfilled = true
    While (Knapsack-overfilled = true) do
      Select the j-th item which is the last item being chosen
       $x_j = 0$ 
      If  $\sum_{i=1}^n w_i x_i < C$  Knapsack-overfilled = false
    End
    While (Knapsack-overfilled = false) do
      Select the j-th item which is the first item being not chosen
       $x_j = 1$ 
      If  $\sum_{i=1}^n w_i x_i > C$  Knapsack-overfilled = true
    End
     $x_j = 0$ 
  
```

All experiments are runned 50 times by IQEA, CGA, QEA, respectively. In these three algorithms, the population size is 10, the maximum number of iteration is 1000. In IQEA,  $\epsilon = 0.01\pi$ ,  $r = 0.005$ . The optimization results of the three algorithms are shown in Table 1-3, and Figs.1-3. show the average results of 50 tests.

**Table 1.** Optimization result with 100 items

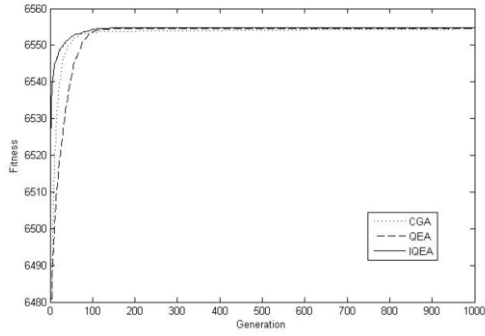
Algorithm	Best result	Average result	Worst result	Standard deviation
CGA	6554.8	6554.5	6552.1	0.8
QEA	6554.8	6554.7	6552.1	0.5
IQEA	6554.8	6554.8	6554.8	0

**Table 2.** Optimization result with 500 items

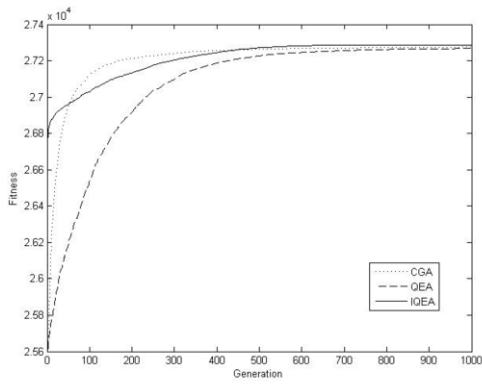
Algorithm	Best result	Average result	Worst result	Standard deviation
CGA	27286	27276	27257	7
QEA	27287	27268	27238	12
IQEA	27288	27288	27285	1

**Table 3.** Optimization result with 1000 items

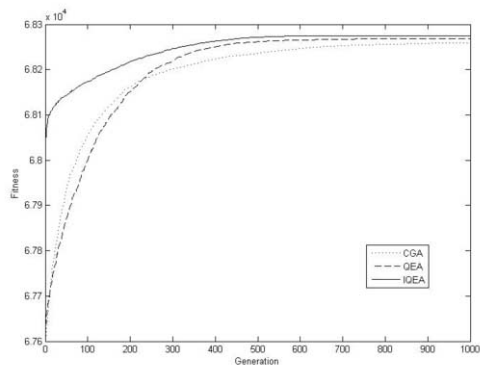
Algorithm	Best result	Average result	Worst result	Standard deviation
CGA	68275	68260	68228	10
QEA	68275	68268	68252	6
IQEA	68275	68275	68270	1



**Figure 1** Performance varying curves (item 100)



**Figure 2** Performance varying curves (item 500)



**Figure 3** Performance varying curves (item 1000)

It can be seen from the table1-3, IQEA has bigger value in all of best, average,

and worst results than CGA and QEA, which indicates that IQEA has better optimal precision. And the smaller standard deviation means it has better robustness. Figure 1-3 show that IQEA has faster convergence speed than CGA and QEA. Totally, its performance is superior to the other two algorithms.

## 6. Conclusion

Knapsack problem is a well-known combinatorial optimization problem. The research on knapsack problem is of great significance both in theory and in practice. However, this problem is a NP complete problem, the difficulty of solving it will increase exponentially as problem size increases. On the basis of QEA, a series of improvements were carried out for knapsack problem. And simulation results show that the improved algorithm has better search ability, and its performance is better than the common genetic algorithm and the quantum evolutionary algorithm.

## References

- [1] LienLienland B, Zeng L. A review and comparison of genetic algorithms for the 0-1 multidimensional knapsack problem[J]. *International Journal of Operations Research and Information systems*. 2015,6(2):21-31.
- [2] Wei Shen, Beibei Xu, Jiangping Huang. An improved genetic algorithm for 0-1 knapsack problems[C]. 2011 Second International Conference on Networking and Distributed computing. 2011, 32-35.
- [3] Tung K T, Li K L, Xua Y M. Chemical reaction optimization with greedy strategy for the 0-1 knapsack problem[J]. *Applied soft computing*, 2013,13(4):1774-1780.
- [4] Wang L, Yang R X, Xua Y. An improved adaptive binary harmony search algorithm[J]. *Information Science*, 2013,232:58-87.
- [5] Farnaz H, Asadollah S, Mohammad H. A parallel solution for the 0-1 knapsack problem using firefly algorithm[C]. 2016 1<sup>st</sup> Conference on swarm Intelligence and Evolutionary Computation, 2016, 25-30.
- [6] K.H. Han, J-H Kim. Quantum- inspired Evolutionary Algorithm for a Class of Combinatorial Optimization. *IEEE Trans Evolutionary Computation*, 2002; 6(6) : 580-593.
- [7] Han K H, Kim J H. Quantum-inspired evolutionary algorithms with a new termination criterion, H<sub>g</sub> gate, and two-phase scheme[J]. *IEEE Trans on Evolutionary Computation*, 2004, 8(2):156-169.
- [8] Han K H, Kim J H. On Setting the Parameters of Quantum-inspired Evolutionary Algorithm for Practical Applications[C] *Proc of the 2003 Congress on Evolutionary Computation*. IEEE Press, 2003, 12: 178-184.
- [9] Qu Zhijian, Liu Xiaohong, Zhang Xianwei. Hamming-distance-based adaptive quantum-inspired evolutionary algorithm for network coding resources optimization[J]. *The Journal of China Universities of Posts and Telecommunications*. 2015, 22(3):92-99.
- [10] Luciano R, Ricardo T, Marley M B R V. Quantum inspired evolutionary algorithm for ordering problems[J]. *Expert Systems with Applications*. 2017,67: 71-83.
- [11] Arpaia P, Maisto D, Manna C. A quantum-inspired evolutionary algorithm with a competitive variation operator for multiple-fault diagnosis[J]. *Applied soft computing*. 2011,11(8): 4655-4666.
- [12] Rui Wang, Ning Guo, Fenghong Xiang. An improved quantum genetic algorithm with mutation and its application to 0-1 knapsack problem[C]. *Proceedings of 2012 International Conference on Measurement, Information and Control*. 2012, 484-488.
- [13] Patvardhan C, Bansal S, Srivastav A. Parallel improved quantum inspired evolutionary algorithm to solve large size quadratic knapsack problems[J]. *Swarm and Evolutionary Computation*. 2016, 26:175-190.
- [14] Lu T, Yu G. An adaptive population multi-objective quantum-inspired evolutionary algorithm for multi-objective 0/1 knapsack problem[J]. *Information science*. 2013,243(10): 39-56.
- [15] Tan L X. *Information Sharing Mechanisms in Quantum Inspired Evolutionary Algorithms*[D]. Hefei: University of Science and Technology of China. 2010: 20-24

# Objects Detection Based on Multi-BING Feature Model

Su Yue<sup>a,1</sup>, Chen Mianshu<sup>a</sup>, Sang Aijun<sup>a</sup>, Li Xiaoni<sup>a</sup> and Li Mengying<sup>a</sup>  
<sup>a</sup>*School of Communication Engineering, Jilin University, Changchun, 130000, China*

**Abstract.** Objects detection is not only an important research direction in computer vision, but also the basis of object tracking and behavior detection. In this paper, we analyze the characteristics and principles of Binarized normed gradients (BING) algorithm, and propose an objects detection algorithm based on multi-BING feature model. At first, the proposed algorithm uses K-means clustering algorithm to cluster the training data, and then trains each category of images to establish the corresponding BING feature model. At the detection stage, multiple BING feature models are respectively used for testing. We collected all the detected results from all models as the final detection results. Experimental results demonstrate that the proposed algorithm effectively improves the object detection rate (DR) under various overlaps at the expense of a small amount of time by generating a small set of high-quality proposals.

**Keywords.** Computer vision, objects detection, BING, K-means clustering

## 1. Introduction

As the name suggests, the purpose of objects detection [1][2] is to determine the location, size and other information about the objects in the image. However, the problems of complex background, object occlusion, intensity change and other issues in the actual application have brought a lot of challenges to objects detection.

In 2006, Dalal and Triggs [1] proposed a model based on Histograms of Oriented Gradients (HOG) feature. This method applies a linear classifier to the classification in the sliding window detection frame [2], which obtain a good effect. However, this model is a global model, which can not well conform to the object deformation. In 2007, Pedro Felzenszwalb et al. [3] put forward the Deformable part model that is composed of a root model and several deformable parts, and its appearance is regarded as a milestone in the field of objects detection. In 2012, professor Hinton [4] first used a model based on the Convolutional Neural Network (CNN) for the first time and achieved unprecedented success, whereas the problem of long training time and large quantities of data still exist. In recent years, the object proposal method has been widely used to improve the speed, which can be mainly divided into two kinds, namely, grouping method [5] and window scoring method [6]. The former refers to dividing an image into many small blocks first and then clustering them based on the distance defined by human. One of the most typical algorithms of the grouping method is the Selective Search [5]. Comparatively, in the latter method, there is no need to cluster small pieces, and a large number of candidate boxes are generated instead to be scored

---

<sup>1</sup> Corresponding Author. Su Yue (suyue15@mails.jlu.edu.cn)



and sorted. Then those windows with low scores will be sorted out. Objectness [6] and binarized normed gradients (BING) algorithm proposed by Cheng Mingming et al. [7] in 2014 are both the typical algorithms of window scoring method.

The major discovery in the BING algorithm [7] is that since the generic object has well-defined closed boundaries, a strong correlation exists between the well-defined closed boundaries of the general object and the gradient norm in normed gradient space. Therefore, objects can be distinguished from non-objects by the utilization of the cascade support vector machines (SVM) [8]. However, the most significant and prominent point of the BING algorithm is that it adopts computer bit operation to replace direct calculation for the sake of speeding up. The final test speed reaches the level of ms on a single laptop CPU on the PASCAL VOC2007 dataset [9], which significantly improves the speed of operation. Nevertheless, only the conditions in which the Intersection over Union (IoU) equals to 0.5 are discussed, without considering the conditions of higher overlap, yet the object detection rate (DR) declines rapidly at the IoU with higher values. So it is essential to further improve DR. In this paper, we propose an objects detection algorithm based on multi-BING feature model to model objects. First, we utilize K-means clustering algorithm to cluster training images into several groups [9][10], and then use each group of training data to train model. Finally, the detection results of each group are merged to generate the end results.

This paper is organized as follows: Section 2 summarizes the BING algorithm. Section 3 describes the proposed objects detection algorithm based on multi-BING feature model. Section 4 carries out the experiment on different overlap and analysis the experimental results. Section 5 draws a conclusion with future work.

## 2. BING Algorithm

Under the influence of the characteristic that human perceives objects before identifying them, Cheng Mingming et al. discovered that there is a similar relationship in normed gradient space, despite the diverse characteristics of objects. Thus, after resizing of their windows to small size, a simple 64D norm of the gradients (NG) feature with the characteristic of scale invariance was proposed. However, the computational efficiency of NG algorithm is still unable to satisfy the requirements of real-time applications. So a binary approximation form of the NG feature, namely BING feature, was proposed. The BING feature uses bit operation to replace the arithmetic operation, which improves the calculation efficiency greatly.

The algorithm adopts two-level linear SVM classifier for training. First of all, in order to find objects, it needs to adjust the images into quantized scales and aspect ratios to obtain images with different size. Then, the model  $\mathbf{w}$  is trained according to the training samples for each size. The filter score for each window is defined as

$$s_l = \langle \mathbf{w}, \mathbf{g}_l \rangle \quad (1)$$

$$l = (i, x, y) \quad (2)$$

where  $s_l$  represents filter score,  $\mathbf{g}_l$  is the NG feature,  $i$  represents the window

size, and  $(x, y)$  refers to the window position. Non-maximum suppression (NMS) is used here to obtain a certain number of windows for each size. Since a  $100 \times 100$  square window, compared with a  $5 \times 148$  rectangular window, is more likely to contain an object, the second level SVM is required to be used to score each of the candidate windows. The objectness score is defined as

$$o_i = v_i \cdot s_i + t_i \quad (3)$$

where  $v_i$  and  $t_i$  represent the learnt coefficient and bias terms for each size  $i$ .

### 2.1. NG Feature and Training

In the NG algorithm, the simple gradient absolute value is selected as the feature. At first, the 1D template  $[-1 \ 0 \ 1]$  is utilized to calculate the gradients  $g_x$  and  $g_y$  of the image in horizontal and vertical directions and the normed gradients are computed through using the  $\min(|g_x| + |g_y|, 255)$ . In the training phase, a method based on the two stages cascaded SVM is used to learn a generic objectness measure. The implementation is as follows:

**Step 1.** Use pre-obtained training samples and (1) to obtain a linear model  $\mathbf{w}$  (as shown in Figure 1), and then conduct coarse filtration on all candidate windows.

**Step 2.** Use the proposals after NMS as training samples and their filter scores as features to obtain  $v_i$  and  $t_i$  for each size  $i$ .



Figure 1. Linear model  $\mathbf{w}$ .

---

#### Algorithm 1: Binary Approximate Model $\mathbf{w}$

---

**Input:**  $\mathbf{w}$ ,  $N_w$

**Output:**  $\{\beta_j\}_{j=1}^{N_w}$ ,  $\{\mathbf{a}_j\}_{j=1}^{N_w}$

**Initialize residual:**  $\varepsilon = \mathbf{w}$

**for**  $j = 1$  to  $N_w$  **do**

$\mathbf{a}_j = \text{sign}(\varepsilon)$

$\beta_j = \langle \mathbf{a}_j, \varepsilon \rangle / \|\mathbf{a}_j\|^2$

$\varepsilon \leftarrow \varepsilon - \beta_j \mathbf{a}_j$  (update residual)

**end for**

---

## 2.2. BING Feature and Training

In order to speed up, a binary approximation strategy is adopted, thus the bit operation in the computer can be effectively used to replace the arithmetic operation. The training process of BING feature is basically the same as that of NG feature.

First, in order to further improve efficiency, the learned model  $\mathbf{w}$  can be transformed into a binarized form with a set of basis vectors using

$$\mathbf{w} \approx \sum_{j=1}^{N_w} \beta_j \mathbf{a}_j \quad (4)$$

where  $\beta_j \in \mathbb{R}$  refers to the corresponding coefficient, and  $\mathbf{a}_j = \{-1, 1\}^{64}$  refers to the basis vector. In order to use the fast bit operations in computer,  $\mathbf{a}_j$  can be represented using  $\mathbf{a}_j = \mathbf{a}_j^+ - \overline{\mathbf{a}_j^+}$ , where  $\mathbf{a}_j^+ \in \{0, 1\}^{64}$ . In this way,  $\mathbf{w}$  can be represented approximately by a group of basis vectors, in which the number of basis vectors is  $N_w$ , and a binarized feature  $\mathbf{b}$  can be defined as

$$\langle \mathbf{w}, \mathbf{b} \rangle \approx \sum_{j=1}^{N_w} \beta_j \left( 2 \langle \mathbf{a}_j^+, \mathbf{b} \rangle - |\mathbf{b}| \right) \quad (5)$$

The next is to transform NG feature into BING feature. Since the range of the NG feature is  $[0, 255]$ , the NG feature of each point on the  $8 \times 8$  window can be saved in a BYTE value. Meanwhile, only the top  $N_g$  bits are taken as the gradient feature for each position, so as to improve the computational efficiency and further save the storage space. Thus a 64D NG feature  $\mathbf{g}_l$  can be approximated by  $N_g$  BING features as

$$\mathbf{g}_l = \sum_{k=1}^{N_g} 2^{8-k} \mathbf{b}_{k,l} \quad (6)$$

At last, on the basis of  $\mathbf{w}$  and  $\mathbf{g}_l$ , the filter score can be rapidly calculated using

$$s_l = \sum_{j=1}^{N_w} \beta_j \sum_{k=1}^{N_g} C_{j,k} \quad (7)$$

where  $C_{j,k} = 2^{8-k} \left( 2 \langle \mathbf{a}_j^+, \mathbf{b}_{k,l} \rangle - |\mathbf{b}_{k,l}| \right)$ .

### 2.3. Detection Phase

In the detection phase, 4952 testing images on the VOC2007 and the corresponding testing image annotations are applied to evaluate the method. The final results indicate that BING method can reach 96.2% DR with 1000 proposals at the overlap threshold of 0.5. However, with the increase of the overlap, DR drops rapidly. Therefore, it is important to improve DR at diverse overlap thresholds.

## 3. Multi-BING Feature Model

BING algorithm models all objects as a pattern. Owing to the diversity of objects, using a single model to represent objects has some limitations. This paper presents a model based on multi-BING feature to model objects. Firstly, through extracting the center-symmetric local binary (CS-LBP) features [12], the training images are divided into multiple groups with the application of K-means clustering, then each group of training images after clustering are used as training data to establish multiple BING feature models. In the detection phase, we use a fusion strategy to generate the final results. The main content of the algorithm are as follows:

- Feature extraction and K-means clustering;
- Multi-BING feature modeling;
- Detection results fusion.

### 3.1. Feature Extraction and K-Means Clustering

#### 3.1.1. CS-LBP Feature

Before the clustering, it is essential to extract and collect the features of the training images. The CS-LBP feature selected in this paper is an improvement of LBP feature [11][12].

Local binary pattern (LBP) is a kind of texture feature, which achieves the characteristic of gray scale invariance [11]. In recent years, it has been successfully applied in license plate recognition, texture classification and other areas. The general idea of the LBP operator is that an arbitrary pixel point is taken as the center point, whose  $3 \times 3$  neighborhood is taken as the detection range, and the gray value of the center pixel is denoted as  $P_c$ . First, the gray values of all the 8 adjacent pixels are compared with that of the center pixel. If the gray value of the adjacent pixels is larger than  $P_c$ , then the binary number is set as 1, else 0. Afterwards, these 8 binary numbers are arranged in a certain order and converted to the decimal form, in which the decimal integer is the LBP value. Figure 2 shows the calculating process of the LBP.

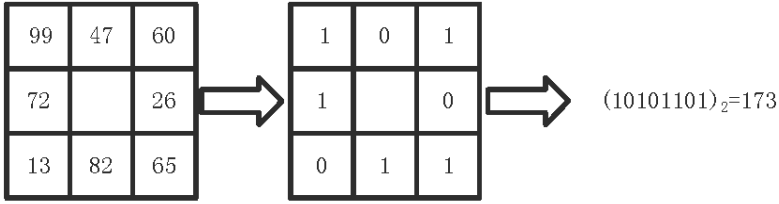


Figure 2. Calculation of the LBP.

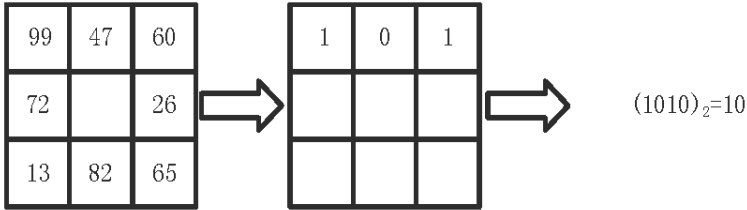


Figure 3. Calculation of the CS-LBP.

After scanning an image with the LBP operator, the whole image can be represented by the histogram of the LBP code, and the description of the characteristics in the whole region can be counted in the form of a histogram.

However, the shortcomings of insufficient operator’s direction along with high computational cost and memory are caused, because the basic LBP operator only considers the relationship between the neighborhood pixel and the central pixel. To handle the above-mentioned shortcomings, Heikkola et al. [12] proposed a center-symmetric local binary pattern (CS-LBP) which basic idea is to introduce the concept of symmetry axis and set the center pixel as the axis. When the number of comparison points is  $N$ , the number of symmetry axes is  $N/2$ , then we can compare the gray values of the pixels at both ends of the axis of symmetry in the prescribed order, thus the dimension of the feature code can be effectively reduced by half. The CS-LBP operator is defined as

$$CS-LBP_{N,R}(x,y) = \sum_{i=0}^{(N/2)-1} \text{sgn}(n_i - n_{i+(N/2)})2^i \tag{8}$$

where  $R$  refers to the detection radius, Figure 3 shows the calculating process of the CS-LBP.

### 3.1.2. K-means Clustering

K-means clustering is one of the most widely known clustering algorithms [10]. Its basic idea is to first specify the number of clusters  $k$ , the number of iterations and convergence conditions. Each sample in the dataset is assigned to the nearest cluster center according to certain similarity measurement criteria, and then each cluster center is recomputed as the average of the points in that cluster. The iterations will continue until the maximum number of iterations is reached or the convergent conditions are met.

Supposed that the samples in the sample space  $X = \{x_1, x_2, \dots, x_n\}$  are divided into  $k$  clusters, the cluster center of each is  $C = \{c_1, c_2, \dots, c_k\}$ ,  $d_{ij}(x_i, c_j)$  refers to the distance between the samples and the corresponding center, and then the sum of the distance between all the samples in the sample space and the corresponding cluster center is defined as the objective function

$$J = \sum_{j=1}^K \sum_{i \in c_j} d_{i,j}(x_i, c_j) \quad (9)$$

By reducing the value of  $J$  to optimize the program, when  $J$  takes the minimum, it is the optimal scheme.

### 3.2. Multi-BING Feature Modeling

After the clustering, we can get multiple groups of training data, and then use each group of training data to train the corresponding BING model. In the detection phase, through changing the size and aspect ratio of the testing images, NMS is utilized to obtain all candidate windows, and corresponding scores are also calculated by loading multiple BING model. So for each model, we can obtain a set of detection results which include all the testing images.

### 3.3. Detection Results Fusion

Since we can collect the detection results from multiple models, hence it is essential to fuse the results of each model to gain the final detection results. Fusion strategy is done as the following steps:

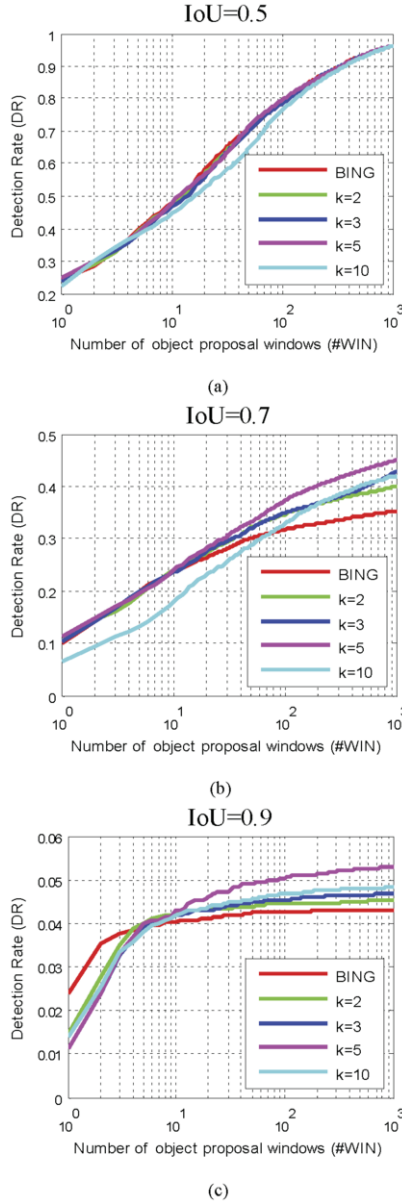
<b>Algorithm 2:</b> Fusion Strategy	
<b>Step 1:</b>	Gather all the proposals obtained from the same image.
<b>Step 2:</b>	Sort the proposals in ascending order according to the corresponding objectness score.
<b>Step 3:</b>	Retain the proposals with the highest objectness score if the duplicate windows occur, and delete the duplicate items.

## 4. Experimental Results

We applied the proposed algorithm on VOC2007 using the detection rate given #WIN proposals (DR-#WIN) [8][9] evaluation metric on the same computer with an Intel i5-6500 CPU. For all experiments, the BING algorithm is considered as a baseline model [7][9]. Since our goal is to effectively achieve a better DR by using high-quality proposals at different values of DR, the situations that IoU is equal to 0.7 and 0.9 are also considered, apart from that IoU equals to 0.5. In addition, the number of clusters

$k$  is respectively selected as 2, 3, 5 and 10, and the IoU thresholds in Figure 4(a), Figure 4(b) and Figure 4(c) are 0.5, 0.7 and 0.9 respectively.

We clustered respectively 2501 training images into 2, 3, 5 and 10 categories by K-means clustering. Then we trained BING model and our multi-BING feature model. At last, we detected objects in all 4952 testing images by BING and our multi-BING feature model. Figure 4 shows the comparison between the BING algorithm and our proposed methods for IoU threshold of 0.5, 0.7 and 0.9 using 1000 proposals.



**Figure 4.** Tradeoff between #WIN and DR for different methods for IoU threshold of 0.5, 0.7 and 0.9 using 1000 proposals.

Figure 4(a) indicates that in the case of  $\text{IoU} = 0.5$ , when  $k$  is equal to 2, 3 and 5, our methods outperform BING in general. When  $k = 10$ , as the number of proposals from our method is much more than that of BING, our results are inferior. From Figure 4(b), we can see that in the case of  $\text{IoU} = 0.7$ , when the number of object windows is between 0 and 10, the methods under  $k$  is equal to 2, 3 and 5 basically has the same DR as BING. Nonetheless, because as the number of clusters increases, the proposals increase sharply, when  $k = 10$ , the performance of our method is no better than that of BING. When the number of windows varies from 10 to 1000 and the value of  $k$  is small, the clustering results become gradually distinct and the characteristics of each type of images are increasingly prominent as  $k$  increases. Thus, DR rises with the increase of  $k$ . Figure 4(c) shows that when  $\text{IoU} = 0.9$  and the number of windows varies from 0 to 10, the results of the proposed algorithm are inferior owing to the increase of proposals. Nevertheless, when the number of windows varies from 10 to 1000, our methods perform better than BING, and when  $k$  is equal to 2, 3 and 5, DR varies directly with  $k$ . When  $k = 10$ , even though the characteristics of each type of images are rather obvious, there are not enough training samples in certain categories to train a reliable model because of too much clusters. Consequently, DR is lower than that when  $k = 5$ .

**Table 1.** Results for our methods compared to BING for IoU threshold of 0.5, 0.7 and 0.9 at 1000 proposals.

IoU	BING	Our Method			
		$k = 2$	$k = 3$	$k = 5$	$k = 10$
0.5	96.20%	96.23%	96.28%	96.33%	96.12%
0.7	35.23%	39.95%	42.74%	44.94%	42.01%
0.9	4.31%	4.53%	4.67%	5.28%	4.82%

Table 1 shows the experimental results from BING and our methods under various IoU thresholds at 1000 proposals. It can be seen that when  $k = 5$ , the proposed algorithm can realize the optimal performance.

## 5. Conclusion

We propose an objects detection method on the basis of multi-BING feature model, which effectively improve DR at the expense of a small amount of time under different overlap thresholds. At first, the proposed algorithm uses K-means clustering to cluster the training images by extracting the CS-LBP features, and then the BING feature model is established by training each category of data. Finally, a simple fusion strategy that includes the steps of merging, de-duplication and sorting is adopted to obtain the final detection results. The results show that the proposed algorithm performs better than BING under various overlaps at the expense of a small amount of time by using small quantities of high-quality candidate windows. When  $k = 5$ , our methods get the optimal performance which is 96.33% DR, 44.94% DR and 5.28% DR respectively under the IoU of 0.5, 0.7 and 0.9. The key direction of future research will focus on the selection of learning approach and the strategy to further reduce the number of object windows while retaining high DR.



## Acknowledgement

The authors thank for the support of the Education Department of Jilin Province “13th Five-Year” Science and Technology project (J.J.K.H.Z. [2016] No. 427) and Project of International Cooperation and Exchange Foundation of Jilin Province, China (No. 20140414013GH).

## References

- [1] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, In: Proceeding of 2005 IEEE Computer society conference on computer vision and pattern recognition, San Diego, CA, USA, pp 886-893, 20-25 June 2005.
- [2] J. Li, J. Liu, W. Zhou, A real-time algorithm for finding frequent items over sliding window, In: Proceeding of 2016 7th IEEE international conference on software engineering and service science, Beijing, China, pp 103-106, 26-28 Aug. 2016.
- [3] P. Felzenszwalb, D. McAllister, D. Ramanan, A discriminatively trained, multiscale, deformable part model, In: Proceeding of Conference on computer vision and pattern recognition, Anchorage, AK, USA, pp 1-8, 23-28 June 2008.
- [4] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, In: Proceedings of the advances in neural information processing systems (NIPS), vol 25, no 2, pp 1097-1105, 2012.
- [5] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, et al. Selective search for object recognition, *International Journal of Computer Vision* **104** (2013), 154-171.
- [6] B. Alexe, T. Deselaers, V. Ferrari, Measuring the objectness of image windows. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34** (2012), 2189-2202.
- [7] M.M. Cheng, Z. Zhang, W.Y. Lin, et al. BING: binarized normed gradients for objectness estimation at 300fps, In: Proceeding of the IEEE conference on computer vision and pattern recognition, pp 3286-3293, June 2014.
- [8] T. Pariwat, P. Seresangtakul, Thai finger-spelling sign language recognition using global and local features with SVM, In: Proceedings of 2017 9th international conference on knowledge and smart technology (KST), Chonburi, Thailand, pp 116-120, 1-4 Feb. 2017.
- [9] M. Everingham, L. Van Gool, C.K.I. Williams, et al. The pascal visual object classes (VOC) challenge, *International Journal of Computer Vision* **88** (2010), 303-338
- [10] T. Kanungo, D.M. Mount, N.S. Netanyahu, et al. An efficient K-means clustering algorithm: analysis and implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** (2002), 881-892.
- [11] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition* **29** (1996), 51-59.
- [12] M. Heikkola, M. Pietikainen, C. Schmid, (2009) Description of interest regions with local binary patterns, *Pattern Recognition* **42** (2009), 425-436.

# Print Defect Detection Based on Image Processing

Mengtao HUANG<sup>1</sup> and Qinyao LI<sup>2</sup>  
*Xi'an University of Science and Technology*

**Abstract.** Aiming at the problem of print defect detection, this paper presents a defect detection system based on image processing. The system collects print image pictures by CCD camera and sends those pictures to the computer. After receiving pictures, the image processing software called VS2010 which is based on OPENCV processes the received pictures and extracts defective pictures by background difference method. The experimental results show that the algorithm can accurately identify the defect print and meet the system design requirements.

**Keywords.** CCD camera, VS2010, OPENCV, background difference method

## 1. Introduction

Newspapers, books, magazines and product packaging and other printed matter has become an important part of people's daily life and work. However, in the print process, due to print equipment, print materials, the operator's technical imperfections and other reasons, the surface of printed products tend to produce a variety of defects, that is, a variety of printed defective products, including the color distortion, misplaced leakage printed, pinhole black spots, text blurred, ink splashing and so on [1–3]. In the past, people do on the production line qualified product testing, mainly by the human eye to identify, that is, visual inspection [4]. The shortcomings of human eye inspection are easy to produce fatigue and continuous detection time is short, and the detection speed is slow and other shortcomings. Especially for small size, character-intensive mark detection rate is high [5]. Compared with the human eye visual inspection, the machine vision detection with defect recognition is accurate, never fatigue, etc; detection system parameters once set to complete, it can work for a long time according to the same standard for automatic detection, and keep detection speed fast from beginning to the end. In addition to the above advantages, the machine vision detection can operate without missing, and can be excellent to complete the defect detection of printed matter. Artificial inspection does not match the superiority.

---

<sup>1</sup> Huang Mengtao (1965–), female, Ph.D., Professor, master tutor, mainly engaged in computer control, intelligent systems and image based measurement and recognition of research; E-mail: 656228336@qq.com.

<sup>2</sup> Li Qinyao (1992–), male, master's degree, research direction: Pattern recognition and intelligent system; E-mail: 635064094@qq.com.

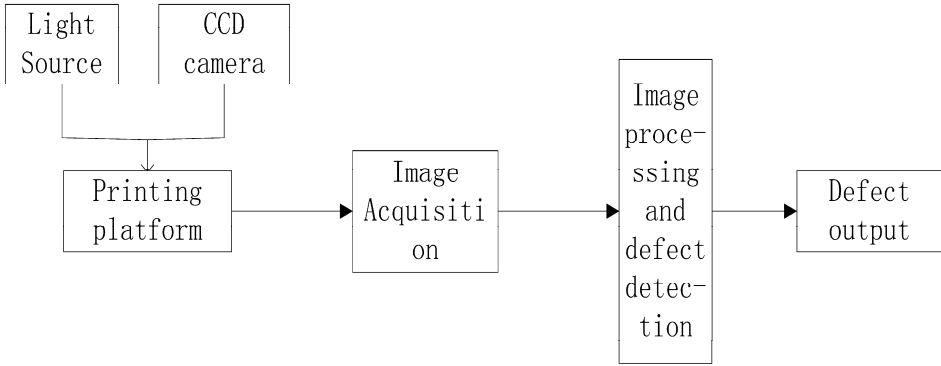


Figure 1. System block diagram.

## 2. Defect Detection System Principle Analysis

Defect detection system consists of lighting system that is light source, imaging system which is optical system, CCD camera, image processing hardware, image processing software, input and output equipment and other components. The imaging system focuses the image of the finished product on the CCD. The CCD camera converts the image into a computer-recognized electrical signal. The image processing hardware collects and processes the signal. The electrical images of the printed images are finally transferred to the computer's memory and the image processing software will analysis and judgment them [6]. Image detection process design principle is that the CCD camera collections non-defective standard print image as the sample, the sample made into a template. Then the standard template image and CCD actual image do acquisition and comparison, in order to detect errors [7].

The system can detect the defects of the surface, mainly relying on image processing software. Image processing software using image processing algorithms, such as gray scale, filtering algorithm background difference method, can eliminate image noise and improve image quality, and identify defects in the image. Figure 1 is a system block diagram.

## 3. Algorithm Design

The algorithm is the core part of the system. through the algorithm design, the system completes the detection of defects. The algorithm mainly includes the following parts: gray scale, smoothing and gradient, background difference and binarization. System algorithm design flow chart is shown in Fig. 2.

### 3.1. Smooth

Because the image may contain noise, it is necessary to smooth the original image. This method of eliminating the noise component of the image is called smoothing or filtering of the image [10]. The energy of the signal or image is mostly concentrated in the low and medium frequency bands of the amplitude spectrum. While in the higher

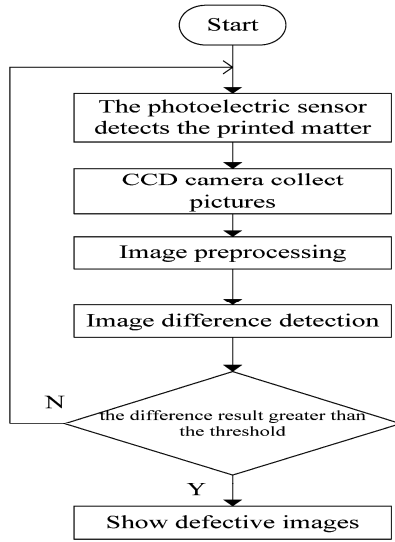


Figure 2. System algorithm design flow chart.

frequency band, useful information is often overwhelmed by noise. So a filter that reduces the amplitude of the high-frequency component can reduce the effect of noise.

A few common filtering methods are: block filter, mean filter, Gaussian filter, median filter, bilateral filtering. This paper uses Gaussian filter as Wave filter.

Gaussian filter is a linear smoothing filter, widely used in image processing of noise reduction process. In general, Gaussian filtering is the process of weighting the entire image, and the value of each pixel is obtained by weighted average of its pixel and the other pixel values of its neighborhood. The specific operation of Gaussian filtering is to scan each pixel in the image with a template (or convolution, mask). Then the filter uses the weighted average gray value of the pixels in the neighborhood determined by the template that takes the place of the value of the central pixel. In image processing, Gaussian filtering generally has two implementations, one called sliding window, and the other called Fourier transform. This paper selects the first sliding window.

### 3.2. Image Grayscale

Those color images that are collected by CCD camera require large storage space. What's more, they need to take up more space in the processing, so the digital image processing computer needs to have a larger memory and faster computing speed. We must change 24-bit true color print image into 8-bit grayscale image, which can greatly reduce the image data and will not affect the image recognition results [8].

Image grayscale has two methods. One is R, G and B three primary colors equal weight method. another is three primary colors unequal weight method [9]. In this paper, we use the inequality method. The transformation formula is as shown in Eq. (1):

$$F(x) = 0.3R_{(x,y)} + 0.59G_{(x,y)} + 0.11B_{(x,y)} \quad (1)$$



Figure 3(a).



Figure 3(b).

$R_{(x,y)}$ ,  $G_{(x,y)}$ ,  $B_{(x,y)}$  are the R, G, B (red, green and blue) tricolor components of the input color image.

The grayscale images are shown in Figs 3(a) and 3(b).

### 3.3. Binarized Images

An image including the object, background and noise. To extract directly out of the target object from the digital image and multiple values, the most commonly used method is to set a threshold of  $T$ . By using  $T$ , image data is divided into two parts: the pixel group is greater than  $T$  and less than  $T$  pixel group. This is the most special method to study the gray level transformation, which is called the two value of the image. There are two ways to binarize, one is global binarization and the other is partial binarization. This paper adopts the global binarization method. The globalization method is to set a global threshold  $T$ , which divides the data of the image into two parts: a pixel group larger than  $T$  and a pixel group smaller than  $T$ . The pixel value of the pixel group larger than  $T$  is set to 255, and the pixel value of the pixel group smaller than  $T$  is set to 0.

The difficulty of binarization lies in the choice of threshold  $T$ ,  $T$  is too small will retain the interference,  $T$  is too large will lose valid data. It is very difficult to determine the parameter  $T$ . After a large number of experimental verification, the system eventually determines the threshold of 80.

### 3.4. Image Background Difference Method to Detect Defects

According to the principle of the image difference method, it can be used to detect the print image quality. The printed image is compared with the standard image by pixel subtraction, so that the difference between the corresponding pixels of the template image and the image those are detected can be obtained intuitively and conveniently. According to the difference between the image and the threshold to judge whether the print image is defective.

The system sets the standard image for  $T_{(i,j)}$  and the image to be checked is  $S_{(i,j)}$ , and the image (called the difference image) is  $D_{(i,j)}$ . We can use the following formula:

$$|D_{(i,j)}| = |S_{(i,j)} - T_{(i,j)}| \tag{2}$$

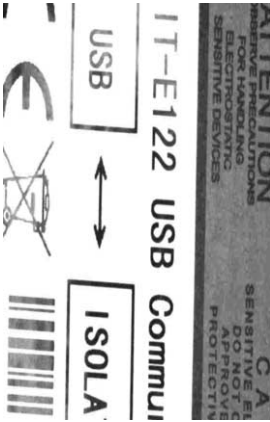


Figure 4(a).

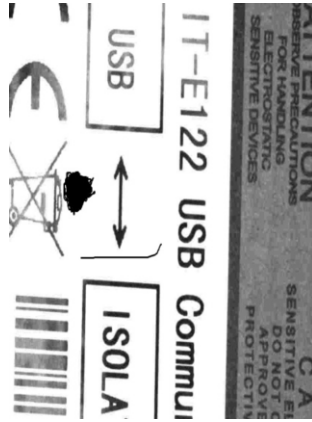


Figure 4(b).

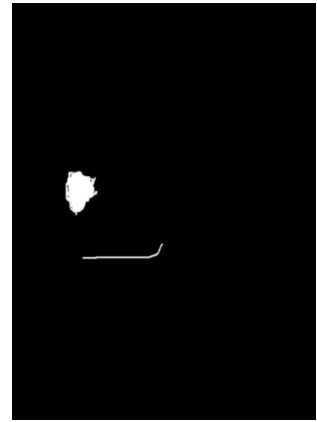


Figure 4(c).

$|D_{(i,j)}|$  represents the absolute value of the gray scale difference between the corresponding pixels of the two images. The smaller the  $D_{(i,j)}$  value is, the smaller the difference between the pixel values of the printed image and the template image is, and the more similar the two images. The image difference method is usually used to detect the image quality as the method is simple and easy to implement. In addition, the observer can visually find the difference between the two pixels of the images according to the difference image, and then determine whether there is a defect in the image.

#### 4. Test and Analysis of Results

In this paper, the experimental platform is used to replace the print platform. The system uses the CCD camera to take pictures on the experimental platform. Figure 4(a) is a template picture, is made from qualified prints. Figure 4(b) is a picture to be detected. The black line in the figure is scratch and the black group in this picture is the ink. There are a total of two defects in Fig. 4(b). Figure 4(c) is the resulting image of the first two pictures using the background difference method. This article uses vs2010 programming, which is based on OPENCV database system. there are two defects in Fig. 4(b), from the results of Fig. 4(c), we can see that the two defects have been detected. Based on the experimental results, the method described in this paper is effective. It can be used to do printed defects detection.

#### 5. Conclusion

Aiming at the problem of print defect detection. This paper presents a print defect detection algorithm based on image processing. The experimental platform which uses the algorithm verified 310 photos. The shows that 302 images were correct detected. It is meant that the correct rate reached 96.1%. The algorithm is basically able to meet the requirements of printed defect detection. Experimental results show that the algorithm is feasible. Of course, there are imperfect aspects of the algorithm. We will continue to improve it in the future.

## References

- [1] Martin T, Herrero J, A soft improves a real time quality assurance machine vision system. Proc. of the 17th International Conference on Pattern Recognition, UK Cambridge, 2004, 301–304.
- [2] Cuenca S, Camara A, New texture descriptor for high-speed Web inspection applications. IEEE International Conference on Image Processing, 2003, 3:537–540.
- [3] You F, Zhang L, Zhang Y, The research of print's image defect inspection based on machine vision [c]. Proc. IEEE International Conference on Mechatronics and Automation 2009, 8:404–2408.
- [4] Gonzalez R C, Woods R E. Digital Image Processing, 2th ed. New Jersey: Prentice Hal, 2002.
- [5] Arica N, Yarman-Vural F. An overview of character recognition focused on off-line handwriting. IEEE Trans on Systems, Man, and Cybernetics Part C: Applications and Reviews, 2001, 31(2):216–233.
- [6] Dai X, Wang G, Detection and recognition of printed matter defects based on machine vision. Application of Optoelectronic Technology.
- [7] Zhao D, Jin H, Print Defect Detection Based on Computer Image Processing. Packaging Engineering, 2008, 29(12):120–121.
- [8] Liu B, The print defect recognition system based on the machine vision. Xi'an, Xi'an University of Science and Technology, 2010.
- [9] Song F, Research on Objective Evaluation Method of Gray Image Gray Image and Its Effect. Qianzhou, Huaqiao University.
- [10] Gao Y, Research on image gray enhancement algorithm. Xi'an, Xi'an University of Science and Technology, 2010.

# New HRM Practices and Innovation Performance; The Moderating Role of Information Technology Ambidexterity

Abdul WAHEED<sup>a,1</sup>, Xiaoming MIAO<sup>a</sup>, Naveed AHMAD<sup>a,2</sup>, Salma WAHEED<sup>a,3</sup>  
and Abdul MAJEED<sup>b</sup>

<sup>a</sup>*School of Management, Northwestern Polytechnical University, Xi'an, China*

<sup>b</sup>*School of Management, University of Lahore, Lahore, Pakistan*

**Abstract.** The aim of this study is to investigate the relationship between new human resource management (NHRM) practices and innovation performance. Further, the moderating role of IT ambidexterity was examined between NHRM and innovation performance. This study selected Pakistan's largest IT based semi-government organization National Database & Registration Authority (NADRA) as a case study. Data was collected from three major cities (Lahore, Gujranwala and Jhelum) of Pakistan. 500 employees of NADRA participated in the survey based study. The empirical results found the positive relationship between NHRM practices and innovation performance. The moderating influence of IT ambidexterity was also found in this study. Employees with high IT ambidexterity are more involved in innovation performance. Continuous adaption of technology enhances long term competitive advantage. Therefore, utilizing new technologies and knowledge consistently, is significant for the enhancement of innovation performance.

**Keywords.** New HRM practices, Innovation performance, IT ambidexterity, IT flexibility, IT standardization

## 1. Introduction

New Human Resource management (NHRM) practices have been focused by researcher from different fields [1,2]. NHRM practices involve the several systems to mobilize employees' proposals for improvement, team-based organizations, continuous learning, and decentralization of decision rights, quality circles and focus on Information dissemination. It is needed to perceive the concept of NHRM practices and innovation performance [3,4].

These practices are clusters that aim to create flexibility with the help of autonomy, greater responsibility and worker's involvement in production, are not newly introduced. But their adaptation in different industries is increasing. Some firms adopts

---

<sup>1</sup> PhD students in School of Management, Northwestern Polytechnical University, Xi'an, China, Waheed\_2506@mail.nwpu.edu.cn.

<sup>2</sup> PhD students in School of Management, Northwestern Polytechnical University, Xi'an, China, naveedahmad@mail.nwpu.edu.cn.

<sup>3</sup> MS students in School of Management, Northwestern Polytechnical University, Xi'an, China, salma@mail.nwpu.edu.cn.



NHRM practices to improve the financial performance, others incorporate NHRM practices in strategic decision making. Managers utilize these practices to achieve organizational goals effectively [5–7]. Indeed, the most recent literature emphasis on the positive relationship between NHRM practices and innovation performance at different stages of Corporate planning [8].

Presently IT has become an important part of every organization which provides a platform to compete in the competitive environment [9]. Firms are using inter-firm governance strategies such as IT ambidexterity to enhance firm performance. IT flexibility and IT standardization are the main compliment of IT ambidexterity [10]. For instance, IT flexibility helps employees in interaction with customers to add other flexible terms that are not being included of their regular work. On the other side, IT standardization can reduce the risk of mistake in existing data of customers such as nationality and religion. Thus, firms with IT ambidexterity can achieve innovation performance by integrating inter-firm IT strategies with NHRM practices. The primary research questions for this study can: 1) What is the NHRM practices effect on the innovation performance? 2) How does IT ambidexterity moderate between NHRM practices and innovation performance? We intend to expand this research link between adopted NHRM practices and innovation performance at semi-government IT organization National Database and Registration Authority (NADRA) with introduction of IT ambidexterity. Further detail about selected organization NADRA as case study was discussed in methodology.

## **2. Theoretical Literature**

### *2.1. New HRM Practices*

NHRM practices are composed of several systems to mobilize employees' proposals for improvement, contemporary changes, team-based organizations, continuous learning, and decentralization of decision rights, quality circles and focus on Information dissemination. However, NHRM practices don't have concern about traditional HRM practices like recruiting, training and development and career paths. While previous studies found a strong influence of NHRM practices on financial performance, productivity and organizational flexibility [4,8,11,12]. Strategists and Corporate entrepreneurs are trying to combine existing solutions to create new combinations and giving new sets of solutions. The significance of individual and systems have recognized by NHRM practices [13]. The adoption of NHRM practices which have been included decentralization of rights, employee participation in decision making, cooperation and evasion of bureaucracy in specific firm [14,15].

### *2.2. Innovation Performance*

Innovation performance emerged as an important tool for measuring the progress and development of any firm. A firm must direct its practices towards operational and managerial areas to achieve innovation performance. In fact, the successful adoption of different types of innovation activities in a firm, it has been important to implement operational practices. Many researchers have discussed the role of HRM in creating innovation performance such as Carnegie and Butlin [16] defined innovation performance "something that is new or improved done by an enterprise to create significantly

added value, either directly for the enterprise or directly for its customer.” Another study stated that innovation performance plays a vital role in today’s competitive business environment for firm’s sustainability. Firms have to be innovative and need to adopt different types of innovation activities in all aspects of the organization rather depend on single innovative activity. Hence innovation provides an opportunity to create new knowledge or absorptive capacity to enhance its competency, to assimilate and exploit external knowledge, then apply this knowledge to create innovative products or services for increasing firm performance [17].

### *2.3. Information Technology Ambidexterity*

The concept of information technology ambidexterity can be described as “Simultaneously adaptation of two opposing concepts, IT flexibility and IT standardization”. The idea behind taking two opposing concepts simultaneously is taken from the research of Mithas and Rust [10]. A dual focus on IT flexibility and IT standardization can create superior firm performance rather than focusing on the single dimension. The capability of any organization to understand, cope and value creation in business operations, selection of modern and promising technologies and proficiently usage of resource are be IT ambidexterity. Surprisingly, Phenomenon of IT ambidexterity is less focused research area in management studies. There are rare calls for IT ambidexterity research at macro-level in management and social sciences research [18,19].

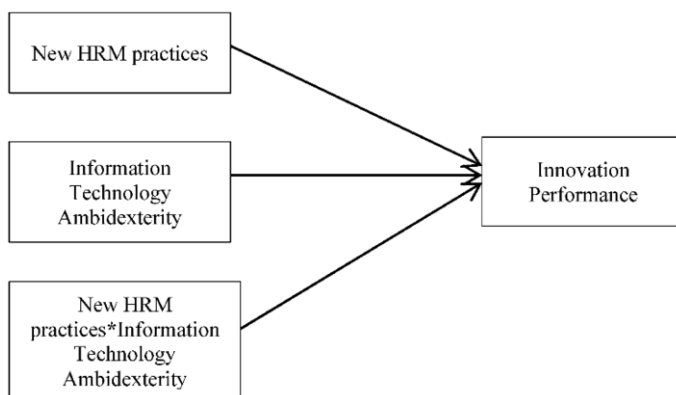
### *2.4. NHRM Practices and Innovation Performance*

The association between HRM practices and innovation performance has been rarely studied. Previously, HRM practices have been associated with large scale organizational issues, strategic issues, and employee performance [20]. Authors from diversified fields explored the relationship of NHRM practices with innovation performance. These practices impact firm’s internal social structures to create a bridge for strengthening the weak bonds in organization, interchange norms and share mental models among employees [21]. In Technology and innovation management research, Michie and Sheehan [22] empirically found an association between NHRM practices and Innovation performance. By adopting NHRM at shopfloor level, Laursen and Foss [3] empirically made a theoretical argument to link NHRM and firm’s financial performance. Further, he recommended to extend existing research framework to explore firm’s innovation performance. From the above literature the proposed hypothesis is:

H1: “There is positive relationship between NHRM practices and innovation performance”.

### *2.5. IT Ambidexterity and Innovation Performance*

Prior literature focused on organizational ambidexterity, considering operational and strategic ambidexterity, Exploitation and exploration are the main strategies adopted to sustain firm performance. There should be a balance between exploitation and exploration to increase the firm innovation performance [23]. Following the idea of Chi, et al. [24], This study conceptualize a broader concept of ambidexterity by introducing ambidexterity between two contradictory dimensions of IT, named IT flexibility and IT standardization. Previous studies investigated the role of organizational IT infrastruc-



**Figure 1.** Model of study.

ture to improve the organizational governance [25]. Cao and Lumineau [26] recommended to devise an IT mechanism to increase the firm innovation performance. Therefore, based on previous literature, the proposed hypothesis is:

H2: “There is positive relationship between IT ambidexterity and innovation performance”.

### 2.6. Moderating Role of Information Technology Ambidexterity

Considering the resource based view of the firm (RBV) [27], IT ambidexterity is one of the major source lead to potentially superior firm performance due to causal ambiguity [10]. The adaptability and scalability of digital technologies is reflected by IT flexibility which could help to solve the emergent conflicts that are not included in exiting data [28]. However, IT standardization is a rigid approach and deals with limited relational governance. If a firm has high level of IT ambidexterity then it will be easier to pursue complementing strategies to increase innovation performance. Therefore, by utilizing IT ambidexterity with NHRM practices, A firms capability to innovate new products and services increases [25]. Based on above literature, the proposed hypothesis is:

H3: “There is moderate effect of IT ambidexterity between NHRM practices and innovation performance”.

## 3. Methodology

NADRA is the largest semi-government IT based organization of Pakistan with the workforce of 15,500 and around 800 data acquisition centers across the country, aiming at to become self-sustaining organization. The rationale behind creation of NADRA was developing to National online data warehouse which are accessible nationwide and registered National identity document for all citizens and non-citizens resident in Pakistan. The data of this study was conducted from a sample of front-service employees who consistently interact with customers. The survey was self-administrated and data was collected by using convenient sampling technique. Data was conducted from the three cities of Pakistan which are Lahore, Gujranwala and Jhelum. Initial stage target

**Table 1.** Employee Respondent's Profile and General IT knowledge information.

Demographics	No of Respondent Employees	Percent (%)
<i>Gender</i>		
Male	180	60
Female	120	40
<i>Age Group</i>		
20–35	130	43.33
35–45	120	40
45–65	50	16.66
<i>Education</i>		
Matric-FA	40	13.33
FA-BA	120	40
BA-MA	130	43.33
Higher Education	10	3.33
<i>IT Knowledge</i>		
Novice Level	90	30
Medium Level	130	43.33
High Level	80	26.66

samples are 500 employees from which 370 employees are returned their survey questionnaires. After the elimination of incomplete questionnaires, the final sample comprised 300 employees with response rate of 60% drawn from the semi-government organization NADRA. A detail of respondents is given in Table 1.

Questionnaire was divided into two sections the first section included demographic questions (Gender, Education, Age group and IT knowledge) and other section included the questions related to main variables of this study. NHRM practices (independent variable) were measured using an eight-item scale developed by Santangelo and Pini [29]. IT ambidexterity (moderator) was measured using a four-item scale developed by [Rai and Tang [30], Saraf, et al. [31]]. Innovation performance (dependent variable) was measured using a four-item scale developed by Li, et al. [32]. All measures employed 5-point Likert-type scale ranges from 1 = strongly disagree to 5 = strongly disagree.

#### 4. Analysis

The analysis was performed by using SPSS 22 (Trial version) then by using factor analysis and reliability analysis the reliability and validity of questionnaire was assessed [33]. All the factor loading were greater than 0.40 that is according to criteria [34]. Reliability was assessed using values of Cronbach's alpha; the reliability of all variables was greater than 0.70 which meets the criteria [35]. The results of reliability and validity are shown in Table 2. There was significant positive correlation in between variables. Correlation between variables ranges from 17% to 26% as shown in Table 3.

##### 4.1. Results of Hypothesis

Following Hayes and Matthes [36], This study used MODPROBE syntax in SPSS, a widely-used technique to check the moderation in a structural model. Introducing

**Table 2.** Factor Loadings & Reliability Analysis.

Variables Name		Items	Factor Loadings
New HRM practices (NHRM) $\alpha = 0.78$	NHRM1	Team Work	0.82
	NHRM2	Total quality projects	0.64
	NHRM3	Job Rotation	0.65
	NHRM4	Autonomy in problem solving	0.54
	NHRM5	Permanent training	0.75
	NHRM6	Involvement of employee	0.54
	NHRM7	Structured channels for workers' suggestions on organizational topics.	0.70
	NHRM8	Structured channels for workers' suggestions on quality topics.	0.60
IT Ambidexterity $\alpha = 0.85$	IF1	IT hardware facilities support the integration of various system platforms (such as online trading, logistics and financial settlement, etc.)	0.56
	IF2	System software or functional components support inter-firm system integration and expansion.	0.80
	IS1	Our firm has standardized business process and data.	0.74
	IS2	Our firm has adopted normalized electronic communication medias (such as Wechat, Alibaba, trade manager) to communicate.	0.57
Innovation Performance $\alpha = 0.92$	IP1	New products or new services reach the expected customer value and sales target.	0.60
	IP2	New products or new services reach the expected market share target	0.65
	IP3	New products or new services reach the expected profit target	0.57
	IP4	New products or new services reach expected sales growth target.	0.71

**Table 3.** Mean, Standard Deviation and Correlation.

Variables	Mean	SD	NHRM	IT Ambidexterity	Innovation Performance
NHRM	2.34	0.658	1		
IT Ambidexterity	2.55	0.542	0.22***	1	
Innovation Performance	3.54	1.065	0.17***	0.26***	1

\*\*\*p<0.001

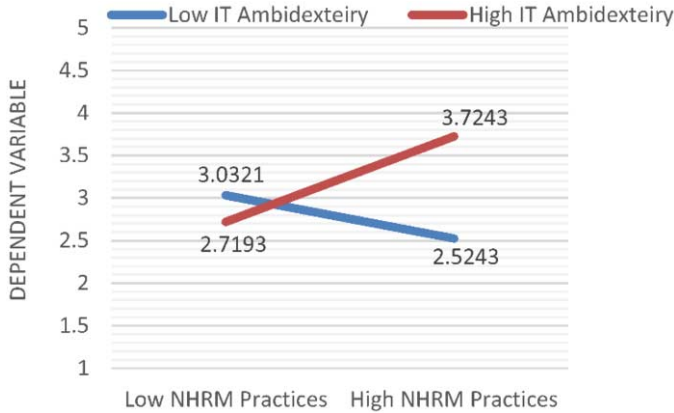
NHRM practices as an independent variable, innovation performance as dependent variable, and IT ambidexterity as moderating variable, statistical output was created using MODPROBE syntax with 5000 bootstrap resamples. The conditional effect of IT ambidexterity was estimated at values of one standard deviation below the mean, the mean, and one standard deviation above the mean [37]. The results are shown in Table 4.

Hypothesis 1 proposed a relationship between NHRM practices and innovation performance in NADRA employees. The results showed significant positive relationship between NHRM practices and Innovation performance (H1:  $\beta = 0.1243$ ,  $t = 8.6817$ ,  $p < 0.001$ ). Hypothesis 2 predicted a relationship between IT ambidexterity and innovation performance. Results revealed a significant positive relationship between IT ambidexterity and innovation performance (H1:  $\beta = 0.2218$ ,  $t = 6.4190$ ,  $p < 0.001$ ).

**Table 4.** Results of Hypothesis (MODPROBE output).

Hypothesis	Variables	$\beta$	SE	t	p
H1	NHRM	0.1243	0.0358	2.6817	0.000***
H2	IT Ambidexterity	0.2218	0.0346	6.4190	0.000***
H3	Interaction (NHRM* IT)	0.3782	0.0325	3.1032	0.000***

\*\*\*p<0.001



**Figure 2.** Two-way Interaction effect.

Hypothesis 3 proposed a moderating effect of IT ambidexterity between NHRM practices and innovation performance in NADRA employees. According to recommendation of Dawson [38], This study used an Excel worksheet to show interaction effect (NHRM practices x IT ambidexterity) on employee’s innovation performance. Slope significance tests were performed using the “Two-way interaction effect” Excel worksheet developed by Jeremy Dawson (available on <http://www.jeremydawson.co.uk/slopes.html>). Figure 2 shows that effect of NHRM practices on innovative performance is stronger when IT ambidexterity is high ( $\beta = 0.3782$ ), and weaker when IT ambidexterity is low (0.1243). The results prove a moderating role of IT ambidexterity in between NHRM practices and Innovation performance. The results are shown in Table 4.

**5. Discussion**

This study found that NHRM practices could affect innovation performance in NADRA employee. The results are consistent with study of Laursen and Foss [3]. Therefore, a major benefit of adopting NHRM practices is to increase innovation performance in employee. These results are also in lined with the recent studies creating firm value by integrating information technology is a major challenge now days. IT ambidexterity was introduced as a moderator in between NHRM practices and innovation performance. The effects of NHRM practices on innovation performance are stronger when IT ambidexterity is high and weaker when IT ambidexterity is low. Previous research gave importance to organizational ambidexterity and neglected the

wide applications IT ambidexterity [23,39,40]. Therefore, this study adopted the wider conceptualization of IT ambidexterity by including IT flexibility and IT standardization. Moderating role of IT ambidexterity was supported in the study and matched with the findings of Chi, et al. [24]. The combined effect of NHRM practices and IT ambidexterity creates an important role in the innovation performance of employees. As NADRA is IT oriented semi-government firm and adopting new methods of digital recognition of a firm. It needs IT oriented professional for innovation performance. NADRA authorities should focus on internal branding through furnishing employees IT skills [41]. However, previous research did not clearly describe about how and when IT ambidexterity generate innovation performance [42]. This study answered the question by identifying an important element (IT ambidexterity) in pursuing innovation performance.

## **6. Conclusion**

Many studies from diversified fields predicted a positive association between NHRM practices and innovation performance. However, when distinguishing between different organizations types, there is lack of research in IT based companies. This study is pioneer attempt to focus on predictors of innovation performance in IT based semi-government organization, NADRA. Our empirical analysis confirmed positive relationship between NHRM practices and innovation performance. Moderating role of IT ambidexterity was also confirmed. These results indicate that by integrating new technologies and knowledge, firms can increase of capability of improving existing products and services. In addition, utilizing new technologies and knowledge consistently, contribute for the enhancement of innovation performance.

## **7. Managerial Implications**

This study suggests that, Organizations should not only focus on the improvement in the existing product and service. But it is need to update the IT infrastructure, and to provide IT based training to increase short term innovation performance. Continuous adaption of technology also enhances long term competitive advantage. Previous research ignores the role of IT ambidexterity in firm performance.

## **8. Limitations and Future Research Directions**

This study is based on a case study in a semi-government firm in Pakistan, named "NADRA" and sample was selected from major cities of Pakistan. Future research can be expanded to other geographical regions and types of firms. Therefore, application of this model needs to be tested in other types of firms. Data obtained in this study is cross sectional. Time series study can be potential research in the future. This study checked the effect of NHRM practices, and IT ambidexterity on innovative performance. However, other emerging variables like work place democracy, organizational politics, and employee relation climate can be potential predictors of innovative performance.

## References

- [1] D. E. Guest, "Human resource management and performance: a review and research agenda," *International Journal of Human Resource Management*, vol. 8, pp. 263–276, 1997.
- [2] K. Jiang, D. P. Lepak, J. Hu, and J. C. Baer, "How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms," *Academy of Management Journal*, vol. 55, pp. 1264–1294, 2012.
- [3] K. Laursen and N. J. Foss, "New human resource management practices, complementarities and the impact on innovation performance," *Cambridge Journal of Economics*, vol. 27, pp. 243–263, 2003.
- [4] H. Mendelson and R. R. Pillai, "Information age organizations, dynamics and performance," *Journal of Economic Behavior & Organization*, vol. 38, pp. 253–281, 1999.
- [5] J. Godard, "What is best for workers? The implications of workplace and human resource management practices revisited," *Industrial Relations: A Journal of Economy and Society*, vol. 49, pp. 466–488, 2010.
- [6] M. Rizov and R. Croucher, "Human resource management and performance in European firms," *Cambridge Journal of Economics*, vol. 33, pp. 253–272, 2009.
- [7] A. Waheed, S. Waheed, J. Karamat, N. Ahmad, and A. Majeed, "Implementation and adoption of E-HRM in Small and Medium Enterprises of Pakistan," 2017.
- [8] A. Waheed, M. Xiaoming, J. Karamat, and S. Waheed, "Comparison of Human Resource Planning and Job Analysis process in banking sector of Pakistan," 2016.
- [9] V. Sambamurthy, A. Bharadwaj, and V. Grover, "Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms," *MIS Quarterly*, pp. 237–263, 2003.
- [10] S. Mithas and R. T. Rust, "How Information Technology Strategy and Investments Influence Firm Performance: Conjecture and Empirical Evidence," *Mis Quarterly*, vol. 40, pp. 223–245, 2016.
- [11] C. Ichniowski, K. Shaw, and G. Prennushi, "The effects of human resource management practices on productivity: A study of steel finishing lines," *The American Economic Review*, pp. 291–313, 1997.
- [12] N. Ahmed, A. K. Khattak, N. Iqbal, O. Farooq, and J. Iqbal, "Role of celebrity endorsement upon consumer vanity with mediating role of materialism: Evidence from business students of Pakistan," *Journal of Business and Management Research*, vol. 4, pp. 98–107, 2014.
- [13] J. C. Hayton, "Promoting corporate entrepreneurship through human resource management practices: A review of empirical research," *Human Resource Management Review*, vol. 15, pp. 21–41, 2005.
- [14] V. Luchsinger and D. R. Bagby, "Entrepreneurship and intrapreneurship: Behaviors, comparisons, and contrasts," *SAM Advanced Management Journal*, vol. 52, p. 10, 1987.
- [15] S. Waheed, A. Waheed, S. Habib, A. Majeed, and I. Ullah, "Impact of Social Media on Effective Recruitment in Business Vendors of Pakistan," 2017.
- [16] R. Carnegie and M. Butlin, *Managing the innovative enterprise: Australian companies competing with the world's best*: Information Australia, 1993.
- [17] R. Chaganti and F. Damanpour, "Institutional ownership, capital structure, and firm performance," *Strategic Management Journal*, vol. 12, pp. 479–491, 1991.
- [18] J. Recker and J. Mendling, "The state of the art of business process management research as published in the BPM conference," *Business & Information Systems Engineering*, vol. 58, pp. 55–72, 2016.
- [19] R. Xie, H. Ling, and C. Zhang, "Effect of business process management on firm performance: An ambidexterity perspective," in *Business Management and Electronic Information (BMEI), 2011 International Conference on*, 2011, pp. 341–345.
- [20] M. Tushman and W. Moore, "Readings in the management of technological innovation," ed: New York: Harper Collins, 1998.
- [21] W. R. Evans and W. D. Davis, "High-performance work systems and organizational performance: The mediating role of internal social structure," *Journal of Management*, vol. 31, pp. 758–775, 2005.
- [22] J. Michie and M. Sheehan, "HRM practices, R&D expenditure and innovative investment: evidence from the UK's 1990 workplace industrial relations survey (WIRS)," *Industrial and Corporate Change*, vol. 8, pp. 211–234, 1999.
- [23] S. Raisch, J. Birkinshaw, G. Probst, and M. L. Tushman, "Organizational ambidexterity: Balancing exploitation and exploration for sustained performance," *Organization Science*, vol. 20, pp. 685–695, 2009.
- [24] M. Chi, J. Zhao, J. F. George, Y. Li, and S. Zhai, "The influence of inter-firm IT governance strategies on relational performance: The moderation effect of information technology ambidexterity," *International Journal of Information Management*, vol. 37, pp. 43–53, 2017.
- [25] A. Tiwana and B. Konsynski, "Complementarities between organizational IT architecture and governance structure," *Information Systems Research*, vol. 21, pp. 288–304, 2010.



- [26] Z. Cao and F. Lumineau, "Revisiting the interplay between contractual and relational governance: A qualitative and meta-analytic investigation," *Journal of Operations Management*, vol. 33, pp. 15–42, 2015.
- [27] G. Piccoli and B. Ives, "Review: IT-dependent strategic initiatives and sustained competitive advantage: a review and synthesis of the literature," *Mis Quarterly*, vol. 29, pp. 747–776, 2005.
- [28] P. P. Tallon and A. Pinsonneault, "Competing perspectives on the link between strategic information technology alignment and organizational agility: insights from a mediation model," *Mis Quarterly*, pp. 463–486, 2011.
- [29] G. D. Santangelo and P. Pini, "New HRM practices and exploitative innovation: A shopfloor level analysis," *Industry and Innovation*, vol. 18, pp. 611–630, 2011.
- [30] A. Rai and X. Tang, "Leveraging IT capabilities and competitive process capabilities for the management of interorganizational relationship portfolios," *Information Systems Research*, vol. 21, pp. 516–542, 2010.
- [31] N. Saraf, C. S. Langdon, and S. Gosain, "IS application capabilities and relational value in interfirm partnerships," *Information Systems Research*, vol. 18, pp. 320–339, 2007.
- [32] X.-l. Li, Y.-w. Yao, and L.-j. Wang, "Research on the Relationship between New Venture Entrepreneurial Orientation and Innovative Performance: Proactive Market Orientation as a Mediator [J]," *China Industrial Economics*, vol. 6, p. 013, 2010.
- [33] F. Shaheen, N. Ahmad, M. Waqas, A. Waheed, and O. Farooq, "Structural Equation Modeling (SEM) in Social Sciences & Medical Research: A Guide for Improved Analysis," *International Journal of Academic Research in Business and Social Sciences*, vol. 7, 2017.
- [34] M. R. Lininger, C. A. Smith, N. J. Chimera, P. Hoog, and M. Warren, "Tuck Jump Assessment: An Exploratory Factor Analysis in a College Age Population," *The Journal of Strength & Conditioning Research*, vol. 31, pp. 653–659, 2017.
- [35] M. Lyvers, J. Karantonis, M. S. Edwards, and F. A. Thorberg, "Traits associated with internet addiction in young adults: Potential risk factors," *Addictive Behaviors Reports*, vol. 3, pp. 56–60, 2016.
- [36] A. F. Hayes and J. Matthes, "Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations," *Behavior Research Methods*, vol. 41, pp. 924–936, 2009.
- [37] P. A. Frazier, A. P. Tix, and K. E. Barron, "Testing moderator and mediator effects in counseling psychology research," *Journal of Counseling Psychology*, vol. 51, p. 115, 2004.
- [38] J. F. Dawson, "Moderation in management research: What, why, when, and how," *Journal of Business and Psychology*, vol. 29, pp. 1–19, 2014.
- [39] N. Haider, N. Ahmad, O. Farooq, I. Rasheed, and S. Parveen, "The role of organizational resources and environment in Organizational performance and customer loyalty; service climate as mediator: A Study of Telecommunication Sector of Pakistan," *Journal of Business and Management Research*, vol. 6, pp. 151–161, 2014.
- [40] W. Yan, Y. Zhu, and N. Ahmad, "Theory of Inventive Problem Solving (TRIZ) based contradiction resolution strategies for Shaanxi Aviation Industrial Upgrading," in *Industrial Engineering and Engineering Management (IEEM), 2016 IEEE International Conference on*, 2016, pp. 1111–1115.
- [41] N. Ahmad, N. Iqbal, R. Kanwal, H. Javed, and K. Javed, "The mediating role of employee engagement in relationship of internal branding and brand experience: Case of service organizations of Dera Ghazi Khan," *International Journal of Information, Business and Management*, vol. 6, pp. 26–41, 2014.
- [42] L. Cao, K. Mohan, B. Ramesh, and S. Sarkar, "Evolution of governance: achieving ambidexterity in IT outsourcing," *Journal of Management Information Systems*, vol. 30, pp. 115–140, 2013.

# The Droplets Constant Culturing Analyzer Based on Microfluidic Chip Control

Qiao Xinyong<sup>a,b</sup>, Cai Qiang<sup>a,b,1</sup>, Huang Cancan<sup>b</sup>, Wang Zhenhua<sup>b</sup>

<sup>a</sup> Hangzhou dianzi University, Hangzhou Zhejiang, China

<sup>b</sup> Yangtze Delta Region Institute of Tsinghua University, Jiaxing Zhejiang, China

**Abstract.** The droplets constant culturing analyzer based on microfluidic chip control is a device that cut the bacterial fluid into 800 microns long columnar droplets in parallel to culture, in order to realize real-time monitor and control parameters like pH and specific chemical factors of each parallel segments. The instrument can perform micro-droplet segmentation, fresh medium droplet generation, electric fusion of new micro-drops function, realizing the cell culture and passage operation of micro-droplet, thus realizing the function of microbial amplification and mutation adaptation evolution of microorganisms.

**Keywords.** Microdroplet, Constancy, bacterial cell, absorbance

## 1. Introduction

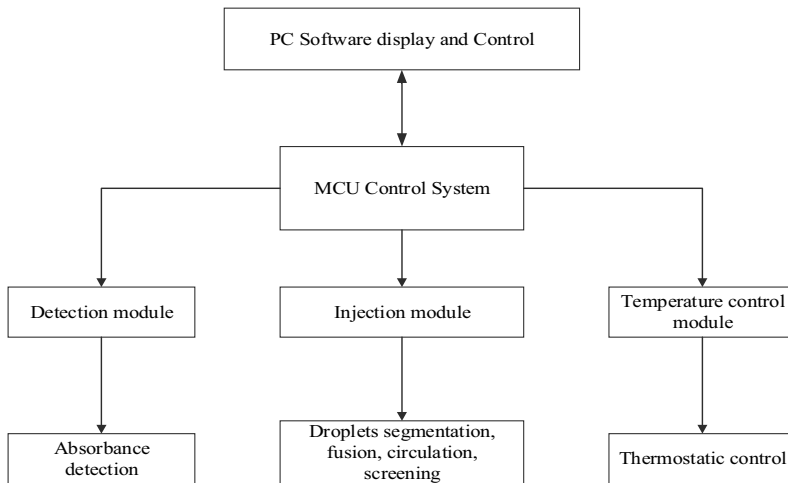
The In-situ microbial cultures and analysis which underpin the microorganisms screening and detection are the key role in the microbial technology industry. The microbial which had the drawbacks of small amount in sample quantum, a wide range of microorganism species, etc. To cover come these defects, the technological to diminish the reagents cost in microbial cultivation, increase the flux of microbial stability training has become a hotspot in microbial technology industry development<sup>[1]</sup>. Therefore, development a microbial cultivation device which has the characterize of efficient, trace amounts will be helpful in building up an industrial scale of micro-organisms selection with high-throughput, rapid detection of targets (such as pathogenic microorganism), el. And will meet the needs of biological industry on the microorganisms with patents' breeding, according with the demands in area of environment and food security, in detecting the trace amount of pathogenic microorganism with the facilities of integration and portable<sup>[2]</sup>. Droplet constant chemical analyzer based on microfluidic chip with the double function of microorganism cultivation and mutagenesis of adaptive evolution. Including "droplet constant/cultivate chips", "high precision flow control system" and "training data real-time acquisition module" three aspects, the key technology and solutions including droplets constant/cultivate chips, micro droplet control system development, droplet constant chemical analyzer based on microfluidic chip development.

---

<sup>1</sup> Corresponding author: Cai Qiang, Professor, Analysis & Measurement Center, JiaXing China, E-mail:caiq@tsinghua.edu.cn.

## 2. System Design

Microbial micro-droplet constant is based on micro-flow control technology and automatic control design of intelligent, informative high-throughput microbial culture apparatus. This design is based on micro-flow control micro-droplet microbial culture technology. Using bio-chips, the micro-droplets of oil-packed water, microorganisms, media, chemical factors, etc. were cultivated in micro-drops. The micro-droplet segmentation, fusion, cycling, screening and other actions are controlled by sensors. The micro-droplet is detected in the area, on-line measuring the Microbial growth status (OD value) of each micro-droplet. The system mainly consists of main control module, detection module, injection module, temperature control module and software part. The system scenario diagram is shown in Figure 1 :



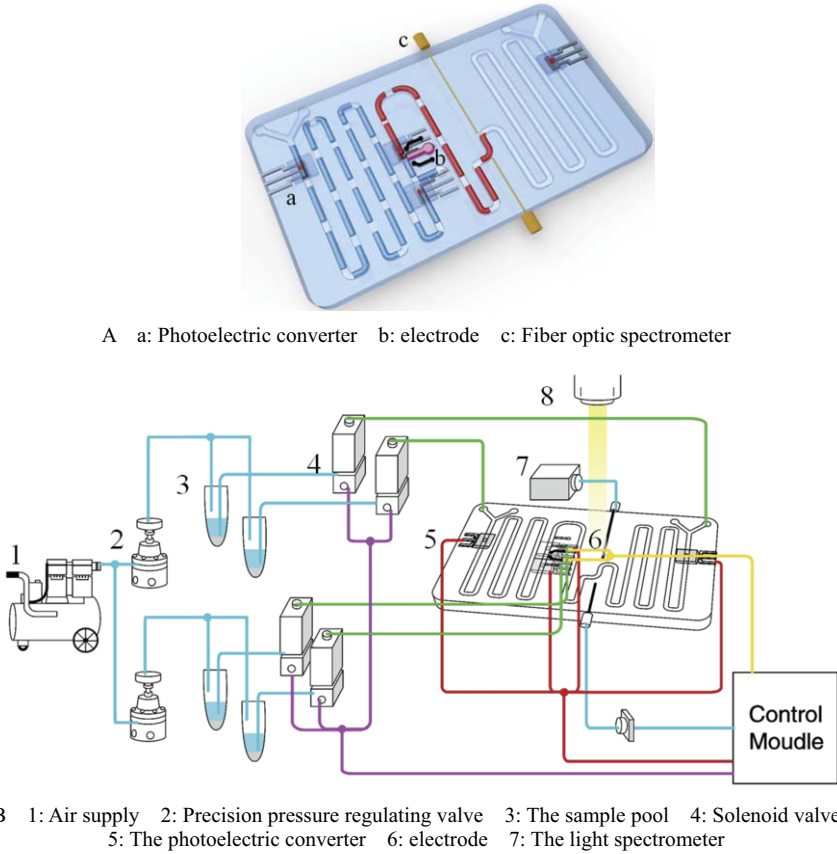
**Figure 1.** System scenario diagram

In the detection module in the graph, the main transmission of microorganism absorbance detection, by measuring OD value to determine the microbial growth of each micro-droplet. In the process of microbial culture, the temperature control will be controlled by temperature. MCU will pass the test parameters through 232 bus to the host computer software, in the upper computer display, and can through the host computer micro-droplet flow control and temperature control.

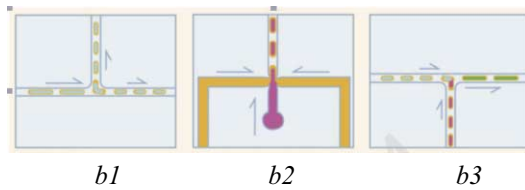
## 3. Micro-Fluidic System Design and Fabrication

In order to ensure the mutant strain in the micro-droplets can be cultivated for a long time and achieve real-time detection of the growth condition of microorganism<sup>[3]</sup>, chip structure of the implementation was designed as following figure 2(A&B). The droplet (length 800  $\mu$  m) of the oil-pack water is generated in the Y-shaped pipe (fuel, water), until the tiny droplet filled the left half of the chip. The long-term cultivation process of microorganisms in the micro-droplet, will consume the nutrients in the medium, for

realize the next generation culture of microbial, fresh medium must supply continually<sup>[4]</sup>. In this project, micro-droplet constant was designed in two parts, which included tiny droplet segmentation and droplet fusion (fig. 3).



**Fig 2.** Cultivate chip structure



b1: Droplet partition b2: Generate fresh medium droplets b3: Droplet fusion

**Fig 3.** Droplet division and integration

#### 4. Hardware circuit design

System hardware circuit design mainly includes: power supply design, OD detection module design and feedback light path design, system temperature control and shading design<sup>[5]</sup>. Aiming at the multi modular control of micro-flow control system and the characteristics of multi-equipment coordination work, the power supply is used in two ways and the filter module is added to ensure the stability of the system. The system adopts the temperature control and the lower computer software dual-isolating power supply, the pump adopts the LAN injection pump to ensure the droplet precision, the narrow band filter of OD detection (the center wavelength 650nm), the bandwidth of 20nm around, guarantees the maximum absorption wavelength of the water item, facilitates the supply OD value detection accuracy. The feedback optical path system ensures the stability of the excited laser. Because the cross-section area of the pipeline is 0.5mm<sup>2</sup>, the installation accuracy of microfluidic chips is proposed, which requires the design precision of the micro-flow control chip installation platform to 0.01mm. The cell needs to be cultivated at a specific temperature, so it is necessary to design a closed temperature control environment, and the temperature control accuracy is 0.5 °C. The droplet position is realized through the photodiode and the associated led lattice, so the shading system needs to be realized<sup>[6]</sup>.

#### 5. Software design

Software design is divided into two parts: the software design of the slave computer and the software design on the PC. PC software mainly realizes the whole training process: The software cleaning pipeline flow of the host computer; droplet generation; the liquid segmentation, droplet fusion, droplet od detection is carried out during the cycle. The main realization of the computer software and the communication of the host computer and related data collection and the pump valve control command forwarding function. The PC software interface is shown in Figure 4.

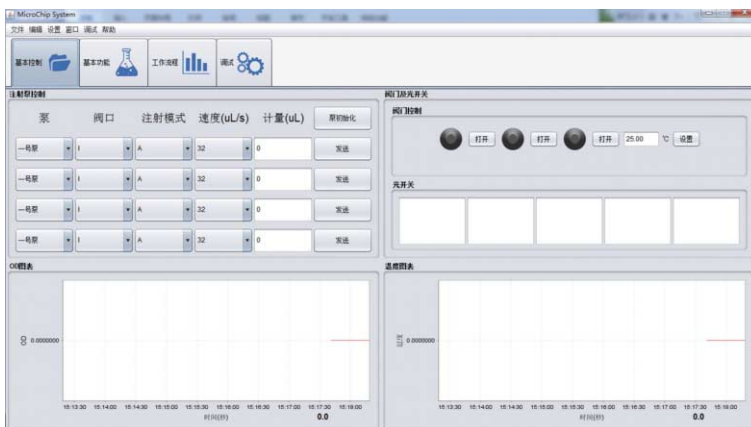
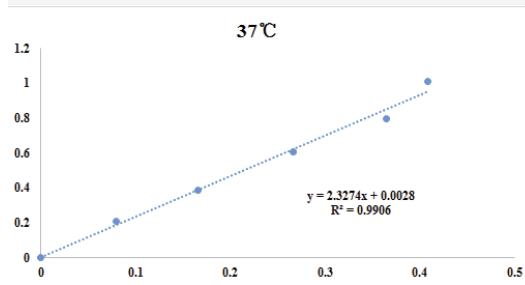


Figure 4. PC software interface

**6. Results**

The final test results can be obtained as follows (Fig. 5), and the microfluidic optical path is 5 mm. the conclusion is as follows, R<sup>2</sup> is greater than 0.99, the most important is that the slope consistency is 2.32, indicating that the results of the test at other temperatures will be reliable.



**Fig 5.** Droplets chemostat test report

Using two kinds of concentration gradient malachite green in this instrument and the standard spectrophotometer to measure the OD value, linearity is more than 0.99%. This system can be concluded that the micro-droplet od detection is more precise. As shown in table 1, two sets of test data:

**Table 1.** Two test data

	voltage		voltage		Standard
concentration	value 1	Absorbance 1	value 2	Absorbance 2	absorbance
0	2.8	0	2.78	0	0
0.001	2.61	0.030518	2.52	0.042644	0.0794
0.002	2.32	0.08167	2.23	0.09574	0.1671
0.003	2	0.146128	1.97	0.149579	0.2629
0.004	1.54	0.259637	1.48	0.273783	0.4226
0.005	1.31	0.329887	1.27	0.340241	0.5601
0.006	1.15	0.38646	1.09	0.406618	0.6793
0.007	1.01	0.442837	0.95	0.466321	0.7558

Microbial micro-droplet constant is based on micro-flow control technology and automatic control design of intelligent, informative high-throughput microbial culture apparatus. The flux of each chip is 200 tiny drops, each micro drops of the volume is about 2ul; Real-time on-line detection of microbial growth (OD value); Automatically replace the medium, complete the secondary culture. The custom program can periodically quantitatively replace the fresh medium and add chemical factor, and can be screened according to the growth condition (od value). Finally, the following indicators can be achieved:

(1) high flux and small size

1 chip can control 200 micro-drops, each micro-drops of the volume of about 2ul, far more than the shaker, deep hole plate flux, low consumption medium.

(2) the real-time detection of microbial growth

Real-time on-line detection of microbial growth in each micro-droplet, far more convenient than the traditional sampling and detection, real uninterrupted cultivation.

(3) Automatic change of the medium

Regular quantitative replacement of fresh medium to add chemical factor. Far more convenient to cultivate the seed liquor than traditional, save time and effort.

(4) intelligent filter

Set the filter OD range, automatically filter the appropriate strains, improve efficiency.

## Acknowledgments

The authors gratefully acknowledge the financial support of this research by Ministry of Science and Technology of the People' s Republic of China, Zhejiang Province Science and Technology Projects(2015C33009,2017F10045) and Jiaying Science and Technology plans (2015AY11008).

## References

- [1] Thorsen, T. , Roberts, R. W., Arnold, F. H., Quake, S. R. *Dynamic pattern formation in a vesicle-generating microfluidic device*. *Phys. Rev. Lett.* 2001, 86, 4163-4166
- [2] Song, H., Tice, J. D, Ismagilov, R. F. *A microfluidic system for controlling reaction networks in time*. *Angew. Chem. Int. Ed.* 2003, 115, 792—796.
- [3] Jakiela, S., Kaminski, T.S., Cybulski, O., Weibel, D.B., Garstecki, P. Bacterial Growth and Adaptation in Microdroplet Chemostats [J]. *Angew. Chem. Int. Ed.* 2013, 52: 8908-8911.
- [4] Balagaddé, F.K., You, L.C., Hansen, C.L., Arnold, F.H., Quake, S.R. Long-Term Monitoring of Bacteria Undergoing Programmed Population Control in a Microchemostat [J]. *Science*, 2005, 309: 137-140.
- [5] Baoyue Zhang, Min-Cheol Kim, Todd Thorsen, Zhanhui Wang. A self-contained microfluidic cell culture system [J]. *Biomedical Microdevices* . 2009 (6)
- [6] Jingdong Chen, Di Chen, Yao Xie, Tao Yuan, Xiang Chen. Progress of Microfluidics for Biology and Medicine [J]. *Nano-Micro Letters*, 2013, (01):66-80

# Research on UAV Online Trajectory Planning Algorithm Based on C/FD-GMRES Method

FANG Bin<sup>1</sup>, YANG Jiang Yong, WANG Yan, XU Shuo

*Department of Traffic Administration and Engineering, Hunan Police Academy, Changsha, 410138, China*

**Abstract.** In order to ensure that the UAV online route planning results meet the terminal constraints of the task, a rolling optimization algorithm based on C/FD-GMRES for UAV online route planning is designed. The C/FD-GMRES is a real-time nonlinear receding horizon control approach; combining the homotopy continuation method differential approximation generalized minimum residual method instead of solving complex Riccati differential equations. At each sampling time, the C/FD-GMRES calculate a residual vector linear equation once and the product of Jacobi matrix and vector product is approximated to the forward difference, the GMRES fast algorithm is used to solve large scale linear equations in the end. The C / FD-GMRES method has a good real-time performance, not sensitive to the selection of initial estimates, and can meet the need of UAV terminal state constraints. Simulation results show that the algorithm can effectively avoid barriers of space and the planned trajectory can converge to a stable terminal state, meeting the real-time requirements.

**Keywords.** UAV; optimal control; online trajectory planning; traffic monitoring; C/FD-GMRES

## 1. Introduction

With the rapid development of UAV, it began to undertake more and more tasks, including remote target monitoring, traffic monitoring, fire monitoring and other tasks. In the process of executing tasks in complex environment, UAV must have the ability of online trajectory planning in order to avoid the obstacles, such as bridges, and buildings etc. Scholars have carried out extensive research on UAV online trajectory planning problem; putting forward the artificial potential field method, roadmap method, sampling method, intelligent calculation method and the optimal control method [1-4]. These methods improve the ability of obstacle avoidance of UAV effectively, while these methods are mostly guide UAV flight to target position safely, but ignored that UAV need to meet some constraints to complete the task, including UAV position and velocity constraints [5-7]. For example, when the UAV is used to carry out the traffic road reconnaissance mission, it is necessary to meet the requirements of reconnaissance altitude and reconnaissance angle, so as to ensure the image-forming condition. Therefore, this paper focuses on the design of the UAV online trajectory planning algorithm which satisfies the terminal state constraints.

---

<sup>1</sup> Corresponding Author.

FANG Bin: EMAIL: aifudi@163.com

This work was supported by Foundation of Hu nan Educational Committee (No:17B087)



## 2. Problem modeling

UAV path planning problem is generally nonlinear, optimal control problems (OCP) with state constraints and control constraints, The path planning method based on the optimal control theory can be easily used to solve the kinematic and dynamic constraints which are difficult to deal with by other methods. The solution is to find a set of optimal control variables, and minimize the general Bolza performance index, as:

$$J = \varphi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t)) dt \tag{1}$$

The first term on the right side of the equation represents the terminal cost:

$$\varphi(\mathbf{x}(t_f)) = \frac{1}{2} \left[ (\mathbf{x}(t_f) - \mathbf{x}_f)^T S_f (\mathbf{x}(t_f) - \mathbf{x}_f) \right] \tag{2}$$

Therein,  $\mathbf{x} = [x, y, z, \phi, \theta, \psi, V, \alpha, \beta, p, q, r]^T$  is the UAV state vector,  $x$ 、 $y$ 、 $z$  are the north component, the East component and the vertical component of the UAV position in the geographic coordinate system; the three variables  $\phi$ 、 $\theta$ 、 $\psi$  are roll angle, pitch angle and yaw angle respectively;  $V$  is UAV airspeed,  $\alpha$ 、 $\beta$  are the angle of attack and sideslip angle respectively;  $p$ 、 $q$ 、 $r$  are the angular velocity of UAV in body coordinate system;  $\mathbf{x}_f$  is the UAV fixed termination state vector, denoted as  $\mathbf{x}_f := [x_{1f} \ x_{2f} \ \dots \ x_{12f}]$ ,  $t_f$  means any termination time.  $S_f$  is a  $12 \times 12$  dimensional terminal cost weighted diagonal matrix, The integrand function in the right second term is Lagrange's function:

$$L(\mathbf{x}, \mathbf{u}) = \frac{1}{2} \left[ (\mathbf{x} - \mathbf{x}_f)^T Q (\mathbf{x} - \mathbf{x}_f) + \mathbf{u}^T R \mathbf{u} \right] \tag{3}$$

In the above equation(3),  $Q$  is a  $12 \times 12$  dimensional state deviation weighted diagonal matrix, diagonal element is denoted as  $Q_i$  ( $i = 1, 2, \dots, 12$ ),  $\mathbf{u} := [u_1, u_2, u_3, u_4]^T$  is control variable, comprise of Engine thrust, elevator deflection, rudder deflection and aileron deflection,  $R$  is a  $4 \times 4$  dimensional control cost weighted diagonal matrix, diagonal element is denoted as  $R_i$  ( $i = 1, 2, 3, 4$ ).

For UAV online trajectory planning, the following constraints are required:

(1) UAV continuous time differential dynamic equations. Without losing generality, the UAV six degree of freedom differential equation is as follows :

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}) \tag{4}$$

Therein,  $f(\mathbf{x}, \mathbf{u}) := [f_1 \ f_2 \ \dots \ f_{12}]^T$  is a nonlinear function of state and control.

(2)Boundary condition

$$\mathbf{x}(t_0) = \mathbf{x}_0, \quad \mathbf{x}(t_f) = \mathbf{x}_f \tag{5}$$

Therein,  $\mathbf{x}_0$  and  $\mathbf{x}_f$  are the initial and terminal states of UAV, respectively.

(3)Path Constraint

Generally, the control variables need to satisfy the input saturation constraints:

$$\begin{cases} u_{1\min} \leq u_1 \leq u_{1\max} , & |u_2| \leq u_{2\max} \\ |u_3| \leq u_{3\max} , & |u_4| \leq u_{4\max} \end{cases} \quad (6)$$

Therein,  $u_{1\min}$  and  $u_{1\max}$  are constants that represent the lower bound and upper bound of the engine thrust amplitude respectively,  $u_{2\max}, u_{3\max}, u_{4\max}$  are constants that represent the amplitude constraints of the elevator, rudder and aileron deflection respectively. In order to convert inequality constraints into equality constraints, dummy input variables  $u_5, u_6, u_7, u_8$  are introduced :

$$\begin{cases} C_1(u_1, u_5) := [(u_1 - u_{1\min})(u_1 - u_{1\max}) + u_5^2] / 2 = 0 \\ C_2(u_2, u_6) := (u_2^2 + u_6^2 - u_{2\max}^2) / 2 = 0 \\ C_3(u_3, u_7) := (u_3^2 + u_7^2 - u_{3\max}^2) / 2 = 0 \\ C_4(u_4, u_8) := (u_4^2 + u_8^2 - u_{4\max}^2) / 2 = 0 \end{cases} \quad (7)$$

According to the theory of optimal control [8], the UAV motion planning problem is equivalent to the first order necessary condition for the optimal control problem, define the hamiltonian function as follows:

$$H(\mathbf{x}, \lambda, \mathbf{u}, \mu) = L(\mathbf{x}, \mathbf{u}) + \lambda^T \mathbf{f}(\mathbf{x}, \mathbf{u}) + \mu^T C(\mathbf{x}, \mathbf{u}) \quad (8)$$

Therein, costate variables is denoted as  $\lambda := [\lambda_1 \ \lambda_2 \ \dots \ \lambda_{12}]^T$ , Lagrange multiplier of associated path constraint is denoted as  $\mu := [\mu_1, \mu_2, \mu_3, \mu_4]^T$ .  $C(\mathbf{x}, \mathbf{u}) = [C_1(u_1, u_5), C_2(u_2, u_6), C_3(u_3, u_7), C_4(u_4, u_8)]^T$  represents path constraint function. Since the Lagrange multiplier is also  $\mu$  included in the optimal control variable  $\mathbf{u}$ ,  $\mu$  is denoted as :

$$\mu = [\mu_1, \mu_2, \mu_3, \mu_4]^T := [u_9, u_{10}, u_{11}, u_{12}]^T \quad (9)$$

Because the sign of the dummy input variable does not affect the optimality of the solution, So the solution set  $\mathbf{u}(t)$  will be divided into two parts of opposite sign, and when the dummy input variable value is zero, it will cause the singularity of C/GMRES to update the optimal solution. Therefore, a small penalty factor is added to the cost function to obtain the modified Lagrange function:

$$\begin{aligned} L(\mathbf{x}, \mathbf{u}) &= \frac{1}{2} \left[ (\mathbf{x} - \mathbf{x}_f)^T Q (\mathbf{x} - \mathbf{x}_f) + \mathbf{u}^T R \mathbf{u} - \sum_{i=5}^8 r_i u_i \right] \\ &= \frac{1}{2} \left( \sum_{i=1}^{12} Q_i (x_i - x_{if})^2 + \sum_{j=1}^4 R_j u_j^2 - \sum_{i=5}^8 r_i u_i \right) \end{aligned} \quad (10)$$

The goal of the optimal control is to minimize the cost, so it is necessary to set the penalty factor to a larger normal number. However, the penalty factor is too large to affect the optimal solution, so choose  $r_i (i = 5, 6, 7, 8)$  as a small normal number. The formula (7), (9), (10) and the UAV differential dynamic equation (4) are plugged into the equation (8).

$$\begin{aligned}
 H(\mathbf{x}, \lambda, \mathbf{u}, \mu) &= L(\mathbf{x}, \mathbf{u}) + \lambda^T \mathbf{f}(\mathbf{x}, \mathbf{u}) + \mu^T C(\mathbf{x}, \mathbf{u}) \\
 &= \frac{1}{2} \left( \sum_{i=1}^{12} Q_i (x_i - x_{if})^2 + \sum_{i=1}^4 R_i u_i^2 - \sum_{i=5}^8 r_i u_i \right) + \sum_{i=1}^{12} \lambda_i f_i \\
 &\quad + \frac{1}{2} u_9 [(u_1 - u_{1\min})(u_{1\max} - u_1) - u_5^2] + \frac{1}{2} u_{10} (u_2^2 + u_6^2 - u_{2\max}^2) \\
 &\quad + \frac{1}{2} u_{11} (u_3^2 + u_7^2 - u_{3\max}^2) + \frac{1}{2} u_{12} (u_4^2 + u_8^2 - u_{4\max}^2)
 \end{aligned} \tag{11}$$

According to the variational method, the first order necessary conditions of UAV online trajectory planning problem are derived, which are the optimal control variable, the Lagrange multiplier and the costate variables satisfy the equation as follows:

$$\frac{\partial H}{\partial \mathbf{u}} = \left[ \frac{\partial H}{\partial u_1} \quad \frac{\partial H}{\partial u_2} \quad \dots \quad \frac{\partial H}{\partial u_{12}} \right]^T = 0 \tag{12}$$

$$\dot{\lambda} = -\frac{\partial H}{\partial \mathbf{x}} \tag{13}$$

$$\lambda(t_f) = \frac{\partial \varphi(\mathbf{x}(t_f))}{\partial \mathbf{x}(t_f)} \tag{14}$$

The formula (11) is substituted into formula (12) and formula (13), got the formula as follows :

$$\begin{aligned}
 \frac{\partial H}{\partial \mathbf{u}} &= \left[ \frac{\partial H}{\partial u_1} \quad \frac{\partial H}{\partial u_2} \quad \dots \quad \frac{\partial H}{\partial u_{12}} \right]^T \\
 &= \begin{bmatrix} R_1 u_1 + \sum_{i=1}^{12} \lambda_i \frac{\partial f_i}{\partial u_1} + \frac{1}{2} u_9 [u_{1\max} - 2u_1 - u_{1\min}] \\ R_j u_j + \sum_{i=1}^{12} \lambda_i \frac{\partial f_i}{\partial u_j} + u_{j+8} u_j, (j = 2, 3, 4) \\ u_{i+4} u_i - \frac{1}{2} r_i, (i = 5, 6, 7, 8) \\ \frac{1}{2} [(u_1 - u_{1\min})(u_{1\max} - u_1) - u_5^2] \\ \frac{1}{2} (u_2^2 + u_6^2 - u_{2\max}^2) \\ \frac{1}{2} (u_3^2 + u_7^2 - u_{3\max}^2) \\ \frac{1}{2} (u_4^2 + u_8^2 - u_{4\max}^2) \end{bmatrix} \\
 &= 0
 \end{aligned} \tag{15}$$

$$\lambda_i = -\frac{\partial H}{\partial x_i} = -\left( \sum_{i=1}^{12} Q_i (x_i - x_{if}) + \sum_{i=1}^{12} \lambda_i \frac{\partial f_i}{\partial x_i} \right) \tag{16}$$

The formula (2) is substituted into formula (14), got the following formula:

$$\lambda(t_f) = \frac{\partial \varphi(\mathbf{x}(t_f))}{\partial \mathbf{x}(t_f)} = S_f(\mathbf{x}(t_f) - \mathbf{x}_f) \quad (17)$$

In addition, synthesize equation (4) and (15), the following equation can be obtained:

$$F(\mathbf{U}(t), \mathbf{x}(t), t) := \begin{bmatrix} H_u^T(x_0^*(t), \lambda_1^*(t), u_0^*(t), \mu_0^*(t)) \\ C(x_0^*(t), u_0^*(t)) \\ \vdots \\ H_u^T(x_{N-1}^*(t), \lambda_N^*(t), u_{N-1}^*(t), \mu_{N-1}^*(t)) \\ C(x_{N-1}^*(t), u_{N-1}^*(t)) \end{bmatrix} = 0 \quad (18)$$

Therefore, the equation (16), (17), (18) and the differential equation of state (4) constitute the first order necessary conditions for the constrained optimal control problem. Among them, (4) and (16) can be obtained by using the initial value of iterative calculation, so the key problem is solving equation (18), which involves a large number of nonlinear equations and differential calculation of complicated calculation, put forward a challenge for real time UAV online planning algorithm. For solution of equation (18), Iterative methods such as Newton method are usually used. However, these methods are sensitive to the choice of the initial value and need to calculate the value of the function and its derivative value in each iteration, and there exists problem of big calculation.

Therefore, this paper proposes the Continuation/Forward Difference - Approximation Generalized Minimum Residual Method (C/FD-GMRES). Specifically, Is to solve the equation  $F(\mathbf{U}(t), \mathbf{x}(t), t) = 0$  by solving the time derivative of  $\mathbf{U}$ .

### 3. UAV Online trajectory planning algorithm based on C/FD-GMRES method

C/FD-GMRES is a kind of nonlinear receding horizon control method, combining the homotopy continuation method differential approximation generalized minimum residual method instead of solving complex Riccati differential equations. Unlike any linear search method or Newton iterative method for solving nonlinear equations by iterative control, C/FD-GMRES only at each sampling time calculates a residual vector linear equation and Jacobi matrix vector product using the forward difference approximation. finally, using a linear equation fast algorithm GMRES method to solve large linear equations. GMRES is a class of Krylov subspace method, It reduces the residuals monotonically and converges to the solution in the iteration number equivalent to the dimension of the equation, for large-scale linear equations, it can converge to the specified error tolerance with fewer iterations. C/FD-GMRES has good real-time performance and is not sensitive to the selection of initial estimation value, which is suitable for real-time optimization of discrete time optimal control problem.

The core idea of C/FD-GMRES to first compute the  $\mathbf{U}(0)$  that satisfies  $F(\mathbf{U}(t), \mathbf{x}(t), t) = 0$ , Then calculate the change rate of  $\dot{\mathbf{U}}$  to ensure satisfying the following equation:

$$\dot{F}(\mathbf{U}, \mathbf{x}, t) = A_s F(\mathbf{U}, \mathbf{x}, t) \quad (19)$$

$A_s$  is a stable factor for stable equilibrium position  $F(\mathbf{U}(t), \mathbf{x}(t), t) = 0$ . In order to express  $\dot{F}(\mathbf{U}, \mathbf{x}, t) = A_s F(\mathbf{U}, \mathbf{x}, t)$  as the standard form of GMRES problem :  $Ax = b$ , Firstly, the  $\dot{F}(\mathbf{U}, \mathbf{x}, t)$  is expanded as:

$$D_U F(\mathbf{U}, \mathbf{x}, t) \dot{\mathbf{U}} = \dot{F}(\mathbf{U}, \mathbf{x}, t) - D_x F(\mathbf{U}, \mathbf{x}, t) \dot{\mathbf{x}} - D_t F(\mathbf{U}, \mathbf{x}, t) \quad (20)$$

The formula (19) is substituted into formula (20), got the formula as follows :

$$D_U F(\mathbf{U}, \mathbf{x}, t) \dot{\mathbf{U}} = A_s F(\mathbf{U}, \mathbf{x}, t) - D_x F(\mathbf{U}, \mathbf{x}, t) \dot{\mathbf{x}} - D_t F(\mathbf{U}, \mathbf{x}, t) \quad (21)$$

By means of the transformation, the formula (21) can be transformed into the standard form of the GMRES problem, in which  $\dot{\mathbf{U}}$  is equivalent to  $\mathbf{x}$ ,

$D_U F(\mathbf{U}, \mathbf{x}, t)$  is equivalent to  $A$ ,  $A_s F(\mathbf{U}, \mathbf{x}, t) - D_x F(\mathbf{U}, \mathbf{x}, t) \dot{\mathbf{x}} - D_t F(\mathbf{U}, \mathbf{x}, t)$  is equivalent to  $b$ . However, there are still complex Jacobi calculation  $D_U F(\mathbf{U}, \mathbf{x}, t)$ ,  $D_x F(\mathbf{U}, \mathbf{x}, t)$  and  $D_t F(\mathbf{U}, \mathbf{x}, t)$ . so, the forward difference (FD) is used to approximate the product of Jacobi and vector, as follows:

$$D_{\Delta\tau} F(\mathbf{U}, \mathbf{x}, t; \dot{\mathbf{U}}, \dot{\mathbf{x}}, 1) \approx \frac{F(\mathbf{U} + \Delta\tau \dot{\mathbf{U}}, \mathbf{x} + \Delta\tau \dot{\mathbf{x}}, t + \Delta\tau) - F(\mathbf{U}, \mathbf{x}, t)}{\Delta\tau} \quad (22)$$

$$D_U F(\mathbf{U}, \mathbf{x}, t) \dot{\mathbf{U}} \approx \frac{F(\mathbf{U} + \Delta\tau \dot{\mathbf{U}}, \mathbf{x}, t) - F(\mathbf{U}, \mathbf{x}, t)}{\Delta\tau} := D_{\Delta\tau} F(\mathbf{U}, \mathbf{x}, t; \dot{\mathbf{U}}, 0, 0) \quad (23)$$

$$D_x F(\mathbf{U}, \mathbf{x}, t) \dot{\mathbf{x}} + D_t F(\mathbf{U}, \mathbf{x}, t) \approx \frac{F(\mathbf{U}, \mathbf{x} + \Delta\tau \dot{\mathbf{x}}, t + \Delta\tau) - F(\mathbf{U}, \mathbf{x}, t)}{\Delta\tau} := D_{\Delta\tau} F(\mathbf{U}, \mathbf{x}, t; 0, \dot{\mathbf{x}}, 1) \quad (24)$$

In the literature [9], the method of mixed FD and GMRES is called FD-GMRES. The equation (21) is solved by FD-GMRES and get  $\dot{\mathbf{U}}$ . Then, getting  $\mathbf{U}$  by integrating  $\dot{\mathbf{U}}$ .

In accordance with the above ideas, the real-time online trajectory planning algorithm based on C/FD-GMRES includes two steps: (1) solving the initialization control variable  $\mathbf{u}_0$ , the algorithm flow chart is as shown in the following figure 1.(2)solving control variables  $\mathbf{u}_k, k = 1, 2 \dots N$  by rolling optimization, the algorithm flow chart is as shown in the following figure 2. therein, the output  $\mathbf{u}_0$  of the algorithm (1) will be used as the starting value of the control variable  $\mathbf{u}_k$ .

### 3.1. Solving control variable start value $\mathbf{u}(0)$

A simple way to solving the control variable start value  $\mathbf{u}(0)$  is to select the Time-domain function  $T(t)$  as a smooth function, in which  $T(0) = 0$  and  $T(t) \rightarrow T_f(t \rightarrow \infty)$ . Then,  $\mathbf{u}_i^*(0) = \mathbf{u}(0)$  and  $\mu_i^*(0) = \mu(0)$  ( $i = 0, \dots, N-1$ ),  $\mathbf{x}_i^*(0) = \mathbf{x}(0)$  and  $\lambda_i^*(0) = \varphi_x^T(\mathbf{x}(0))$  ( $i = 0, \dots, N$ ). so,  $\forall i, j, 0 \leq i, j \leq N-1 (i \neq j)$ , Obviously, the initialization of  $\mathbf{u}(0)$  is reduced to solving the  $\mathbf{u}(0)$  and  $\mu(0)$  that satisfy the formula (25).

$$\begin{aligned} & \|F(\mathbf{u}(0), \mathbf{x}(0), 0)\| \approx \\ & \left\| \begin{bmatrix} H_u^T(\mathbf{x}(0), \varphi_x^T(\mathbf{x}(0)), \mathbf{u}(0), \mu(0)) \\ C(\mathbf{x}(0), \mathbf{u}(0)) \end{bmatrix} \right\| \leq \frac{\delta}{\sqrt{N}} \end{aligned} \quad (25)$$

Therein,  $\delta$  is any positive real number.

In the Algorithm, the key is to use GMRES to optimize the solution  $\mathbf{u}(0)$ . First of all, using modified gram Schmidt orthogonalization method to obtain orthogonal vector group; then, using Vince rotation to optimize the control rate, and converges to the residual tolerance within the iteration times. In general, GMRES monotonically decreases the residuals and converges to the solution in the iteration number equivalent to the dimension of the equation, In particular, for large-scale linear equations, GMRES is able to converge to the specified error tolerance with fewer iterations. Therefore, this property can satisfy the real-time requirement of UAV online motion planning. Therefore, this property can satisfy the real-time requirement of UAV online trajectory planning.

In the algorithm of solving control variable starting value  $\mathbf{u}(0)$ , the input is the initial state  $\mathbf{x}$ , start time  $t_0$ , time step  $\Delta t$ , residual tolerance  $rtol$ , and GMRES iteration times  $k_{max}$ , the output is the control variable starting value is  $\mathbf{u}(0)$ .

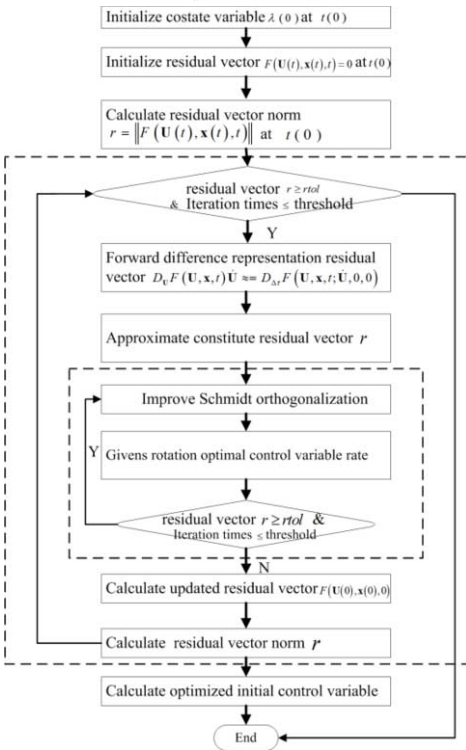


Figure 1. Solve control start variable  $\mathbf{u}(0)$

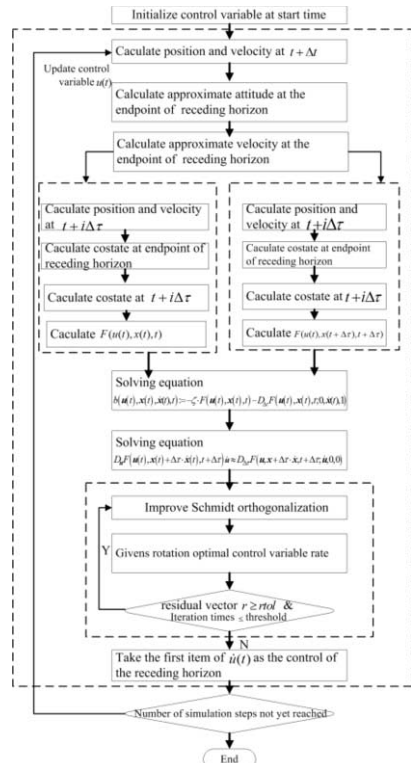


Figure 2. Solve control variable  $\mathbf{u}_k$  with rolling optimization

### 3.2. Solving control variable $\mathbf{u}(t)$ based on rolling optimization

The core of solving control variable  $\mathbf{u}(t)$  is to solve the TPBVPs problem. Firstly, the position and velocity of the discrete time points in the receding horizon are calculated; Secondly, obtaining the common state value at the end point of the receding horizon; Then, the common state values of each discrete point are calculated; finally, calculating the value of  $F(\mathbf{u}(t), \mathbf{x}(t), t)$  for time  $t$  and the value of  $F(\mathbf{u}(t), \mathbf{x}(t) + \Delta\tau \dot{\mathbf{x}}(t), t + \Delta\tau)$  for time  $t + \Delta\tau$ , parameter  $\Delta\tau = T_f (1 - e^{-\alpha(t + \Delta\tau)}) / N$ , and then construct the residual vector.

In the algorithm of solving control variable  $\mathbf{u}(t)$ , the input is the initial state  $\mathbf{x}$ , start time  $t_0$ , end time  $t_f$ , time step  $\Delta t$ , single step receding horizon length  $\Delta\tau$ , receding horizon parameter  $T_f$  and  $\alpha$ , receding horizon discretization steps  $N$ , stability factor  $\zeta$ , residual tolerance  $rtol$ , GMERS iteration times  $k_{\max}$ ; the output is current control variable  $\mathbf{u}(t)$ .

## 4. Simulation experiment

In order to verify the effectiveness and time performance of UAV trajectory planning algorithm based on C/FD-GMRES method, the simulation experiment is carried out under the MATLAB 2012a, the hardware environment of the experiment is Inter Core CPU T2450, memory 2.5GB. Table 1 gives the UAV state and control saturation constraint.

**Table 1.** UAV state and control saturation constraint

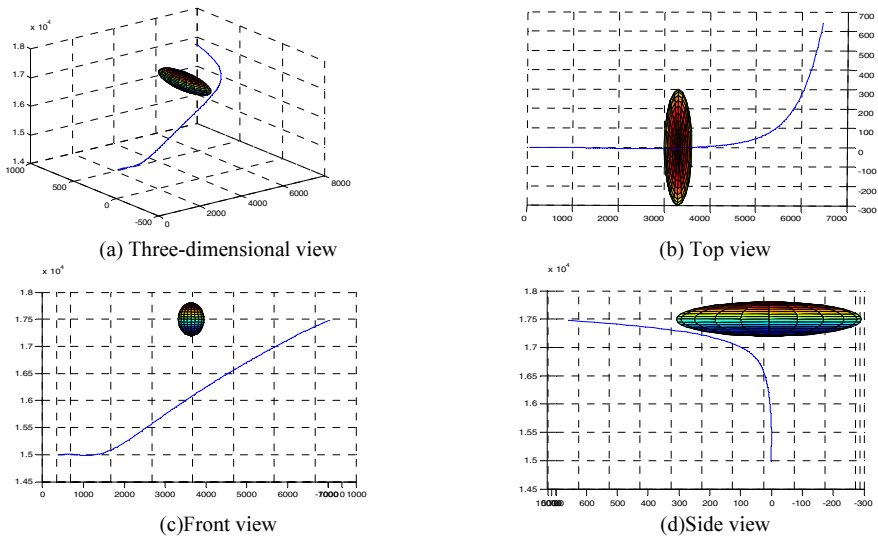
Decision variable	Boundary condition	Decision variable	Boundary condition	Decision variable	Boundary condition
$t_f (s)$	[0,1000]	$\psi (rad)$	$[-\pi, \pi]$	$\alpha (deg)$	[-10,45]
$x(m)$	—	$p(rad/s)$	[-0.2,0.2]	$\beta(deg)$	[-10,10]
$y(m)$	—	$q(rad/s)$	[-0.2,0.2]	$\phi(rad)$	$[-\pi/3, \pi/3]$
$z(m)$	[100,1.2e+4]	$r(rad/s)$	[-0.2,0.2]	$\theta(rad)$	$[-\pi/2, \pi/2]$
$V_f(m/s)$	[82.590]	$T(lb)$	[1.0e3,1.9e4]	$\delta_e(deg)$	[-25,25]

Assume that the threat center is located in the task space (3300m, 0m, 18000m), In order to meet the traffic reconnaissance conditions, the terminal state expectation value of UAV must satisfy certain conditions. Among them, the UAV reconnaissance flight height not less than 17500m, the yaw angle  $\psi$  is 70 degrees (DEG), the simulation with the initial state and the parameters of terminal state quantity setting as shown in the following table, then the simulation results are as follows:

**Table 2.** UAV initial state and end state

state variable	initial value	Terminal state expectation value	state variable	initial value	Terminal state expectation value
x(m)	0	8000	$\psi$ (deg)	0	70
y(m)	0	8000	V(m/s)	500	500
z(m)	15000	17500	p(deg/s)	0	0
$\phi$ (deg)	0	0	q(deg/s)	0	0
$\theta$ (deg)	0	0	r(deg/s)	0	0

Simulation parameters are selected as follows: sampling time  $\Delta t = 0.01s$ , residual tolerance  $rtol = 1.0e-6$ , optimal iteration times  $iternum = 100$ , stability factor  $\zeta = 100$  ( $\zeta\Delta t = 1$ ), stability factor matrix  $A_s = -\zeta I$ ; In the algorithms of initializing control variable start value  $\mathbf{u}(0)$  and solving control variable  $\mathbf{u}(t)$  by rolling optimization, GMRES iteration times  $k_{max}$  are 2 and 15, respectively; the weighting matrix  $Q$ ,  $S_f$  and  $R$  are chosen as the unit matrix; time domain parameter  $T(t) = T_f(1 - e^{-\alpha t})$ ,  $T_f = 1, \alpha = 0.5$ ,  $N = 10$ ,  $\Delta \tau = T(t) / N$ , penalty factor  $r_i = 0.1(i = 5, 6, 7, 8)$ .



**Figure 3.** Three dimensional flight trajectory of UAV



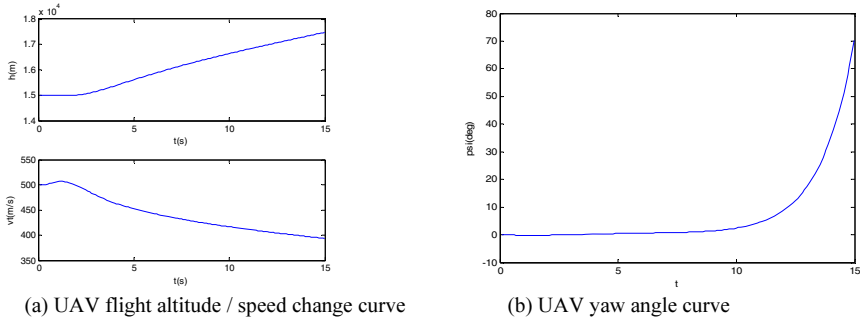


Figure 4. Height and yaw angle of UAV

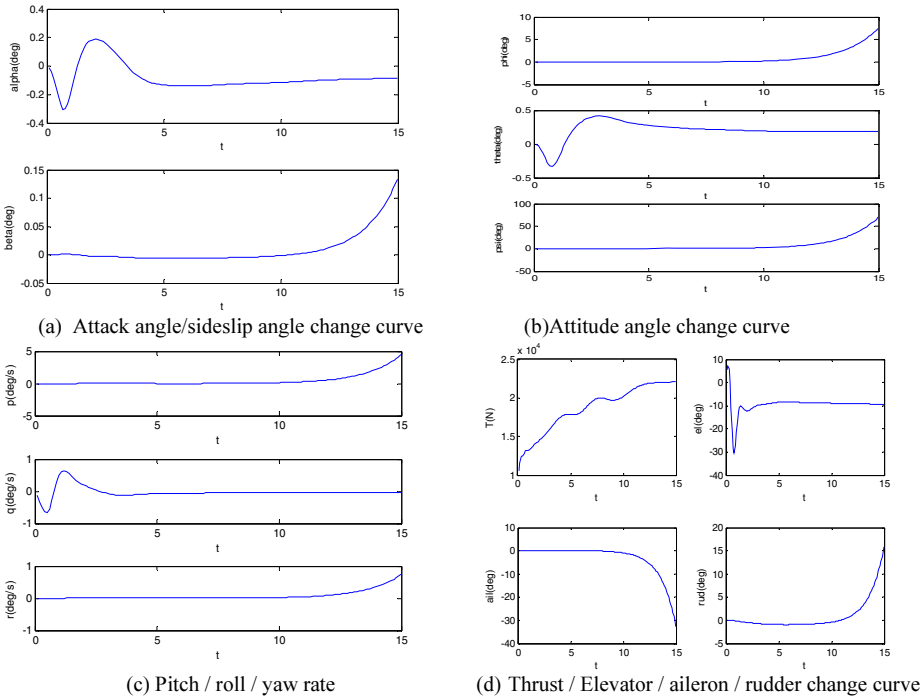


Figure 5. Control and state variables in UAV motion planning

Figure 3 shows the three-dimensional flight trajectory of UAV. It can be seen that the algorithm can effectively plan the trajectory and avoid the obstacles from the starting point to the target point. Figure 4 shows the variation curves of the UAV two terminal constraints, height and heading angle. Obviously, the height and yaw angle gradually converge to the target state of expectation, and the height error is less than 50m, the error of heading angle less than 5 degrees.

Because the control input saturation constraint condition is considered, the control quantity does not exceed the saturation constraint range, which proves the effectiveness of the method. Figure 5 shows the curve of UAV's attitude and control. among them, (a): the angle of attack / sideslip angle change curve; (b): the attitude angle curve; (c): the pitch / roll / yaw rate curve; (d): the control of thrust / Elevator / aileron / rudder change curve.

Table 3 gives the statistical mean of the total planning time and the single step rolling optimization time of the C/FD-GMRES algorithm under the condition of multiple obstacles. It can be seen that the total planning time of the algorithm is about 85 seconds, and the average value of the long time horizon is about 279 milliseconds. It can meet the requirements of UAV real-time trajectory planning .

**Table 3** The total length of planning and the mean value of single step rolling optimization

	The first time	The second time	Third time	Average time
Total time (s)	86.36s	84.01s	85.33s	85.2s
Single step time(ms)	282.48ms	274.02ms	280.79ms	279.10ms

## 5. Conclusion

In this paper, a UAV online trajectory planning algorithm based on C/FD-GMRES method is designed. This method has good real-time performance and is not sensitive to the selection of the initial estimation value. The algorithm can effectively avoid the spatial obstacles, and satisfy the UAV terminal state constraints, such as the height of the UAV, the heading angle, etc, to meet the posture needs of the task implementation. The simulation results show that the proposed algorithm can be used to effectively plan the trajectory of the stable convergence to the stationary state, in the presence of obstacles under the condition that the algorithm can guarantee the UAV effectively avoid obstacles, and rolling optimization time is about 279ms, which can meet the requirement of real-time online planning.

## References

- [1] Goerzen C, Kong, Mettler B. A Survey of Motion Planning Algorithms from the Perspective of Autonomous UAV Guidance. *Journal of Intelligent and Robotic Systems*, 2010, 57(1-4):65-100.
- [2] Guldner J, Utkin V I. Sliding mode control for gradient tracking and robot navigation using artificial potential fields. *IEEE Transactions on Robotics & Automation*, 1995, 11(2):247-254.
- [3] Kavraki L E, Svestka, Petr, Latombe J C, et al. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Transactions on Robotics & Automation*, 1994, 12(4):566-580.
- [4] Karaman S, Frazzoli E. Sampling-based algorithms for optimal motion planning. *International Journal of Robotics Research*, 2011, 30(7):846-894.
- [5] Cheng P, Frazzoli E, Lavalle S. Improving the Performance of Sampling-Based Motion Planning With Symmetry-Based Gap Reduction. *Robotics IEEE Transactions on*, 2008, 24(2):488-494.
- [6] Dong Shiyu, Long Guoqing, Zhu Xiaoping. Research on UAV route planning. *Flight dynamic*, 2004, (3):21-24.
- [7] Wu Shouzhong. *Optimal control theory and Application*. Beijing: China Machine Press, 2008.
- [8] Subchan S, Zbikowski R. *Computational Optimal Control: Tools and Practice*. USA: Wiley, 009.
- [9] Kelley C T. *Iterative methods for linear and nonlinear equations*. Philadelphia: In *Frontiers in applied mathematics*, 1995.

# Time-Varying Formation Control and Collision Avoidance for Unmanned Aerial Vehicles Based on Position Estimation

Yujie Feng<sup>a,1</sup>, Qing Wang<sup>a</sup>, Chaoyang Dong<sup>b</sup>

<sup>a</sup>*School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China*

<sup>b</sup>*School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China*

**Abstract.** Time-varying formation control problems for unmanned aerial vehicle (UAV) systems based on position estimation are investigated. Firstly, a distributed formation controller based on the estimator is proposed using only relative state information of neighbors. The stability can be guaranteed by proper parameters and it can be proven that the control input is able to drive the UAVs to the predefined formation. Secondly, extra item for inner collision avoidance is added to the control strategy, and the stability can be proven by using common Lyapunov approach. Finally, numerical simulations are provided to illustrate the effectiveness of the theoretical results.

**Keywords.** Formation control, cooperative control, unmanned aerial vehicles, position estimation, collision avoidance

## 1. Introduction

Formation control of unmanned aerial vehicle (UAV) systems has drawn great attention in recent years for its broad potential applications in not only civilian but also military fields. Formation is defined as several UAVs' forming a particular configuration, and formation control is an important part of cooperative control for UAVs. In the past few decades, a great many of formation control methods have been proposed, which can be classified as three typical approaches, that is, leader-follower based approach<sup>[1]</sup>, virtual structure based approach<sup>[2]</sup> and behavior based approach<sup>[3]</sup>.

However, as it's pointed in [4], the mentioned methods have their own strengths and weaknesses. With the development of consensus theory, more and more researchers tend to use consensus based approaches to solve formation control problems. Ren<sup>[5]</sup> proposed a consensus protocol to deal with the formation control problems of second-order swarm systems. By applying a variable transformation, the formation problem can be converted into a consensus problem<sup>[6]</sup>. Moreover, the disadvantages of the traditional approaches can be overcome. A formation controller for UAV swarm systems with undirected interaction topologies to achieve time

---

<sup>1</sup> Corresponding Author. [alex.fengyj@gmail.com](mailto:alex.fengyj@gmail.com)

Foundation item: National Natural Science Foundation of China (61374012)

invariant formations in the condition of communication delays was proposed in [7]. Time-varying formation control problem for UAV systems was studied in [8].

It should be pointed out that UAVs must avoid obstacles and remain connected as they move to achieve a desired configuration. [9] proposed a hierarchical control strategy for UAV systems to achieve constant formations while avoiding obstacles. Based on dynamic model, formation control and obstacle avoidance of multi-robot systems was discussed in [10]. Moreover, it has to build a localization platform in order to get the exact location in most of mentioned applications, which is difficult and costly. And because of the transmission fault and the factor of the energy costs, the absolute position is hard to obtain, while the relative position of UAVs can be easily measured by infrared detection and blue-tooth devices.

Motivated by the above observation, this paper investigates the time-varying formation control and collision avoidance for UAVs based on position estimation. On the formation level, a UAV is regarded as a point-mass system, whose dynamics is modeled by a double-integrator system. A position estimation law is proposed, which is applied to the consensus-based formation protocol. And it can be proven that the formation control law based on the estimator can drive the UAVs to the predefined time-varying formation if the estimator converges to stable. And the stability can be guaranteed by properly chosen parameters. Then the collision avoidance law based on the estimator is presented, of which the stability can be proven by a common Lyapunov function.

Compared with the existing works on formation control problems of UAV systems, the new features of this paper are threefold. Firstly, the formation that includes position and velocity, which are specified by time-varying piecewise continuously differentiable vectors. The control law in [11] cannot be applied to the case that velocity vectors are differentiable directly. Secondly, the estimation law is practicable in double-integrator system, while the estimator in [12] is based on single-integrator system. Thirdly, the controller with collision avoidance is robust and distributed and can enable the formation to be realized.

The rest of this paper is organized as follows. In section 2, preliminaries and the problem formulation are provided. In section 3, the main results are presented. In Section 4, simulation examples are presented. Section 5 is the conclusion.

## 2. Preliminaries and Problem Formulation

In this section, several notations in graph theory are introduced, and problem formulation is provided.

### 2.1. Preliminaries

A directed graph  $G = \{V, E, A\}$  consists of the vertex set  $V = \{1, 2, \dots, N\}$ , the directed edges set  $E \subseteq V \times V$ , and the weighted adjacency matrix  $A = [a_{ij}]_{N \times N}$  with nonnegative elements  $a_{ij}$ .  $a_{ij}$  is positive if there is a directed edge between vertex  $i$  and  $j$ ,  $a_{ij} = 0$  otherwise. If  $(i, j)$  is an edge of  $G$ ,  $i$  is defined as the parent vertex and  $j$  is defined as the child vertex. The set of neighbor of  $i \in V$  is defined as  $N_i = \{j \in V : (i, j) \in E\}$ . A directed graph  $G$  is said to have a spanning tree if there exists a vertex, such that it can be connected to all other vertices through paths, and the vertex is called the root of the

spanning tree. The Laplacian matrix of the topology  $G$  is defined as  $L=[L_{ij}]_{N \times N}$ , where  $L_{ii} = \sum_{j \neq i} a_{ij}$  and  $L_{ij} = -a_{ij}$ .

**Lemma 1** [13] Let  $L \in \mathbf{R}^{N \times N}$  be the Laplacian matrix of a directed graph  $G$ ,

- 1) 0 is an eigenvalue of  $L$  with  $1_N$  as the eigenvector.
- 2) 0 is a simple eigenvalue of  $L$  and all the other  $N-1$  eigenvalues have positive real parts if and only if the graph  $G$  has a directed spanning tree.

**Lemma 2** [14] Let  $\text{Re}(\cdot)$  and  $\text{Im}(\cdot)$  represent the real and imaginary parts of a number respectively. Also let

$$\rho_{\pm} = \frac{\gamma\mu - \alpha \pm \sqrt{(\gamma\mu - \alpha)^2 + 4\mu}}{2} \tag{1}$$

$$\gamma > \sqrt{\frac{2}{|\mu| \cos((\pi / 2) - \tan^{-1}(-\text{Re}(\mu) / \text{Im}(\mu)))}} \tag{2}$$

where  $\rho, \mu \in \mathbf{C}$ . If  $\alpha \geq 0$ ,  $\text{Re}(\mu) < 0$ , then  $\text{Re}(\rho_{\pm}) < 0$ .

### 2.2. Problem Formulation

Consider a UAV system with  $n$  UAVs. On the formation control level, a UAV can be regarded as a point-mass system, of which the dynamics can be approximately described by the following double-integrator:

$$\begin{cases} \dot{x}_i = v_i \\ \dot{v}_i = u_i \end{cases} \tag{3}$$

where  $i=1,2,\dots,n$ ,  $x_i \in \mathbf{R}^m$ ,  $v_i \in \mathbf{R}^m$  and  $u_i \in \mathbf{R}^m$  denote the position, velocity and control input of UAV  $i$ , respectively. The interaction topology of the UAV system can be described by a directed graph  $G$ . For  $i, j=\{1,2,\dots,n\}$ , UAV  $i$  can be denoted by node  $i$  in  $G$  and the interaction channel from UAV  $i$  to UAV  $j$  can be denoted by  $(i, j)$ . Let  $x_d = (x_{1d}, \dots, x_{nd})$  and  $v_d = (v_{1d}, \dots, v_{nd})$  be the desire positions and velocities, respectively. In this paper, we give the following assumptions:

- Each possible directed graph has a spanning tree.
- Each UAV  $i$  can't get others' exact position sometimes, but it is able to get the relative position of its neighbors, which can be denoted as  $x_{ij} = x_j - x_i$ ,  $\forall j \in N_i$ , where  $x_{ij}$  is the relative position between  $i$  and  $j$ .
- All of the UAVs can measure their own velocities  $v_i, i=1,2,\dots,n$ , and their own estimated positions denoted as  $\hat{x}_i, i=1,2,\dots,n$ . They can receive the estimated positions and velocities of their neighbors by communication. For all UAVs in the system, we design an estimation law by using  $x_{ij}$  and

$\hat{x}_j (j \in N_i)$  such that,  $\lim_{t \rightarrow \infty} \hat{x}_i(t) = x_i(t) - \tilde{x}_{\infty,c}$ , where  $\hat{x}_{\infty,c} \in \mathbf{R}^m$  is a certain constant vector.

**Lemma 3** [5] For the system (3) with the following consensus algorithm

$$u_i = \dot{v}_d - \alpha(v_i - v_d) - \sum_{j=1}^n a_{ij} [(x_i - x_j) + \gamma(v_i - v_j)] \quad (4)$$

where  $\alpha > 0$ . the system achieves consensus asymptotically as  $t \rightarrow \infty$  if and only if matrix  $\Sigma$  has a simple zero eigenvalue and all the other eigenvalues have negative real parts, where

$$\Sigma = \begin{bmatrix} 0_{n \times n} & \mathbf{I}_n \\ -\mathbf{L} & -\alpha \mathbf{I}_n - \gamma \mathbf{L} \end{bmatrix} \quad (5)$$

Besides,  $x_i(t) \rightarrow x_j(t)$  and  $v_i(t) \rightarrow v_j(t) \rightarrow v_d(t)$  for  $t \rightarrow \infty$ , where  $\mathbf{L}$  is the Laplacian matrix,  $x = [x_1, \dots, x_n]^T$ ,  $v = [v_1, \dots, v_n]^T$ ,  $\mathbf{p}$  is a left eigenvector of  $-\mathbf{L}$  associated with eigenvalue 0 and  $\mathbf{p}^T \mathbf{1} = 1$ .

In order to achieve the formation, a variant of the following distributed controller (4) based on local relative states information of neighbors is proposed, such that the predefined position and velocity can be achieved. To realize collision avoidance among the UAV system, an additional control input item based on the estimator is added to the consensus algorithm.

**Remark 1** The values of  $\hat{x}_i, \hat{x}_j, x_{ji}$  and  $v_i$  are available. The estimated position  $\hat{x}_i$  converges to the value of  $x_i + \tilde{x}_c$  instead of the value of  $x_i$ . And because of the error between  $x_i$  and  $\hat{x}_i$ ,  $x_i$  can't achieve the predefined position  $x_{id}$  exactly. Meanwhile, the velocity  $v_i$  converges to  $v_{id}$  precisely, because it can be measured and transmitted with each other among the UAV system.

### 3. Formation Analysis and Protocol Design

#### 3.1. Position Estimator Design

A distributed position estimation law for UAVs based on model (3) can be designed as:

$$\dot{\hat{x}}_i = v_i + k_0 \sum_{j \in N_i} a_{ij} (\hat{x}_{ji} - x_{ji}) = v_i + k_0 \sum_{j \in N_i} a_{ij} [(\hat{x}_j - \hat{x}_i) - (x_j - x_i)] \quad (6)$$

where  $k_0 > 0$  is a constant,  $\hat{x}_{ji}$  and  $x_{ji}$  are the estimated and real relative position between UAV  $j$  and UAV  $i$ , respectively. The estimator (6) only demands the relative position  $x_{ji}$ , the absolute velocity  $v_i$  and external input  $\hat{x}_j$ , which can be obtained by calculating, sensing and communicating. Besides, the estimation law does not require

the absolute position, which satisfies the mentioned assumption. We define the error of position estimator as  $\tilde{x}_i = x_i - \hat{x}_i$ .

**Lemma 4** [12] For the position estimation law (6), there exists a finite vector  $\tilde{x}_\infty = [\tilde{x}_{\infty,c}, \dots, \tilde{x}_{\infty,c}]^T$ , such that  $\hat{x}$  globally exponentially converges to  $x - \tilde{x}_\infty$  if and only if  $G$  has a spanning tree.

### 3.2. Formation Control Input Based on Position Estimation

Based on the mentioned consensus strategy and the position estimator, a time-varying formation controller for UAVs system is proposed:

$$\begin{aligned}
 u_i = & \dot{v}_{id} + \dot{v}_f - \alpha(v_i - v_{id} - v_f) - \sum_{j=1}^n a_{ij} [(\hat{x}_i - x_{id}) - (\hat{x}_j - x_{jd})] \\
 & - \sum_{j=1}^n a_{ij} \mathcal{N}[(v_i - v_{id}) - (v_j - v_{jd})]
 \end{aligned} \tag{7}$$

where  $v_f(t)$  specifies the nominal formation velocity, and  $\dot{x}_{id} = v_{id}$ ,  $i = 1, \dots, n$ . With controller (7), formation control is achieved if, for all  $i, j = 1, \dots, n$ ,  $x_i - x_{id} \rightarrow x_j - x_{jd}$  and  $v_i - v_{id} \rightarrow v_j - v_{jd} \rightarrow v_f$ , as  $t \rightarrow \infty$ .

**Theorem 1** Let  $L$  and  $p$  be defined in lemma 3. Also let  $\mu_i$  be the eigenvalues of  $-L$ . Based on algorithm (7), the UAVs system can achieve time-varying formation if directed graph  $G$  has a directed spanning tree and

$$\gamma > \max_{\{i | \text{Re}(\mu_i) < 0\}} \sqrt{\frac{2}{|\mu_i| \cos((\pi/2) - \tan^{-1}(-\text{Re}(\mu_i) / \text{Im}(\mu_i)))}} \tag{8}$$

where  $\text{Re}(\cdot)$  and  $\text{Im}(\cdot)$  represent the real and imaginary parts of a number respectively. Therefore, as  $t \rightarrow \infty$ ,  $x_i(t) - x_{id}(t) \rightarrow x_j(t) - x_{jd}(t)$  and  $v_i(t) \rightarrow v_{id}(t) + v_f(t)$ .

**Proof.** Substitute the controller (7) in to the system (3) and let  $\hat{x}_i = x_i - \tilde{x}_i$ ,  $\xi_i = x_i - x_{id}$ , and  $\zeta_i = v_i - v_{id}$ , where  $\dot{x}_{id} = v_{id}$ , the system can be rewritten as

$$\begin{cases} \dot{\xi}_i = \zeta_i \\ \dot{\zeta}_i = \dot{v}_f - \alpha(\zeta_i - v_f) - \sum_{j=1}^n a_{ij} [(\xi_i - \tilde{x}_i) - (\xi_j - \tilde{x}_j) + \gamma(\zeta_i - \zeta_j)] \end{cases} \tag{9}$$

Let  $\xi_i^* = \xi_i - x_f$ , where  $x_f = \int_0^t v_f(\tau) d\tau$ , and  $\zeta_i^* = \zeta_i - v_f$ . From (9) we have

$$\begin{cases} \dot{\zeta}_i^* = \zeta_i^* \\ \dot{\zeta}_i^* = -\alpha \zeta_i^* - \sum_{j=1}^n a_{ij} [(\zeta_i^* - \tilde{x}_i) - (\zeta_j^* - \tilde{x}_j) + \gamma(\zeta_i^* - \zeta_j^*)] \end{cases} \quad (10)$$

Equation (10) can be written in matrix form as

$$\begin{bmatrix} \dot{\zeta}^* \\ \dot{\zeta}^* \end{bmatrix} = \left( \begin{bmatrix} 0_{n \times n} & I_n \\ -L_n & -\alpha I_n - \gamma L_n \end{bmatrix} \otimes I_m \right) \begin{bmatrix} \zeta^* - \tilde{x} \\ \zeta^* \end{bmatrix} \quad (11)$$

where  $\zeta^* = [\zeta_1^{*T}, \zeta_2^{*T}, \dots, \zeta_n^{*T}]^T$ ,  $\zeta^* = [\zeta_1^{*T}, \zeta_2^{*T}, \dots, \zeta_n^{*T}]^T$ , and  $\tilde{x} = [\tilde{x}_1^T, \tilde{x}_2^T, \dots, \tilde{x}_n^T]^T$ . Let

$$\Sigma_1 = \begin{bmatrix} 0_{n \times n} & I_n \\ -L_n & -\alpha I_n - \gamma L_n \end{bmatrix}$$

we have

$$\begin{bmatrix} \dot{\zeta}^* \\ \dot{\zeta}^* \end{bmatrix} = (\Sigma_1 \otimes I_m) \begin{bmatrix} \zeta^* - \tilde{x} \\ \zeta^* \end{bmatrix} = (\Sigma_1 \otimes I_m) \begin{bmatrix} \zeta^* \\ \zeta^* \end{bmatrix} + \begin{bmatrix} 0 \\ (L_n \otimes I_m) \tilde{x} \end{bmatrix} \quad (12)$$

From Lemma 4, we have  $\tilde{x} \rightarrow [\tilde{x}_{\infty,c}, \dots, \tilde{x}_{\infty,c}]^T$ , as  $t \rightarrow \infty$ . Because Laplacian matrix  $L_n$  has the proposition that  $L_n \mathbf{1} = 0$ , we have  $(L_n \otimes I_m) \tilde{x} \rightarrow 0$  as  $t \rightarrow \infty$ .

In order to prove that  $\Sigma$  has a simple zero eigenvalue and all the other eigenvalues have negative real parts, we can solve the equation  $\det(\lambda I_{2n} - \Sigma) = 0$  to find the eigenvalues of  $\Sigma$ .

$$\det(\lambda I_{2n} - \Sigma) = \det(\lambda^2 I_n + \gamma \lambda L_n + \alpha \lambda I_n + L_n) = \det[(\lambda^2 + \alpha \lambda) I_n + (1 + \gamma \lambda) L_n] \quad (13)$$

Also note that

$$\det(\lambda I_n + L_n) = \prod_{i=1}^n (\lambda - \mu_i) \quad (14)$$

where  $\mu_i$  is the  $i$ th eigenvalue of  $-L_n$ . By comparing the above two equations, we see that

$$\det[(\lambda^2 + \alpha \lambda) I_n + (1 + \gamma \lambda) L_n] = \prod_{i=1}^n [\lambda^2 + \alpha \lambda - (1 + \gamma \lambda) \mu_i] \quad (15)$$

which implies that the roots of (13) can be obtained by solving  $\lambda^2 + \alpha \lambda - (1 + \gamma \lambda) \mu_i = 0$ . Therefore, it is straightforward to see that the eigenvalues of  $\Sigma$  are given by



$$\rho_{i\pm} = \frac{\gamma\mu_i - \alpha \pm \sqrt{(\gamma\mu_i - \alpha)^2 + 4\mu_i}}{2} \tag{16}$$

where  $\rho_{i\pm}$  are called eigenvalues of  $\Sigma$  that are associated with  $\mu_i$ .

If the information exchange topology has a directed spanning tree, we know from lemma 1 that  $-L_n$  has a simple zero eigenvalue and all the other eigenvalues have negative real parts. Without loss of generality, let  $\mu_1 = 0$  and  $\text{Re}(\mu_i) < 0, i = 2, \dots, n$ , then we have  $\rho_{1+} = 0$  and  $\rho_{1-} = -\alpha$ . Since that the parameter satisfies (8), combining with the results given by Lemma 2, we know that  $\text{Re}(\rho_{i\pm}) < 0, i = 2, \dots, n$ . Meanwhile,  $\xi_i(t) \rightarrow \xi_j(t)$  and  $\zeta_i(t) \rightarrow \zeta_j(t) \rightarrow v_f(t)$  for  $t \rightarrow \infty$ . As  $\xi_i = x_i - x_{id}$  and  $\zeta_i = v_i - v_{id}$ , we have  $x_i(t) - x_{id}(t) \rightarrow x_j(t) - x_{jd}(t)$  and  $v_i(t) - v_{id}(t) \rightarrow v_j(t) - v_{jd}(t) \rightarrow v_f(t)$ .

In other words, the time-varying formation is achieved.  $\square$

### 3.3. Formation Control Input Based on Position Estimation with Collision Avoidance

In order to avoid collision when UAVs converge to the desired formation, the allowed distance between UAVs should be limited. Collision avoidance constraints can be realized by adding extra control terms to the formation control law. The potential function based on the estimator can be designed as follows:

$$P_{ij}(\hat{x}_i, \hat{x}_j) = \left[ \min \left( 0, \frac{\|\hat{x}_i - \hat{x}_j\|^2 - R^2}{\|\hat{x}_i - \hat{x}_j\|^2 - r^2} \right) \right]^2 \tag{17}$$

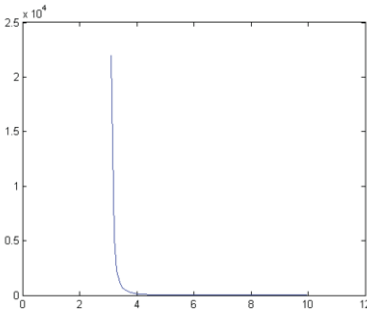


Figure 1. The potential function with  $r = 3$  and  $R = 10$ .

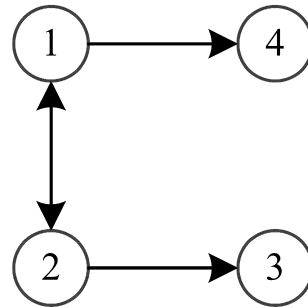


Figure 2. The interaction graph.

where  $R > r$ ,  $R$  and  $r$  denote the radius of the detection region and the safety region, respectively. If the distance between UAV  $i$  and  $j$  is less than  $R$ , which means collisions may occur,  $P_{ij}(\hat{x}_i, \hat{x}_j)$  will be larger than zero and it will be effective in the extra control input. As it depicts in Figure 1,  $P_{ij}(\hat{x}_i, \hat{x}_j)$  tends to  $\infty$  if the distance between two UAVs tends to  $r$ , therefore  $r$  is the inferior limit of the distance between UAVs. The partial derivative of  $P_{ij}(\hat{x}_i, \hat{x}_j)$  with respect to  $\hat{x}_i$  is given by

$$\frac{\partial P_{ij}(\hat{x}_i, \hat{x}_j)}{\partial \hat{x}_i} = \begin{cases} \frac{4(R^2 - r^2)(\|\hat{x}_i - \hat{x}_j\|^2 - R^2)}{(\|\hat{x}_i - \hat{x}_j\|^2 - r^2)^3} (\hat{x}_i - \hat{x}_j)^T, & r \leq \|\hat{x}_i - \hat{x}_j\| \leq R \\ 0, & \|\hat{x}_i - \hat{x}_j\| \geq R \end{cases} \quad (18)$$

The formation control law based on position estimator with collision avoidance is designed as follows:

$$\begin{aligned} u_i^* &= \dot{v}_d + \dot{v}_f - \alpha(v_i - v_d - v_f) - \sum_{j=1}^n a_{ij}[(\hat{x}_i - x_{id}) - (\hat{x}_j - x_{jd})] \\ &\quad - \sum_{j=1}^n a_{ij} \gamma (v_i - v_j) - \sum_{j=1}^n \frac{\partial P_{ij}(\hat{x}_i, \hat{x}_j)}{\partial \hat{x}_i} \end{aligned} \quad (19)$$

where  $v_f(t)$  specifies the nominal formation velocity, and  $\dot{x}_{id} = \dot{x}_{jd} = v_d$ ,  $i = 1, \dots, n$ .

**Theorem 2** For the system (3), based on algorithm (19), the UAVs system can achieve time-varying formation and the collision can be avoided if directed graph  $G$  has a directed spanning tree and the parameter  $\gamma$  is designed satisfying (8).

**Proof.** Let  $\xi_i = \hat{x}_i - x_{id}$ ,  $\zeta_i = v_i - v_{id}$ ,  $i = 1, \dots, n$ . For system (3) with controller (7), a Lyapunov function candidate can be chosen as

$$V = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n [(\xi_i - \xi_j)^T (\xi_i - \xi_j) + (\zeta_i - \zeta_j)^T (\zeta_i - \zeta_j)] \quad (20)$$

Taking derivative of  $V$  along the trajectories of the system, we have

$$\dot{V} = \sum_{i=1}^n \sum_{j=1}^n [(\xi_i - \xi_j)^T (\dot{\xi}_i - \dot{\xi}_j) + (\zeta_i - \zeta_j)^T ((u_i - u_j) - (\dot{v}_{id} - \dot{v}_{jd}))] \quad (21)$$

It have been proven by theorem 2 that  $\xi_i - \xi_j \rightarrow 0$  and  $\zeta_i - \zeta_j \rightarrow 0$  as  $t \rightarrow \infty$ . Therefore, we have  $\dot{V} \leq 0$ , and the equals sign holds only when the formation is achieved.

For system (3) with controller (19), we choose the followed Lyapunov function:

$$V_1 = \frac{1}{4} \sum_{i=1}^n \sum_{j=1}^n [(\xi_i - \xi_j)^T (\xi_i - \xi_j) + (\zeta_i - \zeta_j)^T (\zeta_i - \zeta_j)] + \sum_{i=1}^n \sum_{j=1}^n P_{ij}(\hat{x}_i, \hat{x}_j) \quad (22)$$

Taking derivative of  $V_1$  along the trajectories of the system, we have

$$\begin{aligned}
\dot{V}_1 &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n [(\xi_i - \xi_j)^T (\dot{\xi}_i - \dot{\xi}_j) + (\zeta_i - \zeta_j)^T ((u_i^* - u_j^*) - (\dot{v}_{id} - \dot{v}_{jd}))] \\
&+ \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial P_{ij}}{\partial \hat{x}_i} \dot{\hat{x}}_i + \frac{\partial P_{ij}}{\partial \hat{x}_j} \dot{\hat{x}}_j \right) \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n [(\xi_i - \xi_j)^T (\dot{\xi}_i - \dot{\xi}_j) + (\zeta_i - \zeta_j)^T ((u_i - u_j) - (\dot{v}_{id} - \dot{v}_{jd}))] \\
&- \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left[ (\zeta_i - \zeta_j)^T \left( \sum_{k=1}^n \frac{\partial P_{ik}^T}{\partial \hat{x}_i} - \sum_{l=1}^n \frac{\partial P_{lj}^T}{\partial \hat{x}_j} \right) \right] + \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial P_{ij}}{\partial \hat{x}_i} \dot{\hat{x}}_i + \frac{\partial P_{ij}}{\partial \hat{x}_j} \dot{\hat{x}}_j \right)
\end{aligned}$$

If the estimator converges, we have  $\dot{\hat{x}}_i = v_i$ , and

$$\frac{\partial P_{ij}}{\partial \hat{x}_i} = -\frac{\partial P_{ij}}{\partial \hat{x}_j}, \quad \frac{\partial P_{ij}}{\partial \hat{x}_i} = \frac{\partial P_{ji}}{\partial \hat{x}_i}$$

Therefore, we can obtain

$$\begin{aligned}
&- \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left[ (\zeta_i - \zeta_j)^T \left( \sum_{k=1}^n \frac{\partial P_{ik}^T}{\partial \hat{x}_i} - \sum_{l=1}^n \frac{\partial P_{lj}^T}{\partial \hat{x}_j} \right) \right] + \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial P_{ij}}{\partial \hat{x}_i} \dot{\hat{x}}_i + \frac{\partial P_{ij}}{\partial \hat{x}_j} \dot{\hat{x}}_j \right) \\
&= - \sum_{i=1}^n \sum_{j=1}^n \left[ ((v_i - v_j) - (v_{id} - v_{jd})) \frac{\partial P_{ij}}{\partial \hat{x}_i} \right] + \sum_{i=1}^n \sum_{j=1}^n \left[ (v_i - v_j) \frac{\partial P_{ij}}{\partial \hat{x}_i} \right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \left[ (v_{id} - v_{jd}) \frac{\partial P_{ij}}{\partial \hat{x}_i} \right]
\end{aligned}$$

We assume that  $v_{id} = v_{jd}$ , then

$$\sum_{i=1}^n \sum_{j=1}^n \left[ (v_{id} - v_{jd}) \frac{\partial P_{ij}}{\partial \hat{x}_i} \right] = 0$$

Consequently,  $\dot{V}_1 = \frac{1}{2} \dot{V} \leq 0$

which implies that the system can achieve consensus and the predefined formation  $x_{id}$  can be realized.

On the other hand, for  $i \neq j$ , we have

$$\lim_{\|\hat{x}_i - \hat{x}_j\| \rightarrow r^+} P_{ij}(\hat{x}_i, \hat{x}_j) = \infty, \quad \lim_{\|\hat{x}_i - \hat{x}_j\| \rightarrow r^+} \frac{\partial P_{ij}(\hat{x}_i, \hat{x}_j)}{\partial \hat{x}_i} = \infty$$

We can conclude that the collision between UAVs can be avoided.  $\square$

### 4. Simulation

In this section, simulation examples are provided to illustrate the effectiveness of the above theoretical results. A group of four UAVs is considered. The interaction of communication graph for the four UAVs is shown in Figure 2, we note that the directed topology has a spanning tree.

The time-varying formation trajectories of the UAVs are shown in Figure 3. The formation configuration is a rotating regular square, and we can see that the time-varying formation has been achieved. The trend of the control error  $e_1 = (x_i - x_{id}) - (x_j - x_{jd})$ ,  $e_2 = v_i - v_{id} - v_{jd}$  and estimate error  $\tilde{x}_i = x_i - \hat{x}_i$  are shown in Figure 3, where  $e_1$  and  $e_2$  are the average control error of all UAVs, and  $e_3$  is the estimate error.

In Figure 4 and Figure 5, ‘\*’ stands for the final position and ‘+’ stands for the desired position of each UAV. We note that they are not consistent because of the existence of  $\tilde{x}_\infty$ . Figure 5 depicts that UAVs will collide in their ways to the desired positions. Figure 5 shows that with extra control input, all of the UAVs can make their way round and collision will be avoided. The estimated error and the control errors are also depicted. The convergence rate in Figure 5 is slightly slowed down compared with that in Figure 4.

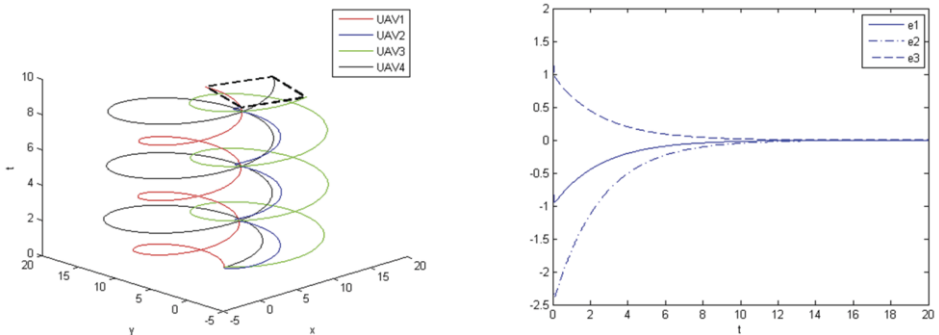


Figure 3. The trajectories of 4-UAVs based on control law, the estimated error and control error

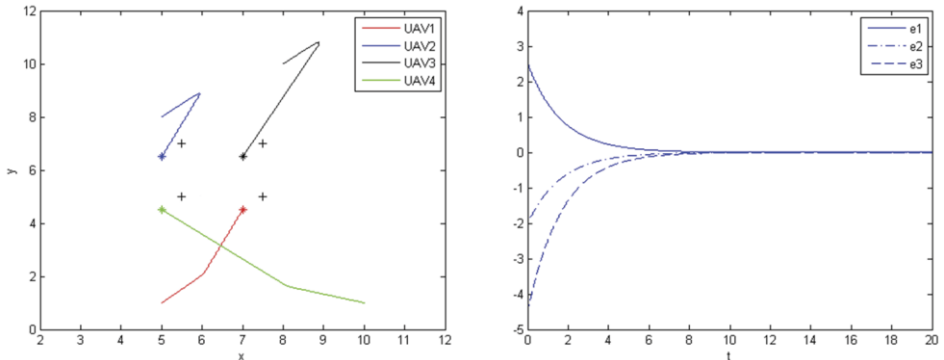


Figure 4. The trajectories of 4-UAVs without avoidance control input, the estimated error and control error

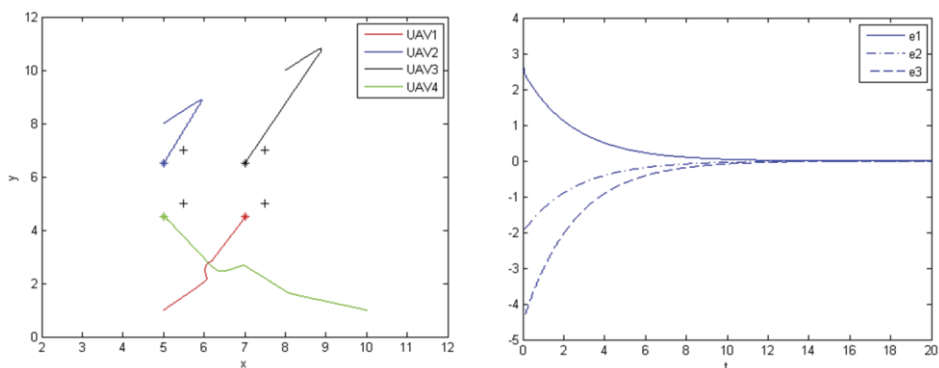


Figure 5. The trajectories of 4-UAVs with avoidance control input, the estimated error and control error

## 5. Conclusion

Time-varying formation control problems for UAV systems based on position estimation are studied. A formation controller based on the estimator is proposed and it can drive the UAVs to the predefined formation. Moreover, extra control input for inner collision avoidance is added to the control law, and the stability can be proven by using Lyapunov function. Finally, simulation examples are presented.

## References

- [1] Z. Y. Lin, W. Ding, G. F. Yan, C. B. Yu, and A. Giua, Leader-follower formation via complex Laplacian, *Automatica*, vol. 49, no. 6, pp. 1900–1906, Jun. 2013.
- [2] M. A. Lewis and K. H. Tan, High precision formation control of mobile robots using virtual structures, *Auton. Robot.*, vol. 4, no. 4, pp. 387–403, Oct. 1997.
- [3] Balch, T., & Arkin, R. C. (1998). Behavior-based formation control for multi robot teams. *IEEE Transactions on Robotics and Automation*, 14(6), 926–939.
- [4] R. W. Beard, J. Lawton, and F. Y. Hadaegh, A coordination architecture for spacecraft formation control, *IEEE Trans. Control Syst. Technol.*, vol. 9, no. 6, pp. 777–790, Nov. 2001.
- [5] Ren, W. (2007). Consensus strategies for cooperative control of vehicle formations. *IET Control Theory & Applications*, 1(2), 505–512.
- [6] Ni, W., & Cheng, D. (2010). Leader-following consensus of multi-agent systems under fixed and switching topologies. *Systems & Control Letters*, 59(3–4), 209–217
- [7] A. Abdessameud and A. Tayebi, Formation control of VTOL unmanned aerial vehicles with communication delays, *Automatica*, vol. 47, no. 11, pp. 2383–2394, Nov. 2011.
- [8] Dong, X., Yu, B., Shi, Z., & Zhong, Y. (2015). Time-varying formation control for unmanned aerial vehicles: Theories and applications. *IEEE Transactions on Control Systems Technology*, 23(1), 340–348.
- [9] Wang, X., Yadav, V., & Balakrishnan, S. N. (2007). Cooperative UAV formation flying with obstacle/collision avoidance. *IEEE Transactions on Control Systems Technology*, 15(4), 672–679.
- [10] Cruz, C. D. L., & Carelli, R. (2008). Dynamic model based formation control and obstacle avoidance of multi-robot systems. *Robotica*, 26(03), 345–356.
- [11] Chen J, Xia Y. Formation control and obstacles avoidance for multi-agent systems based on position estimation. In: 2014 33rd Chinese control conference (CCC), IEEE, Nanjing, China; 2014. p. 1150–55.
- [12] Oh K-K. Formation control of mobile agents based on distributed position estimation. *Autom Control* 2013; 58(3):737–42.
- [13] W. Ren and R. W. Beard, Consensus seeking in multi-agent systems under dynamically changing interaction topologies, *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655–661, May 2005.
- [14] Ren, W., & Atkins, E.M.: Distributed multi-vehicle coordinated control via local information exchange, *Int. J. Robust Nonlinear Control*, 2006, in press

# An Improved Map-Matching Method Based on Hidden Markov Model

Yang Linjian<sup>a,b</sup>, Zhao Xiangmo<sup>a</sup>, Zhang Wei<sup>c</sup>, Meng Fanlin<sup>a</sup>,  
Cheng Xiaodong<sup>d</sup>, An Yisheng<sup>a</sup>

<sup>a</sup> School of Information Engineering, Chang'an University, Xi'an Shaanxi 710064

<sup>b</sup> Bureau of Yunnan Highway Transport Administration, Kunming Yunnan 650031

<sup>c</sup> Xi'an Communication Information Co Ltd, Xi'an Shaanxi 710065

<sup>d</sup> Jilin Provincial Transport scientific Research Institute, Changchun Jilin 130012

**Abstract.** Map-matching is the process to match a sequence of real world coordinates into a digital map, so as to identify the correct segment on which a vehicle is traveling and to determine the vehicle location on the segment. Map matching is one of the key components to model and analyze floating car data, and provide ITS services such as traffic condition analysis and navigation. Complex environment, inadequate attribute information, low sampling frequency, and location deviation exert great influence on the matching performance. This paper presents an improved map-matching algorithm based on Hidden-Markov model. A distance based weighted-average method is applied to improve the quality of the instantaneous GPS data, and a preprocessing and caching method for the shortest paths is used to accelerate the calculation of state transition probability. Comparative analyses show that more than 90% of positions are matched, and computation time is significantly improved.

**Keywords.** Urban traffic; Map matching; Hidden-Markov Model; Floating car

## 1. Introduction

It is commonly realized that Float Car Data (FCD) consists of positional information of moving vehicles and has created exciting opportunities for tracking floating vehicles in large road networks. The traces of vehicles can be mined to derive knowledge about a range of subjects for monitoring traffic conditions, such as obtaining interest point, updating real-time information, traffic condition prediction, and vehicle scheduling and routing [1-3]. However, due to the lack of precision from GPS signal, the error of the electronic map data and organization, network topology characteristics, and interference in the process of signal propagation, there are always certain deviations between the detected and real floating vehicle location [4]. In addition, owing to the sampling interval, a floating vehicle may pass a road segment without being detected.

Map matching is the process of the establishment of the mapping relationship between vehicle GPS position and GIS system based on geographic information, network topology structure, and relevance of time and space position. Map matching is of great importance in precisely analyzing floating vehicle data, so as to obtain real-time, accurate, and reliable traffic information to support ITS services, such as route guidance, fleet management, and accident and emergency responses [5-7].

In literature, map-matching algorithms can be categorized into two groups: geometric analysis and probabilistic algorithms. Geometric analyses are mainly based on the topology of the road network. The connectivity and contiguity of the links are used to amend missing values caused by sampling intervals [8-10]. However, these algorithms are mainly based on the local information and often fall into a local optimum. Probabilistic algorithms consider the position change of a floating vehicle as a sequence, and model the position change of vehicles utilizing appropriate probabilistic model [11-12]. Hidden-Markov model (HMM) is a typical probability model for map-matching algorithms [13]. However, for a large road network, the repeating calculation of the shortest paths makes HMM less efficient and not suitable to be applied to the real-time traffic monitoring. This paper presents an improved Hidden-Markov map-matching (IHMM) algorithm, and accelerates the shortest paths calculation through a preprocessing and caching method, and improves the matching precision through an outlier recognizing algorithm. Numerical results show that IHMM improves the overall efficiency of the algorithm, and reaches a matching accuracy of more than 90%.

## 2. The model

### 2.1. Map matching algorithm based on Hidden-Markov model

#### 2.1.1. Observation probability and state transition probability

A hidden Markov model is defined as a discrete-time stochastic process with a set of states,  $R = r_1, r_2, \dots, r_N$  and where  $Q = q_1, q_2, \dots, q_T$  is a state sequence for the time  $t = 1, 2, \dots, T$ . The initial state distribution is denoted  $\pi = \pi_1, \pi_2, \dots, \pi_N$ . The state of the process cannot be directly observed, instead, some sequences of observation symbols,  $O = o_1, o_2, \dots, o_T$  are measured. According to Ref. [14], GPS data has shown a clearer pattern of a Gaussian distribution, therefore, the observation probability matrix B can be determined as,

$$b_j(k) = P(o_t = o_k | q_t = r_j) = e^{-(b^2/4a-c)/2\sigma^2} \left[ \Phi\left(\frac{2a+b}{2\sigma\sqrt{a}}\right) - \Phi\left(\frac{b}{2\sigma\sqrt{a}}\right) \right] \quad (1)$$

Where  $\sigma$  is the assumed standard deviation of the GPS measurements.  $\Phi$  is the standard cumulative distribution function for the Gaussian distribution. Suppose  $(x_a, y_a)$  and  $(x_b, y_b)$  are the coordinates of nodes defining the straight road segment S, and  $p_1, (x_1, y_1)$  is the GPS position to be matched, then constants a, b, c are:

$$a = (x_b - x_a)^2 + (y_b - y_a)^2 \quad (2)$$

$$b = 2[(x_a - x_1)(x_b - x_a) + (y_a - y_1)(y_b - y_a)] \quad (3)$$

$$c = (x_a - x_1)^2 + (y_a - y_1)^2 \quad (4)$$

On the other hand, if there were a large number of links, it would take plenty of time to calculate observation probability. Therefore, a threshold is used to control the number of links, while the number of links is above the threshold, observation probability is obtained by the following equation [15].

$$b_j(k) = P(o_t = o_k | q_t = r_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{dist_{o_k \rightarrow r_j}^2}{2\sigma^2}} \tag{5}$$

where  $dist_{o_k \rightarrow r_j}^2$  is the distance of the kth observation point  $o_k$  and the  $i$ th link.

Elements in the state transition matrix A can be obtained by the following equation, where  $\beta$  is a constant.

$$a_{ij} = P(q_{t+1} = r_j | q_t = r_i) = \frac{1}{\beta} e^{-\frac{|EuDist_{i \rightarrow j} - SP_{i \rightarrow j}|}{\beta}} \tag{6}$$

2.1.2. Filtering of candidate link set

Generally, the candidate link set is determined by a circle around the GPS position to be matched. The circle should have a big enough radius so as not to miss the real link on which the vehicle is driving. However, if the circle is too big, there will be extra computing resources consumption; consequently, the algorithm efficiency will be reduced. This paper determines the radius according to  $\sigma_0$ , the standard deviation of GPS measuring errors, which can be obtained from statistic data. If the GPS measuring error follows a bivariate irrelevant Gaussian distribution, and the radius  $R$  of the circle is  $5\sigma_0$ , the probability to miss the real link will be less than or equal to 0.0001% [16]. Considering the effect of the deviation data, this paper uses a doubled radius  $R = 10\sigma_0$ .

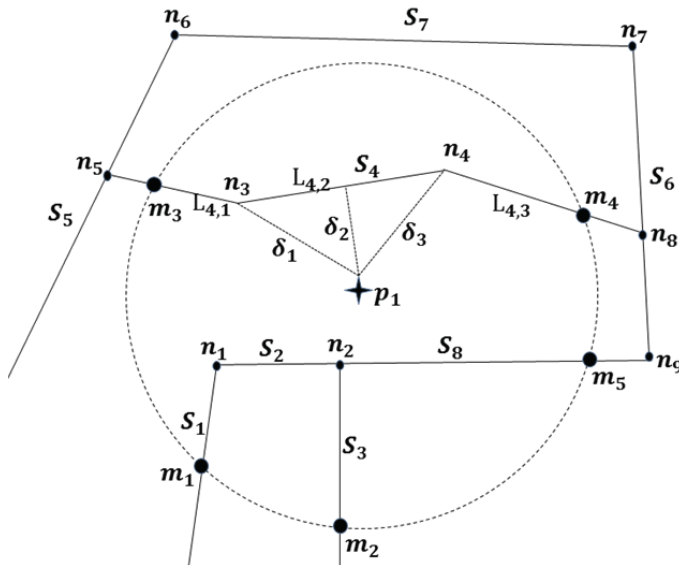


Figure 1. Circular error region defined around GPS point.



Since the requirement for analyzing the links of a segment is also needed in a shortest-distance-based weight formulation, the distance from all the links of that segment to the point to be matched would need to be found. As is shown in Figure 1,  $\delta_1$  and  $\delta_2$  need to be found to calculate the shortest distance from  $S_4$  to  $p_1$ . And coordinates of  $n_1$  and  $n_2$  will be used to find weights of  $S_2$ , and coordinates of  $m_3$  and  $n_3$  will be used for  $L_{4,1}$ . So, simple algebraic line-circle intersection formulas are used to find the intersection points of road links with error regions to get a new weight.

### 2.1.3. Online Viterbi optimization

Viterbi algorithm is a dynamic programming algorithm for solving HMM. For a hidden Markov model  $\lambda = (a_{ij}, b_j, \pi)$ . The most likely state sequence to produce  $O$  can be determined, using the classical Viterbi optimization by defining,

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, \dots, q_{t-1}, q_t = r_i, o_1, \dots, o_t | \lambda) \quad (7)$$

In which,  $N$  and  $T$  are the sizes of the candidate set and the observation set, respectively. For  $t = 1$ ,  $\delta_0(i)$  is  $P(q_1 = r_i, o_1)$ , which can be calculated as  $\delta_0(i) = \pi_i b_{i0}$ , and for  $t > 0$  it follows

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, \dots, q_{t-1}, q_t = r_i, o_1, \dots, o_t | \lambda) = \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}] b_i(o_t) \quad (8)$$

Then, the maximum probabilistic path for candidate state  $i$  at time  $t$  can be determined by (9).

$$\psi_t(i) = \arg \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}], \quad i = 1, 2, \dots, N \quad (9)$$

## 2.2. Improved algorithm

This paper improves matching precision by introducing the elimination of the speed outlier to increase the data quality, and increases execution efficiency by modifying the traditional process of the shortest paths calculation, which is the key process of solving Hidden-Markov model.

### 2.2.1. Recognizing and processing of speed outlier

Speed outlier is a number of unreasonably sharp speed changes, such as peak top, deep valley, and "0" speed with the vehicle position change at the same time. Because of the braking and resistance factors, vehicle acceleration has an upper bound, which can be used to identify speed outlier, as is shown in Figure 2, where  $v_k$  is the  $k$ th speed data, and  $\Delta_v$  is the instant speed change.

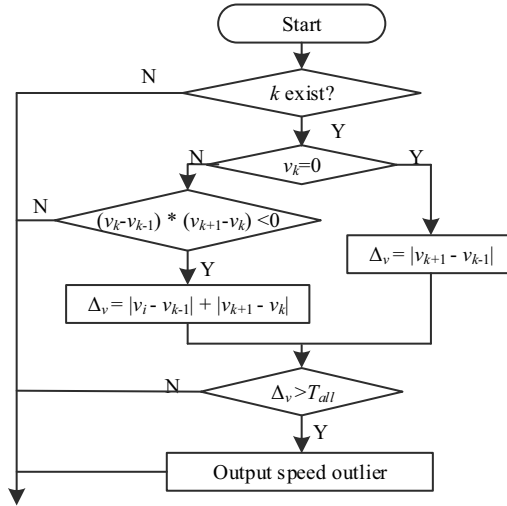


Figure 2. Recognizing speed outliers.

By observing the floating vehicle data, it is known that the ratio of speed outlier is far less than 10%, which means we can correct the speed outlier by using the neighboring speed data. These neighboring speed data that we choose to make a modification to a speed outlier should be close enough to the outlier so that these data can precisely show the speed variation and estimate the normal speed value. Therefore, a distance-based weighted average method is used to process the speed outlier, in which the normal speed data is closer to the outlier in distance which has a higher weight. Figure 3 shows the process in details.

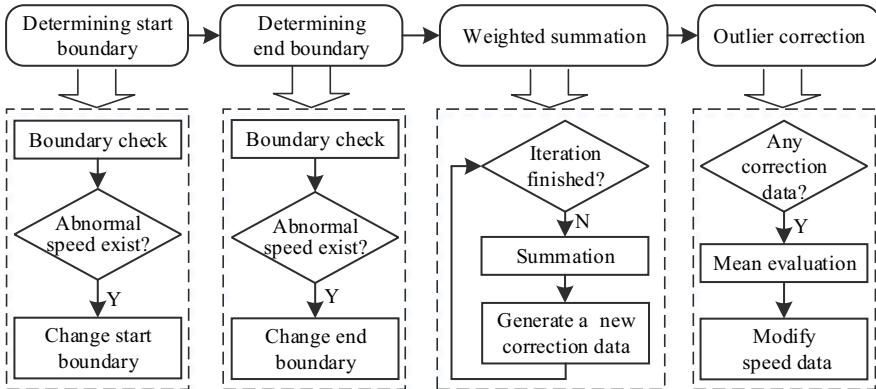


Figure 3. Processing of speed outliers

2.2.2. Preprocessing of the shortest path

When determining the transfer probabilities between links, it is necessary to calculate a large number of the shortest paths, which brings about huge computational resources consumption. Since the time complexity of map-matching algorithm itself is in a log-

linear form, the calculation of the shortest paths becomes the bottleneck of the map-matching algorithm when the map is complicated.

In this paper, the shortest paths between each pair of nodes are pretreated offline in advance, and stored in the cache table with indexing. Therefore, when calculating the transfer probability, the shortest path can be obtained directly by querying the catch table. As common sense, it is almost impossible for a vehicle to pass four or more intersections within one sample period (about 10s-45s). Therefore, to store the shortest paths from an original node to all the destinations within an area covering three intersections in each direction from the origin is enough. Experiments have shown that for a big city, memory consumed in storing the shortest paths is at the level of Gigabyte (GB). For the case of Xi'an, there are totally 20350 nodes, and the memory used to store all the shortest paths is about 2GB-4GB, which is acceptable memory consumption in practice for a modern computer. At the same time, the computing time is at a level of only several milliseconds by directly querying the catch table. Therefore, this method of pretreating the shortest paths to reduce the computing time is able to improve the response speed significantly.

This paper introduces a path length and section quantity constraints based on single origin diffusion algorithm for calculating the shortest paths. Figure 4 shows the processing details of the diffusion from a single origin S. Data structure of each node consists of "len", which is the shortest path length, and "num", which is the number of visited links, and the origin S.  $E_{OQ}$  is the link between node O and Q, including the length of the link.

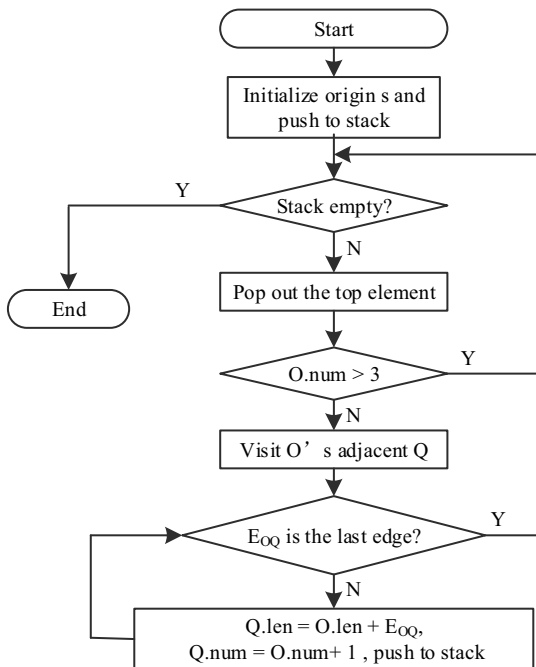


Figure 4. Diffusion from a single origin S

### 3. Experimental analysis

#### 3.1. A numeral example

This paper uses the real floating car data from Xi'an, China, where about 11000 taxis serve as floating cars. In order to show the map-matching result and compare with different algorithms, this paper also implements the simple geometric algorithm (SGA), and the traditional HMM map matching algorithm, along with our IHMM map matching algorithm. We record the total amount of data and computation time of each algorithm. Figure 5 is the 24-hour matching result of IHMM for float car whose license number is "At0078". Figure 6 is the comparison of different algorithms.

Figure 6 shows that by considering the topology structure of electronic map and probabilities of state transition, matching methods based on Hidden-Markov model can effectively reduce the case of mismatches near the intersections compared with the simple geometric algorithm. The results of HMM and IHMM are similar in general. Nonetheless, after manually testing the match result for each point, it is shown that IHMM is more precise than HMM. Nearly 80% of the mismatches from HMM algorithm is revised. This is mainly because IHMM modifies the speed outliers in the GPS data.

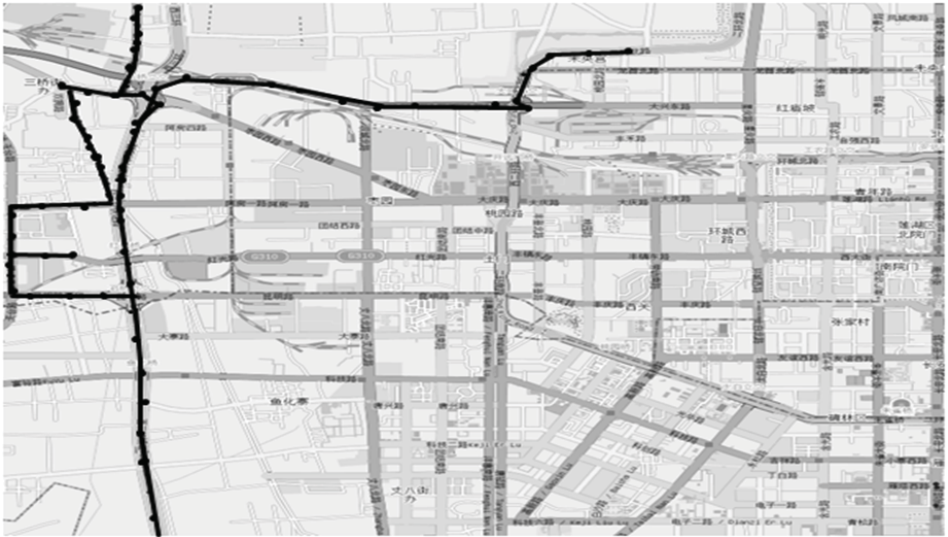


Figure 5. Map matching result of Taxi "AT0078"

#### 3.2. Time complexity and computation time

The time complexity of IHMM mainly lies in the generation of candidate link set and on-line Viterbi algorithm. By preprocessing the shortest paths and storing them in the cache, the time complexity of the generation of candidate link set and calculation of the transition probability are mainly determined by the number of nodes to be matched, which is  $O(n)$ .

The complexity of the Viterbi algorithm is mainly determined by two processes, which are the calculation of forward probability and the process of backward backtrack. The latter is directly based on the result of the former, which means that if the forward probability is known, the time complexity of backtrack process would be  $O(1)$ . While the complexity of the calculation of forward probability mainly depends on the number of GPS positions to be matched and the complexities of the probability calculations, with the overall time complexity to be  $O(n)$ .



Figure 6. Matching results of different algorithm

Therefore, by introducing the cache mechanism to preprocess the shortest paths, this paper improves the map-matching time complexity and effectively reduces the computation time. To draw a final conclusion, we also compare the SG, HMM and IHMM algorithm, as is depicted in Figure 7, where the x-axis is the data size and the y-axis is the computation time. We tested those three algorithms with different data set and different network structures, and calculated the average value. Due to the variance of complexity and topological structure in a road network, Figure 7 shows fluctuant curves. It is shown that SG algorithm has the best computation efficiency, but this is largely associated with the poor matching precision. For IHMM and HMM, IHMM is superior to HMM in computation efficiency as well as matching precision.

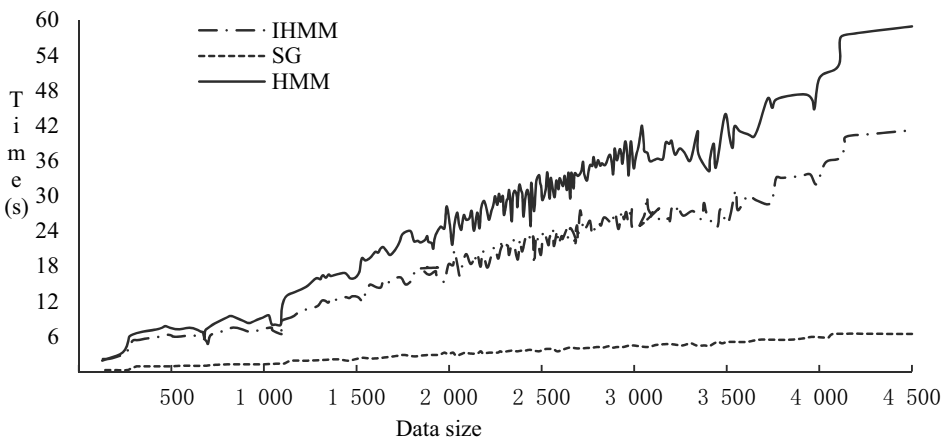


Figure 7. Variation curve of execution time via data volume

#### 4. Conclusion

This paper proposes an improved map-matching algorithm based on Hidden-Markov model. The matching precision is prompted by introducing a distance-based weighted-average method to promote the quality of the instantaneous GPS data. The computational resources consumption is reduced remarkably by implementing a preprocessing and caching method for the shortest paths which are applied to calculate the transition probability. Experiments have shown that this algorithm has a higher computational efficiency as well as a relatively better matching precision. A further step is being made into the fusion model of the fuzzy set correlated to the vehicle direction and angle, so as to obtain a better matching precision.

#### References

- [1] D. S. Guo, X. Zhu, H. Jin, et al. Discovering Spatial Patterns in Origin - Destination Mobility Data[J]. *Transactions in GIS*. 2012, 16(3):411-429.
- [2] J Li, Q. M. Qin C. Xie, et al. Integrated use of spatial and semantic relationships for extracting road networks from floating car data [J]. *International Journal of Applied Earth Observations and Geoinformation*.2012, 19(10):238-247.
- [3] L.L.Lu, Y. Tian. Weak GPS signal acquisition under multiplicative noise[J]. *Computer Applications and Software*, 2013, 30(5): 260-263.
- [4] Z.Y.Zhu, M.Cui, L.Liu. Research on information collection method of real-time road condition based on GPS terminal[J]. *Computer Engineering*, 2013, 39(7): 21-25.
- [5] G.Y.Jiang, A.D. Chang, Q.Li, et al. Estimation models for average speed of traffic flow based on GPS data of taxi [J]. *Journal Of Southwest JiaoTong University*, 2011, 46(4):638-644.
- [6] G.X.Guan, Y.B.Liu, D.H.Qiu. The vehicle tracking algorithm based on GPS/map-matching [J]. *Computer Applications and Software*, 2008, 25(3):219-220+277.
- [7] W.P.Jiang, C.Y.Xia, Z.Li, et al. Analysis of environmental loading effects on regional GPS coordinate time series[J]. *Acta Geodaetica et Cartographica Sinica*, 2014, 43(12):1217-1223.
- [8] M. A. Quddus, W. Y. Ochieng, R.B. Noland. Current map-matching algorithms for transport applications: State-of-the art and future research directions[J]. *Transportation Research Part C*, 2007, 15:312-328.
- [9] N.R. Velaga, M.A. Quddus, A.L. Bristow. Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems[J]. *Transportation Research Part C*,2009, 17:672-683.
- [10] K.Li, Y.Yang, X.S.Qiu. An improved map matching algorithm based on D-S evidence theory in city vehicle navigation [J]. *Acta Geodaetica et Cartographica Sinica*, 2014, 43(2): 208-213, 220.
- [11] M.Wang, H.H.Wei, Y.L.Bao. Map-matching algorithm in GPS navigation system[J]. *Computer Engineering*, 2012, 38(14):259-261.
- [12] P.Newson, J.Krumm J. Hidden markov map matching through noise and sparseness [C]. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York: ACM Press, 2009.
- [13] Y.Lou, C.Zhang, Y.Zheng, et al. Map-matching for low-sampling-rate GPS trajectories[C]//*Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2009: 352-361.
- [14] A. Oran, P.Jaillet. An HMM-based map matching method with cumulative proximity-weight formulation [C]. *ICCV: International Conference on Connected Vehicles and Expo*, New York: IEEE, 2013.
- [15] R.Raymond, T.Morimura, T.Osogami, et al. Map matching with Hidden Markov Model on sampled road network [C]. *ICPR: 21st International Conference on Pattern Recognition*, New York: IEEE, 2012.
- [16] J.J.Tang, F.Liu. A driver route prediction based map-matching algorithm integrating uncertain reasoning [J]. *Acta Geodaetica et Cartographica Sinica*, 2010, 39(5):546-550.

# Study on Application of Wavelet Transform Reconstruction Acceleration in Vehicle Vibration Displacement Measurement

Xiao-feng Feng <sup>a,1</sup>, Bin Fang <sup>b</sup> and Jiang-hong Liu <sup>c</sup>  
<sup>a,b,c</sup> *Hunan Police Academy, Changsha, China;410138*

**Abstract.** In order to solve the problem of noise error in the process of vehicle vibration displacement, a new method of multi-scale wavelet transform reconstruction acceleration is proposed. this method makes full use of characteristics of wavelet transform in multi-scales. According to frequency range of low-frequency noise to determine the number of decomposition, while to removal high-frequency noise using the correlation between wavelet scales, reconstruct acceleration which has removed the signal noise, then integrate to get the vibration displacement. The simulation analysis results show that the method can be used to measure the vehicle vibration displacement accurately.

**Keywords.** multi-scale wavelet transform; acceleration; vibration displacement

## 1. Introduction

It is necessary to calculate the vibration velocity and displacement of the vehicle in order to get the excitation characteristics of the vehicle directly from the acceleration response of the vehicle<sup>[1]</sup>. As the influence of vehicle movement and limitation of installation position and environment, it is unable to establish a unified and unchanged measuring basis, so is difficult to measure directly by displacement sensor, sometimes even impossible. With the development of sensor technology, a new type of acceleration sensor with high precision, small volume and light weight has appeared, the velocity and displacement can be obtained by integration of the acquired data directly, so it is more widely used to obtain the velocity and displacement by the method of acceleration integral<sup>[2-3]</sup>.

Acceleration integral can be realized by hardware integration circuit. In practical application, the integral precision is low as the limit of hardware device and circuit itself, in addition it has a higher signal frequency requirements, cannot be used to measure those signals contains more frequency components such as vibration signal. Another method is using software integration to achieve<sup>[4-5]</sup>. This method include time domain integral method, frequency domain integral method, time frequency mixed integral method, exact information reconstruction integration method and so on. most of these methods are with a purpose of removing trend. in addition to trend, the signal still contains middle and high frequency noise, so there will bring great deviation to the

---

1. Corresponding author

measurement of velocity and displacement. The method can separate signal and noise completely in case that the signal spectrum and the noise spectrum are not overlapped, But in the actual signal spectrum and noise spectrum are often overlapping, the noise is almost distributed in the whole frequency domain, if there is a high demand for the noise smoothing effect, the signal will be blurred and the high frequency of the signal will be lost<sup>[6-8]</sup>.

As a multi-scale analysis tool, wavelet transform can decompose the different frequency components of signal to the non overlapping frequency bands, and the low-frequency trend can be completely separated from signal<sup>[9]</sup>. At the same time, according to the different propagation characteristics of random noise and signal, through the correlation algorithm between the scales, we can eliminate the random noise, and reconstruct the signal after obtaining the accurate information to solve disturbance of the trend term and noise in the integration process.

## 2. Acceleration integral error analysis

The error of accelerometer mainly includes two parts: deterministic error and random drift error. The deterministic error includes the scale factor error, installation error and the fixed part of drift error, this part can be effectively compensated by experimental calibration; The other part includes random constant error, correlation error and white noise component, the error model of the accelerometer is assumed to be:

$$a_c(t) = a(t) + a_z(t) + \varepsilon \tag{1}$$

Where,  $a_c(t)$  、  $a(t)$  are measured value and true value of accelerometer separately at t moment,  $\varepsilon$  is a random constant error of accelerometer,  $\dot{\varepsilon}=0$  . as the relative error component of the accelerometer is relatively small, so it is ignored.  $a_z(t)$  is white noise,  $E[a_z(t)a_z(\tau)] = q\delta(t - \tau)$  .  $q$  is the variance of the white noise of accelerometer.

Integrate acceleration signal and we can obtain velocity and displacement as follows:

$$v(t) = \int_0^t a_c(t)dt = \int_0^t a(t)dt + \int_0^t a_z(t)dt + \varepsilon t + \eta \tag{2}$$

$$s(t) = \int_0^t v(t)dt = \int_0^t a(t)dt + \int_0^t a_z(t)dt + \varepsilon t^2 / 2 + \eta t + \sigma \tag{3}$$

The velocity signal after integration includes the linear trend term and the constant term. The linear trend term and the constant term error can be removed effectively by the least square method and polynomial fitting method, but as the noise will make the integral results cumulative and generates a larger trend (not simple two times trend), we need to know the type of trend, and it is difficult to achieve the desired effect in the absence of prior knowledge. In view of this, some scholars use the method of frequency domain integration to research. Transform the acceleration by FFT, we can get:

$$A(n) = \sum_{n=0}^{N-1} a_n e^{-j2\pi kn/N} \tag{4}$$



$$V(n) = \sum_{n=0}^{N-1} \frac{1}{j2k\pi\Delta f} H(k)A(k)e^{j2\pi kn/N} \tag{5}$$

$$S(n) = \sum_{n=0}^{N-1} \frac{-1}{(2k\pi\Delta f)^2} H(k)A(k)e^{j2\pi kn/N} \tag{6}$$

Where, 
$$H(k) = \begin{cases} 1, & f_d \leq k\Delta f \leq f_u \\ 0, & \text{other} \end{cases} \tag{7}$$

$f_d$  and  $f_u$  are the lower and upper cutoff frequencies, respectively,  $A(k)$  is the Fourier transform of acceleration  $a(n)$ ,  $\Delta f$  is frequency resolution,  $j$  is the imaginary unit.

By formula (5)、(6), the frequency domain integral directly using the integral frequency sine and cosine of the exchange relationship, avoid the cumulative amplification of small errors in the integration process of the time-domain signal, so the result is more accurate. However, from the formula (6), as the Fourier transform coefficient is inversely proportional to  $(2k\pi\Delta f)^2$ , it is very sensitive to low frequency signal, so the low-frequency trend is an important error source in the frequency domain. In addition, the high pass filter does not deal with the middle and high frequency noise components is also one of the important causes of error.

### 3. Principle of wavelet transform signal reconstruction

#### 3.1. Remove trend term by wavelet transformation

As a multi-scale analysis tool, wavelet transform can decompose the different frequency components in the signal into non-overlapping frequency band, and can separate low-frequency trend from signal completely. At the same time, according to the propagation characteristics of the random noise and the signal at different scales, the random noise can be eliminated by the correlation algorithm between the scales. reconstruct the signal after obtaining the accurate information, and the disturbance of the trend term and the noise is removed completely.

Set  $\{V_{J-1}\}_{J \in \mathbb{Z}}$  is an orthogonal multi-resolution analysis of scaling function  $\varphi(x)$ , the following two scaling relations are established:

$$\varphi(x) = \sum_{k \in \mathbb{Z}} h_k \varphi(2x - k) \tag{8}$$

Where,  $h_k$  is scale filter coefficient.

The wavelet function  $\psi(x)$  can be generated by the above two scale relation of the scale function, and the two scale relation of wavelet function is:

$$\psi(x) = \sum_{k \in \mathbb{Z}} g_k \varphi(2x - k) \tag{9}$$

$\psi(x)$  can generate linear space sequence  $\{W_{J-1}\}_{J \in \mathbb{Z}}$ , for square integrable space  $L^2(R)$ , Its subspace  $V_J$  satisfies:

$$V_J = V_{J-1} \oplus W_{J-1}, \text{ for given signal } f(t) \in V_J, \text{ there are}$$

$$f(t) = f_{J-1}(t) + g_{J-1}(t) \tag{10}$$

As  $V_J = V_{J-1} \oplus W_{J-1} = V_j \oplus W_j \oplus W_{j+1} \dots \oplus W_{J-1}$ , so, the signal is represented as a sum of several frequency bands:

$$f(t) = \sum_{k \in Z} c_{j,k} \varphi_{j,k}(t) + \sum_{j \leq j' < J} \sum_{k \in Z} d_{j',k} \psi_{j',k}(t) \tag{11}$$

Where,  $c_{j,k} = \langle f, \varphi_{j,k} \rangle$ ,  $d_{j',k} = \langle f, \psi_{j',k} \rangle$ .

the former part in formula (11) represents the low frequency part after wavelet decomposition, and the latter represents the high frequency part of the wavelet decomposition.  $c_{j,k}$  and  $d_{j,k}$  are scale coefficients and wavelet coefficients respectively.

The more layers of decomposition, the more narrow the frequency band of the signal is, the easier it is to separate the low-frequency trend items that do not overlap with the frequency bands. It can be seen that if the decomposition layer is selected reasonably, the low-frequency trend signal can be removed.

Frequency space of signal  $f(t)$  is  $V_j = [0, f_s]$ . due to frequency space characteristic of dyadic wavelet transform, the frequency space after the first level wavelet decomposition is  $[0, f_s / 2]$ , and after the  $j$  level wavelet decomposition is  $[0, f_s / 2^j]$ . so,

$$f \leq \frac{f_s}{2^j} \tag{12}$$

Where,  $f_s$  is the sampling period.

For vehicles, the main vibration frequency is 0~2 Hz, the frequency is in the low frequency range. Frequency below 0.5 Hz is often produced by the test instrument itself or environmental interference. Therefore, the 0 ~ 0.3 Hz parts is considered as the frequency band corresponding to the trend item and should be eliminated.

### 3.2 Theory of wavelet transform to remove noise

Due to medium and high frequency noise in the signal, the accumulation of noise in the integration process will seriously affect the accuracy of the integration, so it should be filtered. For the medium and high frequency noise in the signal, its propagation characteristics are different with signals at different levels, it is uniform and dense on the finer scales and has poor correlation. Therefore, after wavelet multi-scale decomposition, multiply wavelet transform coefficients on two adjacent scale layers, calculate the correlation of different scale signals, according to the principle that the noise has no correlation in each scale, eliminate the noise correlation. The correlation coefficient at  $k$  point on the scale  $J$  is defined as follows:

$$C_{W_{j_1,k}} = W_{j_1,k} \bullet W_{j_2,k} \tag{13}$$

Where,  $W_{j,k}$  is the wavelet coefficient of  $j$  scale after signal decomposition. In order to compare the correlation coefficients and wavelet coefficients of each scale, normalize the energy of  $C_{W_{j,k}}$ , and the normalized correlation coefficients is:

$$w_{j,k} = C_{W_{j,k}} \bullet [P_{W_j} / P_{C_{W_j}}]^{1/2} \tag{14}$$

Where,  $P_{W_j} = \sum_k W_{j,k}^2$ ,  $P_{CW_j} = \sum_k C_{W_{j,k}}^2$  .

When  $|w_{j,k}| > s|W_{j,k}|$ , it is considered that the wavelet coefficients belong to the wavelet coefficients of the useful signal decomposition and preserve it. repeat the process until the energy of non-extracted point in  $W_{j,k}$  is less than a preset energy threshold. after edge extracting in each decomposition scale layer, the filtered signal can be got by reconstruct the data.

3.3 Process of signal's wavelet decomposition and reconstruction

After removing the trend term and removing the medium and high frequency noise, reconstruct the signal and restore the signal which can reflect the vibration of the vehicle body, improve the precision of the acceleration signal integration. process of signal's wavelet decomposition and reconstruction is as follows:

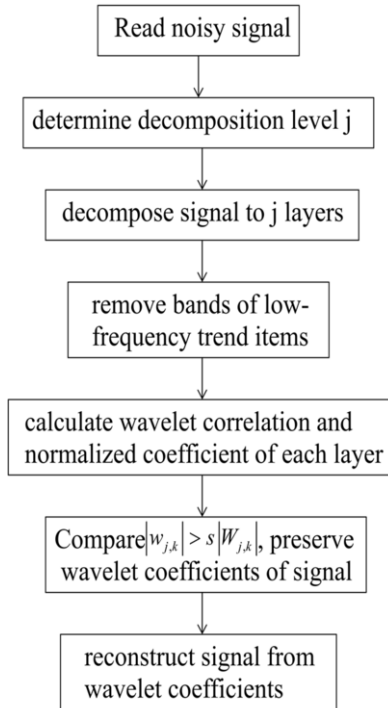


Fig. 1 decomposition and reconstruction process of signal with wavelet

4. Simulation analysis

The sampling frequency of the simulation signal is  $f_s = 50H_z$ , the amplitude of vibration acceleration is  $1\text{ mm/s}^2$ , frequency is superposition of  $2H_z$  and  $3H_z$ , Adding random noise to the signal, and adding DC trend term with a amplitude of  $0.08\text{ mm/s}^2$ , simulate in MATLAB. Figure 2 is the original acceleration signal and acceleration signal with random noise and trend term; Figure 3 and Figure 4 show the results of theoretical signal integration and error signal integration. As can be seen from the

figures, due to the effect of noise and trend term, in one second, the velocity deviates from the theoretical value 1 mm, and the displacement deviates from the theoretical value of 0.4 mm .

After signal denoising by wavelet and trend removing, the reconstructed acceleration signal is shown in Figure 5. It can be seen that the reconstructed acceleration is almost coincident with the original acceleration signal. the velocity signal and displacement signal obtained after acceleration integration are almost coincident with the theoretical value in Figure 6 and Figure 7. It can be seen from the results that the reconstructed signal based on wavelet multi-scale wavelet transform has high accuracy and reliability in the integral of the vibration displacement.

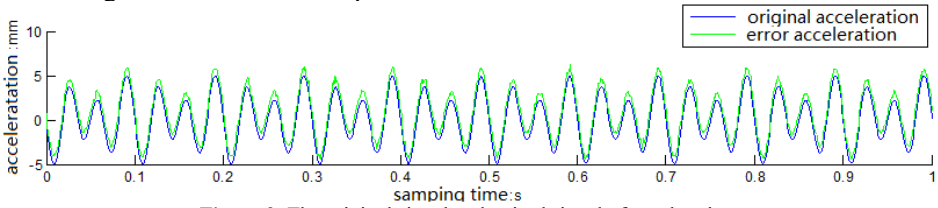


Figure 2. The original signal and noised signal of acceleration

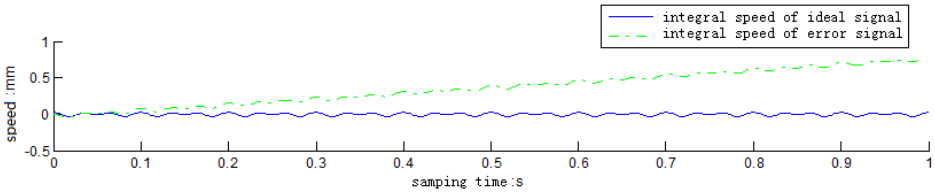


Figure 3. Integral speed of the ideal signal and error signal

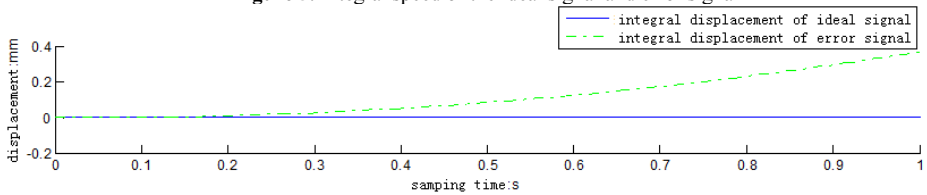


Figure 4. Integral displacement of the ideal signal and error signal

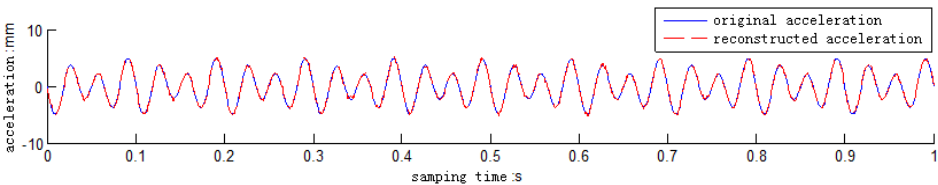


Figure 5. Original acceleration and reconstructed acceleration

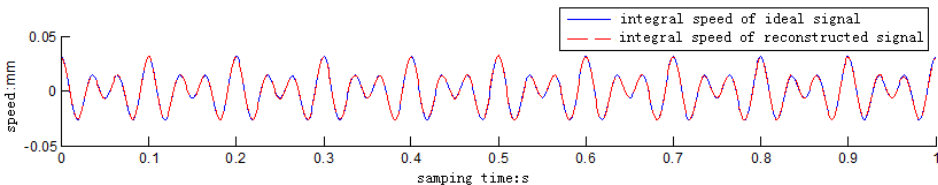


Figure 6. Integral speed of ideal signal and reconstructed signal

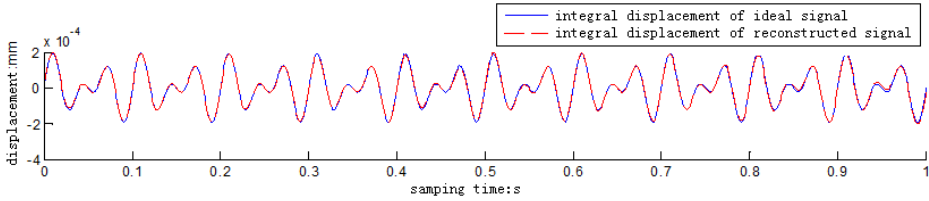


Figure 7. Integral displacement of ideal signal and reconstructed signal

### 5. Integral error evaluation index

In order to evaluate the difference between the displacement curve obtained by acceleration integral and the actual displacement curve, the concept of peak error, difference error and absolute error is introduced.

The peak error is average difference of integral displacement time history  $S'(t)$ 's peak relative to measured displacement time history  $S(t)$ , it is defined as follows:

$$Errp = \frac{1}{2} \times \left\{ \left| \frac{\max[S'(t)] - \max[S(t)]}{\max[S(t)]} \right| + \left| \frac{\min[S'(t)] - \min[S(t)]}{\min[S(t)]} \right| \right\} \tag{15}$$

The difference error is the mean value of the peak value of the time history  $S'(t) - S(t)$  with respect to the peak value of  $S(t)$ , it is defined as follows:

$$Err = \frac{1}{2} \times \left\{ \left| \frac{\max[S'(t) - S(t)]}{\max[S(t)]} \right| + \left| \frac{\min[S'(t) - S(t)]}{\min[S(t)]} \right| \right\} \tag{16}$$

The absolute error is defined as follows:

$$Ers = \frac{\sum_{i=1}^N |S'(t) - S(t)|}{\sum_{i=1}^N |S(t)|} \tag{17}$$

Tab.1 Error of integral speed and integral displacement

Name of error type	Average peak error	Difference value error	Absolute error
Integral velocity error	0.0124	0.03123	0.02135
Integral displacement error	0.2326	0.04602	0.02682

It can be seen from the results that the method proposed in this paper has a high accuracy in the calculation of vibration displacement, and fully meets the needs of the actual displacement measurement.

### 6. Conclusions

Based on the study on the traditional integration method obtaining the vibration displacement, a new method to reconstruct the acceleration using multi-scale wavelet transform is proposed. this method makes full use of characteristics of wavelet transform in multi scales. reconstruct the acceleration signal which is removed noise, then solve vibration displacement by integration, it can effectively distinguish the noise and trend from signal. The results of simulation and experiment show that the method can be used to measure vibration displacement accurately.

## Acknowledgements

All members of the subject actively participated in the research, and we hereby thank them all.

## Funding

The work was supported by a program of Hunan Department of Education outstanding youth project(17B086), and a program of Hunan Police Academy Doctoral project(2016ZX02) .

## References

- [1] Kang Xiong, Zeng Yuqing. Speed-Frequency Domain Method for the Response Analysis of Vehicle Vibration Acceleration [J]. *China Railway Science*. 2012, 33(1):60-70.
- [2] Chen Weizhen, Wang Bingwen, Hu Xiaoya. Acceleration signal processing by numerical integration [J]. *J.Huazhong Univ. of Sci. &Tech.*, 2010, 38 (1) : 1-4. *Sci.&Tech.*, 2010, 38 (1) : 1-4.
- [3] Wang Jianfeng, Ma Jian, Ma Ronggui, Song Hongxun. Study on calculation of dynamic displacement from time-frequency integration of acceleration [J]. *Computer Science*. 2010, 37 (12) : 201-202.
- [4] Wen Guangrui, Li Yang, Liao Yuhe, He Qing. Faulty rotor system vibration acceleration signal integration method based on precise information reconstruction [J]. *Journal of Mechanical Engineering*. 2013, 9 (8) :1-9.
- [5] Discussion of vibration velocity and displacement signal identification method based on the vibration acceleration measurement [J]. *Mechanical Science and Technology*. 2011, 30(4) : 522-526.
- [6] Zhou Yingjie. A study on integral algorithm for acceleration test to get displacement and application [M]. Chong Qing University, 2013, 19-30.
- [7] Ma Yue, Li Song, Li Ying, Weng Yinkan. Modeling and simulation of acceleration signal processing [J]. *Computer Simulation*. 2012, 29 (3) : 351-354.
- [8] Dong Li, Liao Mingfu, Yang Kunji. Filtering modification of vibration signal integration algorithm [J]. *Machinery Design & Manufacture*, 2010(1) : 46-48.
- [9] Wu Zhicheng, Wang Chongyang, Ren Aijun. Optimal selection of wavelet base functions for eliminating signal trend based on wavelet analysis [J]. *Transactions of Beijing Institute of Technology*. 2013,33 (8) :811-814.

# dSPACE Based HIL Simulation Platform for Missile Control System

Yun CHEN <sup>a,1</sup>, Qinghua MA <sup>b</sup>, Gen WANG <sup>b</sup> and Haochun MIAO <sup>b</sup>

<sup>a</sup> College of Astronautics, Northwestern Polytechnical University, Xi'an, 710072 China

<sup>b</sup> No.203 Research Institute of China Ordnance Industries, Xi'an, 710065 China

**Abstract.** dSPACE based hardware-in-the-loop(HIL) missile control system simulation platform with the key benefits of powerful operational capability and convenient hardware interface is provided. The system structure design and development steps were analyzed. In the platform, the compiled simulation model, established in the MATLAB/Simulink, runs on the real-time hardware platform, and the actuator as the real physical part. This platform realized the HIL design and simulation, and the simulation process is similar to the real environment. Experimental results showed that the dSPACE based real-time simulation system is an excellent platform for missile control system design which enhanced the efficiency and reliability.

**Keywords.** Missile control system, hardware-in-the-loop simulation, dSPACE

## 1. Introduction

Missile control system is one of the key subsystems of the missile, with the characteristics of multi-function, complexity construction, and high reliability. With the rapid development of computer hardware and software, the technique of missile control system design and simulation has been improved to a new stage. Comparing to the mathematical simulation, hardware-in-the-loop simulation is a more reliable method of missile control system design, because the real physical parts are in the simulation model.

dSPACE based hardware-in-the-loop(HIL) missile control system simulation platform is provided in this paper. The hardware platform comprises components including a host computer and the dSPACE hardware. MATLAB/Simulink has the powerful modeling and operational capability; on the other hand, dSPACE provides abundant hardware support and efficient code compilation. Missile control system can be designed efficiently on the platform with high quality.

---

<sup>1</sup> Corresponding Author, College of Astronautics, Northwestern Polytechnical University, No.127 Youyi West Road, Xi'an, 710072, China, E-mail: 7601753@qq.com.

## 2. dSPACE Based HIL Simulation Platform

### 2.1. System Construction

dSPACE is a single-source solution for HIL simulation with a comprehensive hardware and software tool chain. The processing unit is the core hardware component of a typical dSPACE testing system, for computing large and complex simulation models and for connecting the real-time model with the relevant I/O. dSPACE's software consisted of implementation software and experiment software. A wide range of I/O interface can be configured and easily integrated into the function model by means of a flexible graphical I/O block library, the Real-Time Interface (RTI) software. For evaluation and calibration, parameters of the control function can be changed and signals can be captured on-the-fly by using ControlDesk, the experiment software of dSPACE. dSPACE based HIL simulation platform for missile control system is shown in Figure 1.

The MATLAB/Simulink software (including Real-Time Workshop, RTW) is installed in the host computer, fulfilling the task of offline model design and simulation. With dSPACE's implementation software RTI, models designed with MATLAB/Simulink can be implemented on the dSPACE hardware automatically. Synchronously, the experiment software ControlDesk is run in the host computer. The hardware platform of dSPACE is connected with the host computer by the means of local network. Actuator, the real physical part of the missile is led into the system by connecting with the dSPACE A/D, D/A interfaces, realizing the HIL simulation.

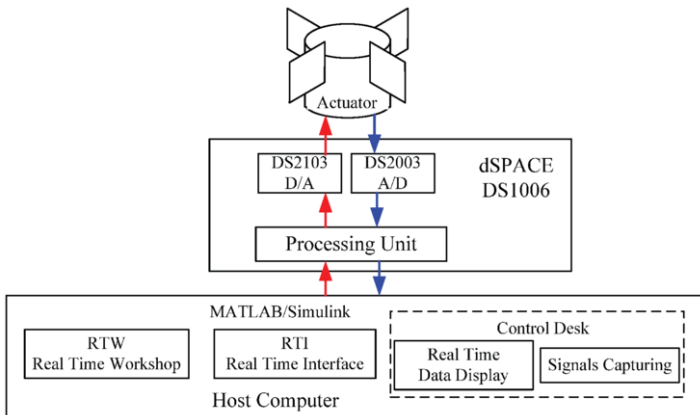


Figure 1. dSPACE based HIL simulation platform

### 2.2. HIL Missile Control System Simulation Development Steps

HIL missile control system simulation development steps are shown in Figure 2.

- Missile control system modeling.
- Control design and Simulink simulation.
- Algorithm and interface definition by using Real-Time-Workshop (RTW) of MATLAB and Real-Time-Interface (RTI) of dSPACE.



- Real-time code can be generated and downloaded to the dSPACE system, when the RTI model of missile control system is completed;
- Realizing the HIL simulation. ControlDesk, dSPACE’s well-established experiment software, provides all the function of experiment manager, real-time monitor, on-line parameter editor, data capture, data display and real-time communication with the upper monitor.

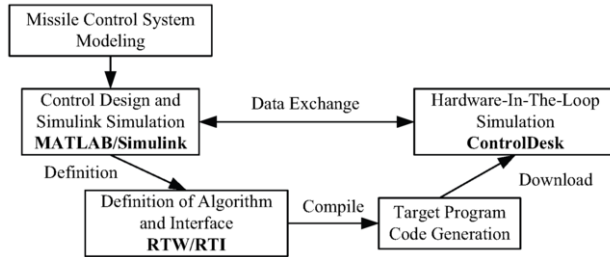


Figure 2. HIL Missile Control System Development Steps

### 3. HIL Missile Control System Design and Simulation

#### 3.1. Missile Control System Modeling

Take some missile pitch channel control system design as an example, illustrating the platform of HIL missile control system design and simulation. The structure block of the pitch channel acceleration control system is shown in Figure 3, it is a classical two-loop acceleration autopilot, realized depending on the measurements of rate gyro and inertial navigator.

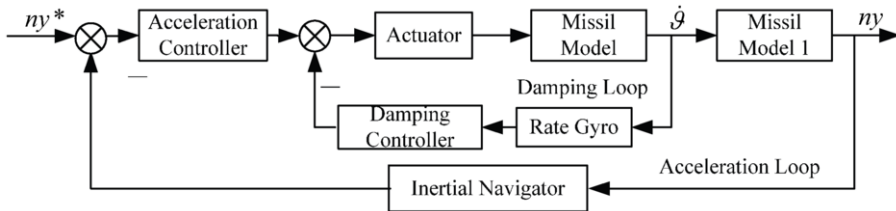


Figure 3. The Structure Block of The Pitch Channel Acceleration Control System

The control system is composed by the damping loop and the acceleration loop. The measurement of the rate gyro is the angle rate of the missile, which is the input of damping controller, and the feedback signal of the damping loop (the inner loop) is the output of the damping controller. The feedback signal of the acceleration loop is the measurement of the inertial navigator. The acceleration controller is the one we designed in the paper.

The transfer functions of the missile are shown in Eqs. (1):

$$G_d(s) = \frac{K_d(T_{1d}s + 1)}{(T_d^2s^2 + 2\xi_d T_d s + 1)} \quad G_{a1}(s) = \frac{1}{T_{1d}s + 1} \cdot \frac{V_d}{57.3g} \quad (1)$$

### 3.2. Control Design and Simulink Simulation

At this stage of development the offline simulation is performed in Simulink. In a general way, the controllers of different loops are designed dividedly, and firstly the inner loop, then the outside loop. Since the space of the whole page is limited, this paper focuses on the acceleration controller (Figure 3) design and simulation.

The frequency domain indexes of the missile control system are: the magnitude-frequency margin is greater than 6dB, and the phase angle margin is greater than 45°.

Acceleration control system Simulink model is shown in Figure 4, ignoring the delay and the measurement errors of the rate gyro and inertial navigator. Mathematical model of the Actuator is 'Gdlt' in the figure. The transfer function of  $Gdlt(s)$  is shown in Eq. (2), including a treble mathematical model and the delay of  $T_\delta$  :

$$Gdlt(s) = \frac{K_\delta (s + \omega_n) e^{-T_\delta s}}{(s + \omega_{n2}) (s^2 + 2\xi\omega_n s + \omega_n^2)} \tag{2}$$

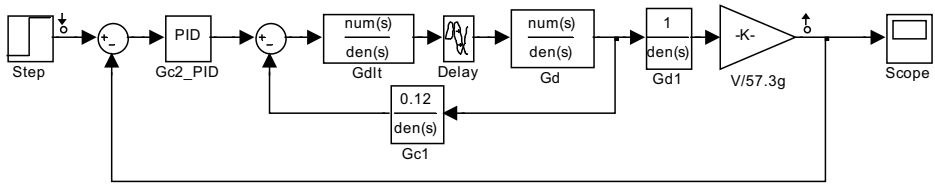


Figure 4. Acceleration Control System Simulink Model

By using MATLAB Control System Toolbox, the acceleration controller is designed. Step response curve and open-loop bode plots is shown in Figure 5 and Figure 6. According to the results shown in the figures, the magnitude-frequency margin is 9.14dB, and the phase angle margin is 54.4°, meeting the demands of the system performance indexes.

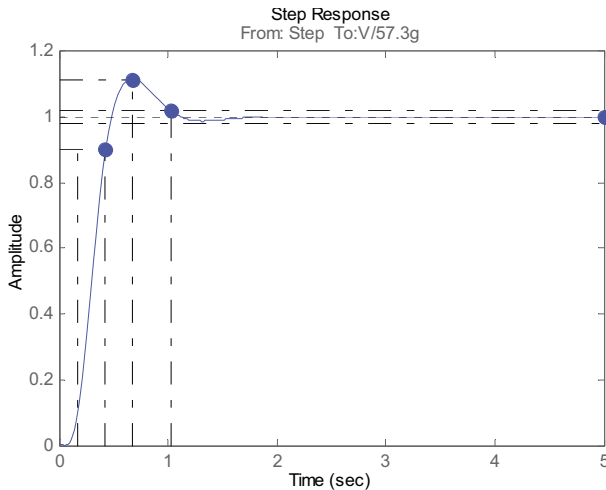


Figure 5. Step Response Curve

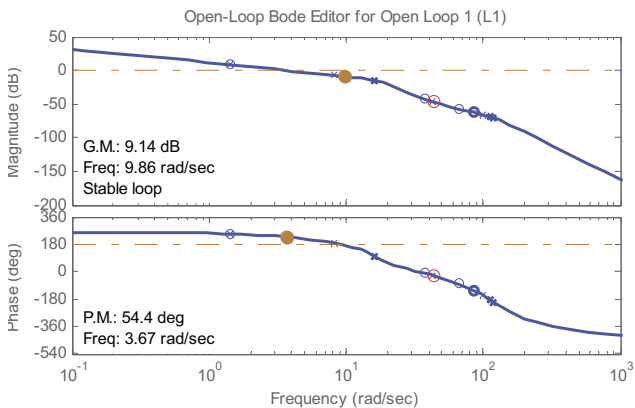


Figure 6. Open-Loop Bode Plots

### 3.3. RTI Model Build and Download

The input and output signal of the actuator are analog. The RTI model of actuator is shown in Figure 7. dSPACE simulation platform receives the digital signal from the controller, and converts to the analog one, transmitting to the actuator, at the same time, receives the output of the actuator potentiometers, and converts to the digital one, transmitting to the missile model. By connecting the actuator RTI model with the relevant I/O, the real physical part instead of the mathematical one is led into the system. Acceleration control system RTI model is shown in Figure 8.

When the RTI model is complete, real-time code can be generated and downloaded to the dSPACE system.

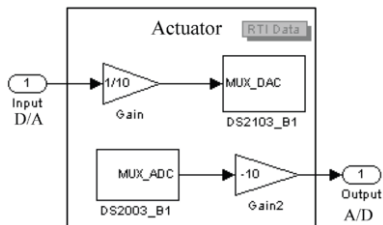


Figure 7. Actuator RTI Model

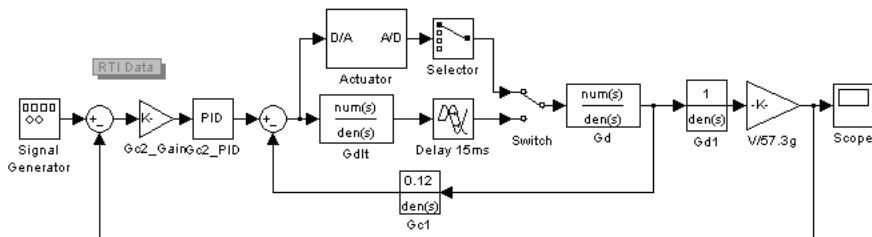


Figure 8. Acceleration Control System RTI Model

### 3.4. dSPACE Based HIL Simulation

ControlDesk is the dSPACE experiment software for seamless HIL development. It performs all the necessary tasks and gives you a single working environment, from the start of experimentation right to the end.

ControlDesk interface of dSPACE based HIL simulation is shown in Figure 9. The major functions achieved in the platform including: experiment manager, signal capturing, variables tuning on-the-fly, real-time response curves displaying.

Controller 1 is the controller given in 3.2, designed based on the actuator mathematical model. Since the actuator RTI model is the real physical part, the step response is different from the response of the mathematical model, shown in Figure 5. The parameters of the controller are tuned during run time.

Analyzing the real-time response curves in Figure 9, the response of Controller 1 and Controller 2 is fast and the overshoot is high, and the response of Controller 4 is slow. Controller 3 is the right one for the control system, satisfying the requirements of rapidity and stability.

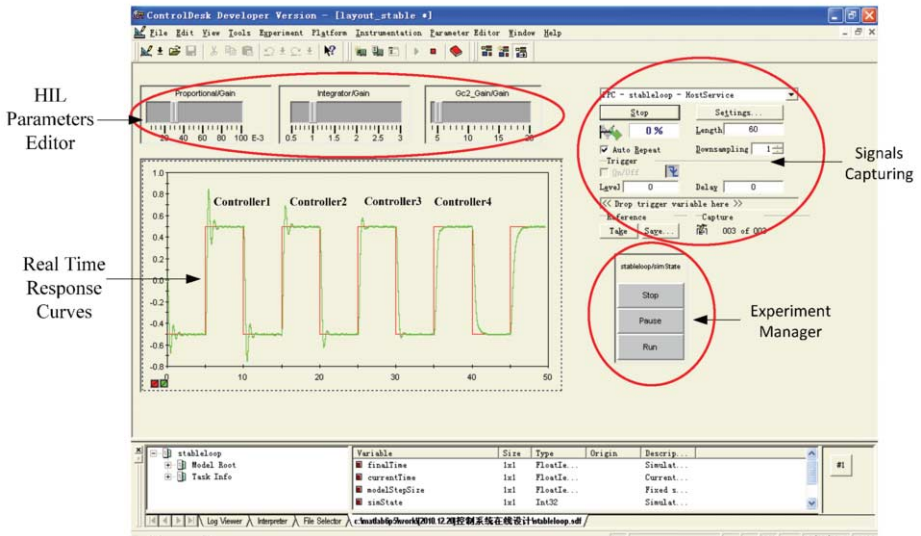


Figure 9. ControlDesk Interface of dSPACE Based HIL Simulation

## 4. Conclusions

dSPACE based HIL simulation platform for missile control system is presented in this paper. The development steps of the platform including missile control system modeling, control design and Simulink simulation, RTI model build and download, and finally the HIL controller parameters tuning and simulation. The real physical part actuator is led into the system with the relevant I/O, avoiding the inauthenticity of the mathematical model, enhancing quality and efficiency of the missile control system design.

## References

- [1] *RTI and RTI-MP Implementation Guide*. dSPACE GmbH, Germany, 2003.
- [2] *ControlDesk Experiment Guide*. dSPACE GmbH, Germany, 2003.
- [3] Lang Pengfei. *Research of dSPACE-Based HIL Simulation System*, South China University of Technology, 2013.
- [4] Qian Xingfang, Lin Ruixiong, Zhao Yanan, *Missile Flight Dynamics*, Beijing Institute of Technology Press, Beijing, 2000.
- [5] Zhu Hanqing, Gu Liangxian, Parameter Optimization Design of Missile Control System Based on Simulink, *Aerospace Control*, (2008), 79-82.
- [6] Liu Xiaoming, Zhou Hao, Chen Wanchun, Simulink/RT-LAB based HIL missile guidance system simulation platform research, *Infrared and Laser Engineering*, (2007), 366-369.
- [7] Song Ke, Liu Weiguo, Luo Guangzhao, dSPACE Implementation of Hardware in Loop Real-Time Simulation Laboratory Plants for Motion Control, *Drive and Control*, (2008): 28-31.

# Toward Intelligent Traffic Light Control with Quality-of-Service Provisioning

Lei MIAO<sup>a,1</sup> and Lijian XU<sup>b</sup>

<sup>a</sup>*Mechatronics Engineering, Middle Tennessee State University*

<sup>b</sup>*Dept. of Electrical and Computer Engineering Technology,  
Farmingdale State College*

**Abstract.** Today's fixed-cycle traffic signaling is highly suboptimal and aggregates traffic congestion and waste of energy in urban areas. In addition, it offers no quality-of-service guarantee and makes travel time prediction extremely hard. While existing traffic light control research primarily focuses on improving the average wait time of cars, we study in this paper how traffic light scheduling affects the worst-case wait time. In particular, we derive the time a car spends at an intersection in the best-case and the worst-case, respectively. Using the theoretical results, we propose a simple but effective controller and run simulation to verify its performance. The result shows that it works much better than fixed-cycle controllers in both light and heavy traffic scenarios.

**Keywords.** Intelligent transportation systems, traffic light control, real-time systems

## 1. Introduction

Traffic congestion is a serious problem in many urban areas all over the world and has become a severe challenge to the sustainability of economic growth and urbanization. A report from the Texas Transportation Institute says that in 2011, traffic congestion costs drivers \$121 billion in the U.S. Americans also spent 5.5 billion additional hours sitting in traffic in the same year. That comes down to \$818 per commuter in lost time and wasted fuel. Traffic congestion, to a large extent, is caused by the inefficiency of today's traffic signal systems, which primarily use fixed-cycle scheduling algorithms. Fixed-cycle approaches are derived off-line based on historical traffic data. The major drawback of fixed-cycle approaches is that they cannot adapt to traffic fluctuation in real-time and may aggravate the congestion when the traffic pattern changes. Therefore, fixed-cycle traffic signaling is highly suboptimal and needs to be replaced with a better approach.

Another problem of today's traffic control system is that it does not provide any quality-of-service (QoS) guarantee. For example, it cannot ensure that each car only spends up to certain amount of time at an intersection. As a result, it is hard to estimate how long it would take a car to travel from point A to point B; this is especially true in case of traffic jam.

Transportation systems are analogous to communication networks in many aspects: cars travel from one place to another whereas packets are sent from one device to another; roads are like communication links where packets are transmitted in a first-come-

---

<sup>1</sup> Corresponding author. Box 19, Middle Tennessee State University, 1301 E Main St, Murfreesboro, TN 37132, USA; E-mail: lei.miao@mtsu.edu.

first-served fashion; and intersections are very much similar to routers and switches. A question then naturally arises: *why is QoS readily available in communication networks but very lacking in transportation systems?* The answers to it are multi-fold, but the major reason is that historically, transportation systems simply do not have the necessary information of the cars, such as where and how fast they are and where they are headed.

Fortunately, things have changed dramatically in the past ten years or so: most drivers now have smart phones which are capable of reporting the GPS locations and velocities of the vehicles; inexpensive and embedded electronics with similar functionalities are also being built into automobiles; and cars are becoming autonomous and have started to have wireless communication capabilities. All these evidences indicate that transportation systems are ready to be smarter and provide QoS provisioning to travelers. Yet, another complication is that traffic systems contain both time-driven and event-driven dynamics: cars continuously accelerate and decelerate and have to obey Newton's law; and traffic lights switch between green and red at discrete time instances. The hybrid nature makes it hard to model and control in applications related to intelligent transportation systems.

In this paper, we present some preliminary work regarding how a traffic intersection can provide QoS to the cars using it. The contributions of the paper include: (i) Different from existing work in the literature, we incorporate car dynamics into our analysis and derive the best-case and worst-case deadline information for a simple two-way intersection; (ii) We propose a simple but effective open-loop controller that utilizes the vehicle information to perform control; and (iii) We run simulation to verify the effectiveness and improvement of our approach over the existing fixed-cycle traffic signaling.

The organization of the paper is as follows: in Section 2, we summarize related work; in Section 3, we introduce our system model; in Section 4, we present the main results; in Section 5, we show simulation results; finally, we conclude and discuss future work in Section 6. Due to page limitation, all the proofs are omitted in this paper; nonetheless, they can be found in [1].

## 2. Related Work

Perturbation Analysis (PA) and stochastic fluid models [2–5] are used to find the performance sensitivity with respect to certain parameters of traffic light scheduling. Combined with gradient-based algorithms, such approaches are able to iteratively converge to the optimal solution in an on-line setting. Its powerfulness lies in the fact that it does not rely on the stochastic assumptions of the vehicles' arrival and departure process; in addition, the PA estimator requires very light computation such as counting the number of traffic light switching between certain events. Fuzzy logic controllers are proposed in [6–8], in which the theory of fuzzy sets and linguistic control are used to model traffic intersections where analytical methods are in general lacking. Li et al. formulate a traffic light control problem that reduces the stop-and-go time and  $CO_2$  emissions [9]. They achieve this goal by using a three-tier structure as well as a branch-and-bound based real-time algorithm. Ahmad et al. [10] considered a scenario that each vehicle has its specific deadline that needs to be satisfied. The performance of two algorithms: Earliest Deadline First (EDF) and Fixed Priority (FP), in terms of the average delay, is compared with that of fixed-cycle scheduling. Reinforcement learning are used in [11] and [12] to adaptively reduce the average cost. Mixed integer linear programming is used in [13] to perform adaptive control on an isolated intersection. In [14], Khamis et al. use Bayesian

probability interpretation and the Intelligent Driver Model (IDM) to reduce the average trip waiting time. In [15], the authors use thresholds of the vehicle waiting queues to adjust the traffic light scheduling; specifically, they use stochastic optimization and Q-learning to dynamically change the thresholds. Protschky et al. use the Kalman filter and a generic statistical prediction model to perform traffic light control [16].

### 3. System Model

We consider an isolated intersection of two one-way streets in this paper. Nonetheless, the results can be easily extended to more complex intersections. Figure 1 shows an illustration of the system model. In particular, we denote:

$L_C$ : the length of the car. We assume that all the cars have equal length.

$L_I$ : the length of the intersection. We assume that the intersection has equal length in both directions.

$L_Q$ : the length of the vehicle queue. Assuming that each car is capable of communicating with the traffic light controller via wireless technologies, this is essentially the maximum distance of reliable wireless transmission.

$V_{\max}$ : the speed limit of the roads near the intersection. It is the maximum speed each car can have.

$v(t)$ : velocity of a vehicle as a function of time.

$d_s$ : the safe distance between two adjacent cars when they are both static. It is obvious that the safe distance depends on the velocities of the cars. Here, we are only concerned with the safe distance when both cars are static.

$V_s$ : the safe velocity. When two adjacent cars are both initially static,  $d_s$  meters away, and the first one starts to accelerate, it is the velocity of the first car at which the second one starts to accelerate. Here, we assume that the two cars are able to learn each other's information, including velocity, via wireless communications. Note that  $V_s \leq V_{\max}$ . We also would like to emphasize that this safe velocity can be translated into another safe distance by doing an integral, as shown later in the paper.

$N$ : the maximum number of cars in each queue. For notation easy, we assume that  $N = \frac{L_Q}{L_C + d_s}$  is an integer. This can be satisfied by simply choosing  $L_Q$  to be the integer multiple of  $L_C + d_s$ .

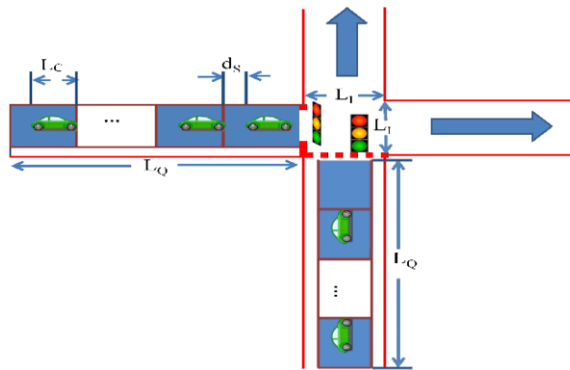


Figure 1. System Model Illustration.



$T_Y$ : the duration of yellow light. We assume that  $T_Y$  is greater than one second.

$D_i^j$ : the amount of time the  $i$ -th car spends in the  $j$ -th queue.

After introducing the system model of the intersection, we now discuss the car model. Since we are interested in providing QoS provisioning to the cars using the intersection, we need a simple model which can estimate the time it takes for the cars to travel through the intersection. We adopt a model used in [17]. Specifically, we assume that each car has the same mass  $m$  and identical driving force  $F$ . The road friction  $F_1(t)$  and wind dragging force  $F_2(t)$  can be calculated as:

$$F_1(t) = c_1v(t) \quad \text{and} \quad F_2(t) = c_2v^2(t),$$

respectively, where  $c_1$  and  $c_2$  are constant coefficients. Applying Newton’s Law, we obtain

$$m\dot{v}(t) + c_1v(t) + c_2v^2(t) = F \tag{1}$$

Because the vehicles are subject to  $V_{\max}$  near an intersection and are typically operated at low speed, we omit the wind’s dragging force, which is only large when the velocity is high. (1) can then be reduced to:

$$m\dot{v}(t) + c_1v(t) = F$$

Taking Laplace transform on both sides and assuming zero initial condition, we get the transfer function, which is first-order:

$$G(s) = \frac{V(s)}{F(S)} = \frac{1}{ms+c_1}$$

The velocity  $V(s)$  is:

$$V(s) = \frac{F}{m} \times \frac{1}{s(s+\frac{c_1}{m})}$$

Let  $K = F/c_1$  and  $a = c_1/m$ , the inverse Laplace Transform yields:

$$v(t) = K(1 - e^{-at})$$

#### 4. Main Results

Our long term goal is to design an adaptive traffic light controller which is capable of providing QoS provisioning. In this paper, we first answer the following question: once a car enters a queue, what is the best and worst-case time that it will spend at the intersection? We use  $D_{\min}$  and  $D_{\max}$  to denote these two time, respectively. Only after they are determined, we can choose the right vehicle deadline  $D$ .  $D_{\min}$  corresponds to the best case scenario where a car enters a queue and travels through the intersection at velocity  $V_{\max}$ . Therefore,

$$D_{\min} = \frac{L_Q + L_I}{V_{\max}}$$

Note that  $D_{\max}$  depends on the traffic light control algorithm. We emphasize that we are looking for the maximum delay introduced by the best scheduling algorithm, i.e.,

$$D_{\max} = \min_{\text{All traffic controllers}} \max_{\text{All traffic conditions}} \max(D_i^j)$$

Note that  $\max(D_i^j)$  is the worst-case delay among all vehicles, and it depends on both the traffic condition and the traffic light controller. Obviously, it is maximized when the traffic is extremely heavy in both directions so that even if the green light is constant on in one direction, the intersection will never idle. To obtain  $D_{\max}$ , we first establish the following result.

**Lemma 4.1.** Suppose that  $N$  cars are sitting in a queue and waiting for the green light. Assume that  $L_I$  is wide enough for the first car to accelerate to  $V_{\max}$  before leaving the intersection. If the green light is turned on and stays on, it takes  $T_i$  seconds for the  $i$ -th car to leave the intersection,  $i \in \{1, \dots, N\}$ :

$$T_i = T_{i,w} + T_{i,a} + T_{i,t}, \text{ where}$$

$T_{i,w} = [(i-1)/a][\ln K - \ln(K - V_S)]$  is the time for the  $i$ -th car to reach the safe velocity  $V_S$ ;

$T_{i,a} = (1/a)[\ln K - \ln(K - V_{\max})]$  is the time for the  $i$ -th car to accelerate to  $V_{\max}$ ;

and  $T_{i,t} = [(i-1)(L_C + d_S)/V_{\max}] + [L_I - (KT_{1,a} - V_{\max}/a)]/V_{\max}$  is the time for it to travel through the rest of the intersection at maximum speed.

We now use the results established in Lemma 4.1 to find  $D_{\max}$ . It is obvious that  $D_{\max}$  is achieved when the traffic is heavy in both directions. Next, we will not give a precise definition of “heavy”; instead, we give one scenario that will be guaranteed to yield  $D_{\max}$ . Let us assume that the maximum service rate the intersection provides to the cars is  $\mu_{\max}$ ; this is obtained when cars in one queue are traveling through the intersection at velocity  $V_{\max}$  with the right safe distance between adjacent ones. Then, the traffic condition can be considered as heavy or congested when the vehicle arrival rate to each queue is at least  $\mu_{\max}$  at all times. This traffic condition will lead to the maximum delay  $D_{\max}$ .

**Lemma 4.2.** In heavily congested conditions, the optimal traffic control policy that yields  $D_{\max}$  is to let all  $N$  cars in each queue pass the intersection during a single green light period.

**Lemma 4.3.** Suppose that in heavily congested conditions, the optimal traffic control policy is used, i.e.,  $N$  cars in each queue are allowed to pass the intersection during a single green light period. In the  $N$  cars sitting in the queue and waiting for the green light to be turned on, the cars closer to the intersection spend not less time than those further away from the intersection.

Lemma 4.3 actually shows something very intuitive: in a queue packed with cars waiting for the traffic light to turn green, the cars closer to the intersection tend to be the unlucky ones that spend longer time in the queue. They would be luckier if they passed the intersection in the previous green light interval.

**Corollary 4.1.**  $D_{\max} = 2T_N + 2T_Y - N\Delta T_w + T_1$ .

Since we now have  $D_{\min}$  and  $D_{\max}$ , adaptive traffic light controllers can be built. Intuitively, the traffic controller should use the state of the system (number of existing cars in the queues), arrival rate, etc., to intelligently assign the deadline for each queue, based on which, the traffic lights are toggled between green and red. In this paper, we present a simple open-loop controller, which uses the theoretical results above to perform traffic light scheduling:

---

**Algorithm 1.** Adaptive open-loop traffic light controller

---

```

1: Deadline  $\leftarrow D_{\max}$ 
2: while True do
3:   if Intersection is not busy then
4:     Turn on green light for the closest car
5:   else
6:     if The deadline of any car waiting for the green light is about to be missed
       then
7:       Toggle the traffic lights
8:     end if
9:   end if
10: end while

```

---

## 5. Simulation Results

We run simulation to verify the theoretical results and the performance of the above adaptive traffic controller. In our experiments, we let the numbers of vehicles in the two queues be  $M_1$  and  $M_2$ , respectively. Initially, all vehicles are at least  $L_Q$  meters away from the intersection, and the distance between two adjacent cars in queue 2 is uniformly distributed between an interval  $[d_1, d_2]$ . The  $M_1$  cars in queue 1 are uniformly distributed between the locations of the first and the last cars in queue 2. The initial velocities of all cars are uniformly distributed between 0 and  $V_{\max}$ , which is set to 13.3 m/s, i.e., 30 miles per hour. Other parameters' values are:  $m = 1500$  kg,  $F = 44444$  N,  $c_1 = 1000$ ,  $L_C = 4$  m,  $L_1 = 25$  m,  $N = 20$ ,  $d_s = 1$ ,  $L_Q = 100$  m, and  $T_Y = 3$  s. Using these values, we calculate that  $D_{\min} = 9.40$  s and  $D_{\max} = 31.05$  s.

Traffic congestion depends on the vehicle arrival rate or car density in each queue: if one queue has zero or very few cars, then it is likely that there is no traffic jam at the intersection, provided that an adaptive traffic control mechanism is in place. For this reason, we define  $R$ , the ratio between the numbers of cars:  $R = M_1/M_2$ . In our simulation,  $M_1 \leq M_2$  and  $M_2$  is fixed at 200. Therefore,  $R \leq 1$ . Note that traffic congestion depends on not only  $R$ , but also the values of  $d_1$  and  $d_2$ .

We compare the performance of the adaptive controller specified in Algorithm 1 with two other controllers:

*Fixed-cycle controller 1:* the green light and the red light cycles of both queues are  $T_N$  seconds.

*Fixed-cycle controller 2:* the green and the red light cycles of queue 2 are  $T_N$  and  $R \times T_N$ , respectively.

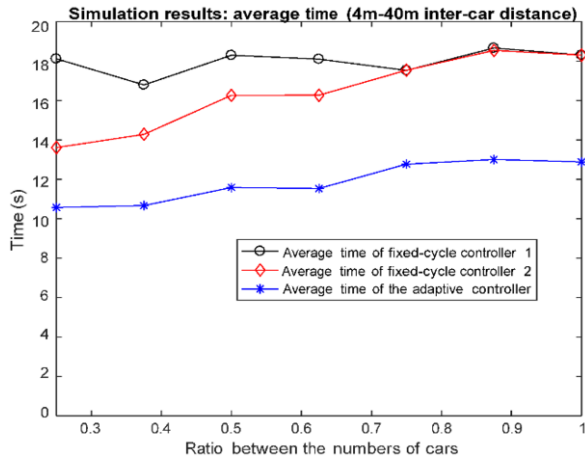


Figure 2. Average time a car spends at the intersection (4 m–40 m inter-car distance).

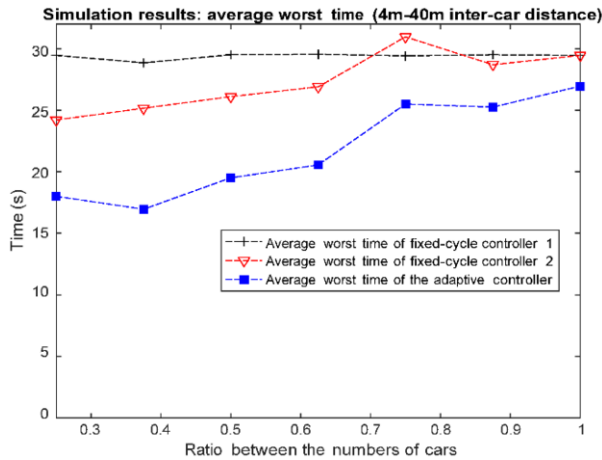


Figure 3. Average of the worst time in each queue (4 m–40 m inter-car distance).

Figures 2 and 3 show the results when  $d_1 = 4$  m and  $d_2 = 40$  m. In this case, the traffic is light, and the adaptive controller we propose reduces the average time by a large percentage (around 40% when  $R = 0.25$  in Fig. 2). The reason why the large deadline (31.05 s) does not affect things is that when the traffic is light, the intersection tends to become idle long before a waiting car’s system time reaches this deadline. This will turn on the green light for the cars waiting in the queue. Another observation is that fixed-cycle controller 2 is better than controller 1. However, it is not as good as the adaptive controller.

Nonetheless, the large deadline does help in the heavy traffic scenario, as shown in Figs 4 and 5, where  $d_1 = 4$  m and  $d_2 = 20$  m. In this case, it is less likely for the intersection to be idle, as  $R$  gets closer to 1. As a result, the large deadline used in the adaptive controller prevents the traffic light from toggling too often; this, in turn, helps with lowering the average time. We also point out that when  $R$  approaches 1, there is not much optimization to be done since the optimal scheduling is simply divide the green light time

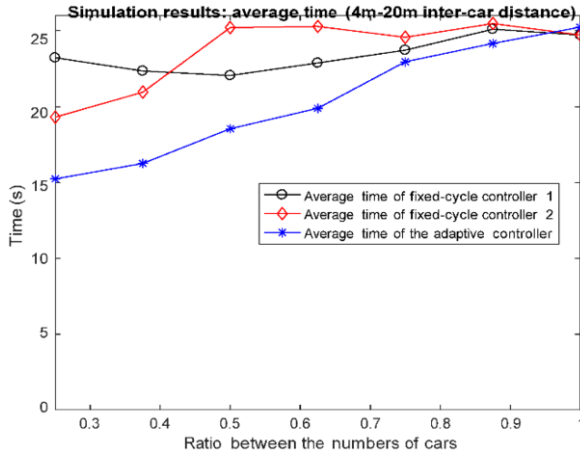


Figure 4. Average time a car spends at the intersection (4 m–20 m inter-car distance).

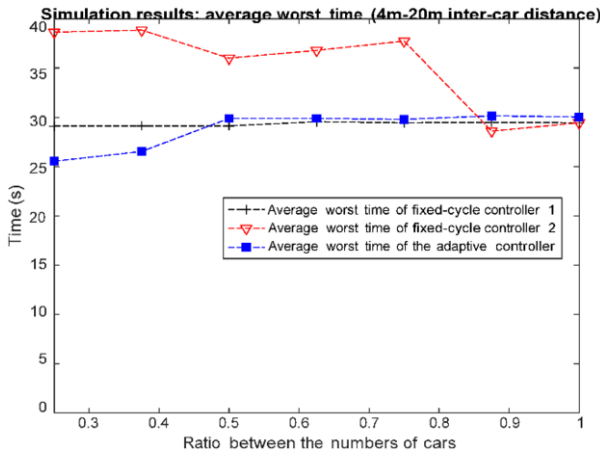


Figure 5. Average of the worst time in each queue (4 m–20 m inter-car distance).

equally between the two queues, as indicated in Lemma 4.2. In addition, when the traffic is moderate ( $R \leq 0.75$ ), fixed-cycle controller 2 is worse than controller 1, indicating that simply modifying the cycle length of fixed-cycle scheduling does not always help.

Overall, the adaptive controller significantly outperforms the other two fixed-cycle ones when the traffic is not very heavy and maintains roughly the same performance when the traffic is extremely heavy.

## 6. Conclusions and Future Work

In this paper, we model the process of cars passing through an intersection and come up with the shortest and the longest time they could possibly spend at the intersection.

The results are significant because they help people better understand how traffic light controllers can be designed in order to provide quality-of-service (QoS) to travelers

in intelligent transportation systems. We also propose a very simple adaptive controller, which is shown to be much better than two other fixed-cycle ones in most scenarios.

Our future work involves improving the adaptive controller and providing QoS guarantee (e.g., the worst-case waiting time at an intersection) to each individual car. It has been shown in our simulation that when the traffic is light and the controller is adaptive, it is possible to improve the average time and worst time of both queues. However, we have also seen that when the traffic is no longer light, the improvement of wait time over one queue actually hurts the wait time of the other. We are interested in studying how the social aspects affect the decision making in adaptive traffic light control.

## References

- [1] <http://arxiv.org/abs/1705.05440>.
- [2] C. G. Panayiotou, W. C. Howell, and M. Fu, "Online traffic light control through gradient estimation using stochastic fluid models," *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 90–95, 2005.
- [3] Y. Geng and C. G. Cassandras, "Traffic light control using infinitesimal perturbation analysis," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 7001–7006.
- [4] Y. Geng and C. G. Cassandras, "Quasi-dynamic traffic light control for a single intersection," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 880–885.
- [5] Y. Geng and C. G. Cassandras, "Multi-intersection two-way traffic light control with blocking using infinitesimal perturbation analysis," in *Control Applications (CCA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 382–387.
- [6] C. P. Pappis and E. H. Mamdani, "A fuzzy logic controller for a traffic junction," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 7, no. 10, pp. 707–717, 1977.
- [7] Y. S. Murat and E. Gedizlioglu, "A fuzzy logic multi-phased signal control model for isolated junctions," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 1, pp. 19–36, 2005.
- [8] M. Collotta, L. L. Bello, and G. Pau, "A novel approach for dynamic traffic lights management based on wireless sensor networks and multiple fuzzy logic controllers," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5403–5415, 2015.
- [9] C. Li and S. Shimamoto, "An open traffic light control model for reducing vehicles' CO<sub>2</sub> emissions based on etc vehicles," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 97–110, 2012.
- [10] A. Ahmad, R. Arshad, S. A. Mahmud, G. M. Khan, and H. S. Al-Raweshidy, "Earliest-deadline-based scheduling to reduce urban traffic congestion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1510–1526, 2014.
- [11] L. Prashanth and S. Bhatnagar, "Reinforcement learning with average cost for adaptive control of traffic lights at intersections," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE, 2011, pp. 1640–1645.
- [12] S. El-Tantawy, B. Abdulhai, and H. Abdelgawad, "Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (marlin-atsc): methodology and large-scale application on downtown toronto," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1140–1150, 2013.
- [13] Y. Dujardin, F. Boillot, D. Vanderpooten, and P. Vinant, "Multiobjective and multimodal adaptive traffic light control on single junctions," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE, 2011, pp. 1361–1368.
- [14] M. A. Khamis, W. Gomaa, A. El-Mahdy, and A. Shoukry, "Adaptive traffic control system based on bayesian probability interpretation," in *Electronics, Communications and Computers (JEC-ECC), 2012 Japan-Egypt Conference on*. IEEE, 2012, pp. 151–156.
- [15] L. Prashanth and S. Bhatnagar, "Threshold tuning using stochastic optimization for graded signal control," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 9, pp. 3865–3880, 2012.
- [16] V. Protschky, K. Wiesner, and S. Feit, "Adaptive traffic light prediction via Kalman filtering," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*. IEEE, 2014, pp. 151–157.
- [17] L. Xu, L. Y. Wang, G. Yin, and H. Zhang, "Communication information structures and contents for enhanced safety of highway vehicle platoons," *IEEE Transactions on vehicular Technology*, vol. 63, no. 9, pp. 4206–4220, 2014.

# $H_\infty$ Anti-Interference Control of EPS System for Distributed Driving Electric Vehicle

Zhang NI, Jing-bo ZHAO and Xin FAN

*School of Automobile and Traffic Engineering, Jiangsu University of Technology, Changzhou 213001, China*

**Abstract.** Distributed drive electric vehicle steering system affect the vehicle handling stability, active safety and ride comfort and other performance. Electric power steering (EPS) is a new vehicle power steering technology, in line with the automotive energy conservation, environmental development theme, consistent with the theme of the development of automotive electronics, intelligent vehicles, and become the focus of attention for domestic and foreign auto industry experts and scholars. The dynamic model of EPS is established, and the influence of the perturbation of parameters such as parameter perturbation, sensor measurement noise and external disturbance in the dynamic model of electric power steering control system is analyzed by using  $H_\infty$  control theory. In the process of solving controller, the weight function modeling and linear fractional transformation theory deals with these uncertainties and designs the  $H_\infty$  controller of the steering assist process. The performance of the controller is simulated by Matlab. The simulation results show that the controller has good robust stability and robust performance.

**Keywords.** Dynamic model, EPS,  $H_\infty$ , distributed drive electric vehicle, vehicle handling stability

## 1. EPS system dynamics model

The vehicle steering system with EPS is simplified, as shown in Fig. 1. According to the use conditions and idealized assumptions of the steering system, some minor factors are neglected. The simplified EPS system mainly includes five main parts of steering wheel and steering column, motor, deceleration structure and gear, rack. According to the Newton's law of motion, the differential kinematics equation of each part is established, and the system state vector, output vector and input vector are selected rationally to get the whole system model.

$T_d, T_a, T_m$  is the driver's input torque, the actual power of the motor torque, and the theoretical output torque of the motor respectively,  $\theta_c, X_r, \theta_m$  is the angle of steering column, the displacement of the rack and power motor rotation respectively,  $J_c, K_c, B_c$  is steering column inertia, stiffness and damping respectively,  $M_r, B_r, K_r$  is the rack mass, rack damping and rack equivalent stiffness,  $r_p$  is the steering gear pitch circle radius,  $G$  is reduction ratio for the motor to the steering column,  $K_m, J_m, B_m$  is

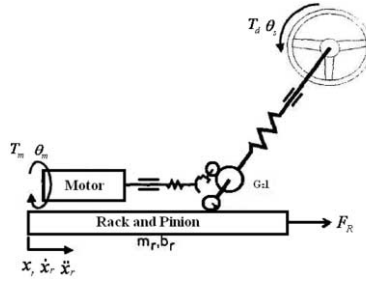


Figure 1. Electric power steering system model.

the motor torsion stiffness and the motor moment of inertia and the motor damping coefficient respectively.

The steering wheel and the input shaft are analyzed, the inertia of the steering wheel is taken into account, and also the viscous damping of the input shaft, the following equation of motion can be obtained:

$$T_d = J_c \ddot{\theta}_c + B_c \dot{\theta}_c + K_c (\theta_c - \frac{X_r}{r_p}) + f_c(\theta_c, \dot{\theta}_c) \quad (1)$$

Where,  $J_c$  is the moment of inertia for steering column,  $Kg \cdot m^2$ ;  $B_c$  is steering column viscous damping coefficient,  $N \cdot m / (rad / s)$ ;  $K_c$  is torsional stiffness of steering column  $N \cdot m / rad$ ;  $\theta_c$  is steering column angle,  $rad$ ;  $T_d$  is steering wheel input torque,  $N \cdot m$ ;  $X_r$  is rack displacement,  $m$ ;  $r_p$  is steering gear pitch circle radius,  $m$ ;  $f_c(\theta_c, \dot{\theta}_c)$  is friction torque of steering column and coupling,  $N \cdot m$ .

The rack and pinion can be modeled by dynamic modeling of the rack and pinion:

$$M_r \ddot{x}_r = \frac{K_c}{r_p} \theta_c - \frac{K_c}{r_p^2} x_r - B_r \dot{x}_r + \frac{K_m G}{r_p} \theta_m - \frac{K_m G^2}{r_p^2} x_r - K_t x_r - F_{TR} + f_r(x_r, \dot{x}_r) \quad (2)$$

Where in (2),  $M_r$  is rack equivalent mass,  $Kg$ ;  $B_r$  is rack equivalent damping coefficient,  $N \cdot m / (rad / s)$ ;  $x_r$  is rack displacement,  $m$ ;  $r_p$  is pinion pitch radius,  $m$ ;  $K_m$  is motor torsional rigidity,  $N \cdot m / rad$ ;  $\theta_m$  is power motor rotation angle,  $rad$ ;  $G$  is motor reduction mechanism reduction ratio;  $K_t$  is rack equivalent stiffness,  $N / m$ ;  $K_c$  is steering column torsional stiffness,  $N \cdot m / rad$ ;  $\theta_c$  is steering column angle,  $rad$ ;  $f_r(x_r, \dot{x}_r)$  is nonlinear frictional force acting on the rack,  $N$ ;  $F_{TR}$  is interference force for road conditions change,  $N$ .

The system uses a permanent magnet DC motor. The relationship between the terminal voltage of the motor  $U$  and the inductance  $L$ , armature resistance  $R$ , back electromotive force constant  $K_b$ , speed  $\dot{\theta}_m$ , current  $I$  and time  $t$  is as follows:

$$U = L \dot{I} + RI + K_b \dot{\theta}_m \quad (3)$$



The electromagnetic torque generated by the motor is

$$T_m = K_a I \tag{4}$$

Where  $K_a$  is the torque coefficient of the motor,  $N \cdot m \cdot A^{-1}$ .

The mechanical analysis of the parts of the motor can be obtained,

$$T_m - K_m \left( \theta_m - \frac{Gx_r}{r_p} \right) + f_m(\theta_m, \dot{\theta}_m) = J_m \ddot{\theta}_m + B_m \dot{\theta}_m \tag{5}$$

Where  $J_m$  is the motor moment of inertia,  $Kg \cdot m^2$ ;  $B_m$  is the motor viscous damping coefficient,  $N \cdot m / (rad / s)$ ;  $\theta_m$  is the angle of the motor,  $rad$ ;  $T_m$  is the motor electromagnetic torque,  $N \cdot m$ .

According to the above dynamic equation to establish the system simulation model of the state equation, where the state variable is  $x = [\theta_c \quad \dot{\theta}_c \quad x_r \quad \dot{x}_r \quad \theta_m \quad \dot{\theta}_m \quad I]^T$ , control input is  $u = [U \quad F_{TR}]$ , the output is  $y = [T_c \quad T_a \quad \theta_c \quad \dot{\theta}_m \quad I]^T$ .

The state equation and the output equation of the system are shown in Eq. (6)

$$\begin{cases} \dot{X} = AX + BU \\ Y = CX + DU \end{cases} \tag{6}$$

Where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -\frac{K_c}{J_c} & -\frac{B_c}{J_c} & \frac{K_c}{J_c r_p} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{K_c}{M_r r_p} & 0 & -\frac{K_c + K_m G^2}{M_r r_p^2} - \frac{K_t}{M_r} & -\frac{B_r}{M_r} & \frac{K_m G}{M_r r_p} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{K_m G}{J_m r_p} & 0 & -\frac{K_m}{J_m} & -\frac{B_m}{J_m} & \frac{k_a}{J_m} \\ 0 & 0 & 0 & 0 & 0 & -\frac{k_b}{L} & -\frac{R}{L} \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L} \\ 0 & 0 & 0 & -\frac{1}{M_r} & 0 & 0 & 0 \end{bmatrix}^T$$

$$C = \begin{bmatrix} K_c & 0 & \frac{-K_c}{r_p} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{K_m G}{r_p} & 0 & K_m & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The form of its transfer function is:

$$\begin{bmatrix} T_c \\ T_a \\ \theta_c \\ \dot{\theta}_m \\ I \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \\ G_{41} & G_{42} & G_{43} \\ G_{51} & G_{52} & G_{53} \end{bmatrix} \begin{bmatrix} T_d \\ U \\ F_{TR} \end{bmatrix}$$

That is  $y = G(s) \bullet u$ , where  $G(s) = C(sI - A)^{-1} B$  is the transfer function from input to output.

The model has three main inputs: the steering wheel torque, the motor voltage and the random resistance of the road. The output of the system mainly includes the torque sensor signal, the actual power supply, the steering column angle, the motor speed and the current.

## 2. Representation of Parameter Uncertainty in EPS System

With the increase of vehicle running time, the vehicle parts will inevitably wear out, and due to the maintenance status of the vehicle is different, some parameters  $M_r$ ,  $B_s$ ,  $B_r$ ,  $B_m$  in the model will change accordingly, these changes will directly affect the parameters of the vehicle steering system performance.

These parameters are assumed to be uncertain, including a nominal value and a range of variation:

$$M_r = \bar{M}_r(1 + d_{mr} \delta_{mr}) \tag{7}$$

$$B_s = \bar{B}_s(1 + d_{Bs} \delta_{Bs}) \tag{8}$$

$$B_r = \bar{B}_r(1 + d_{B_r}\delta_{B_r}) \tag{9}$$

$$B_m = \bar{B}_m(1 + d_{B_m}\delta_{B_m}) \tag{10}$$

$\bar{M}_r, \bar{B}_s, \bar{B}_r$  is the nominal value of each parameter,  $d_{m_r}, d_{B_s}, d_{B_r}, d_{B_m} \in [0,1]$  is scalar quantity, indicates the percentage of parameters that are allowed to change around the nominal value,  $\delta_{m_r}, \delta_{B_s}, \delta_{B_r}, \delta_{B_m} \in [-1,1]$ . The change  $\delta$  determines the deviation of the actual parameters.

Parameter perturbation linear fractional change processing, written in LFT form, parameters  $M_r, J_s, J_m$  in the above equation are in the form of denominator, the LFT form are:

$$\begin{aligned} \frac{1}{M_r} &= \frac{1}{\bar{M}_r(1 + d_{m_r}\delta_{m_r})} = \frac{1}{\bar{M}_r} - \frac{d_{m_r}}{\bar{M}_r} \delta_{m_r}(1 + d_{m_r}\delta_{m_r})^{-1} \\ &= F_1 \left[ \begin{array}{cc} \frac{1}{\bar{M}_r} & -\frac{d_{m_r}}{\bar{M}_r} \\ 1 & -d_{m_r} \end{array} \right], \delta_{m_r} = F_1(M_{M_r}, \delta_{m_r}) \end{aligned} \tag{11}$$

Parameters  $B_s, B_r, B_m$  are in the form of molecules, the LFT form are:

$$B_s = \bar{B}_s(1 + d_{B_s}\delta_{B_s}) = F_1 \left[ \begin{array}{cc} \bar{B}_s & \bar{B}_s d_{B_s} \\ 1 & 0 \end{array} \right], \delta_{B_s} = F_1(M_{B_s}, \delta_{B_s}) \tag{12}$$

$$B_r = \bar{B}_r(1 + d_{B_r}\delta_{B_r}) = F_1 \left[ \begin{array}{cc} \bar{B}_r & \bar{B}_r d_{B_r} \\ 1 & 0 \end{array} \right], \delta_{B_r} = F_1(M_{B_r}, \delta_{B_r}) \tag{13}$$

$$B_m = \bar{B}_m(1 + d_{B_m}\delta_{B_m}) = F_1 \left[ \begin{array}{cc} \bar{B}_m & \bar{B}_m d_{B_m} \\ 1 & 0 \end{array} \right], \delta_{B_m} = F_1(M_{B_m}, \delta_{B_m}) \tag{14}$$

(7) to (11) are substituted into the formulas (1) to (5) respectively then

$$J_s \ddot{\theta}_s + \bar{B}_s(1 + d_{B_s}\delta_{B_s}) \dot{\theta}_s + K_s \theta_s = T_h + K_s \frac{x_r}{r_p} \tag{15}$$

$$J_m \ddot{\theta}_m + \bar{B}_m(1 + d_{B_m}\delta_{B_m}) \dot{\theta}_m + K_m \theta_m = T_m + GK_m \frac{x_r}{r_p} \tag{16}$$

$$\bar{M}_r(1 + d_{m_r}\delta_{m_r}) \ddot{x}_r + \bar{B}_r(1 + d_{B_r}\delta_{B_r}) \dot{x}_r + K_r x_r = GK_m \frac{\theta_m}{r_p} + \frac{K_s}{r_p} \theta_s - F_\delta \tag{17}$$

Introduce the variables corresponding to perturbation,

$$z = [z_1 \ z_2 \ z_3 \ z_4 \ z_5 \ z_6 \ z_7 \ z_8 \ z_9]^T$$

$$v = [v_1 \ v_2 \ v_3 \ v_4]^T$$

Where  $z_1 = x_1$ ,  $z_2 = x_2$ ,  $z_3 = \dot{x}_2$ ,  $z_4 = x_3$ ,  $z_5 = x_4$ ,  $z_6 = \dot{x}_4$ ,  $z_7 = x_5$ ,  $z_8 = x_6$ ,  $z_9 = \dot{x}_6$ ,  $v_1 = \delta_{Bz_2}$ ,  $v_2 = \delta_{Bmz_8}$ ,  $v_3 = \delta_{Bz_5}$ ,  $v_4 = \delta_{Mrz_6}$ .

Thus, the state space model of the system can be collated into the lower linear fractional transformation which consisting of the M-matrix and the pachulum array  $\Delta$  shown in Eq. (18).

$$\begin{bmatrix} \dot{x} \\ y \end{bmatrix} = F_l(M, \Delta) \begin{bmatrix} x \\ u \end{bmatrix} \quad (18)$$

Where the corresponding input and output relationship of the array M is

$$\begin{bmatrix} \dot{x} \\ y \\ z \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} x \\ u \\ v \end{bmatrix} \quad (19)$$

The M-matrix is shown below:

$$M_{11} = \begin{pmatrix} a_{11} & \cdots & a_{16} \\ \vdots & \ddots & \vdots \\ a_{61} & \cdots & a_{66} \end{pmatrix}, M_{12} = \begin{pmatrix} b_{11} & \cdots & b_{13} \\ \vdots & \ddots & \vdots \\ b_{61} & \cdots & b_{63} \end{pmatrix}, M_{13} = \begin{pmatrix} c_{11} & \cdots & c_{14} \\ \vdots & \ddots & \vdots \\ c_{61} & \cdots & c_{64} \end{pmatrix};$$

$$M_{21} = \begin{pmatrix} d_{11} & \cdots & d_{16} \\ \vdots & \ddots & \vdots \\ d_{51} & \cdots & d_{56} \end{pmatrix}, M_{22} = 0, M_{23} = 0;$$

$$M_{31} = \begin{pmatrix} e_{11} & \cdots & e_{19} \\ \vdots & \ddots & \vdots \\ e_{91} & \cdots & e_{96} \end{pmatrix}, M_{32} = \begin{pmatrix} f_{11} & \cdots & f_{13} \\ \vdots & \ddots & \vdots \\ f_{91} & \cdots & f_{93} \end{pmatrix}, M_{33} = \begin{pmatrix} g_{11} & \cdots & g_{14} \\ \vdots & \ddots & \vdots \\ g_{91} & \cdots & g_{94} \end{pmatrix}.$$

Where,  $a_{12} = 1$ ,  $a_{21} = -\frac{K_s}{J_s}$ ,  $a_{22} = -\frac{B_s}{J_s}$ ,  $a_{23} = \frac{K_s}{J_s r_p}$ ,  $a_{34} = 1$ ,  $a_{41} = \frac{K_s}{M_r r_p}$ ,

$$a_{43} = -\frac{K_r}{M_r}, a_{44} = -\frac{B_r}{M_r}, a_{45} = \frac{K_m G}{M_r r_p}, a_{56} = 1, a_{63} = \frac{K_m G}{J_m r_p}, a_{65} = -\frac{K_m}{J_m}, a_{66} = -\frac{B_m}{J_m},$$

$$b_{21} = \frac{1}{J_s}, b_{43} = -\frac{1}{M_r}, b_{62} = \frac{1}{J_m}, c_{21} = -\frac{\bar{B}sdBs}{J_s}, c_{43} = -\frac{\bar{B}rdBr}{M_r}, c_{44} = -d_{Mr},$$

$$\begin{aligned}
c_{32} &= -\frac{\bar{B}m dBm}{Jm}, \quad d_{11} = K_s, \quad d_{13} = -\frac{K_r}{r_p}, \quad d_{23} = -\frac{K_m G}{r_p}, \quad d_{25} = K_m, \quad d_{35} = 1, \quad d_{46} = 1, \\
d_{63} &= 1, \quad e_{11} = 1, \quad e_{21} = 1, \quad e_{31} = -\frac{K_s}{J_s}, \quad e_{32} = -\frac{B_s}{J_s}, \quad e_{33} = \frac{K_s}{J_s r_p}, \quad e_{43} = 1, \quad e_{54} = 1, \\
e_{61} &= \frac{K_s}{M_r r_p}, \quad e_{63} = -\frac{K_r}{M_r}, \quad e_{64} = -\frac{B_r}{M_r}, \quad e_{65} = \frac{K_m G}{M_r r_p}, \quad e_{75} = 1, \quad e_{86} = 1, \quad f_{31} = \frac{1}{J_s}, \\
f_{63} &= -\frac{1}{M_r}, \quad f_{92} = \frac{1}{J_m}, \quad g_{31} = -\frac{\bar{B}_s dB_s}{J_s}, \quad g_{63} = -\frac{\bar{B}_r dB_r}{M_r}, \quad g_{64} = -d_{Mr}, \quad g_{92} = -\frac{\bar{B}m dBm}{J_m}, \text{ the} \\
&\text{rest of the matrix is zero.}
\end{aligned}$$

### 3. EPS Anti-Interference Control

The system needs to measure the signals which are the driver torque signal, the steering wheel angle signal, the motor speed signal, when using sensors for measurement, the system performance will inevitably be affected by the sensor noise, thus base on the introduction of the weight function,  $W_{T_c}, W_{\theta_c}, W_{\dot{\theta}_m}$  which represent the sensor noise amplitude frequency characteristics,  $W_{\text{noise}} = \text{diag}\{W_{\theta_c}, W_{\dot{\theta}_m}, W_{T_c}\}$ , and take the sensor noise weight function  $W_{T_c} = W_{\theta_c} = W_{\dot{\theta}_m} = 0.01$ .

In the rack stress analysis, the lateral force of the tire  $F_t = K_t X_r + F_r$ . For the actual tire model, the model does not reflect the high frequency dynamic characteristics of the tire, so there is a high frequency unmodeled dynamic model error. The multiplication perturbation can be used to characterize the amplitude and frequency characteristics of the model error. Assume that the actual tire model  $\tilde{F}_t$  is expressed

$$\text{as: } M(F_t, W_{K_t}) = \left\{ \tilde{F}_t \left\| \frac{\tilde{F}_t(j\omega) - F_t(j\omega)}{F_t(j\omega)} \right\| \leq |W_{K_t}(j\omega)| \right\}. \text{ The weight function } W_{K_t} =$$

$\frac{0.125s + 0.25}{0.125s + 1}$  shows that the difference between the actual model of the lateral force of the tire and the nominal model is 25% in the low frequency mode, and the difference of the high frequency mode module is 100%.

In order to design the  $H_\infty$  controller  $K_s$ , the evaluation vector must be introduced according to the control target of the whole control system so that the description of the problem conforms to the standard control mode of the  $H_\infty$  control theory. The control objectives of the control system can be qualitatively described as:

(1) to ensure the performance, the error should as small as possible.

The EPS performance can be described by the dynamic compensation helper rule, which reflects the relationship between the ideal torque of the motor and the steering torque of the steering wheel. Since the passive torque of the steering wheel can not be directly detected, it can only be estimated by the output torque of the torque sensor mounted on the torsion bar, causing the system to generate delay on the input of the steering wheel's passive torque, besides plus the road interference and the torque sen-

sor to measure the noise influence, so that there is a big difference between the actual value and the ideal value, affecting the tracking performance of the system. Therefore, the primary goal of the EPS control system is to design the controller to minimize the boost torque deviation  $e_a = T_a - T_a^*$  within a certain frequency range.

Design the weighting function  $W_1(s)$  as a low-pass filter;

$$W_1(s) = k_1 \frac{a_1s + 1}{b_1s + 1} \quad (20)$$

Where  $k_1$  is low-frequency gain factor, so the controlled output is

$$e_1 = W_1(s)(T_a - T_a^*) \quad (21)$$

Design goals are minimized  $\|e_1\|_2$ .

(2) can suppress external interference, thus get the best way to turn.

To ensure the stability of the control system and steering sense, the control system can inhibit external interference.

Design the weighting function  $W_2(s)$  as a low-pass filter:

$$W_2(s) = k_2 \frac{a_2s + 1}{b_2s + 1} \quad (22)$$

In the formula,  $k_2$  for the low-frequency gain coefficient, by setting the parameters, so that the high-frequency components are filtered to meet the wheel side of the steering wheel caused by the torque transmission smooth requirements, and avoid steering wheel jitters. So the controlled output is:

$$e_2 = W_2(s)(T_c - K_f F_{TR}) \quad (23)$$

In the formula,  $T_c - K_f F_{TR}$  means that the steering road error can be defined as the intensity of the steering road  $K = K_f \frac{dT_d}{dF_{TR}}$ , and the meaning is the change value of the driver's hand when the load is increased.

Design goals are minimized  $\|e_2\|_2$ .

(3) effectively suppress the motor torque ripple.

EPS system in the power process, the motor speed constantly changing due to the magnetic field distribution of non-uniformity, coupled with the current sensor measuring the impact of noise, it will cause the motor torque fluctuations, causing the steering wheel vibration. Design the weighting function

$$W_3(s) = k_3 \frac{a_3s + 1}{b_3s + 1} \quad (24)$$

The controlled output is

$$e_3 = W_3(s)(U - Ri) \tag{25}$$

Design goals are minimized  $\|e_3\|_2$  .

(4) having the most economical energy consumption. Select the controlled output  $e_4 = \rho u$  , the design goal  $\|e_4\|_2$  is minimized.

According to the above-mentioned system control target and the selected weighting function, a generalized controlled object including external disturbance and weighting function is established to form the standard  $H_\infty$  control problem.

The design goal of the  $H_\infty$  controller is to keep the effect of interference  $w$  on the output of the controlled object  $z$  as small as possible and to keep the output of the actuator under a certain amount of control energy. The solution of the optimal  $H_\infty$  control problem can be expressed as follows: finding a stable feedback controller from the control object  $P(S)$  to stabilize the closed-loop system and make the  $H_\infty$  norm of the transfer function matrix  $H_{zw}$  extremely small, that is  $\min_K \|H_{zw}\|_\infty = \gamma$  . The size of  $\gamma$  direct impact on the performance of the control system, the smaller  $\gamma$  the better the performance of the control system. EPS system parameters are shown in Table 1.

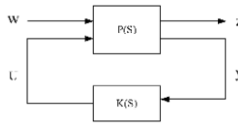


Figure 2. Standard  $H_\infty$  control block diagram.

Table 1. EPS system parameters.

$M_r$ -rack equivalent mass	$M_r = 28 \text{ kg}$
$J_c$ Moment of inertia of steering column	$J_c = 0.04 \text{ kg.m}^2$
$J_m$ -Moment of inertia of motor	$J_m = 0.0004707 \text{ kg.m}^2$
$K_a$ -Motor torque factor	$K_a = 0.02 \text{ v.s / rad}$
$K_c$ - Steering rigidity of steering column	$K_c = 115 \text{ Nm / rad}$
$K_m$ -Torsional rigidity of motor	$K_m = 90 \text{ Nm/rad}$
$K_t$ -rack equivalent stiffness	$K_t = 91061 \text{ N/m}$
$K_b$ -back electromotive force constants	$K_b = 0.02$
$B_c$ -Viscous damping coefficient of steering column	$B_c = 0.362 \text{ N.m.s/rad}$
$B_r$ -rack equivalent damping coefficient	$B_r = 650.5 \text{ N.m.s/rad}$
$B_m$ -viscous damping coefficient of motor	$B_m = 0.00334 \text{ N.m.s/rad}$
$r_p$ -pinion pitch circle radius	$r_p = 0.0078 \text{ m}$
$G$ -Reducer ratio of motor reduction mechanism	$G = 16.5$
$L$ -inductance	$L = 0.0015 \text{ H}$
$R$ -armature resistance	$R = 0.15 \Omega$

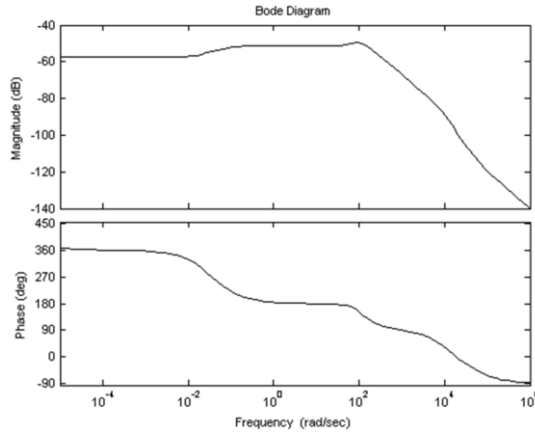


Figure 3. Controller transfer function bode chart.

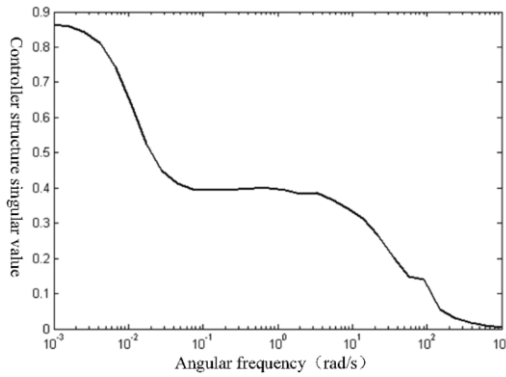


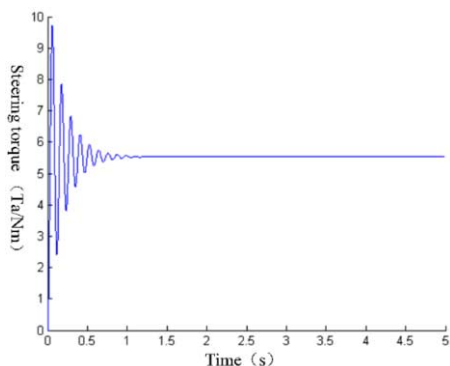
Figure 4. Singular value of  $H_\infty$  controller structure.

The bode diagram of the controller from the torsion bar torque to the motor control voltage is shown in Fig. 3.

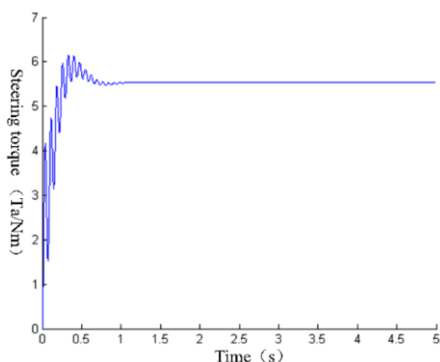
#### 4. $H_\infty$ Controller Anti-Interfere Performance Analysis

Figure 4 shows the singular value of the system performance of the  $H_\infty$  controller. It can be seen from the figure that the maximum value of the singular value of the system structure is  $0.863 < 1$ . According to the  $H_\infty$  control theory, it can be deduced that the designed  $H_\infty$  controller has good performance robustness and robust stability under the influence of external noise such as sensor noise and high frequency torque of the road, and can meet the performance requirements of EPS in low frequency band and high frequency band, which reflecting the good anti-interfere performance.





**Figure 5.** PD Control System Step Response.



**Figure 6.**  $H_\infty$  Control system step response.

Join the white noise to the steering wheel torque measurement, steering wheel angle measurement and assist motor speed measurement value of the covariance of 0.2. Respectively, as the external interference noise input, driver torque, road interference torque input for the unit step signal. Using the  $H_\infty$  robust controller designed above and the PD controller proposed in reference, the simulation results of the motor boost torque are shown in Figs 5 and 6. It can be seen that the controller stabilizes at 0.6 s and the settling time of the PD controller is 1 s under the influence of sensor noise and external disturbance. The maximum boosting torque of the two controllers is 6.05 Nm and 9.6 Nm, The  $H_\infty$  controller's anti-interfere effect is better than the PD controller. The performance of the two controllers is shown in Table 2.

Under the interference of road interference torque and sensor noise, and considering the influence of tire non-modeling error, compared with the PD controller, the  $H_\infty$  controller has better performance under the influence of external disturbance such as sensor noise and high frequency moment of road Rod stability and robust stability, to meet the performance requirements of EPS, reflecting the good anti-interfere performance.

**Table 2.** The performance of the two controllers.

	Power stabilization time	Power torque overshoot
$H_\infty$ Controller	0.6 s	109%
PD controller	1 s	175%
Performance comparison	$H_\infty$ controller settles the settling time by 0.4 s for the PD controller	The overshoot of the $H_\infty$ controller is significantly smaller than the PD controller

## 5. Conclusion

The  $H_\infty$  control theory is used to analyze the internal uncertainties and external disturbances of the EPS system. The corresponding weight function is designed to solve the  $H_\infty$  controller and compared with the PD control. The results show that the  $H_\infty$  controller has better robust stability and performance robustness.

## Acknowledgment

Funding from the National Natural Science Foundation of China (Grant No: 61503163), the “333 project” of Jiangsu Province (Grant No: BRA2016440) and the six talent peaks project in Jiangsu Province (Grant No: ZBZZ-024) are gratefully acknowledged.

## References

- [1] Bobier C G, Gerdes J C. Sliding surface vehicle envelope control: A cooperative design between controller and envelope [C]//American Control Conference, June 27–29, 2012. Fairmont Queen Elizabeth, Montréal, Canada: 2012:6521–6526.
- [2] Jiang Qingyun. Global chassis control based on ESP and EPS [D]. BeOing:Tsinghua University, 2011.
- [3] Choi J, Yi K, Suh J, et al. Coordinated control of motor-driven power steering torque overlay and differential braking for emergency driving support [J]. IEEE Transactions on Vehicular Technology, 2014, 63(2):566–579.
- [4] Parmar M, Hung J Y. A sensorless optimal control system for an automotive electric power assist steering system [J]. Industrial Electronics, IEEE Transactions on, 2004, 51(2):290–298.
- [5] Xiang Dan, Li Wu-bo, Yang Yong. Study on return-to-center control and simulation for electric power steering system [J]. Machinery Design and Manufacture, 2012(8):115–117.
- [6] Brunnsccheiler D. Modeme lenksysteme [J]. ATL, 2005(2):243–250.
- [7] Wang Jun, Yang Shengbing. Adaptive yaw angular velocity control of electric power steering [R]. SAE, 2011-01-0759.
- [8] Hsu Y H J, Laws S, Gerdes J C. Estimation of tire slip angle and friction limits using steering toeque [J]. IEEE Transactions on Control Systems Technology, 2010, 18(4):896–907.
- [9] Hao Chen, Yali Yang, Ruoping Zhang. Study on Electric Power Steering System Based on ADAMS [J]. Procedia Engineering, 2011, 15.
- [10] Kemmetmüller Wolfgang, Müller Steffen, Kugi Andreas. Mathematical modeling and nonlinear controller design for a novel electrohydraulic power-steering system. IEEE ASME Transactions on Mechatronics. 2007.

# Research on Vehicle Navigation System Based on Low-Cost Sensors

Jiankui Du<sup>a,1</sup>, Kanghua Tang<sup>a</sup>, Xiaoping Hu<sup>a</sup>

<sup>a</sup>College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha, China

**Abstract.** Although the vehicle navigation system based on GNSS/MIMU can effectively improve the reliability and precision of the single navigation system, its application is still restricted by many problems such as expensive equipment, unstable satellite signals and other issues. In order to solve these problems, a vehicle navigation system based on low-cost sensors is designed in this paper. Firstly, the hardware platform of vehicle navigation system based on microcontroller, low-precision MIMU and low-cost GNSS receiver is designed. Then, the model of sensor's error, navigation system's error and vehicle velocity constraint is analyzed, and a vehicle navigation algorithm with dual modes is proposed. When satellite signal is valid, the navigation system works at GNSS/SINS integrated navigation mode based on Kalman filter. While the GNSS is unable to provide navigation information, the system switches to SINS mode with the constraint of vehicle velocity. Finally, the feasibility and practicability of vehicle navigation system designed in the paper is verified by vehicle test. The results show that the dynamic accuracy of horizontal position is better than 1.5 meters, the dynamic accuracy of horizontal attitude is better than 0.2 degrees, the dynamic accuracy of azimuth attitude is better than 1.5 degrees and the dynamic accuracy of velocity is better than 0.15 meters per second. In addition, this system can work on the situation of satellite short-term failure, so as to meet the requirements of vehicle navigation in changeable driving environment.

**Keywords.** vehicle navigation, low-cost sensors, Kalman filter, vehicle velocity constraint

## 1. Introduction

In recent years, the technology of vehicle navigation is developing rapidly and many algorithms have been used for precise navigation of vehicle. However, high-precision navigation technology often relies on expensive equipment, complex calculations and specific conditions, which greatly constrains its application. With the rapid development and application of GNSS and MEMS, the GNSS/SINS integrated navigation system using low-cost sensors is becoming a new research direction of vehicle navigation.

Among numerous vehicle navigation methods based on low-cost sensors, the two-dimensional dead reckoning method is widely used [1-3], which obtains the current position through the past position, velocity and heading information. Although this

---

<sup>1</sup> Corresponding Author, College of Mechatronics Engineering and Automation, National University of Defense Technology, No.109, Deya Road, Kaifu District, Changsha, Hunan Province, China; E-mail: dujiankui@foxmail.com.

method is easy to implement, it can not get all navigation information of vehicle, such as height, pitch angle and roll angle which is necessary for the precise navigation of vehicle. Comparing to 2D Dead Reckoning, the combination of SINS and GNSS can obtain comprehensive navigation information and some data fusion algorithms based on it can improve the accuracy of navigation to a certain extent [4-8]. Nevertheless, some complex algorithms are usually difficult to be applied in practice because of its long-time calculation and bad real-time performance. In recent years, some scholars use complementary filtering, gradient descent and other methods for the vehicle navigation research [9, 10]. The advantage of such methods is that it does not need to establish the error drift model of sensor, but it is difficult to determine the control parameters when using these methods in vehicle navigation.

Based on previous research in the application of vehicle navigation, a vehicle navigation system based on low-cost sensors is designed in this paper. This system can provide a full range of navigation information which is essential to the precise navigation and state monitoring of the vehicle. And it is easy-implement and cost-effective. In the first section of this paper, the introduction and related works of vehicle navigation based on low-cost sensors are elaborated. In the second section, some low-cost components are used to design the hardware platform of the vehicle navigation system. Then, a vehicle navigation algorithm with dual modes is proposed. Finally, the feasibility and practicability of vehicle navigation system is verified by vehicle test.

## 2. Hardware Platform Design

Considering the cost, volume, quality, power consumption and reliability, some components are selected for the hardware platform of vehicle navigation system as follows. The STM32F427 of STMicroelectronics is chose as control element of the hardware platform. The 3DM-GX3- 15 with MEMS accelerometer and gyro is used as motion sensor whose specific parameters shown in Table 1. The STA8090FG of STMicroelectronics is selected to receive satellite data, the position accuracy of which is better than 1.8 meters (RMS). The total cost of the hardware platform is no more than RMB 300 and it is easy-implement. The structure and hardware diagram of the navigation system are illustrated in figure 1 and figure 2.

Table 1. 3DM-GX3-15 specifications

	Accelerometer	Gyro
<b>Range</b>	±5g	±300deg/sec
<b>Non-Linearity</b>	±0.1%fs	±0.03%fs
<b>Bias Instability(10 sec)</b>	0.04mg	18deg/h
<b>Noise Density</b>	80 μg / √Hz	0.03 deg/sec/√Hz
<b>Bandwidth</b>	50Hz	50Hz

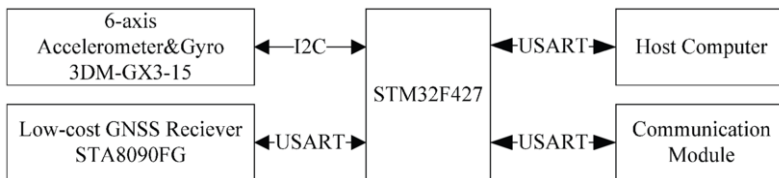


Figure 1. Structure of low-cost vehicle navigation system hardware platform

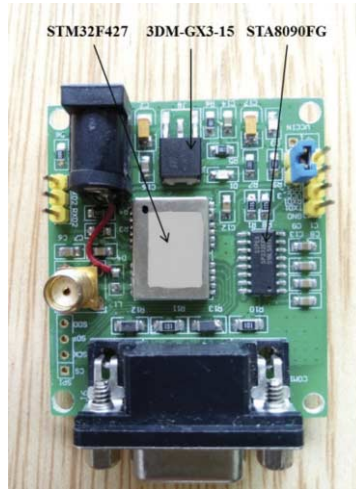


Figure 2. Hardware diagram of low-cost vehicle navigation system

### 3. System Error Analysis

#### 3.1. Error Model of Inertial Sensor

The navigation coordinate system in this paper is North-Up-East, and the vehicle coordinate system is Front-Up-Right.

Assuming that the stochastic model of gyro drift is a combination of random constants and first-order Markov processes [11]:

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_b + \boldsymbol{\varepsilon}_r, \quad \dot{\boldsymbol{\varepsilon}}_b = 0, \quad \dot{\boldsymbol{\varepsilon}}_r = -\boldsymbol{\beta}\boldsymbol{\varepsilon}_r + \mathbf{w}_g \quad (1)$$

Assuming that the stochastic error model of the accelerometer is the first-order Markov process [11]:

$$\dot{\boldsymbol{\xi}} = -\boldsymbol{\mu}\boldsymbol{\xi} + \mathbf{w}_a \quad (2)$$

Where  $\boldsymbol{\beta}$  and  $\boldsymbol{\mu}$  are the inverse correlation time constant,  $\mathbf{w}_g$  and  $\mathbf{w}_a$  are white noise.

#### 3.2. Error Model of Strapdown Inertial Navigation System

According to the mathematic model of SINS [11], the error model of navigation parameters is shown as below.

The error equation of position is:

$$\left\{ \begin{aligned} \delta \dot{L} &= \frac{1}{R_N + h} \delta V_N - \frac{V_N}{(R_N + h)^2} \delta h \\ \delta \dot{\lambda} &= \frac{1}{(R_E + h) \cos L} \delta V_E + \frac{V_E \sin L}{(R_E + h) \cos^2 L} \delta L - \frac{V_E}{(R_E + h)^2 \cos L} \delta h \\ \delta \dot{h} &= \delta V_U \end{aligned} \right. \quad (3)$$

The error equation of attitude is:

$$\dot{\boldsymbol{\phi}} = -\boldsymbol{\omega}_m^n \times \boldsymbol{\phi} + \delta\boldsymbol{\omega}_m^n - \mathbf{C}_b^n \delta\boldsymbol{\omega}_{ib}^b \quad (4)$$

The error equation of velocity is:

$$\dot{\delta\mathbf{V}}^n = \mathbf{f}^n \times \boldsymbol{\phi} - (2\delta\boldsymbol{\omega}_{ie}^n + \delta\boldsymbol{\omega}_{en}^n) \times \mathbf{V}^n - (2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n) \times \delta\mathbf{V}^n + \mathbf{C}_b^n \delta\mathbf{f}^b + \delta\mathbf{g}^n \quad (5)$$

### 3.3. Model of Vehicle Velocity Constraint

The velocity of the vehicle in the vehicle coordinate system is [12]:

$$\mathbf{V}^b = \mathbf{C}_n^b \mathbf{V}^n \quad (6)$$

The differential of the velocity is:

$$\delta\mathbf{V}^b = \mathbf{C}_n^b \delta\mathbf{V}^n + \delta\mathbf{C}_n^b \mathbf{V}^n \quad (7)$$

In Eq. (7),  $\delta\mathbf{C}_n^b$  is the deviation of the calculated attitude matrix and the real attitude matrix. It is assumed that the misalignment angle of the mathematical calculation platform is  $\boldsymbol{\phi} = [\varphi_N \quad \varphi_U \quad \varphi_E]^T$ . Then we can get:

$$\delta\mathbf{C}_n^b = [\boldsymbol{\phi} \times] \mathbf{C}_n^b \quad (8)$$

Where:

$$[\boldsymbol{\phi} \times] = \begin{bmatrix} 0 & -\varphi_E & \varphi_U \\ \varphi_E & 0 & -\varphi_N \\ -\varphi_U & \varphi_N & 0 \end{bmatrix} \quad (9)$$

Take Eqs. (9) and (10) into Eq. (8):

$$\delta\mathbf{V}^b = \mathbf{C}_n^b \delta\mathbf{V}^n + [\boldsymbol{\phi} \times] \mathbf{V}^b = \mathbf{C}_n^b \delta\mathbf{V}^n - [\mathbf{V}^b \times] \boldsymbol{\phi} \quad (10)$$

If the slide and bump during vehicle driving can be ignored, we can get:

$$V_y^b = 0, \quad V_z^b = 0 \quad (11)$$

$$\delta\mathbf{V}^b = \mathbf{C}_n^b \begin{bmatrix} \delta V_N^n \\ \delta V_U^n \\ \delta V_E^n \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & V_x^b \\ 0 & -V_x^b & 0 \end{bmatrix} \begin{bmatrix} \varphi_N \\ \varphi_U \\ \varphi_E \end{bmatrix} \quad (12)$$

## 4. Navigation Algorithm Design

Aiming at the changable vehicle driving environment, a vehicle navigation algorithm with dual modes is proposed. When satellite signal is valid, the algorithm works at GNSS/SINS integrated navigation mode (Mode 1) which can provide the precise position, attitude and velocity information of the vehicle. While the GNSS is unable to provide navigation information, the algorithm switches to SINS mode with the constraint of vehicle velocity (Mode 2). In Mode 2, the algorithm assumes that the driving vehicle does not slide and bump. Although this assumption affects the estimation of some navigation parameters, it greatly improves the position accuracy of the vehicle on the situation of satellite short-term failure. The main difference between Mode 1 and Mode 2 is the measurement equation of Kalman filter which is shown in figure 3.

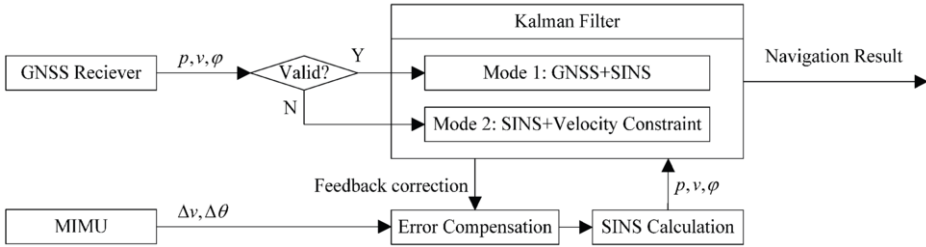


Figure 3. Diagram of vehicle navigation algorithm with dual modes

#### 4.1. Design of State Equation

The system state vector of Kalman filter is selected as: attitude error, velocity error, position error of vehicle, bias of gyro and accelerometer.

$$\mathbf{X} = [\varphi \quad \delta \mathbf{V} \quad \delta \mathbf{P} \quad \boldsymbol{\varepsilon}_b \quad \boldsymbol{\varepsilon}_r \quad \boldsymbol{\xi}]^T \quad (13)$$

According to chapter 3.1 and 3.2, we can acquire the state equation of the navigation system:

$$\dot{\mathbf{X}} = \mathbf{F}\mathbf{X} + \mathbf{G}\mathbf{W} \quad (14)$$

Where:

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{F}_{13} & -\mathbf{C}_b^n & -\mathbf{C}_b^n & \mathbf{0}_{3 \times 3} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{23} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{C}_b^n \\ \mathbf{F}_{31} & \mathbf{F}_{32} & \mathbf{F}_{33} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ & & & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ & \mathbf{0}_{9 \times 9} & & \mathbf{0}_{3 \times 3} & -\boldsymbol{\beta} & \mathbf{0}_{3 \times 3} \\ & & & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & -\boldsymbol{\mu} \end{bmatrix} \quad (15)$$

$$\mathbf{F}_{11} = \begin{bmatrix} 0 & -\frac{V_N}{R_N + h} & -\omega_{ie} \sin L - \frac{V_E}{R_E + h} \tan L \\ \frac{V_N}{R_N + h} & 0 & \omega_{ie} \cos L + \frac{V_E}{R_E + h} \\ \omega_{ie} \sin L + \frac{V_E}{R_E + h} \tan L & -\omega_{ie} \cos L - \frac{V_E}{R_E + h} & 0 \end{bmatrix} \quad (16)$$

$$\mathbf{F}_{21} = \begin{bmatrix} 0 & -f_E & f_U \\ f_E & 0 & -f_N \\ -f_U & f_N & 0 \end{bmatrix}, \mathbf{F}_{31} = \mathbf{0}_{3 \times 3} \quad (17)$$

$$\mathbf{F}_{12} = \begin{bmatrix} 0 & 0 & \frac{1}{R_E + h} \\ 0 & 0 & \frac{\tan L}{R_E + h} \\ -\frac{1}{R_N + h} & 0 & 0 \end{bmatrix}, \mathbf{F}_{32} = \begin{bmatrix} \frac{1}{R_N + h} & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & \frac{1}{(R_E + h) \cos L} \end{bmatrix} \quad (18)$$

$$\mathbf{F}_{22} = \begin{bmatrix} -\frac{V_U}{R_N+h} & -\frac{V_N}{R_N+h} & -2(\omega_{ie} \sin L + \frac{V_E}{R_E+h} \tan L) \\ \frac{2V_N}{R_N+h} & 0 & 2(\omega_{ie} \cos L + \frac{V_E}{R_E+h}) \\ 2\omega_{ie} \sin L + \frac{V_E}{R_E+h} \tan L & -2\omega_{ie} \cos L - \frac{V_E}{R_E+h} & \frac{V_N \tan L - V_U}{R_E+h} \end{bmatrix} \quad (19)$$

$$\mathbf{F}_{13} = \begin{bmatrix} -\omega_{ie} \sin L & \frac{V_E}{(R_E+h)^2} & 0 \\ \omega_{ie} \cos L + \frac{V_E}{(R_E+h) \cos^2 L} & \frac{V_E \tan L}{(R_E+h)^2} & 0 \\ 0 & -\frac{V_N}{(R_N+h)^2} & 0 \end{bmatrix} \quad (20)$$

$$\mathbf{F}_{23} = \begin{bmatrix} -V_E(2\omega_{ie} \cos L + \frac{V_E}{(R_E+h) \cos^2 L}) & -\frac{V_E^2 \tan L}{(R_E+h)^2} - \frac{V_N V_U}{(R_N+h)^2} & 0 \\ -2\omega_{ie} V_E \sin L & \frac{V_N^2}{(R_N+h)^2} + \frac{V_E^2}{(R_E+h)^2} & 0 \\ 2\omega_{ie}(V_N \cos L + V_U \sin L) + \frac{V_N V_E}{(R_E+h) \cos^2 L} & \frac{V_E}{(R_E+h)^2}(V_N \tan L - V_U) & 0 \end{bmatrix} \quad (21)$$

$$\mathbf{F}_{33} = \begin{bmatrix} 0 & \frac{V_N}{(R_N+h)^2} & 0 \\ 0 & 0 & 0 \\ \frac{V_E \tan L}{(R_E+h) \cos L} & \frac{V_E}{(R_E+h)^2 \cos L} & 0 \end{bmatrix} \quad (22)$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0}_{12 \times 3} & \mathbf{0}_{12 \times 3} \\ \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} \end{bmatrix} \quad (23)$$

$$\mathbf{W} = [w_{gx} \quad w_{gy} \quad w_{gz} \quad w_{ax} \quad w_{az}]^T \quad (24)$$

#### 4.2. Design of Measurement Equation

When the algorithm works at Mode 1, the measurement vector is the difference between SINS and GNSS in the aspects of yaw angle, velocity and position:

$$\begin{aligned} \mathbf{Z}_1 &= [\varphi_{yaw}^{SINS} - \varphi_{yaw}^{GNSS} \quad \mathbf{V}^{SINS} - \mathbf{V}^{GNSS} \quad \mathbf{P}^{SINS} - \mathbf{P}^{GNSS}]^T \\ &= \mathbf{H}_1 \mathbf{X} + \mathbf{V}_1 \end{aligned} \quad (25)$$

$$\mathbf{H}_1 = \begin{bmatrix} 0 & 1 & 0 & \mathbf{0}_{1 \times 15} \\ \mathbf{0}_{6 \times 3} & \mathbf{I}_{6 \times 6} & \mathbf{0}_{6 \times 9} \end{bmatrix} \quad (26)$$



When the algorithm works at Mode 2, the measurement vector is the change of vehicle's upward velocity and lateral velocity:

$$\begin{aligned} \mathbf{Z}_2 &= [\delta V_y^b \quad \delta V_z^b]^T \\ &= \mathbf{H}_2 \mathbf{X} + \mathbf{V}_2 \end{aligned} \tag{27}$$

$$\mathbf{H}_2 = \begin{bmatrix} 0 & 0 & V_x^b & C_{21} & C_{22} & C_{23} & \mathbf{0}_{1 \times 12} \\ 0 & -V_x^b & 0 & C_{31} & C_{32} & C_{33} & \mathbf{0}_{1 \times 12} \end{bmatrix} \tag{28}$$

Where  $C_{ij}$  is the corresponding elements of  $\mathbf{C}_n^b$ ,  $\mathbf{V}_1$  and  $\mathbf{V}_2$  are white noise.

### 5. Test Results and Analysis

Many tests about the vehicle navigation system designed in this paper have been carried out in different areas including campus, urban area, suburbs and so on. Through the process and analysis of data, we compare the result of designed navigation system costing hundreds RMB with high-precision navigation system costing millions RMB. Taking driving areas of urban and tunnel as example, the performance of navigation system is specified as follows.

Test vehicle and equipment are shown in Figure 4. The test vehicle is equipped with two sets of GNSS receiver: one is the low-cost STA8090FG whose position accuracy is better than 1.8 meters (RMS), the other is the Novatel's high-precision OEM4 star enhanced system with the position accuracy better than 0.2 meters (RMS), which is regarded as reference of position and velocity. The test vehicle is also equipped with two sets of IMU: one is the low-precision MIMU whose specifications show in Table 1, the other is high-precision laser gyro inertial navigation system with the accuracy of horizontal attitude better than 0.05 degrees and azimuth attitude better than 0.15 degrees, which is regarded as reference of attitude.



**Figure 4.** Test vehicle and equipment

The navigation results of test vehicle when it was driving in the urban area and the navigation system was working at Mode 1 are shown from figure 5 to figure 8. We can find that the navigation results of the designed navigation system are basically consistent with the reference one so as to meet the navigation precision requirements of

common vehicles. In addition, it can be seen from Table 2 to Table 4 that the dynamic accuracy of horizontal position is better than 1.5 meters, the dynamic accuracy of horizontal attitude is better than 0.2 degrees, the dynamic accuracy of azimuth attitude is better than 1.5 degrees and the dynamic accuracy of velocity is better than 0.15 m/s.

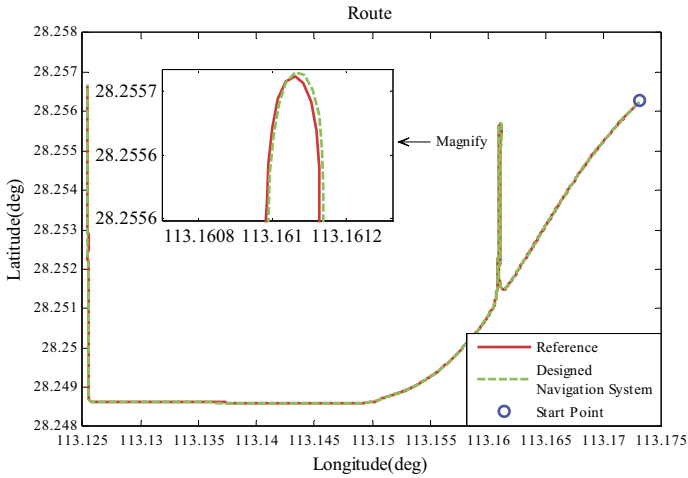


Figure 5. The route of test vehicle in urban area

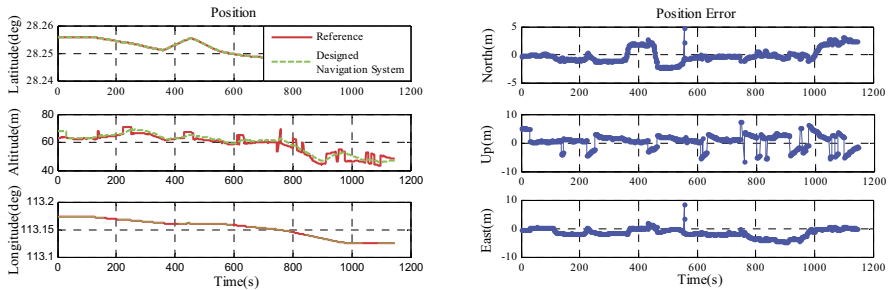


Figure 6. The position and position error of test vehicle in urban area

Table 2. Statistics of position error in urban area

	Mean(m)	Standard Deviation(m)	Maximum(m)
<b>Northward Position Error</b>	-0.07	1.24	4.67
<b>Upward Position Error</b>	0.75	2.34	7.29
<b>Eastward Position Error</b>	-1.43	1.45	8.43

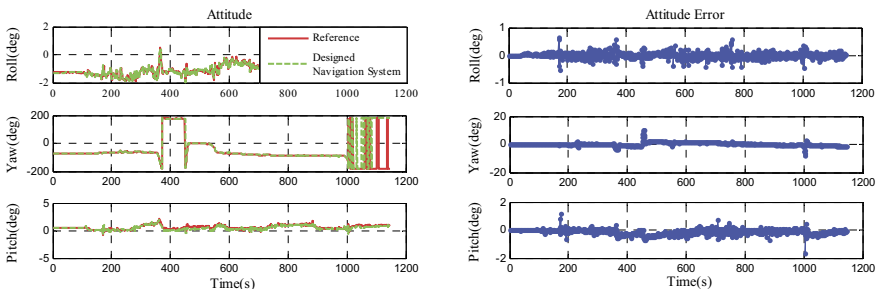
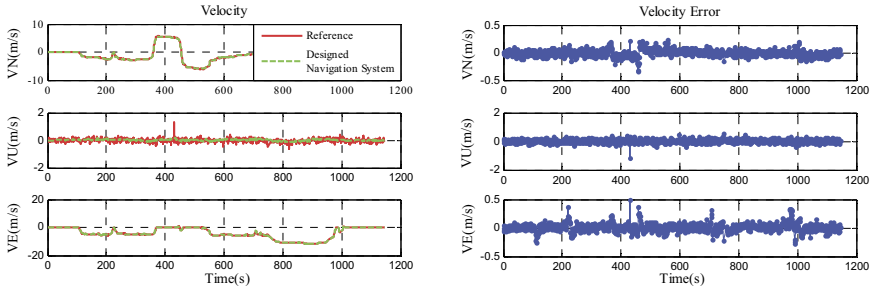


Figure 7. The attitude and attitude error of test vehicle in urban area

**Table 3.** Statistics of attitude error in urban area

	Mean(degree)	Standard Deviation(degree)	Maximum(degree)
<b>Roll Angle Error</b>	0.0007	0.10	0.63
<b>Yaw Angle Error</b>	0.1901	1.23	10.44
<b>Pitch Angle Error</b>	-0.0936	0.18	1.14

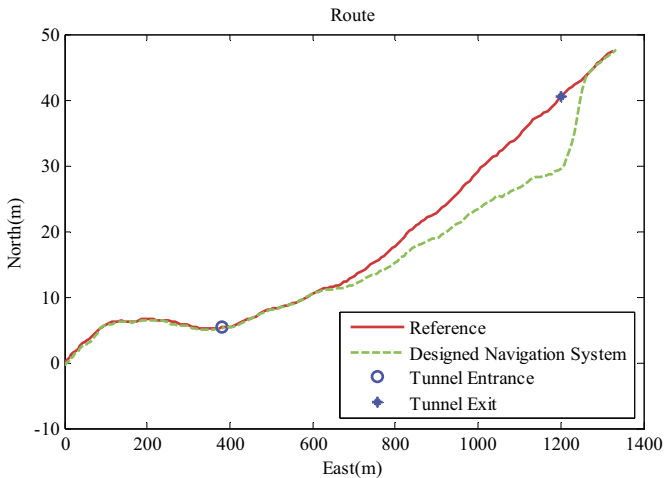


**Figure 8.** The velocity and velocity error of test vehicle in urban area

**Table 4.** Statistics of velocity error in urban area

	Mean(m/s)	Standard Deviation(m/s)	Maximum(m/s)
<b>Northward Velocity Error</b>	-0.0039	0.05	0.22
<b>Upward Velocity Error</b>	0.0184	0.11	0.52
<b>Eastward Velocity Error</b>	0.0128	0.06	0.48

Figure 9 illustrates the route of the test vehicle passing through a tunnel about 850 meters. When the vehicle entered the tunnel, the navigation system switched to Mode 2, which effectively suppressed the rapid offset of the position. Within 120 seconds of the GNSS blocked in tunnel, the maximum position deviation of the vehicle is 10.78 meters. After the test vehicle left the tunnel, the navigation system switched back to Mode 1 and the position was corrected quickly. It can be seen that the designed navigation system with dual modes can adapt to the changeable driving environment.



**Figure 9.** The route of test vehicle in tunnel

## 6. Conclusion

Aiming at the problems such as expensive equipment and unstable satellite signals in the practical application of vehicle navigation technology, the research on vehicle navigation system based on low-cost sensors is conducted in this paper. Specifically, a hardware platform of low-cost vehicle navigation system is designed and a vehicle navigation algorithm with dual modes is proposed. The test results show that the designed navigation system can meet the navigation precision requirements of common vehicle and can adapt to the changeable driving environment. At the same time, the system is easy-to-implement and cost-effective.

It can be found from many tests that the accuracy of the yaw angle is generally poor which has a great impact on the navigation accuracy. Therefore, the further research will be settled in how to use low-cost MEMS magnetometer or other devices to provide heading assistance and how to utilize the satellite signals to estimate and correct the error of these devices in real time to improve the accuracy of vehicle navigation.

## References

- [1] Leccadito, Matthew, et al, A Kalman Filter Based Attitude Heading Reference System Using a Low Cost Inertial Measurement Unit, *ALAA Guidance, Navigation, and Control Conference* 2017.
- [2] Valenti, Roberto G., I. Dryanovski, and J. Xiao, Keeping a Good Attitude: A Quaternion-Based Orientation Filter for IMUs and MARGs, *Sensors* 15.8(2015), 19302-19330.
- [3] Ning, Yunkun, et al, A quaternion-based Attitude Estimate System Based on a Low Power Consumption Inertial Measurement Unit, *Eai International Conference on Wireless Mobile Communication and Healthcare ICST*, 2015, 113-119.
- [4] Sasani, S, J. Asgari, and A. R. Amiri-Simkooei, Improving MEMS-IMU/GPS integrated systems for land vehicle navigation applications, *GPS Solutions* 20.1(2016), 89-100.
- [5] Finally, A New Technique for Integrating MEMS-Based Low-Cost IMU and GPS in Vehicular Navigation, *Journal of Sensors* 2016.4(2016), 1-16.
- [6] Navidi, N, and R. J. Landry, A New Survey on Self-Tuning Integrated Low-Cost GPS/INS Vehicle Navigation System in Harsh Environment, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W4.58(2015)*, 75-81.
- [7] Li, Xu, and Q. Xu, A Reliable Fusion Positioning Strategy for Land Vehicles in GPS-denied Environments Based on Low-cost Sensors, *IEEE Transactions on Industrial Electronics* PP.99(2016), 1-1.
- [8] Navidi, Néda, and R. J. Landry, Low-Cost GPS/INS Integrated Land-Vehicular Navigation System for Harsh Environments Using Hybrid Mamdani AFIS/KF Model, *Journal of Traffic & Transportation Engineering* 4.1(2016).
- [9] Gama, Jose, V. Maximov, and S. O. H. Madgwick, Data Fusion Filters for Attitude Heading Reference System (AHRS) with Several Variants of the Kalman Filter and the Mahoney and Madgwick Filters, (2015).
- [10] Sheng, Hanlin, and T. Zhang, MEMS-based low-cost strap-down AHRS research, *Measurement* 59(2015), 63-72.
- [11] Groves, Paul D, Principles of GNSS, inertial, and multisensor integrated navigation systems, 2nd edition [Book review], *IEEE Aerospace & Electronic Systems Magazine* 30.2(2015), 26-27.
- [12] Maklouf, O. M, et al, GPS Aiding Low Cost Inertial Navigation System for Land Vehicle Navigation Application, *Abstract of Applied Sciences & Engineering* 12(2016).

# Fault Diagnosis Method for High-Speed Train Lateral Damper Based on Variational Mode Decomposition and Multiscale Entropy

Changxi Li<sup>a1</sup>, Xiantai Gou<sup>a</sup>, Xiu Li<sup>b</sup> and Weidong Jin<sup>a</sup>

<sup>a</sup>*School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China*

<sup>b</sup>*Chengdu College of University of Electronic Science and Technology, Chengdu, China*

**Abstract.** In the fault period of high-speed train lateral damper, the vibration signal is non-linear and nonstationary, and features extracting is relatively difficult. In order to save this problem, a method of features extracting based on variational mode decomposition and multiscale entropy was proposed. The original signal was decomposed into several intrinsic mode function components after being processed by the variational mode decomposition method. Then, the best component was selected by the mutual information index. The feature matrix was constructed through the multiscale entropy of the best component, and removed redundant features using feature evaluation algorithm. The fault type of lateral damper was judged by transforming in the best subset of feature matrix in support vector machine. Experimental results show that the proposed method can extract the feature and judge the fault type of lateral damper effectively, which proves the feasibility of this mechanical fault diagnosis method.

**Keywords.** variational mode decomposition, multiscale entropy, high-speed train lateral damper, fault diagnosis

## 1. Introduction

High-speed trains belong to typical non-linear complex systems, and the vibration signals produced by train operation are obvious nonlinearity and non-stationary. Therefore, it is difficult to extract the effective fault signature by using the traditional signal processing method [1]. In order to save this problem, a number of scholar have studied. Qin et al. [2] put forward a feature extraction method based on ensemble empirical mode decomposition (EEMD) and permutation entropy, and it can solve mode mixing and end effect effectively. Jin et al. [3] proposed a feature extraction method based on Copula function. Although EEMD is an improved algorithm based on EMD and overcome mode mixing effectively, it is Inefficient. Variational mode decomposition (VMD), first proposed by Dragomiretskiy and Zosso, is newly adaptive

---

<sup>1</sup> Corresponding Author.

Tel: +86 15528127718

E-mail:changxili429@163.com

and quasi-orthogonal signal decomposition method, and decompose signal into a set of band-limited intrinsic modal functions (IMF) by recursively solving the variational problem. VMD can separate the frequency of component and overcome the mode mixing and the lack of mathematical theory, and the algorithm is efficient highly [4].

The initial approximate entropy (ApEn) [5] and sample entropy (SpEn) [6] are the methods to measure the complexity of time series and the statistical quantization. Because ApEn compares the data with its own and measures the new information generation rate, it is meaningless. Although SpEn has the characteristics of strong anti-noise and anti-interference ability and good consistency within the range of parameters, it measures single-scale complexity of the time series and cannot characterize multiscale complexity [7]. Aiming at the problem, Costa et al. [8] put forward a method based on multiscale entropy (MSE) that is multiscale SpEn of time series. MSE measures the complexity of signal as a whole, but also extracts multiscale detail feature. Therefore MSE can recognize different signals from both qualitative and quantitative perspectives effectively and has obvious advantages comparing with SpEn and ApEn.

In order to save the problems that vibration signal is non-linear and nonstationary, features extracting is relatively difficult in the fault period of high-speed train lateral damper and the effective fault signature are mainly low frequency [9]. In this paper, we combine with VMD and MSE (VMD-MSE) to extract feature of vibration signal of high-speed trains' lateral damper. The results show that VMD-MSE can extract the effective fault signature and accurately judge fault type.

## 2. Variational Mode Decomposition

As a recently developed method for adaptive signal process, VMD can non-recursively decomposes a multi-component signal into defined number of band-limited intrinsic mode function components. After performing VMD on a given real valued signal  $f$ , the IMF component  $u_k(t)$  with the specific sparsity property of its bandwidth in the spectral domain could be acquired. Each component is compact around a central frequency  $\omega(t)$  and its bandwidth is estimated using  $H^1$  Gaussian smoothness of the shifted signal. The VMD algorithm is written as a constrained variational problem

$$\left\{ \begin{array}{l} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 \right\} \\ \text{subject to } \sum_k u_k = f \end{array} \right. \quad (1)$$

Where  $\{u_k\} := \{u_1, \dots, u_K\}$ ,  $\{\omega_k\} := \{\omega_1, \dots, \omega_K\}$  and  $K$  denotes the number of the IMF component.

The above equation can be addressed by introducing a quadratic penalty and Lagrangian multipliers. The augmented Lagrangian is described as follows

$$L(\{u_k\}, \{\omega_k\}, \lambda) := \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \tag{2}$$

Eq. (2) is solved with the alternate direction method of multipliers (ADMM). All the IMF components gained from solutions in spectral domain are described as Eq. (3)

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3}$$

Where  $\omega(t)$  is computed at the center of gravity of the corresponding component's power spectrum. And the update strategy of  $\omega(t)$  is based on the Eq. (4)

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{4}$$

Perform the inverse Fourier transform on  $\hat{u}_k^n(\omega)$ , then the corresponding IMF component could be obtained. The complete algorithm of VMD in detail can be found in literature [4].

On the while, the VMD finally can realize adaptive decomposition according to the frequency-band analysis of the original signals in the process where IMFs constantly update their central frequencies and bandwidths. The detailed steps are shown in Figure 1.

### 3. Fault Diagnosis and Analysis of High-speed Train Lateral Damper

#### 3.1 Data Source

The simulation data is achieved through the nonlinear dynamic model of EMU vehicle system that is established with multi-body dynamics analysis software (SIMPACK). L-MA wheel tread and CN60 rail are applied in dynamic model, and inner moment of wheel chooses Chinese standard. Nonlinearity of wheel-rail contact, wheel-rail creep and suspension are taken into consideration fully. The vehicle model is constituted with one body, two bogie frames, four wheelsets, eight tumblers and two traction rods. The track excitation spectrum, which was measured in Wuhan-Guangzhou passenger line, is adopted during simulation experiment. The position of the lateral damper in the bogie is shown in Figure 2.

From the Fig. 2, there are four lateral dampers of the whole vehicle. In this paper, seven working conditions are investigated for four lateral dampers including normal, left lateral damper of first bogie fault (B1LF), right lateral damper of first bogie fault (B1RF), left lateral damper of second bogie fault (B2LF), right lateral damper of first bogie fault (B2RF), left lateral damper of first bogie and right lateral damper of second bogie fault (B1LF+ B2RF) as well as all lateral dampers of first bogie fault (B1LF+B1RF).

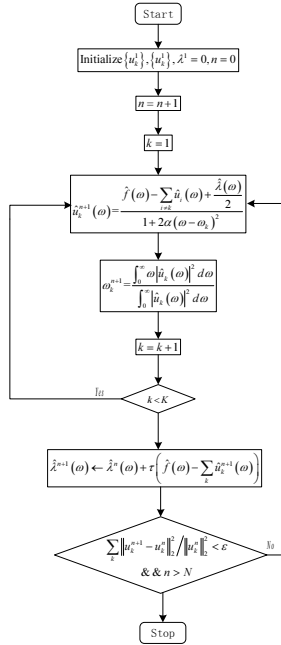


Figure 1. Flowchart of VMD

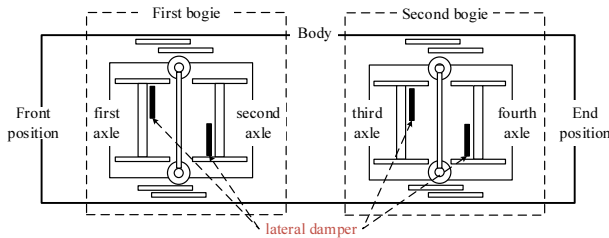


Figure 2. The position of lateral dampers in the bogie

The simulation data collected by 58 sensors installed on different positions of high-speed train bogie, including the lateral displacement, vertical displacement, vertical displacement, lateral acceleration, vertical acceleration and longitudinal acceleration. The running speed of vehicle is 200km/h, the sampling frequency is 243Hz and the total sampling time is 3.5 minutes. Figure 3 displays the waveforms and frequency spectrums of the time domain signal that is from eighth channel sensor. We can find that there are obvious difference about the waveforms and frequency spectrums of the vibration signal between high-speed train fault condition and normal, and it is difficult to distinguish the different conditions through using the time domain waveforms and frequency spectrums. Therefore, it is very essential to perform an



effective method to identify the different working conditions of lateral dampers accurately. The major challenge for condition identification is to acquire the reliable and sensitive features from the vibration signals. And the proposed feature extraction method based on VMD-MSE is applied.

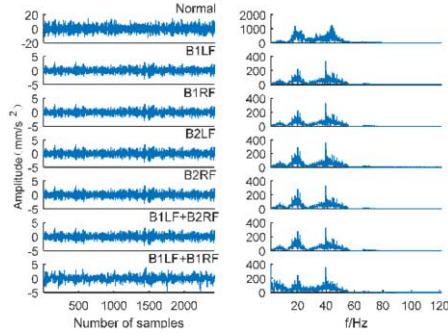


Figure 3. Time domain waveforms and frequency spectrums of seven working conditions

### 3.2 Processing Vibration Signal by VMD

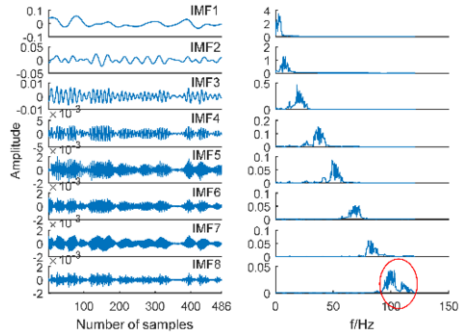
In the VMD algorithm, the number of the IMF component  $K$  need to be selected in advance. If this parameter is set too large, there will be exist excessive decomposition. Contrarily, if this parameter is set too small, the feature information contained in the original sample could not be extracted effectively. In this paper, we set this parameter though observing the frequency spectrum of the IMF component. Using VMD decomposes vibration signal collected by first channel sensor that contains 486 data points. The results processed by VMD under different  $K$  shown in Figure 4 (a) to (c), where we set the balancing parameter of the data-fidelity constraint  $\alpha$  to 30 and time-step of the dual ascent  $\tau$  to zero.

From the Figure 4 (a) to (c), we can find that the eighth IMF component includes multiscale information meaning that decomposition is not sufficient when we set  $K = 8$ , and the frequency spectrum of the seventh IMF component and the eighth IMF component is coincident when we set  $K = 10$ . Therefore, we set  $K = 9$ . In order to verify the advantage of proposed method the same vibration signal in Figure 4 is decomposed by EMD and EEMD, and the results are shown in Figure 5 (a)-(b). AS can be seen, there is a mode mixing problem of IMFs processed by EMD. Although the third to seventh IMF processed by EEMD are normal, the first and second IMF obtaining relatively high frequency information are mode mixing.

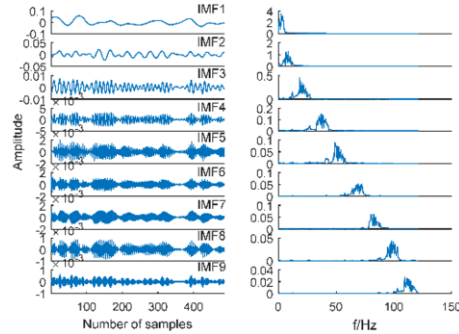
Ideally, each IMF component is a simple and stationary signal that represents one of the characteristic components of the original signal. Due to factors such as parameter selection, there is inevitably a false component in the decomposition result. Hu et al. [11] studied the HHT false component identification based on mutual information. The results show that the mutual information method is more accurate than the correlation coefficient method and can better identify the false component. Considering the number of IMFs decomposed by each algorithm, this paper adopts the mutual information method to select the top four effective IMFs decomposed by VMD and EEMD algorithm and the top three effective IMFs decomposed by EMD algorithm.

### 3.3 Multiscale Entropy Feature Extraction

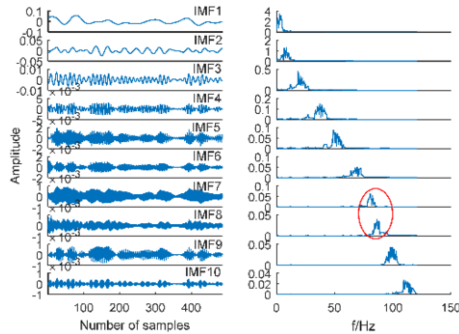
In the calculation of multi-scale entropy, the choice of parameter value has an effect on the calculation of entropy. The similarity within tolerance  $r$  is generally related to the standard deviation of the original sequence. Generally, similarity within tolerance  $r = (0.1 - 0.25)SD$ , where  $SD$  represents the standard deviation. Therefore, the similarity within tolerance reflects the complexity of the original sequence on different scales. In addition, the study shows that the dependence between the accuracy of calculation results and the length of sequence is minimal when the embedded dimension  $m$  is set to two, and the scale factor  $\tau$  is set  $\tau_{max}$ .



(a) K=8

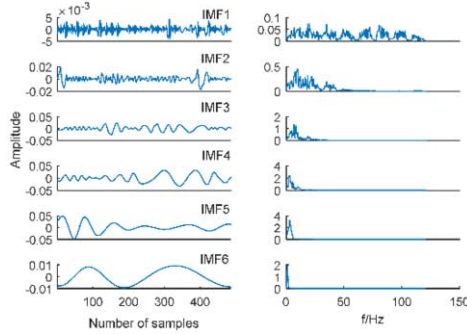


(b) K=9

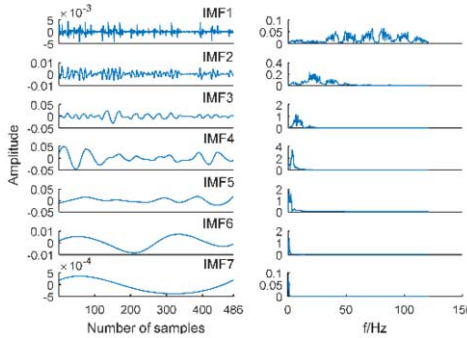


(c) K=10

Figure 4. Processing results by VMD under different K values



(a) Processing results by EMD



(b) Processing results by EEMD

**Figure 5.** Processing results of vibration signal by EMD and EEMD

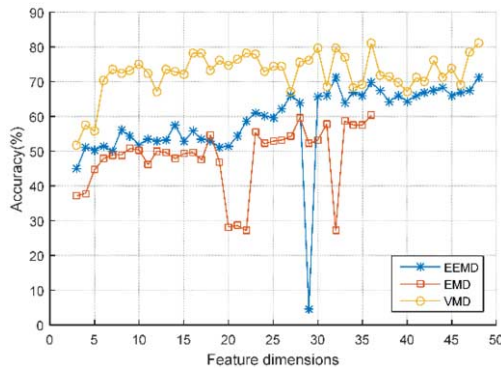
In the experiment, we intercept 486 data points as one sample and remove the abnormal point. Finally, each working condition has 90 samples. Since multiple degrees of freedom can represent the movement state of the train and the multi-channel data has a strong correlation, we use the serial way to data collected by seventh to 18th channel sensors for data fusion when extract feature. The fusion mode is shown in Equation (5).

$$F(x_i) = [f_1(c_7), f_2(c_7), \dots, f_n(c_7), f_1(c_j), f_2(c_j), \dots, f_n(c_j), \dots, f_1(c_{18}), f_2(c_{18}), \dots, f_n(c_{18})], 1 \leq i \leq 90, 7 \leq j \leq 18 \quad (5)$$

Where  $x_i$  is sample  $i$ ,  $c_j$  represents data of channel  $j$ ,  $n$  is the number of effective IMF,  $f_n(c_j) = MSE_{IMF_n}$  represents multiscale entropy of  $x_i$  the effective IMF component  $n$  of  $c_j$ , and  $F(x_i)$  represents multiscale entropy of. Therefore, the size of feature matrix based on VMD and EEMD algorithm is  $90 \times (12 \times 4)$  and the size of feature matrix based on EMD algorithm is  $90 \times (12 \times 3)$  in this paper.

### 3.4 Feature Selection and Fault Identification

We use the feature evaluation method proposed by Du et al. [10] to the feature matrix, and the output of feature evaluation method is a feature ranked list. For the feature ranked list, each time one-dimensional redundancy feature is removed and a number of nested feature subsets  $F_1 \subset F_2 \subset \dots \subset F$  are obtained to train the SVM. Then we evaluate the merits of the feature subset with the predictive accuracy of SVM to obtain the optimal feature subset. In the experiment, the 630 samples of seven working conditions are normalized and 50 samples are randomly selected as the training samples and 40 samples of each working condition were used as the test samples. When using SVM fault identification, the input feature dimension is not less than 3, and Figure 6 is the classification accuracy of different feature subsets of different algorithms.



**Figure 6.** Classification accuracy of different feature subsets

It can be observed from Fig.6 that the classification accuracy of the feature matrix based on VMD-MSE algorithm is greater than the classification accuracy of the feature matrix based on EMD-MSE and EEMD-MSE algorithms, and the classification accuracy curve tends to be stable when the feature subset dimension is more than 6. However, the classification accuracy curve of the feature matrix based on EMD-MSE and EEMD-MSE algorithm has big fluctuation, which proves the superiority of VMD-MSE algorithm in feature extraction. In addition, the optimal feature subset dimension of EMD-MSE algorithm is 28, the optimal feature subset dimension of EEMD-MSE algorithm is 32 and the optimal feature subset dimension of VMD-MSE algorithm is 36.

We use the optimal feature subset to recognize the fault type of lateral damper, Table1 (a) to (b) are the confusion matrix of optimal feature subset of each algorithm. The confusion matrix can describe the relationship between the true properties of the fault samples and the recognition results and is a common method for evaluating the classification effect.

As can be seen, aiming at the feature extraction of the vibration signal of the high-speed train, the classification accuracy based on VMD-MSE algorithm is more than that based on EMD-MSE and EEMD-MSE algorithm. The classification accuracies based on EEMD-MSE and VMD-MSE algorithm of normal working condition and multiple lateral damper fault is 100%, which proves the feasibility of these two feature extraction methods for fault diagnosis identification of multiple lateral damper fault. However, the classification accuracy of signal lateral damper fault is low. The reason is

that the positions of two lateral dampers on the same bogie are relatively close to each other, and the effect of anyone fault on the vibration signal is not so great that it is difficult to recognize the specific position of the two lateral dampers on the same bogie. If the two lateral dampers on the same bogie are classified as a class, the fault recognition accuracy greatly is improved. This also indirectly shows that the impact of multiple lateral damper fault on the body is greater than the impact of a single lateral dampers fault.

**Table 1.** Confusion matrices of optimal feature subset of different algorithms

(a) Classification results of feature extraction method based on EMD-MSE

		Predicted class						
		Working condition	Normal	B1L	B1R	B2L	B2R	B1L+B2R
Actual class	Normal	39	1	0	0	0	0	0
	B1L	0	21	8	2	4	4	1
	B1R	0	20	8	2	6	3	1
	B2L	1	4	0	12	19	3	1
	B2R	0	4	0	3	28	4	1
	B1L+B2R0	9	0	1	7	23	0	0
	B1L+B1R0	0	0	0	2	0	0	38

(b) Classification results of feature extraction method based on EEMD-MSE

		Predicted class						
		Working condition	Normal	B1L	B1R	B2L	B2R	B1L+B2R
Actual class	Normal	40	0	0	0	0	0	0
	B1L	0	28	9	0	2	1	0
	B1R	0	21	16	1	0	2	0
	B2L	0	0	0	20	20	0	0
	B2R	0	1	1	16	22	0	0
	B1L+B2R	0	1	2	2	2	33	0
	B1L+B1R	0	0	0	0	0	0	40

(c) Classification results of feature extraction method based on VMD-MSE

		Predicted class						
		Working condition	Normal	B1L	B1R	B2L	B2R	B1L+B2R
Actual class	Normal	40	0	0	0	0	0	0
	B1L	0	29	8	1	2	0	0
	B1R	0	6	33	0	1	0	0
	B2L	0	1	1	28	7	3	0
	B2R	0	1	0	17	20	2	0
	B1L+B2R	0	0	0	2	1	37	0
	B1L+B1R	0	0	0	0	0	0	40

#### 4. Conclusion

EMD and EEMD are essentially binary filter groups. The frequency band center and bandwidth of the signal are uncertain during the process of decomposition. Moreover, the recursive pattern decomposition adopted by EMD and EEMD will continue to propagate the envelope estimation error and the vibration signal contains noise or intermittent signal, which leads to the problem of the mode mixing. VMD is a non-recursive decomposition model, which determines the center frequency and frequency band of each modal component through iteratively searching the optimal solution of

variational model, and can adaptively realize the frequency domain segmentation of variational signals and the effective separation of each component. In this paper, we apply the feature extraction method based on VMD-MSE to the fault diagnosis of the high-speed train lateral damper. The experimental results show that the proposed feature matrix can effectively recognize the fault type.

## References

- [1] Xiaoliang W, Li L, Weihua Z, Fuzzy self-tuning based washout algorithm for simulator of high-speed locomotive vertical vibration, *Vibration, Measurement & Diagnosis* 29(1) (2009), 101-104+121-122.
- [2] Na Q, Peng J, Yongkui S, Fault diagnosis of high-speed train bogie based on EEMD and permutation entropy, *Vibration, Measurement & Diagnosis* 35(05) (2015), 885-891+991.
- [3] Weidong J, Qiangyong L, Yongkui S, Extracting fault features of high-speed train bogies using copula function, *Southwest Jiao Tong University* 50(04) (2015), 676-682.
- [4] Dragomiretskity K, Zosso D, Variation-al mode decomposition, *IEEE Trans Signal Processing* 62(3) (2014), 531-544.
- [5] Pincus S M, Approximate entropy as a mea-sure of system complexity, *Proceedings of the National Academy of Sciences of the United States of America* 88(6) (1991), 2297-2301.
- [6] Richman J S, Moorman J R, Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of Physiology Heart & Circulatory Physiology* 278(6):H2039-H2049.
- [7] Jinde Z, Junsheng C, Siyu H. (2013) Rotor fault diagnosis based on multiscale entropy, *Vibration, Measurement & Diagnosis* 33(02) (2000), 294-297+342.
- [8] Costa M, Goldberger AL, Peng CK, Multiscale entropy to distinguish physiologic and synthetic RR time series, *29th Annual Conference on Computers in Cardiology* 29(2002), 137-140.
- [9] Zhimin L, Xiantai G, Qin N, Vibration monitoring signals frequency feature of high-speed train, *Instrument Technique and Sensor* (05) (2015), 99-103.
- [10] Jing D, Zhenzhen C, Peng J, A new feature evaluation algorithm and its application in fault of high-speed railway, *Computer Technology and Its Applications* 41(09) (2015), 153-156.
- [11] Aijun H, Research on the application of Hilbert-Huang transform in vibration signal analysis of rotating machinery, Baoding, North China Electric Power University (2008).

# The Influence of Using XBRL on Quoted Company's Information Quality

Yu-chen YANG<sup>a</sup>, Yan-li CHEN<sup>b,1</sup>, Lei WANG<sup>b</sup> and Chen TANG<sup>b</sup>

<sup>a</sup>*College of Internet of Things Engineering, Hohai University,  
Changzhou 213022, China*

<sup>b</sup>*College of Business Administration, Hohai University, Changzhou 213022, China*

**Abstract.** Due to the Extensible Business Reporting Language (XBRL) have high compatibility and other features, it is widely used in finance field, as an information disclosure and retrieval tool in the Shanghai and Shenzhen Stock Exchange at present. This article mainly uses normative research and empirical research to discuss the influence of XBRL on quoted accounting information quality. Firstly, this article theoretically discusses the positive effect on the relevance and reliability of quoted company's accounting information, which based on XBRL standard. This article brings into the concept of earnings response coefficient in the process of empirical research, and uses event study method to research the reaction degree of market excess return on unexpected factors of company's financial report. This article uses stratified sampling to select annual report disclosure information of 501 quoted companies in the Shanghai stock market from 2012 to 2015 and uses these samples to do empirical research. The results are as follows, unexpected accounting surplus has positive correlation with earnings response coefficient; the persistence of company surplus has positive correlation with earnings response coefficient; the growth rate of company has positive correlation with earnings response coefficient; the risk variation of company has positive correlation with earnings response coefficient. From the western accounting theory perspective, the earnings response coefficient is larger and the higher of accounting information quality. From the empirical research, to some extent, the implementation of XBRL made a positive effect on our country's quoted company's accounting information quality.

**Keywords.** XBRL, accounting information quality, earnings response coefficient

## 1. Introduction

XBRL-Extensible Business Reporting Language is an international extensible business reporting language which bases on XML – Extensible Markup Language. It isn't restricted to special platform to make operation. Using this kind of language to tokenize data can store, deal with, exchange financial and business report data truly, timely and effectively. XBRL specification file includes XBRL Specification, XBRL documents, XBRL Taxonomy and Style sheet.

Enterprise is a basic unit of market and mainly uses publishing financial report to guide investors to get relative accounting information. So the quality of financial report

---

<sup>1</sup> Corresponding author.

directly influences the stakeholders' decision making. And it is not difficult to infer that information users have high demand on accounting information quality.

Accounting main objective is to provide information users accounting information. FASB highlights that accounting information must be provided with two important quality features. They are relativity and reliability.

Only accounting information has the ability of "causing difference", can we confirm its relativity. The main reason is that accounting information should have relationship with the economic predictions of the information users. It should help information users to make predictions on the past, moment and future event, and this may prove the previous expectation and make them have the ability of "causing difference" on decision making. "Causing difference" is to decrease or increase the differences of information, and reducing the information users' uncertainty on the event to increase the quality on decision making.

Relativity and reliability are two most important indexes to satisfy whether the decisions are effective. In general, we always assume that the higher relativity and reliability of the accounting information it will be more likely to satisfy users' demands.

Reliability means that accounting information should truly express the object which they want to reflect. Reliability also means that accounting information can express the company operation activities' process and result impartially, and companies avoid making the unfair disclosure for the specific needs of the specific group. Reliability includes objectivity, verifiability and authenticity. Accounting information reliability reflects moderate thought. Moderatism requires investors to obtain a careful reflect attitude towards the factors which can't be absolutely confirmed. For example, we can't undervalue liability and overstate revenue. "Accountability" held the view that reliability is the first important factor which should be considered in the accounting information quality. If accounting information loses reliability, it will also lose the relativity. We can consider this idea in this view, reliability is the basic guarantee for accounting information quality equipped with other characters.

## **2. Research Assumptions**

Due to FASB highlighted that accounting information should equip with two important quality features, relativity and reliability. We can come to a conclusion that XBRL will have an effect on relativity and reliability of accounting information quality. So first of all, we argue that financial report which based on XBRL will make an effect on accounting information quality.

Since XBRL itself is a kind of markup language and accounting information quality is a kind of non-quantitative objective standard, this article aims to research on the influence of financial report which based on XBRL. In order to build a bridge between them, we decide to bring into the Efficient Market Hypothesis which is led by the Scott western positive accounting theory. In this theory, when the market are strong efficient, the accounting information contents will be all reflected on the share price perfectly. And share price will roughly reflect the proportion of the social resources on the capital market. At the moment, the share price will provided some references when the information users make decisions.



Western positive accounting theory also refers that, on the assumption efficient market, the higher quality of accounting information quality the market has larger earnings response coefficient on the accounting surplus. The related literature which is published by Ball and Brown in 1968, "empirical evaluation of accounting data" refers that if the market is efficient, the share price will reflect all of the company's historical information and the unexpected information of financial report will lead to the investors make unbiased response quickly. The research shows that, in the strong or semi-strong capital market, the surplus has information content and the change of share price will timely reflect the investors' right expectation on profitable information. So after the information has been disclosed, if the market makes unbiased response on the information, share accumulative abnormally return will be kept on a stable level and don't change.

On account of above views, this article brings into the concept of earnings response coefficient in the empirical analysis. Earnings response coefficient is a tool to measure the extent of response. In this article, earnings response coefficient is to measure the response extent of the abnormal returns on the unexpected factors which is shown in the surplus part of the financial report. So this article treats the financial report disclosure as a specific event, and studies the change of earnings response coefficient during window phase so as to judge whether the surplus information of the quoted company is in high quality in this time.

In terms of above statement, we propose the following assumption:

The share price will reflect all of the company's historical information and the unexpected information of financial report will lead to the investors make unbiased response quickly. So in the empirical research, we mainly choose the company's surplus which is in high quality in the financial report to do analysis. We treat the absolute value of the unexpected accounting surplus as hypothetical variable and assume hypothetical variable has positive correlation with earnings response coefficient. As a result, this article has the following assumption:

*Assumption 1: Earnings response coefficient has positive correlation with the absolute value of the unexpected accounting surplus.*

If the financial report which based on XBRL standard can improve the accounting information quality, the earnings persistence variable which can measure the accounting robustness will have positive correlation with earnings response coefficient. As a result, this article has the following assumption:

*Assumption 2: Earnings response coefficient has positive correlation with the company's earnings persistence.*

In the financial report information, main business growth rate and Tobin Q (TBQ) value can reflect the growth of company. If the company's growth has any expected information changes, the rational information users will make related decisions. So we assume that share accumulative abnormal return rate has positive correlation with the company's growth, and in this way we can verify the decision makers' response extent to the financial information which based on XBRL standard.

*Assumption 3: Earnings response coefficient has positive correlation with the company's growth.*

In the CAPM model, only the uncontrolled systematic risk which lead to the decrease of share will influence company's expected return rate, so as to reflect on the share price to influence the earnings response coefficient, because we have considered the risk variables in the calculation of dependent variable – accumulative abnormally return rate, so this article choose the debt-to-equity ratio as proxy variable of the growth. As a result, this article has the following assumption:

*Assumption 4: Earnings response coefficient has positive correlation with the company's risk variable.*

### 3. Empirical Analysis

This article adopt event study to do empirical research, which is found by Ball & Brown (1968) and Famaetal (1969). Its principle is to choose a specific event in terms of the research objective. This method is to research the change of the shareholder return rate before and after this event happened, so as to explain the influence of the specific event on the share price and shareholder return. This method is mainly to examine the change of price or the price response extent on disclosure information

Event study is based on the effective market assumption. It means that the share price will reflect all of the public and published information. The theory assumes the investors are rational and they are rational to the response on unexpected information. As a result, in the actual return of the sample share, it eliminates the event that investors can receive abnormal return when the assumption event doesn't happen and estimated normal return. Abnormal return can measure the extent of share price response on the event happen and the abnormal information disclosure.

In terms of the above theory, this article partly uses daily return of window phase and estimation period when it calculates the accumulative abnormally return, and it adopts expected return to calculate the market index return. Then it calculate the abnormal return by calculate the difference of actual return and normal return. At last, it adopts statistical method to verify its significance level.

#### 3.1. The Selection of Sample and Data

Extensible Business Reporting Language came into service lately when we compare with developed country. We begin to use in 2008, and on the actual operation Shenzhen and Shanghai stock exchange are different. And Shanghai is more standardization on the implementation of the XBRL. As a result, this article's samples came from A and B share markets in the Shanghai stock exchange from 2009 to 2011, and choose this three years annual report as initial sample. In order to elimination the industry interference, we choose the real estate industry among the 13 industries. In order to avoid the interference of information leakage, we draft the window phase as  $[-2, 2]$ . After preliminary selection, the number of the quoted company is 857. And in order to make the empirical research more convenient, we do a stratified sampling with the reference of the past three months share closing price before annual report disclosure average date-April 30, and select maximum 50, minimum 50, intermediate value 100, in total 200 companies as sample to do analysis. On the process of empirical analysis, we

eliminate the following samples: 1. The companies which have disclose material event in the window phase; 2. The companies which have incomplete share price or whose stock has been suspended in the window phase; 3. The companies whose accumulative abnormal return rate greater than 50% or less than -50%; 4. The companies which don't make stock trading last for 10 days; 5. The companies whose data are incomplete in the research years; 6. The ST or PT companies. At last, we get 168 effective samples. These data came from Tai'an database. Data handling is based on Microsoft Excel2012 and STATA12.0.

### 3.2. Define Variables

#### 3.2.1. Dependent Variable

We add up the abnormal return in the window phase to get the dependent variable, named accumulative abnormal return (CAR). This article uses market risk adjustment method to calculate CAR.

(1) select closing price ( $P_i$ ) in the first 90 days of the window phase to calculate the daily return rate ( $R_{i,j}$ ) of every sample share.  $R_{i,j} = (P_{i,j} - P_{i,j-1}) / P_{i,j}$  ( $P_{i,j}$  means the closing price on  $j$  of share  $i$ ), and at the same day select the 300 market indexes of Shanghai and Shenzhen to calculate the indexed daily returns of sample shares:

$$R_{m,j} = (Index_{i,j} - Index_{i,j-1}) / Index_{i,j} \tag{1}$$

Index  $i,j$  means the closing index on  $j$ .

(2) In terms of the CAPM model  $R_{i,j} = \alpha_i + \beta_i \times R_{m,j} + \varepsilon_i$ , using the value of  $R_{i,j}$  and  $R_{m,j}$  in the first 90 days of window phase, regard each company as a unit to do regression analysis successively and get the parameter  $\alpha_i$  and  $\beta_i$ , then in terms of  $\alpha_i$ ,  $\beta_i$  and  $R_{m,j}$  in the window phase to calculate the expected return  $ER_{i,j} = \alpha_i + \beta_i \times R_{m,j}$  of share  $i$  during the window phase.

(3) Abnormal shareholder daily return during the window phase:  $AR_{i,j} = R_{i,j} - ER_{i,j}$ , and get the accumulative abnormal shareholder daily return during the window phase:  $CAR_{i,j} = \sum AR_{i,j}$ .

#### 3.2.2. Explain Variable

Explaining variable is the absolute value of unexpected accounting surplus (ABSUE), unexpected accounting surplus means the difference between the company actual surplus and the expected surplus of investors. This article treats the Earning Per Share (EPS) which mainly influence the investors' decision making as the proxy variable of accounting information content, and treats the last actual surplus level as the unbiased estimator of current expected surplus level:

$$UE_{i,t} = Y_{i,t} - EY_{i,t-1} = Y_{i,t} - Y_{i,t-1} = EPS_{i,t} - EPS_{i,t-1} \tag{2}$$

$Y_{i,t}$  means share  $i$ 's surplus level on  $t$ .

### 3.2.3. Earning Persistence Variable

Earning persistence variable is the robustness index to measure accounting information-EP. Earning persistence means the surplus maintain a moderate situation during a long time, and it also means the possibility of last surplus continuing on next surplus. If companies profit mainly comes from main business income which is a permanent surplus, and as a result, the surplus quality is considered very high and the future cash flow is sufficient.

$$EP = EPS_{i,t} - P_{i,t-1} \quad (3)$$

$EPS_{i,t}$  means the every share surplus reserve of share I on year  $t$ ,  $P_{i,t-1}$  means share I's price at the beginning of year  $t$ .

### 3.2.4. Companies Growth Variable

Companies growth variable is expressed by the main business growth rate (MBIRI) and Tobin Q (TBQ).

### 3.2.5. Companies Risk Variable

Since the calculation of CAR which used market risk adjustment method has involved the calculation of risk coefficient, so we choose debt-equity ratio to be the risk control variable. Debt-equity ratio equals to liability divided by the equity.

### 3.2.6. Basic Regression Equation

$$CAR_{i,t} = \alpha + \beta_1 ABSUE_{i,t} + \beta_2 EP + \beta_3 MBIRI + \beta_4 TBQ + \beta_5 DE + \varepsilon_{i,t} \quad (4)$$

## 4. Analysis Outcome

### 4.1. Descriptive Statistical Analysis

This article makes basic statistical analysis on the respondent's UE, EP, MBIRI, Tobin Q, debt-equity ratio, CAR, and understands the roughly structure of the respondent. By making basic statistical analysis on data, we can understand the whole distribution of the data.

**Table 1.** Descriptive analysis.

	Sample capacity	Minimum	Maximum	Average	Standard deviation
CAR	501	-0.9563	0.6223	0.0309	0.0991
UE	501	-0.4906	0.5148	0.0169	0.0839
EP	501	0.0003945	0.3472	0.0386	0.0417
MBIRI	501	-1.2121	0.9761	-0.5316	2.5628
Tobin Q	501	0.7476	13.2072	3.7003	2.3916
D/E ratio	501	-2.6232	11.6517	-2.4445	20.0979

Firstly, we make the descriptive statistic for the whole samples. As shown in form 1, the average CAR is 0.0309385, the average UE is 0.0168996, the average EP is 0.0386088 yuan, the average of MBIRI -0.531561, the average of Tobin Q is 3.700288, the average of debt-equity ratio is -2.444517. From the Standard deviation, maximum and minimum we can see: the Standard deviation of MBIRI, TBQ, debt-equity ratio are larger, and the differences of the TBQ and debt-equity ratio between maximum and minimum. From the statistical level, it indicates that those companies' variables are in large fluctuation. It is possible that it didn't have relationship with the dependent variables. As a result, this article will make further efforts to do correlation analysis to eliminate irrelevant factors.

4.2. Correlation Analysis

There are relationships between events but we can't directly make the explanation of causal relationship, we call this relationship correlation. Correlation analysis is to use some reasonable indexes to make statistical analysis on the observed value of correlative events.

Correlation analysis can judge the close degree of the dependency relation between variables. When the article is researching questions, we use Pearson correlation analysis. In the correlation analysis, the value of correlation coefficient (r) is between positive and negative 1. If the absolute value of correlation coefficient is closer to 1, it indicates there are closer relationships among these variables. If the absolute value of correlation coefficient is between 0-0.1, it indicates there are weak relationships among these variables. If the absolute value of correlation coefficient is between 0.1-0.4, it indicates there are low relationships among these variables. If the absolute value of correlation coefficient is between 0.4-0.7, it indicates there are high relationships among these variables.

Table 2. Correlation analysis.

	CAR	UE	EP	MBIRI	TBQ	D/E ratio
CAR	1					
UE	0.014**	1				
EP	0.1326**	0.0696**	1			
MBIRI	-0.0171	-0.0172**	-0.0533**	1		
TBQ	0.0132**	0.0219**	0.0227*	-0.0069*	1	
D/E ratio	0.0141**	0.0145**	0.1062*	-0.0031*	0.0062*	1

The above form is the correlation analysis coefficient between CAR and independent variables. The outcome shows: the correlation coefficient between UE, EP, MBIRI, TBQ, D/E ratio and dependent variable CAR are respectively -0.014, 0.1326, -0.0171, -0.0132, 0.0141, and the corresponding value of P are less than 0.01 and the correlations are all significant. And the correlation coefficient between UE, EP, TBQ, D/E ratio and dependent variable CAR are positive, it indicates that there are significant positive correlations between UE, EP, TBQ, D/E ratio and CAR. It means that the higher of UE, EP, MBIRI, TBQ and D/E ratio, the higher of CAR. There are negative

correlation coefficient between MBIRI and CAR, it indicates that there are negative correlations between MBIRI and CAR. It means that the higher of MBIRI, the lower of CAR. But the corresponding value of P is larger than 0.05 and the correlations are not significant. Since MBIRI and TBQ are both companies' growth variables, in order to avoid influencing the integrity among the variables, so in the following regression analysis we will also bring MBIRI into regression analysis.

### 4.3. Regression Analysis

The objective of regression analysis is to analysis massive sample data to find the interaction relationship among variables and confirm the mathematical relationships among variables. And make statistical test on the trusted program of the mathematical relationships, so as to distinguish the variables which have significant effects on a specific variable and other variables which have no significant effects. By using the confirmed mathematical relationships, we can predict another specific variable value in terms of one or several variable values, and give the accuracy of the prediction and control.

Treat CAR as dependent variable, and treat UE, EP, MBIRI, TBQ, D/E ratio as independent variable to do regression analysis.

Empirical analysis mainly uses the multiple linear regression models, so does this article. The article will establish the multiple regression models as follows:

$$Y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + c + \delta \tag{5}$$

In this model, *C* is constant term; *a*<sub>1</sub>, *a*<sub>2</sub>, *a*<sub>3</sub>, *a*<sub>4</sub>, *a*<sub>5</sub> are regression 为 coefficient which show the influence degree of other factors to *Y*;  $\delta$  is random error term.

In this article there are three assumptions in the linear regression: normal distribution assumption, linear assumption, the same variance assumption. In the large number of sample, the normal distribution will meets conditions; using t test to make the linear assumption; calculating correlation coefficient to test the same variance assumption. Only when the variable passes all the tests can it be bring into this model.

The followings are the specific process of the test: using sample data to estimate the regression coefficient; bringing t test into the significant level of the equation's variable coefficient; using the value of F to test the significant level, using R2 to test the fitting degree of the model.

**Table 3.** Indexes in this model.

Sample capacity	501
F(5, 493)	11.91
Prob > F	0
Determination coefficient	0.658
Adjusted determination coefficient	0.645
Residual standard deviation	2.987

The above form shows: determination coefficient of the model is 0.658; adjusted determination coefficient is 0.645; it indicates that the explanation degree of the selection independent variable has reached on 64.5%. The value of F is 11.91, and its significant probability value is 0.000, which is less than 0.01, so we refuse the original assumption that total regression coefficient are all 0. As a result, the fitting effect is good when we build the regression equation.

**Table 4.** Regression analysis outcomes.

CAR	Coefficient	Standard deviation	t	P > t	95% confidence interval	Interval
UE	0.28598	0.052792	5.4	0	0.07512	0.4323
EP	0.01322	0.006783	3.02	0.003	0.00215	0.0125
MBIRI	-0.0016	0.000029	-0.55	0.582	-7.3E-05	0.000041
TBQ	0.009	0.002406	3.66	0	0.00563	0.0383
D/E ratio	0.001	0.000554	2.01	0.046	0.00109	0.0109
Constant term	0.043406	0.006134	7.08	0	0.03135	0.0555

In the above form, it treats CAR as dependent variable and treats UE, EP, MBIRI, TBQ, D/E ratio as independent variables to get the outcomes of multiple regression analysis. It shows that: the regression coefficient of UE, EP, TBQ, D/E ratio are respectively 0.28598, 0.01322, 0.009, 0.001; and the value of T are respectively 5.40, 3.02, 3.66, 2.01; And the value of P are all less than 0.05 and meets the condition that  $P > t$ . It has notable statistical significance. As a result, the UE, EP, TBQ, D/E ratio will all have a significance influence on CAR. As shown above, the regression coefficient of UE, EP, TBQ, D/E ratio are positive, and it indicates that UE, EP, TBQ, D/E ratio will have a positive notable effect on CAR. It means that if the value of UE, EP, TBQ, D/E ratio is higher, CAR will be higher. But the value of P is larger than 0.05. It shows that the outcome is not significant and MBIRI will have no effect on CAR.

*4.4. Collinearity Diagnosis*

If there are any collinearity diagnosis problems; and we carry on multiple regression analysis, it will bring about the following problems: 1. The regression system will change violently; 2. the value and direct of regression coefficient are different from other research; 3. the variables which have affect on the outcome in fact will be excluded of the model. As a result, we should carry on collinearity diagnosis on every variable before we conduct the multiple regression analysis. If the tolerance of every variable is larger than 0.1 and VIF is less than 10, it indicates that there is no multicollinearity among variables. After we carry on the test, it is found that the tolerance of every variable is larger than 0.1 and VIF is less than 10, it indicates that there is no multicollinearity among variables. They are retained to do multiple regression analysis (See form 5).

**Table 5.** Collinearity diagnosis.

Variables	VIF	Tolerance
X2	1.02	0.980486
X5	1.01	0.98853
X1	1.01	0.994126
X3	1	0.996618
X4	1	0.998832

In terms of above analysis outcomes, we can build the following model:

$$\text{CAR} = 0.28598 \times \text{UE} + 0.01322 \times \text{EP} + 0.009 \times \text{TBQ} + 0.001 \\ \times \text{D/E ratio} + 0.043406$$

## 5. Conclusions and Suggestions

### 5.1. Conclusions

After drawing lessons from predecessors' research and studying the existing literature and combining related theoretical basis, from the view of theoretical analysis, this article used normative research and empirical research to find that the business reports which are based on XBRL will increase the relativity and reliability of quoted companies' accounting information. After that, we brought into the empirical research and selected annual report data in realty industry which are published in Shanghai stock market to carry on research. And we used event study method and brought into the concept of earnings response coefficient, for this reason to measure the response extent of the abnormal returns on the unexpected factors which is shown in the surplus part of the financial report. In order to discuss the effect of other variables on CAR for a further step, we adopted the multiple regression analysis to make an intensive study. And the outcomes show that: ABSUE, EP, TBQ, D/E ratio have positive relationships with CAR. The value of P is larger than 0.05. It shows that the outcome is not significant and MBIRI will have no effect or little on CAR.

In summary, because the market has higher earnings response coefficient, the accounting information quality is higher. And in terms of empirical research outcomes, we can get the following hypothesis test outcome:

- Assumption 1: Earnings response coefficient has positive correlation with the absolute value of the unexpected accounting surplus.
- Assumption 2: Earnings response coefficient has positive correlation with the company's earnings persistence.
- Assumption 3: Earnings response coefficient has positive correlation with the company's growth.
- Assumption 4: Earnings response coefficient has positive correlation with the company's risk variable.



## 5.2. Suggestion to Implement XBRL

Though Shanghai and Shenzhen stock exchange market both compulsively introduced XBRL standard financial report from 2009; In 2010 the Ministry of Finance also published <Extensible Business Reporting Language (XBRL) Technical Specification Series Standard> and <General Accounting Classification Standards for Enterprise> and regulated that part of the companies and securities and futures certified accounting firm which are listed on New York Stock Exchange should carry on XBRL standard financial report from Jan 1st 2011. But the step is slower than America and other countries.

In terms of this article research and some literatures and news which related to XBRL, we conclude the following points that the XBRL has not gave rise to intensive market response: (1) The period that the spread and use of Extensible Business Reporting Language is very short. In the actual operation, companies will upload tradition financial report in the form of PDF, and then the exchange will submit the financial report which is based on XBRL. In this way, the information users did not totally change the way to obtain the information, and as a result they did not make the best of XBRL to make decisions. (2) Though the two exchanges both compulsively adopted XBRL standard, they did not unify the standards and are needed to improve. And the XBRL standard reports are mainly used in securities and fund industry. The coverage area is too small.

In terms of implement of XBRL we propose the following suggestions: (1) from country perspective: on the process of XBRL implementation, the government and relative regulator should compulsively push this standard. At the moment, it is imperative to enhance the propaganda of XBRL and adopt allowance to accelerate the development of XBRL. (2) From enterprise perspective: whether in the process of implementation or research, we should stand in the angle of enterprise. By this means, the enterprise will have motivation to apply XBRL. (3) From software perspective: on the process of XBRL implementation, we need to guide related software developers to do research. On the process of research, we can organize domestic developers (i.e. UF and Golden Butterfly) to make contact with users(accounting firm) in order to develop a perfect and applicative system. (4) In order to guarantee the financial information which based on XBRL is true and fair, we need to rely on the audited company's effective internal control and management. We can't use XBRL to judge the internal control statement. The perfect and effective internal control system is imperative to guarantee the effectiveness of XBRL. (5) Accelerate the speed to raise the inter-disciplinary talent. At the moment, there are not many accounting employees understand XBRL. Many accountants and auditors understanding of XBRL are limited and they did not know this system or what the system can provide. It is indispensable to develop the XBRL education. We can set up the corresponding lessons in the specific academic school, and we can also organize corresponding social training courses. At the same time, with the implementation of XBRL around the world, we can set up corresponding grade authentication courses to encourage public to apply XBRL.

## References

- [1] Zheng Jixiao. The research on the influence of XBRL standard financial report to Chinese equity market [J]. *Financial Research*, 2015, (12):194–206.
- [2] Du Wei, Wu Zhongsheng, Zhang Tianxi. XBRL Classification standard, The research on Voluntary disclosure of information and synchronism of share price-based on the view of financial information value chain [J]. *Investment Research*, 2016, (04):23–36.
- [3] Meng Xiaojun, Ying Yafei, Li Dongyan, Zhang Yupu. The research on application of XBRL financial report-based on Earning Response Coefficient [J]. *Chinese Market*, 2016, (42):64–67.
- [4] Zhou Rui. The research on the disclosure level of XBRL financial report which based on Earning Response Coefficient [J]. *The Managers*, 2014, (05):191–192.
- [5] Weng Binyu, Qiu Yufen. The improvement of XBRL on accounting information quality in the view of accounting information disclosure [J]. *Ji Lin Agriculture Science and Technology College Academic Journal*, 2015, (04):76–79.
- [6] Qin Gu, Xin Chaoqun. The research on quality and quantity of accounting information based on XBRL standard. [J]. *Bei Jing Economic Management Cadre College Academic Journal*, 2014, (04):46–50.
- [7] Xu Jinye, Wang Menglin. Price drift phenomenon after XBRL financial report and earnings announcement-based on the empirical research on quoted companies which are listed on Shang Hai stock market. [J]. *Monthly Journal of Financial Accounting*, 2015, (03):3–6.
- [8] Li Fengui. Using empirical research to discuss the impact of XBRL on accounting information quality [D]. *Changsha University of Science and Technology*, 2012.
- [9] Wang Lin, Gong Xin. The application of XBRL financial report and accounting information quality in China-based on the empirical research of Shang Hai and Shen Zhen market data [J]. *Research on Financial and Economic Issues*, 2012, (11):124–129.
- [10] Shi Yong, Zhang Longping. The research on the implementation of XBRL financial report-based on the view of stock price synchronicity [J]. *Accounting Research*, 2014, (03):3–10+95.
- [11] Jap Efendi, Jin Dong Park, Chandra Subramaniam. Does the XBRL Reporting Format Provide Incremental Information Value? A Study Using XBRL Disclosures During the Voluntary Filing Program [J]. *Journal of Business Research*, 2016, 69(11):4793–4797.
- [12] El Ansary M, Oubrich M. State of the Art and Trends of the Research on XBRL Bibliometric Analysis from 2000–2014 [J]. *Colloquium in Information Science and Technology*, 2016, 243–250.
- [13] Troshani Indrit, Parker Lee D, Lymer Andy. Institutionalising XBRL for financial reporting: resorting to regulation [J]. *Accounting and Business Research*, 2015, 45(2):196–228.
- [14] Liu Chunhui, Yao Lee Jian, Sia Choon Ling. Research of accounting information quality – Based on XBRL taxonomy [J]. *Electronic Markets*, 2014, 1319–1325.
- [15] Seele Peter. Digitally unified reporting: how XBRL-based real-time transparency helps in combining integrated sustainability reporting and performance control [J]. *Journal of Cleaner Production*, 2016, 136:65–77.

# Big Data Analysis for Output of Chinese Research Papers Based on SCI

Li Yang<sup>a,1</sup> and Fang Xu<sup>b</sup>

<sup>a</sup>School of Arts and Communications, Anhui University

<sup>b</sup>School of Arts and Communications, Anhui University

**Abstract.** The paper is carried out based on the problem that is not solved by the current Knowledge Inventory systemically. It means science and technology output of Chinese researchers and their international influence of the academy output. To study the scientific issue feasibly, choose the data base from 2007 to 2016 of the most important scientific and technological literature retrieval system of scientific statistics and scientific evaluation with international recognition, SCI as data source and collect the paper data of Chinese researchers which is inspected by SCI. Made use of information to explore the general situation of the science and technology output of Chinese researchers from quantitative to qualitative scientific practice based on the data analysis. Some feasible suggestions are provided to the strategic allocation of scientific talents.

**Keywords.** Information; Chinese research papers; quantitative analysis

## 1. Introduction

Academy Output of researchers is an important topic of science and academy because it has global and regional perspective at the same time. With the appearance of the evaluation index system of science and technology talent output, the study of academic ability and evaluation of science and technology talents has been developed. It has a rapid extension in these 30 years. A few of relevant articles are released in world's top journals, *Science* and *Nature*. The researches in these ten years have been the highlight of domestic researchers. Researchers of Tsinghua University, Chinese Academy of Sciences and University of Science and Technology of China have carried out the study on academic evaluation of science and technology talents and relevant jobs. In a word, present domestic and foreign researches have been a certain level and they have positive academic influence. With the technological strength of country has been stronger in recent years, more and more researchers have been in international platform to publish important paper in international core journals, which has positive effect for the researchers who work on basic study to have academy exploration and participate in the advanced region of international. It has a profound meaning to strengthen the international influence of Chinese technology. Chinese international technology paper is not separated as an independent research orientation in the statistics in the authority. To meet the demand of research management and academic media tracking reports, it is necessary to have analysis and statistics on the international papers of the science

---

<sup>1</sup> Corresponding Author.

published by domestic scholars to know the position and development status of our science and technology in the world.

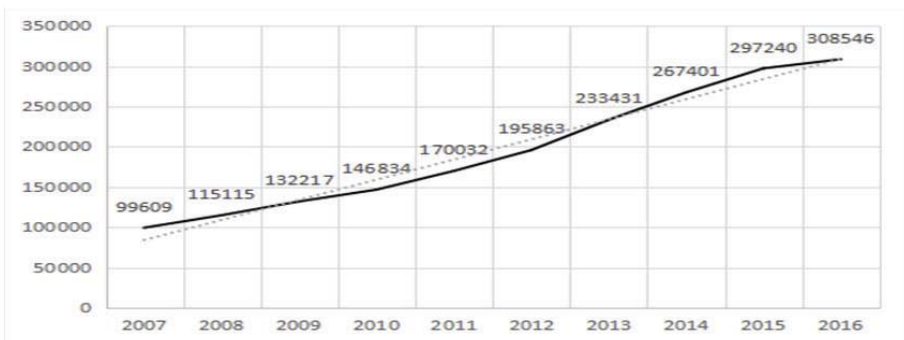
## 2. Method

The research makes the SCI network version of the original database data released by Web of Science 2007—2016 as statistical source. Input CU=China to inspect the technology theory released by Chinese writer from 2007 to 2016. Having been screening and classifying data, there is critical element analysis on SCI paper released by Chinese writers with the methods of bibliometrics, classification statistics and social networks to have a brief knowledge of general releasing trend, paper classification, languages of the releasing, fund agency and releasing organization. The aim is to master the study trend of domestic technology presently. Generally speaking, the studies of theory, parameter design and empirical analysis with scientific metrology based on SCI data base from micro, meso and macro three levels, which is based on the analysis and processing of large data. Construction of logical and rigorous argumentation system and enriching the knowledge of the status of international influence on academic output of science and technology talents could not only have breakthrough and progress on quantitative data analysis; they also apply the data analysis result into actual application. With the help of metrology theory, new suggestions and solutions which are benefit for the promotion of academic level of scientific research personnel are mentioned.

## 3. Results

Inspect the SCI academy paper released by Chinese researchers according to different sciences from 2007 to 2016 these ten years in the platform of WOS (Web of Science). Have a quantitative analysis on such critical elements as historical trend, science categories and types of literature. The scientific and accurate general situation and development context are gained. However, there are differences in the output of all sciences, so there is unbalanced development of sciences.

Figure 1. Trend of releasing on Chinese paper output



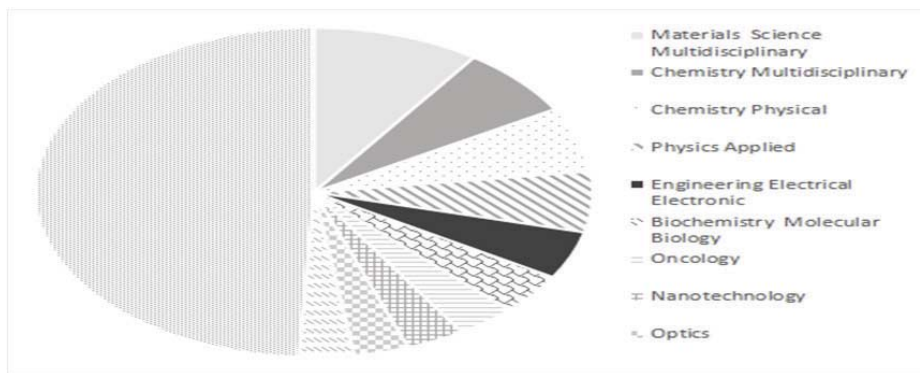
### 3.1. Trend of Releasing

Carding the data of in these 10 years and inspect 1966288 papers of Chinese writer in SCI.As shown in figure 1 above, it is primarily analyzed that the releasing percentage of 2007 and 2016 is 1: 3.1 and the average growth rate of paper output in ten years is 209.757% with more than 1/5 of average annual growth rate. Seen from the trend dash, the releasing from 2007 to 2008 was in a benign development trend; it was the same from 2008 to 2009; the increase is relatively slow although the output from 2009 to 2013 was increased annually; the increase trend from 2013 to 2015 was strong and the output was soaring but it was a slight fall in the output of 2016. It could be judged that the overall situation of SCI papers released is good. During these ten years, although there are slight fluctuated in the releasing of different all year sections and it is fluctuated surrounding the trend dash, there is a stable increase trend of Chinese paper output. It means that the academic output is one the way of increase and the international influence of research output in China are strengthened annually. It also disclosed the scenery that science in China is prosperous.

### 3.2. Classification of WOS Paper

According to the WOS paper classification, there are many subdivisions on the classification of sciences. Among these, Materials Science Multidisciplinary has the strongest academic output, which has 191082 papers with the percentage of 9.718 %; the next one is Chemistry Multidisciplinary which has 135107 papers with the percentage of 6.871%. The Religion has the weakest academic output.

Figure 2. Classification of WOS on China SCI paper



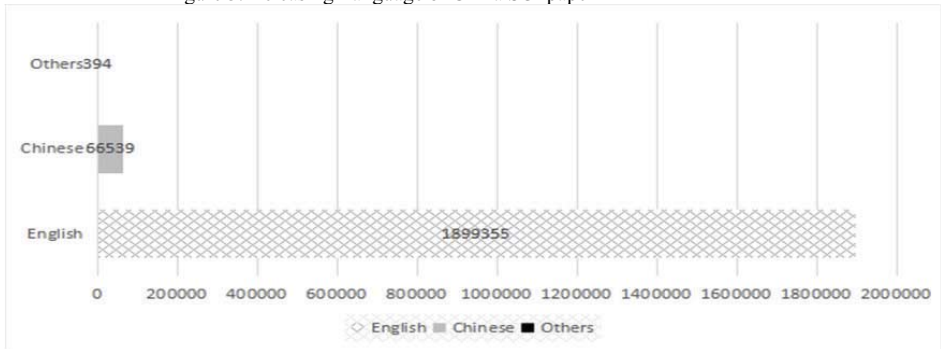
It can be seen from figure 2 that the top ten of classification papers with highest proportion are Materials Science; Multidisciplinary; Chemistry Multidisciplinary; Chemistry Physical; Physics Applied; Engineering Electrical Electronic; Biochemistry Molecular Biology; Oncology; Nanotechnology; Optics and Multidisciplinary Sciences. The total classification proportion of these sciences is 51% and the rest of them are 49%. The proportions of top three sciences, Materials Science Multidisciplinary, Chemistry Multidisciplinary, Physics Multidisciplinary, are 9.718%, 6.871% and 6.381%, which have advantage of study; there are 54 study classifications with the proportions of more than 1%. The research categories are popular. The classifications

with the later ranking are History, Social History of Science, Musicology and Religion because they are unpopular. It could be judged that the paper inspected by SCI is based on natural science and there are seldom of the development of social science. Physics and Chemistry are two basic sciences of natural science and they derivative many interdisciplinary. Science has a development with depth level in recent ten years and science pays more attention to the subject edge or finds new theory and inventions from the intersection. Hence, interdisciplinary is popular in academy. It shows the trend that modern science develops from single science to the comprehensive one, which meets the law of natural science development and have important international academic influence.

### 3.3. Releasing Language

As shown in fig.3, the language of SCI written by Chinese writers releasing is mainly English. There are 1899355 papers with the proportion of 96.596%. The second is Chinese which has 66539 papers with the percentage of 3.384%. There are 394 papers of other languages.

Figure 3. Releasing Language of China SCI paper



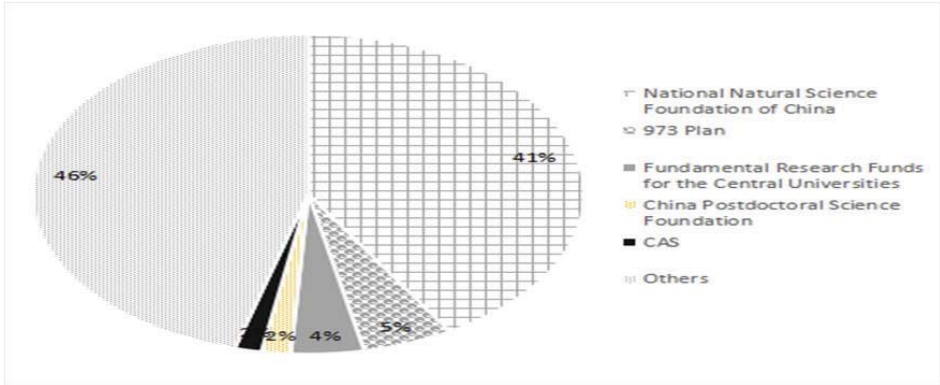
The top ten languages are English, Chinese, French, German, Japanese, Spanish, Korean, Portuguese, Russian and Danish. It could be concluded that there were many languages as the communication tool of technology and academy while English has been the mainstream language in the globalization development. It could be seen that English is the most important language received by international scientific and technological circles and academic circles. It takes extremely important position and it has been a ruled language in the communication of science and technology. It should be mentioned that there are Chinese periodical in the region of chemistry and material in SCI data base, so the Chinese papers could be inspected in SCI but the quantity is less. The SCI paper written by Chinese writers in other languages are even slightly less.

### 3.4. Fund Organization

With the observation of the fund organization of SCI papers of Chinese writer, it could be seen from figure4 that National Natural Science Foundation of China is in leading position with the proportion of 41%; the next are 973 Plan, Fundamental Research Funds for the Central Universities, China Postdoctoral Science Foundation, Chinese Academy of Sciences and other fund organizations with the proportion of 46%. It could

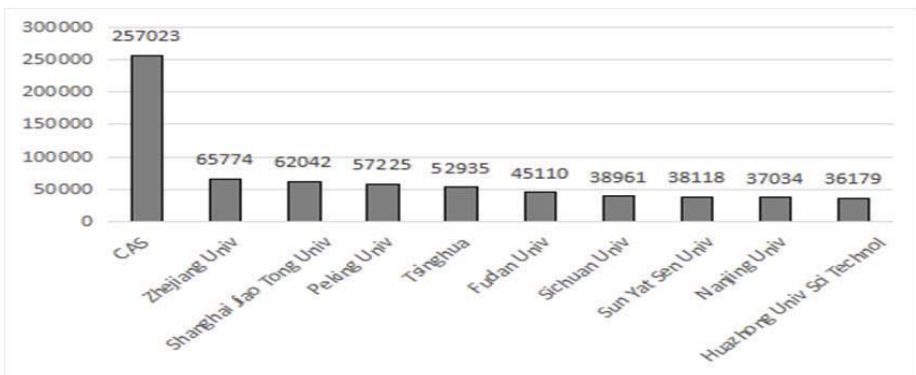
be concluded that National Natural Science Foundation of China has a full play of guiding role of national independent innovation ability according to the policies and plan of the country. It supports academic research of science and technology effectively and subsidizes large researchers to have basic research.

Figure 4. Fund Organization of China SCI paper



It funds researchers to write lots of SCI academic paper largely. 973 Plan is national basic research program of China. It supports researchers to explore the scientific frontier and have high-end research activities as well as releasing high-level paper surrounding with national strategic demand. It promotes the development of natural science powerfully. The Basic Research Funds of the Central University is set by central finance with special fund for basic scientific research of Central University. It funds high level talents in Colleges and Universities to work on research and supports the releasing of SCI paper with the aim of promoting the university independent innovation ability. China Postdoctoral Science Foundation also plays a role in the aspect of aiding Chinese postdoctoral young creative talents releasing SCI paper.

Figure 5. Releasing Organization of China SCI paper



### 3.5. Releasing Organization

According to the statistics of releasing organization of SCI, it could be found that

Chinese Academy of Sciences has most of releasing which has 257023 papers with the proportion of 13.071%. The second is Zhejiang University which has 65774 papers. The third is Shanghai Jiao Tong University which has 62042 papers. The next are Peking University, Tsinghua University, Fudan University, Sichuan University, Sun Yat Sen University, Nanjing University, Huazhong University of Science and Technology (see figure 5). It is reflected clearly from the picture that national the senior academic institution, Chinese Academy of Sciences, has a strong academic output in science and technology. As a national comprehensive research and development center, it has excellent academic output level of natural science and high technology. Zhejiang University, locating in the Hangzhou with the nickname of Heaven of Human, has a rapid development in academic output in these ten years and it is the second one. Shanghai Jiao Tong University, one of the famous top universities in China, has a good performance to be the third one. These reflect that economically affluent Zhejiang and Shanghai devote vast research capital to support researchers to have scientific research and the academic outputs are strong. The fourth and the fifth, Peking University and Tsinghua University, locate in the Chinese political center Beijing with abundant academic resource and strong academic platform to have obvious academy advantages.

#### **4. Conclusion**

##### *4.1. Total quantity of SCI paper has a rapid increase*

Since from 2007, Chinese SCI paper output has rapid increase and the average increase rate of paper quantity in these ten years is 20.976%. Chinese research activity is in a status of active development annually. The rapid increase of Chinese SCI technology paper is closely related with the strengthening in the breadth and depth of international cooperation in scientific research. The academic cooperation with Europe, America, Asia and other countries promotes the academic output of China largely.

##### *4.2. Wide Distribution of Science*

Seen from the science structure of Chinese SCI releasing, the range of the sciences are very vast and the interdisciplinary is very common. Compared with other target research institute, China has obvious advantage in Multidisciplinary Synthesis, Materials Science Multidisciplinary and Chemistry Multidisciplinary and other natural sciences.

##### *4.3 Strong Academy Output of Coastal and Political Center*

Zhejiang province and Shanghai such coastal cities have strong economic strength and they invest vast research capital to support researchers to have scientific research with strong academy output. Peking University and Tsinghua University locate in politics center with powerful science and technology policy support and strong academy platform, so they have obvious academy advantage.

#### **5. Discussion**

The paper makes SCI data base as evidence. It has a macro analysis and evaluation on the Chinese research level and international academic influence with the aspect of



science paper, the main form of academic output. It provides reference for the research management department having further profound analysis and making important technology decision as well as promoting research system reform. It is a topic needed to be discussed in the future that how to have a comprehensive evaluation on Chinese international academic influence with patents, monographs, and important academic awards at home and abroad such factors.

A study of science and technology output of Chinese researchers and their international influence of the academy output by the detail analysis of vast data. It provides some feasible suggestions to the cultivation of future technology talents and the balanced development of researchers of all science. It also provides evidence for education, science and technology and related departments to make decision of training and mobility of scientific and technical personnel.

The paper is helpful to promote the internationalization of science and technology output and research of our researchers in China with the aim to increase the influence and recognition of the science and technology output of our researcher in the world. The SCI data base in the study is one of the major scientific and technological literature retrieval systems in the world and the data usage obeys the academic evaluation criteria of scientific research personnel which is recognized by the world. Hence, the study hopes to be as reference data in the customization of the uniform scientific research ability and academic evaluation standard.

## **Acknowledgement**

This research was supported by Natural Science Foundation of Anhui Province grants KJ2017A946; Outstanding Young Talent Support Program grant gxyqZD2016374 from Anhui Province.

## **References**

- [1] Price DJD. Networks of scientific papers, *Science*, 1965, 149:510-515.
- [2] Ball P. Index aims for fair ranking of scientists, *Nature*, 2005, 436(7053):900.
- [3] Peter V. Subfield problems in applying the Garfield factors in practice. *Scientometrics*, 2002, 53(2):267-279.
- [4] Hirsch JE. An index to quantify an individual's scientific research output. *Processing of the national academy of sciences of the United States of America*. 2002, 231-232.
- [5] Noyons EC, et al. Combining mapping and citation for evaluation bibliometric Purpose: A Bibliography study, *J Am Soc Inform*, 1999, 50.
- [6] Egghe, L. Theory and practice of the g-index in function of time. *Scientometrics*, 2006, 69(1):131-152.
- [7] Glanzel, W. On the H-index-a mathematical approach to a activity and citation impact. *Scientometrics*, 2006, 67(2):315-321. new measure of publication
- [8] Glanzel, W. and Persson, O. H-index for Price medalists. *ISSI Newsletter*, 2005, 1(4):5-18
- [9] Antonis Sidiropoulos. Generalized Hirsch h-index for disclosing latent facts in citation networks, *Scientometrics*, 2007, 72(2):253-280.
- [10] Tibo Braun, Wolfgang Glanzel, Andras Schubert. A Hirsch-type index for journals. *Scientometrics*, 2006, 69(1):169-173.

# ZLCC: Vehicle Detection and Fine-Grained Classification Based on Deep Network Responses and Hierarchical Learning

Chen Joya<sup>a</sup> and Shunxi Li<sup>b,1</sup>

<sup>a</sup>*School of Automotive engineering, Wuhan University of Technology, Wuhan, China*

<sup>b</sup>*School of Automotive engineering, Wuhan University of Technology, Wuhan, China*

**Abstract.** Vehicle detection and recognition is the research focus in Intelligent Transportation System (ITS) with many challenges. Based on the success of Convolutional neural networks (CNN) in object detection and image classification, we propose a ZLCC to locate vehicle and classify its maker & model & shape. Our framework focuses on two new perspectives: (i) how to generate a small number of high quality region proposals, (ii) how to improve vehicle classification accuracy rate by hierarchical learning policy. We use deep network's responses to generate aware-map in detection, and train network with multiple candidates softmax regression. We demonstrate the success of ZLCC on Stanford Cars for using the deep VGG16 architecture.

**Keywords.** Convolutional neural network, Object detection, Region proposal, Fine-grained classification, Aware-map, multiple candidates softmax regression

## 1. Introduction

Vehicle recognition is an important part of ITS, it has been widely concerned as a complementary technology for license plate recognition. While general image classification has achieved impressive success within the last few years [4], [19], [20], it is still very challenging to recognize vehicle in an image with fine granularity, since vehicles' large yield lead to similar appearance each other, which make it is hard to classify. The previous work [1], [2], [3] for vehicle recognition depends on designing feature extractor manually, the performance of the algorithm is affected by environment easily.

As the AlexNet [4] won the ILSVRC-2012 championship, CNN [5] began to develop rapidly in computer vision, new theories for vehicle recognition emerged in large numbers. Gao Y et al. [6] proposed an architecture to recognize vehicle's maker by CNN, DONG Zhen et al. [7] use an unsupervised CNN to realize six shapes (Truck, Bus, Sedan, Microbus, Minivan, SUV) classification. However, their algorithm only focus on vehicle's single information, which limit their contribution to ITS. Recently, the research for vehicle fine-grained classification have made some progress. Bin Hu et al. [8] introduced a multi-task CNN to learn vehicle's view-point and fine-grand class, but they simply feed the whole picture and composite label to CNN, lacking of more

---

<sup>1</sup> Corresponding Author.

Email address: [ChenJoya@foxmail.com](mailto:ChenJoya@foxmail.com) (Chen Joya), [lsx@whut.edu.cn](mailto:lsx@whut.edu.cn) (Shunxi Li)

purposeful expression. J Sochor et al. [9] proposed BoxCars, try to represent richer features by estimating 3D vehicle orientation and information in CNN, boosts the recognition performance considerably. However, it is expensive to predict the 3D information of the vehicle from the single plane image.

We try to regard vehicle recognition as a series of different but connected processes, for example, when we glance at a vehicle, we firstly know its location (bounding box), then we can acquire its maker & shape (e.g. Audi Sedan), lastly we can get more details (e.g. Audi A8 SUV 2011) by more observation. Inspired by it, we design ZLCC with the idea of global to local, coarse to fine. The architecture of ZLCC is shown in Fig. 1.

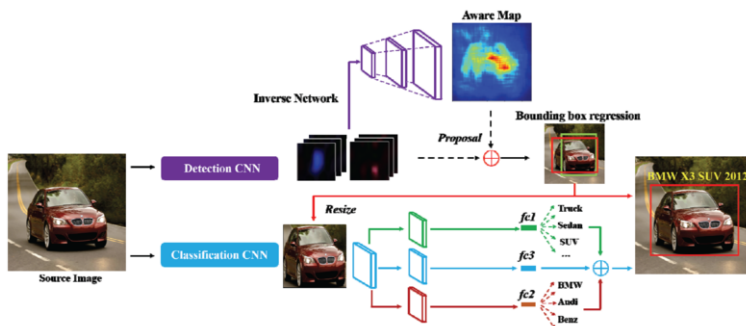


Figure 1. Architecture of ZLCC.

Our first contribution is to propose a new region proposal approach based on deep network responses, which can generate a small number ( $\sim 90$ ) but high quality proposals. Recent state-of-the-art object detection [10], [11], [12], [13] using CNN can be divided into two-stage strategy: (i) region proposal, (ii) region classification and bounding box regression. Though methods which do not rely on region proposal (e.g. [14], [15]) has advantage in speed, region proposal systems lead higher accuracy on several benchmarks, so the quality of region proposal is important. Traditional methods [16], [17] in can generate high quality proposal, however they consume much running time in detection task. RPN which was proposed in Faster R-CNN is better in predicting object position, but it replaces traditional method by sliding window in feature map, and it will generate a large number of proposals ( $2k\sim 20k$ ). To remedy the issue, we research on the relationship between original image and CNN feature map, inverting the network to generate aware-map, and do proposals on it. It turns out that our method performs well on vehicle detection.

Our second contribution is to propose a hierarchical learning technique with multiple candidates softmax regression. Although there are a large number of frameworks for fine-grained classification, some of them depend on specific region location. For instance, vehicle's frontal face position is needed in [6] for recognizing maker, bird's head and lower body position are needed for multiple description in [22]. Other methods use multi-task learning to classify, however they simply combine different loss function rather than more human-like and logical recognition step. Our proposed framework use 3 networks with different task to solve the problem, each of them cares about different aspect about the vehicle, but make final decision based on them. It is demonstrated by our experiments in Stanford Cars dataset [18] that our method is more advantageous for vehicle fine-grained classification.

The rest of this paper is organized as follows: The detailed explanations of ZLCC are described in Sect. 2 and Sect. 3, the training approach and our experiment results are given in Sect. 4.

## 2. Vehicle Detector in ZLCC

This section describes the vehicle detection approach in ZLCC, and it includes three-stage strategy: (i) generating aware-map by inverting network, (ii) region proposal on aware-map, (iii) bounding box regression. Note that all the strategies accomplish in a CNN. The number of proposals is the mainly differences between a series of R-CNN ( $2k\sim 4k$ ) and our method ( $\sim 90$ ). Besides, the region proposal which is recommended by ZLCC often have higher quality.

### 2.1. Inverse network for detection

The first step for detection is generating aware-map in ZLCC, and it means get the local areas which need to pay attention. Popular convolution neural network have similar structures generally, consisting of a series of layers as convolutional layer (*conv*), rectified linear units layer (*relu*), pooling layer (*pool*), fully-connected (*fc*) layer, etc. We take the CaffeNet [23] as an example to illustrate. When the network forward, the size of input data should be resized to  $R^{227\times 227\times 3\times N}$ , and we will get the feature map  $F^{6\times 6\times 256\times N}$  after *pool5*. The feature map can reflect activated special area in different channels, as Fig. 2(b) shows. According to the computing rule of receptive field, it is easy to map the area to original image, then find its location. However, such the method has very low IOU (intersection-over-union), because vast quantities of spatial information have been lost after *conv* and *pool*. Therefore we inverse the network to get  $R^{227\times 227\times 3\times N}$  from  $F^{6\times 6\times 256\times N}$ , then generating aware-map  $A^{227\times 227\times 1\times N}$  based on  $R^{227\times 227\times 3\times N}$ . Here are the key steps:

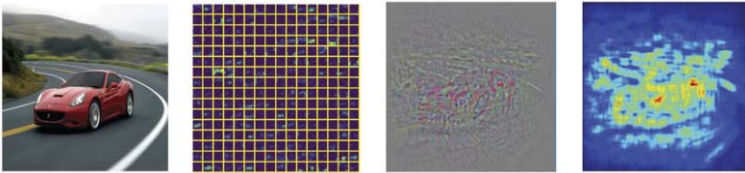


Figure 2. (a~d) A series of images from inverse network.

**Flipping filter and memorizing position for inverting convolution and max-pooling.** Let's start with the convolution in 2D matrix  $M^{w\times h}$ . The filter can be expressed as  $f^{t\times t}$ , and the process is described in formula (1):

$$M^{t'w'\times h'}(i, j) = \sum_{k, l} M^{w\times h}(i+k, j+l) f_R^{t\times t}(k, l) \quad (1)$$

Where  $f_R$  represents  $180^\circ$  rotation for original filter. If we want to get  $M^{m\times n}$  with the same size of  $M^{w\times h}$  by inverse convolution, we only need to rotate the  $f_R^{t\times t}$

again since  $(f_R^{1 \times t})_R = f^{1 \times t}$ . Therefore in some CNN toolboxes just like Caffe, we can flip the *conv* layer’s filters horizontally and vertically, then copy them to *deconv* layer. However, such of the idea may not be applied to inverse max-pooling, as there is no way to acquire the location of in original map. To solve the issue, we define *pooling-memorize* layer, Fig. 3 gives a general description of it. We can use the index to get the appropriate information as Fig. 3 shows.

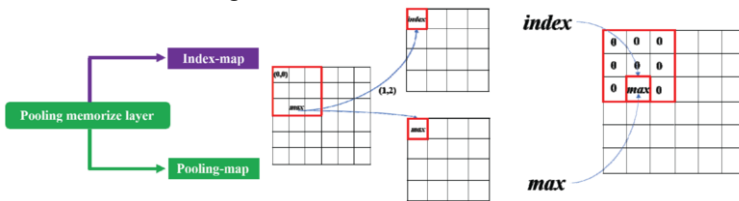


Figure 3. Pooling-memorizing layer.

**Restraining irrelevant pixels in aware-map.** The output data can be directly seen in Fig. 2(c). We can see that the information around the vehicle is well preserved, while other background areas are suppressed. To more powerful expression, we norm the aware-map and do mean filtering on it, which can be seen in Fig. 2(d). Based on that, we explore the effects on aware-map with different classification task. As shown in Fig. 4, four types’ CaffeNet model have been trained. The network which is trained in ImageNet [24] is a generic model for 1000-classes classification in Fig. 4(b), and networks in the Fig. 4 (d, e) are fine-tuned from pre-trained ImageNet-1000 model in Stanford Cars dataset. Fig. 4 (c) is from the binary classification (vehicle or non-vehicle) network. It is obvious that aware-map which is generated by generic model is hard to locate vehicle, and the active region depends on the classification task. For example, 49 maker-classes network cares more about vehicle’s front face in Fig. 4(d). Here we use aware-map from 27 shape-classes network to do region proposal, since it focus more on vehicle’s whole surface rather than local, performs better when frontal face is not clear as Fig. 4(e) shows.

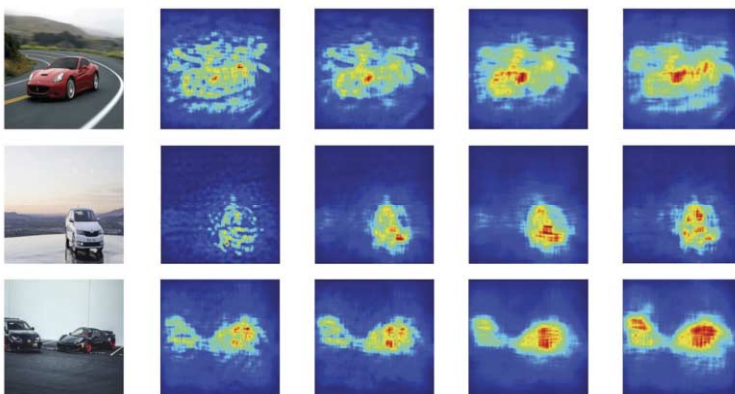


Figure 4. (a~e) Aware-map from different classification task network.

## 2.2. Region proposal on aware-map

Benefiting from generated aware-map, we can get more purposeful candidate region. Inspired by RPN anchor in Faster-RCNN, we set different scales and aspect ratios proposals in pivotal parts, which is illustrated in Fig. 5 (b). The selective parts depends on aware-map response as Fig. 5(a) shows.

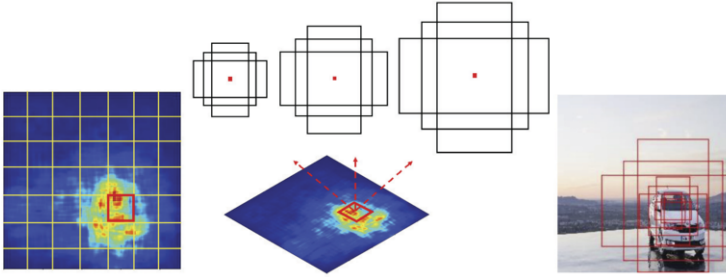


Figure 5. (a~c) Region proposal

Our region proposal algorithm is simple but efficient. At the beginning, we put a  $32 \times 32$  cell on each of them, which can be seen in Fig. 5(a). Then we compute the internal sum of the pixels for each cell, and sort all the sum value, set top  $i$  cells to generate proposals. Inspired by Faster-RCNN, we set 3 scales ( $32^2, 64^2, 128^2$ ) and 3 aspect ratios (1:1, 1:2, 2:1). The proposals from highest score cells can be seen in Fig. 5(c). We have the advantage of generating a small number proposals ( $\sim 90$ ) with more purposeful selection. Note that we employ bounding-box regression in Fast R-CNN [12] to refine proposals lastly.

## 3. Vehicle fine-grained classifier in ZLCC

In this section, we present the proposed framework for tackling on vehicle fine-grained classification. There are many previous works [21], [22], [25], [26] about image fine-grained classification, which can be divided into two subnetworks: (i) recognizing based on key part detection, (ii) augmenting data or policy to more effective learning. Here we don't use first method, since we think key part recognition is not suit for every image (e.g. we can't see vehicle's logo if the vehicle backs to us). Instead, we improve training policy as hierarchical learning, which means training different network separately, but make the final decision based on them.

### 3.1. Maker & shape classification

Though multi-task network can forward at high speed by sharing with parameter, we find that it will limit the potential of CNN and its performance is worse than multi single networks. Therefore, we build two single network for maker and shape classification, each of them use traditional softmax regression for back propagation. Taking the one of the network as an example. The training sample can be seen as  $\{x_i, y_i\}$ , where  $y_i$  is the maker label for data  $x_i$ ,  $y_i \in \{1, 2, \dots, k\}$ . Providing the last  $fc$

layer output is  $f(x_i)$ , which has  $k$  channels. In softmax regression, the probability that the label of  $x_i$  is  $j$  can be denoted as:

$$P(y_i = j | f(x_i)) = \frac{e^{f_j(x_i)}}{\sum_{j=1}^k e^{f_j(x_i)}} \quad (2)$$

We hope to maximize the probability of right label  $P(y_i = r | f(x_i))$ , so the loss function will be:

$$L = -\log P(y_i = r | f(x_i)) = \log \sum_{j=1}^k e^{f_j(x_i)} - f_r(x_i) \quad (3)$$

Our goal is to minimize  $L$ , therefore we take the derivative of  $f(x_i)$ :

$$\frac{\delta L}{\delta f(x_i)} = \frac{e^{f_r(x_i)}}{\sum_{j=1}^k e^{f_j(x_i)}} - 1 = P(y_i = r | f(x_i)) - 1 \quad (4)$$

According to formula (4), we can use batch gradient descent to update parameter to minimize  $L$ . In fact, the significance of building maker & shape classifier is that CNN can achieve higher classification in such the two tasks than building fine-grained classifier directly, and we hope to use maker & shape information to guide CNN to find more details, which can be seen in Fig. 6.

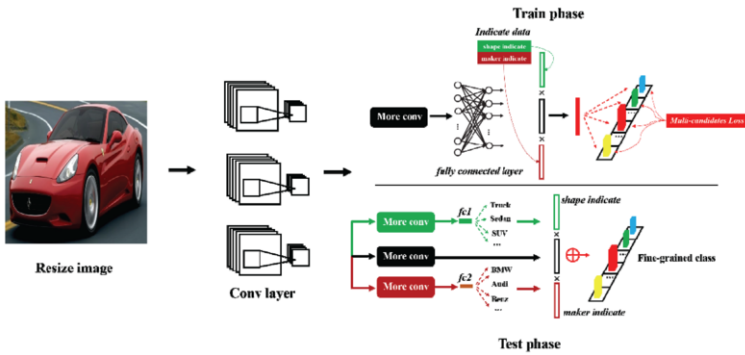


Figure 6. Classification network structure

### 3.2. Multiple candidates softmax regression

After the maker and shape have been predicted, our fine-grained classifier can use these results to do a final decision. We hope that the prediction should have the following two characteristics: (i) The items which contain predicted maker and shape should have a higher score than the other in predicted probability array (e.g. **BMW M1 Sedan** should have higher score than **BMW X3 SUV** or **Chevrolet Camaro Sedan**, and their score should all be higher than **Volkswagen 3 SUV**). (ii) Correct class should have

the highest score. To achieve it, we improve softmax regression for multiple candidates. Here we define indicate array  $I$  which has  $k$  channels,  $I[u] \in [0,1]$ ,  $u \in \{1,2,\dots,k\}$ .

$I$  denotes multiple candidates (e.g.  $I[3]=1$ ,  $I[4]=1$  suggest 3<sup>rd</sup> or 4<sup>th</sup> class should have higher probability than other classes). By indicate array  $I$ , we can express maker or shape information in fine grained classification. Firstly, we revise traditional softmax regression's probability to formula (5):

$$H(y_i = uf(x_i); I) = \frac{P(y_i = uf(x_i))I[u]}{\sum_{l=1}^k P(y_i = lf(x_i))I[l]} = \frac{e^{f_u(x_i)}}{\sum_{j=1}^k e^{f_j(x_i)}} \frac{I[u]}{\sum_{l=1}^k \left( \frac{e^{f_l(x_i)}}{\sum_{v=1}^k e^{f_v(x_i)}} I[l] \right)} \quad (5)$$

We assume  $t$  is the right label for  $x_i$ , therefore we he to maximize  $H(y_i = tf(x_i); I)$ , and the loss function can be expressed as:

$$L = -\log H(y_i = tf(x_i); I) = \log \sum_{l=1}^k P(y_i = lf(x_i))I[l] - \log P(y_i = tf(x_i)) - \log I[t] \quad (6)$$

We take the derivative of  $f(x_i)$ :

$$\begin{aligned} \frac{\delta L}{\delta f(x_i)} &= \frac{I[t]P(y_i = tf(x_i)) - I[t]P(y_i = tf(x_i))^2}{\sum_{l=1}^k f_l(x_i)I[l]} + P(y_i = tf(x_i)) - 1 \\ &= (1 - P(y_i = tf(x_i)))(H(y_i = tf(x_i); I) - 1) \end{aligned} \quad (7)$$

Based on the gradient in formula (7), we can train a new network to find more details when we have already know maker or shape. The indicate array is generated by maker & shape network in testing phase as Fig. 6 shows.

## 4. Training and testing

In this section, we briefly introduce our method of training our detector and classifier by taking example on Stanford Cars. We use Caffe to train CNN, and we fine-tuned from pre-trained ImageNet model because it always achieved better result.

### 4.1. Comprehensive Labels Separation

We take Stanford Cars as an example. The dataset contains 16,185 images of 196 classes of cars, and labels are typically at the level of make, model, year, e.g. 2012 Tesla Model S or 2012 BMW M3 coupe. We counted each comprehensive labels to single label, some of them are depicted in Table 1.

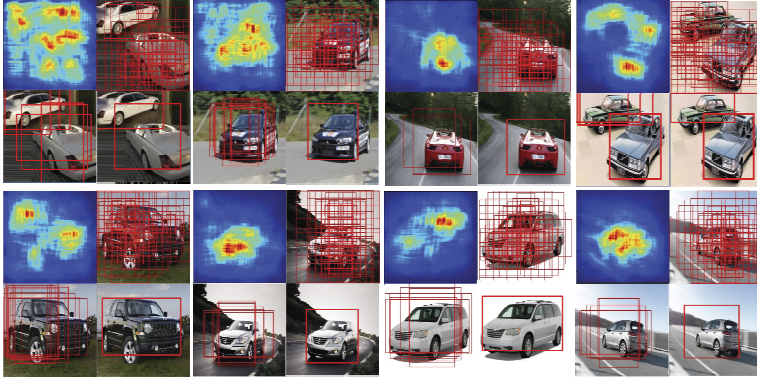


**Table 1.** Partial maker and shape labels for Stanford Cars. The comprehensive label can be divided into 49 classes of maker, 27 classes of shape.

Comprehensive labels	Maker	Shape
Acura RL Sedan 2012	Acura	Sedan
BMW M3 Coupe 2012	BMW	Coupe
MINI Cooper Roadster Convertible 2012	MINI	Convertible
Chevrolet Express Van 2007	Chevrolet	Van
Ford F-150 Regular Cab 2012	Ford	Cab

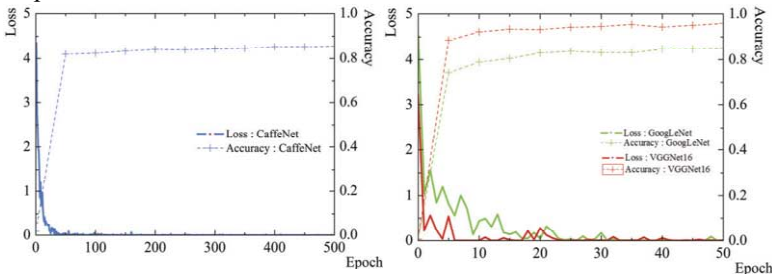
#### 4.2. DetectionNet

Firstly, we trained a 27-classes shape classifier. For faster detection, we only use a tiny network as CaffeNet here. Then we set conv layer's learning rate to 0, and training RCNN which means to add bounding box regression for the network. We change the classify output to 2 classes (vehicle/non-vehicle). Fig. 7 shows that our bounding box regression for our region proposal (after NMS).

**Figure 7.** Effects of region proposal and bounding box regression

#### 4.3. Maker & shape ClassificationNet

We train network for maker and shape classification, here we use 3 types architecture as CaffeNet, VGG16, GoogLeNet. (note that we only train 50 epoches for VGG16 & GoogLeNet because of our GPU memory limit). The training result is illustrated in Fig. 8, 9. We can learn that VGG16 performs better. The accuracy of them in test sample reach to 95.8% and 94.9%.

**Figure 8.** Maker classification network training curve

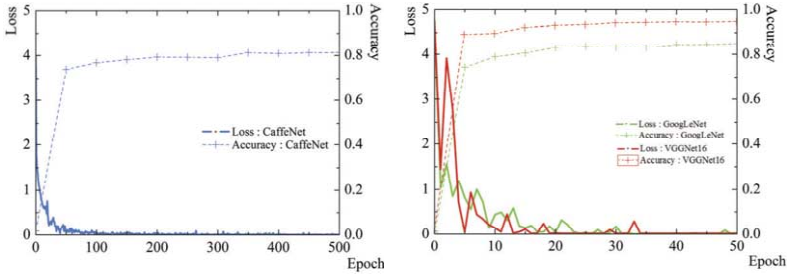


Figure 9. Shape classification network training curve

#### 4.4. Fine-grained ClassificationNet

**Indicate array.** The indicate array is a 196-channels array for each maker/shape class. For example, if there is a car whose maker is *Acura*, then the fine-grained label which contain *Acura* maker will be set to 1. We use the array to generate indicative function in Sect. 3.1. In a word, we hope to punish those predictive classes which have wrong maker or shape.

**Network Training.** Based on the previous analysis in Sect. 4.2, Sect. 4.3, we directly train fine-grained classification network by finetuing VGGNet. Our network find more subtle features in the premise of giving of the vehicle's real maker and shape, and the classification accuracy reaches to a very high value (nearly 97%) which is shown in Fig. 10.

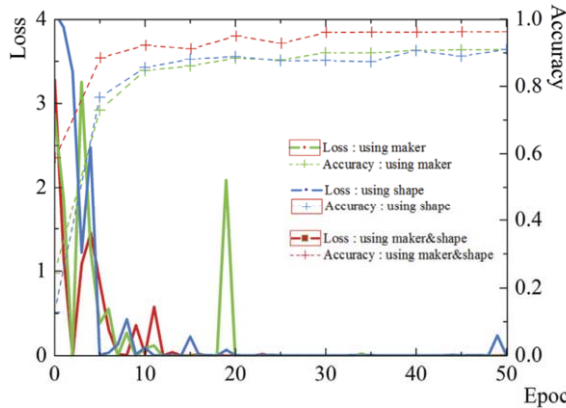


Figure 10. Fine-grained network training curve by different indicate array

#### 4.5. Experiment on Stanford Cars dataset

We use our trained maker & shape and fine-grained classification network to evaluate on *Stanford Cars*, the probability array from maker & shape network will be converted to 196 channels firstly, and then we multiply them as the indicate array for fine-grained classification network. We compare ZLCC with other algorithm while giving true bounding box or not giving true bounding box, the result is in Table 2.

**Table 2.** Comparison on the Stanford car dataset. **Test with Bbox** means to classify with ground truth bounding box. The numbers represent the accuracy(%).

Methods	Test with true bbox	Test without true bbox
LLC [27]	69.5	---
ELLF [28]	73.9	---
BB [29]	63.6	---
BB-3D-G [18]	67.6	---
HAR-CNN [26]	80.8	---
CoSeg [25]	<b>92.8</b>	<b>92.6</b>
Bilinear [30]	88.2	---
R-CNN [10]	---	57.4
<b>ZLCC(ours)</b>	<b>89.3</b>	<b>86.1</b>

From the experiments, we observe that ZLCC can significantly improved the accuracy rate of recognizing. However, our method does not perform well at state-of-the-art method [25] while testing both in with and without true bounding box. It may show that detection for key parts may be important for fine-grained classification.

## 5. Conclusion

This paper presents a framework for vehicle detection and fine-grained classification. We use feature map to get aware-map by inversing network, and generating a small number of high quality region proposals on it. Then, we put forward the method of building fine-grained classifier through hierarchical learning policy, including separate training and multiple candidates softmax regression. Compared with lots of algorithms' testing result on Stanford Cars, ZLCC can significantly improve the accuracy rate, reaching to 89.3% with true bounding box and 86.1% without any annotation. We believe that our approach is generalizable, and we plan to apply it to other fine-grained classification tasks.

**Acknowledgments** This work was supported by the National Natural Science Foundation of China (Nos. 51508432), National College Students Innovation and entrepreneurship training program in Wuhan University of Technology (No. 20161049707007).

## References

- [1] Dlagnekov L, Belongie S. Recognizing cars[J]. Ecological Modelling, 2005, volume 113(13):71-81(11).
- [2] B. Zhang, Y. Zhou, and H. Pan, "Vehicle classification with confidence by classified vector quantization," Intelligent Transportation Systems Magazine, IEEE, vol. 5, no. 3, pp. 8–20, 2013.3.
- [3] Varga M, Radford J. Automatic car model classification[C]// International Conference on Acoustics, Speech, and Signal Processing. IEEE Xplore, 1991:2369-2372 vol.4.
- [4] A Krizhevsky , I Sutskever , GE Hinton. ImageNet Classification with Deep c Neural Networks[C]//Proc of International Conference on Neural Information Processing, 2012, 25(2):1097-1105.
- [5] Lecun Y, Boser B, Denker J S, D Henderson, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.
- [6] Gao Y, Lee H J. Vehicle Make Recognition Based on Convolutional Neural Network[C]// International Conference on Information Science and Security. 2015:1-4.

- [7] DONG Zhen, PEI Mingtao, HE Yang, LIU Ting, DONG Yanmei, and JIA Yunde. Vehicle Type Classification Using Unsupervised Convolutional Neural Network. IEEE International Conference on Pattern Recognition (ICPR), 2014.
- [8] Bin Hu, Jian-Huang Lai, Chun-Chao Guo, Location-Aware Fine-Grained Vehicle Type Recognition Using Multi-Task Deep Networks, *Neurocomputing* (2017), doi:10.1016/j.neucom.2017.02.085
- [9] Sochor J, Herout A, Havel J. BoxCars: 3D Boxes as CNN Input for Improved Fine-Grained Vehicle Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:3006-3015.
- [10] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. 2013:580-587.
- [11] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-16.
- [12] Girshick R. Fast R-CNN[C]// IEEE International Conference on Computer Vision. IEEE, 2015:1440-1448.
- [13] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016:1-1.
- [14] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:779-788.
- [15] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. 2016.
- [16] Van d S K E A, Uijlings J R R, Gevers T, et al. Segmentation as selective search for object recognition[C]// International Conference on Computer Vision. IEEE Computer Society, 2011:1879-1886.
- [17] Zitnick C L, Dollár P. Edge Boxes: Locating Object Proposals from Edges[J]. 2014, 8693:391-405.
- [18] Krause J, Stark M, Deng J, et al. 3D Object Representations for Fine-Grained Categorization[C]// IEEE International Conference on Computer Vision Workshops. IEEE Computer Society, 2013:554-561.
- [19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. CoRR, abs/1409.4842, 2014. 1, 2, 3, 5 (source 20)
- [20] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014. 1, 3, 5
- [21] Zhou F, Lin Y. Fine-Grained Image Classification by Exploring Bipartite-Graph Labels[J]. 2015, 9(2):1124-1133.
- [22] Wang D, Shen Z, Shao J, et al. Multiple Granularity Descriptors for Fine-Grained Categorization[C]// IEEE International Conference on Computer Vision. IEEE, 2015:2399-2406.
- [23] Y Jia, E Shelhamer, J Donahue, S Karayev, J Long. Caffe: Convolutional Architecture for Fast Feature Embedding[C]//Proc of Acm International Conference on Multimedia, 2014: 675-678.
- [24] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database[J]. 2009:248-255.
- [25] Krause J, Jin H, Yang J, et al. Fine-grained recognition without part annotations[J]. 2015:5546-5555.
- [26] Xie S, Yang T, Wang X, et al. Hyper-class augmented and regularized deep learning for fine-grained image classification[C]// Computer Vision and Pattern Recognition. IEEE, 2015:2645-2654.
- [27] M. S. Jonathan Krause, Jia Deng and L. Fei-Fei. Collecting a large-scale dataset of fine-grained cars. The Second Workshop on Fine-Grained Visual Categorization, 2013. 7
- [28] J. Krause, T. Gebru, J. Deng, L.-J. Li, and L. Fei-Fei. Learning features and parts for fine-grained recognition. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 26–33. IEEE, 2014.
- [29] J. Deng, J. Krause, and L. Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 580–587, 2013.
- [30] T.-Y. Lin, A. R. Chowdhury, and S. Maji. Bilinear CNN models for fine-grained visual recognition. In *ICCV*, 2015. 2, 5, 6

# Health Care Management Using Knowledge Management and Information Technology

Jawad KARAMAT<sup>a,1</sup>, Shurong TONG<sup>a</sup>, Abdul WAHEED<sup>a,2</sup> and Kashif MAHMOOD<sup>b</sup>

<sup>a</sup>*School of Management, Northwestern Polytechnical University, Xi'an, China*

<sup>b</sup>*School of Management, Bahria University, Islamabad, Pakistan*

**Abstract.** The health care industry is a vital part of any country and its economy. The health care industry seek increasingly to improve quality by improving clinical outcomes, patient experiences and a business case that support movement toward 'patient-centered care'. The health care industry by becoming knowledge based community. This can be done by creating communication between many hospitals, clinics, pharmacies, and customers. The sharing of knowledge should be increased, the reduction of administrative costs and improvement in the quality of care. The health care will be able to do this by using knowledge management (KM) and information technology (IT). This research study seeks to come up with a framework that investigates the different dimensions of knowledge management and IT and how it will help in improving health care. In order to come up with a framework the three elements of KM which are knowledge acquisition, knowledge application and knowledge dissemination and use of IT to make it efficient. If properly used IT can accelerate knowledge-sharing capabilities in both time and space dimensions. Locality, timing, and relevancy factors determine the expediency and the strength of IT's role in KM initiatives.

**Keywords.** Knowledge management, information technology, health care, health care management, knowledge management system

## 1. Introduction

Health care is one of the most important part of any civilized community. It can be said to be a system that tends to meet the health care needs of a target population. The system of health care in different countries is different. There are some countries where the health care is generally in private sector. Whereas there are also countries in which the government provides the hospitals. The definition of health care according to Minkova and Stockwell [1] "The prevention, treatment, and management of illness and the preservation of mental and physical well-being through the services offered by the medical and allied health professions." The main aim of health care is the same, the prevention of illness.

Health care industry these days has become a profit making business in the private sector. It is for this reason most of the private sector health care providers are resorting

---

<sup>1</sup> PhD students in School of Management, Northwestern Polytechnical University, Xi'an, China; E-mail: jawad@mail.nwpu.edu.cn.

<sup>2</sup> PhD students in School of Management, Northwestern Polytechnical University, Xi'an, China; E-mail: waheed\_2506@mail.nwpu.edu.cn.

to improved patient administration or quality health services. In the 1990s, the health care industry has increasingly tried to embrace new information technologies as well as software applications, as it searched for opportunities for efficiency and higher-quality care [2]. Also during the last decade, health care organizations started to use information systems for clinical purposes to improve patient care [3]. Computerized decision support or expert system, which is targeted at assisting health care providers and administrators with such decision tasks as information retrieval, data analysis, diagnosis and test, procedure and case management recommendation has been one of the critical information technologies heavily deployed to transform health care [4,5]. All the improvement can be brought about with knowledge.

As Francis Bacon said, "Knowledge is power". This power should be utilized accordingly and with great effect. This power can be used to preserving valuable heritage. On the basis of our previous knowledge we can solve problems and learn new things and come up with newer and better ideas. This power can also become the basis for creating new core competences.

The importance of knowledge has finally been realized. That is what gave birth to knowledge management. Knowledge Management (KM) has developed into different areas in the study of firms and is alleged to play an important part in attaining sustainable competitive advantage in the present day business and academic arena [6]. According to Scarbrough et al., (OCED, 2003), KM can be described as "covering any intended and methodological process or put into practice the knowledge of acquiring, capturing, sharing and using knowledge, wherever it resides in, to improve the learning capability and performance of firms" [4].

During the past few years many health care organizations have started to use IT related systems to improve patient care [3]. These IT systems are targeted at assisting health care providers and administrators with information retrieval, data analysis, diagnosis and test, procedure and case management recommendation [4]. If IT is properly and effectively utilized it can accelerate the capability of sharing knowledge in both space and time. Locality, timing, and relevancy factors determine the expediency and the strength of IT's role in KM initiatives [7].

## **2. Problem Statement**

Can knowledge management be paired up with information technology to improve the health care industry of Pakistan?

A lot of work has been done on it in the health care sector but it is mostly on the medical perspective. This study will try to design a framework to be implemented in the hospital as to improve patient administration perspective. A combination of KM and IT will provide great advantage to the patient administration. As it is said that, "The health care industry will turn into a more cost-effective, error averse, and accountable public resource" [8]. Knowledge management in health care is "aligning people, processes, data and technologies to optimize information, collaboration, expertise, and experience in order to drive organizational performance and growth" according to a current knowledge management special interest subgroup of the Healthcare Information & Management Systems Society (HIMSS).

### 3. Literature Review

#### 3.1. Health Care Management

Health care is one of the most essential components of any civilized community. It can be said to be a system that tends to meet the health care needs of a target population. The system of health care in different countries is different. There are some countries where the health care is generally in private sector. Whereas there are also countries in which the government provides the hospitals. At the end it does not matter in which sector the health care is being provided. It is the delivery, the quality and the accuracy of the service that matters. Providing health care to the patients is no easy task. The health care organizations depend on a lot of information. They need to know about that particular patient, the kind of care to be provided, the previous treatments provided, the results of previous treatment, the type of performance expected of them, and how they can effectively coordinate amongst departments. Like human, material, and financial resources, information is a resource that must be managed effectively by the health care managers [9,10]. The traditional single physician–patient relationship is increasingly being replaced by one in which the patient is managed by a team of health care professionals each specializing in one aspect of care. Such seamless and shared care depends critically on the ability to share information easily between care providers. Hence, the ability to access and use electronic health care record (EHCR) of the patient is fundamental [11].

People are living longer and thus demand more and higher quality preventive and long-term care. The demand for health-care workers is expected to grow faster than the average rate for all occupations between 2000 and 2010. In particular, the demand for home care aides, registered nurses, physician assistants, nurse practitioners, physical therapists, nontraditional health aides, and physicians will continue to increase at a healthy pace. The health care industry as a whole is all business these days. Hospitals, nursing homes, home health care, specialized clinics are being run increasingly like any other major for-profit organization. Health care today is all about big business-with its focus sharpening on driving profits higher.

#### 3.2. Health Care in Pakistan

In Pakistan there are two types of health care delivery system, public and private. They provide secondary and tertiary healthcare services. According to the Bureau of Statistics (2009) [12], the facilities were as follows,

**Table 1.** Health Care facilities in Pakistan.

Total Health Facilities	13,937	103,708 beds
Hospitals	968	84,257 beds
Dispensaries	4,813	2,845 beds
Rural health centers	572	9,612 beds
Tuberculosis clinic	293	184 beds
Basic health units	5,345	6,555 beds
M.C.H. centers	906	256 beds

Despite the fact that Pakistan has both private and public hospitals, the private are more frequently used. According to *Pakistan social and living standards measurement (PSLM)* Survey, 67.4% of Pakistani households consult private sector health consultants for their health related problems [13]. Most of these private sector hospitals provide primary health care and are located in urban areas only. Despite vast public sector network, it is observed that many of the health facilities remain non-functional due to non-availability of health staff. Due to rapidly growing population, the existing number of public and private health institutions, facilities and services are not sufficient to cope with the health care provision to all people [12,14].

**Table 2.** People Employed in Health Care, Pakistan.

Doctors (2009)	139,555
Dentists (2009)	9,822
Nurses (2009)	69,313
Midwives (2009)	26,225
Health visitors (2009)	10,731
Registered vets (2009)	4,800

Due to these reasons it is important to implement the knowledge management and IT system so that the patients can be handled quickly and in a cost effective manner.

### 3.3. Knowledge Management

The American Productivity & Quality Center (APQC), a nonprofit education and research organization which fostered the creation of the Malcolm Baldrige National Quality Award, defines knowledge management as “the strategies and processes of identifying, capturing and leveraging knowledge” to enhance competitiveness [15]. According to Gartner Group findings for 1998, implementation of knowledge management systems has begun, and is currently being deployed by most large companies. Senior managers described knowledge management as the core of the consulting strategy and the concept was pervasive in the company’s internal and external documents [16]. The health care industry by becoming knowledge based community. This can be done by creating communication between many hospitals, clinics, pharmacies, and customers. The sharing of knowledge should be increased, the reduction of administrative costs and improvement in the quality of care. Despite the fact that the health care institutes should be connected and sharing constant knowledge but that is not the case. Information in the medical community is not frequently shared. This results in errors in the treatment of the patient. From the health sector point of view the knowledge management can be thought of ensuring the right information is available to the right people and practiced by the right people at the right time [4]. The health care will be able to do this by using knowledge management (KM) and information technology (IT).

### 3.4. Knowledge Management System

Knowledge and knowledge management are complex and multi-faceted concepts. These days knowledge management has given birth to information systems, referred to as knowledge management systems (KMS). The objective of KMS is to support crea-



tion, transfer, and application of knowledge in organizations. These days a lot of research work is being done on various field of management. There are researches on how the organization learns and then how it uses that knowledge to its advantage to gain competitive advantage. There is also a lot of research that is going on in the field of knowledge management, information technology and systems. What we need to do is to gather all of them together. This will give birth to knowledge management system (KMS). From the organizational perspective we will have to determine the type of organization it is. Then from the learning perspective we have to find out the knowledge it will generate. With knowledge management the use and flow and generating of new knowledge will be controlled. The information technology and system will help storing it and providing it when required.

Knowledge management has become a very important topic these days. Many businesses are moving towards it, and investing heavily in it. The three basic tools are,

- The text should be digitized and available at all times.
- This is the time of globalization. Information should be readily available anywhere to improve the product.
- The right information should be available at the right time to minimize risk of making mistakes.

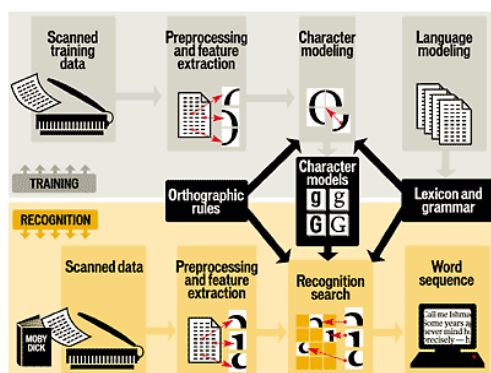


Figure 1. BBN Technologies Inc. and GTE Internetworking (now Genuity Inc.).

### 3.5. Information Technology

These days the importance of IT cannot be ignored. It is used for several purposes. The most important is that it can improve the knowledge-sharing capabilities in both time and space. This is why it makes the role of IT crucial for role in KM initiatives. IT is used for document management and storage. Groupware contributes to knowledge sharing and management to improve organizational effectiveness [17]. IT is used to achieve two basic goals to reduce the uncertainties of knowledge and loss of knowledge by depending on memory of concerned individuals. IT is used to improve knowledge creation, searching, and diffusion. KM is useless for competitive ends unless communication and application systems are well developed to permeate every aspect of business operations. In this study IT plays a major role. In this study the main role of IT is knowledge-sharing capabilities, document management and storage. The knowledge is in document called electronic patient service (EPS) and electronic health record (EHR).

### 3.6. EPS and HER

The EPR as defined by National Health Services (NHS) is a periodic record of patient's care provided by an acute hospital. This hospital is providing HC treatment for that patient during a specific period of time [18]. In every hospital there are many departments. They all have their different systems. These systems consist of patient records. Now what the EPR does is keeps the record of one patient in one location instead of many different locations. This EPR helps in time saving and there is no duplication of efforts. This will also help the healthcare professionals in making better decisions [18]. The EPR can also be referred to as electronic medical record (EMR) [19]. The EHR is a collection of the electronic patient records (EPR) over the life of the patient. The implementation of the EPR would require the linking of Knowledge bases, Rules and guidelines and Expert systems. These three components should be integrated with the clinical modules and document imaging. If all this is done it can be said that the stage one of EHR has been completed [20]. The EHR will be available on line for those who are authorized to view the information. These are generally four types of people the hospital, the patient, the doctor and nurses. These people can use this information for decision making. The EHR consists of patient information, history, past treatment and result. The EHR helps the healthcare people in understanding the complications the patient might face and how to overcome them. This will help in providing better care and improved quality.

The patient goes through many treatments. So the EHR might be a very long file to study. In order to make it easy to understand the EHR is divided into 3 parts,

- Primary care, the care provided by a family doctor or a general physician.
- Secondary care, the care provided by a specialist.
- Tertiary care, the treatment provided by major hospitals.

The information from the nursing department will be in this record. This record will keep track of the nursing treatment, test results and patient progress. In this record the patient can also put in some information himself. This will help the physicians in better understanding the patient position and improved communication.

## 4. Research Methodology

Before conducting the research the theory of KM and its linkage with IT was studied to the greatest extent to understand the possible implications and advantages it has to offer to the health care industry. The research method adopted was mainly interpretive and analytical. This research mostly depends on inductive reasoning. The data gathered is qualitative in nature. The main reason is that the information mostly gathered was through interview. 100 interviews were conducted with doctors in private and public sector. Since knowledge management is a concept not well known to the doctors in Pakistan they were given a little material to get an idea KM and why we want to conduct our interview. All the doctors considered for interview were holding administration position at their respective hospitals. These doctors were randomly selected. The data that was gathered by the interviews is given below, they were asked if they were currently satisfied with the level of IT in their hospital their response was.

**Table 3.** The present status of their IT.

Satisfaction Level of Doctors with Current IT Status		
	Satisfied	Unsatisfied
Public Sector Doctors	23%	77%
Private Sector Doctors	68%	32%

From the result it is clear that the level of IT in the public sector is not as good as in the private sector. The system of keeping patient record in the public sector is almost nonexistent. The same patient will always be treated as a new patient every time due to lack of record.

**Table 4.** The potential of their extension into IT.

Extension in Current IT Status		
	Agreed	Non-Agreed
Public Sector Doc	81%	19%
Private Sector Doc	14%	86%

The doctors were then asked if they wanted improvement in their IT. The private sector doctors were quite satisfied with their current situation whereas public sector doctors were not. They did not expect that the government will be willing to spend on IT. The public sector doctors were however of the opinion if they got the funds for extension in IT they would spend heavily since the existing IT is almost nonexistent. The current budget allocation is however showing that the private sector is not willing to spend on IT since they are already short on funds. The private sector is not spending highly because they are satisfied with their current level of IT.

**Table 5.** The budget allocated for the extension of IT.

Budget Allocated for Extension in IT	
Public Sector	05%
Private Sector	13%

**Table 6.** Do you think your current management can handle this change?

Satisfaction with Current Management to handle Changes		
	Satisfied	Not Satisfied
Public Sector	35%	65%
Private Sector	73%	27%

The public sector doctors are of the opinion that their current staff will not be able to cope with the change in current IT. The private sector doctors are of the opinion that their staff will be able to manage the change.

### 5. Recommendation

In Pakistan the major source of health care provider is the public sector, it is due to the lack of its resource that the people go to private sector hospitals. In order for our model to be successfully implemented the government hospitals should increase their spending in IT. They must set up a data base in order to store patient’s records. The model works in the following steps.

1. The private sector hospitals must be given certain area to cover. In the area covered there may be many small clinics and other private hospitals.
2. The residents of those area must register them self with that main hospital. This registration should be done and the person be handed a card. This card must be proof of his registration and to which area he belongs.
3. This card will contain three types of information, the treatment given by personal doctor, the treatment given by specialist, the lab test results, allergies and general information. The information on this card must be constantly updated by the clinic the patient frequently visits. The data on this card can be viewed by any doctor that the patient wants to share with.
4. The information on this card will not be limited to the area where the patient currently is registered. It can be accessed by other hospitals of other area. This will be done at the prior request of the patient.
5. In care of emergency the hospital where the patient is taken, using his identity card number the hospitals can withdraw patient records.

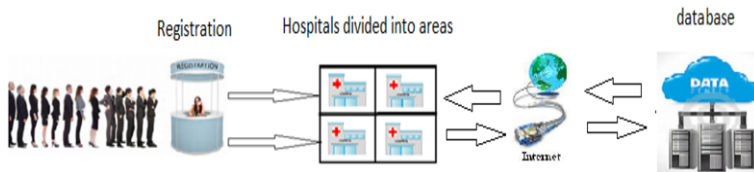


Figure 2.

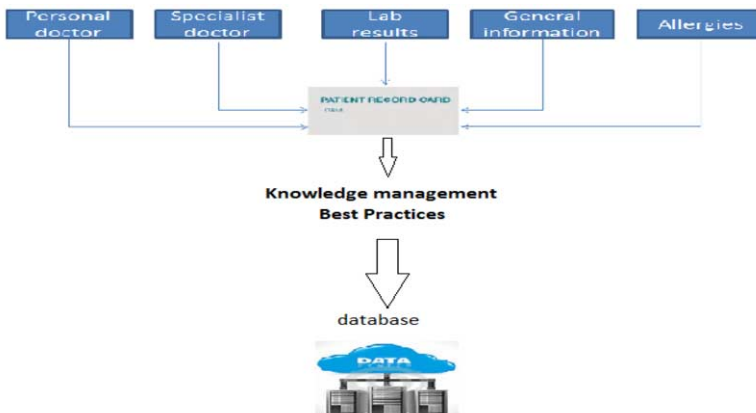


Figure 3.

## 6. Conclusion

Knowledge management has emerged in the recent times as a phenomenon which has various implications. It helps in innovation and competitiveness in many fields. It is also argued by many experts that firms effectively use KM establish long term internal strengths, which lead to competitive advantage. The concept of knowledge management (KM) is growing quickly these days. The main focus of KM is basically to utilize the knowledge that is available to the firm and make the most of it to gain competitive advantage. KM can be implemented in every organizational discipline. It can most definitely be used in the health care industry as well. In the recent times communication has improved significantly. This is mainly due to the advancement of communication networks and internet access. The world has become a global village. The knowledge can be accessed quickly and at lower cost. This is all possible due to the rapid advancement in technology.

If this model is followed it will certainly help in dealing with the patient better. All the records will be available to the doctor with a swipe of a card. The new doctor can then make informed decisions based on his prior records. This will help the doctor to give good diagnosis and deal with the patient quickly. Since in Pakistan the doctor to patient ratio is very low it will make the patient turnover much better. The added advantage of this model is that it will help in the reduction of medical error. If a medical error does occur then it will be easy to find out which doctor or nurse made the mistake. The existence of fake doctors and medical centers will be eliminated. More evidences can be found by adopting some new methodology to evaluate health care system of Pakistan like SEM [21].

## References

- [1] D. Minkova and R. Stockwell, *English words: History and structure*: Cambridge University Press, 2009.
- [2] W. Raghupathi and J. Tan, "Strategic uses of information technology in health care: a state-of-the-art survey," *Topics in Health Information Management*, vol. 20, pp. 1–15, 1999.
- [3] J. G. Anderson and C. E. Aydin, "Evaluating the impact of health care information systems," *International Journal of Technology Assessment in Health Care*, vol. 13, pp. 380–393, 1997.
- [4] S. L. Achour, M. Dojat, C. Rieux, P. Bierling, and E. Lepage, "A UMLS-based knowledge acquisition tool for rule-based clinical decision support system development," *Journal of the American Medical Informatics Association*, vol. 8, pp. 351–360, 2001.
- [5] A. Waheed, M. Xiaoming, N. Ahmad, S. Waheed, and A. Majeed, "New HRM Practices and Innovation Performance; the moderating role of Information Technology Ambidexterity," 2017.
- [6] I. Nonaka and H. Takeuchi, "The knowledge creation company: how Japanese companies create the dynamics of innovation," *New York*, 1995.
- [7] M. Mohamed, M. Stankosky, and A. Murray, "Knowledge management and information technology: can they work in perfect harmony?" *Journal of Knowledge Management*, vol. 10, pp. 103–116, 2006.
- [8] J. Guptill, "Knowledge management in health care," *Journal of Health Care Finance*, vol. 31, pp. 10–14, 2005.
- [9] H. Heathfield and G. Louw, "New challenges for clinical informatics: knowledge management tools," *Health Informatics Journal*, vol. 5, pp. 67–73, 1999.
- [10] A. Waheed, S. Waheed, J. Karamat, N. Ahmad, and A. Majeed, "Implementation and adoption of E-HRM in Small and Medium Enterprises of Pakistan," 2017.
- [11] R. OEVER, "Supporting High Technology Medicine Through Insurer-Provider Partnership," *Pacing and Clinical Electrophysiology*, vol. 22, pp. 363–368, 1999.
- [12] A. Sherin, "Obesity: how to prevent Pakistani people from getting heavier?" *Khyber Medical University Journal*, vol. 5, pp. 59–60, 2014.
- [13] H. S. Rout, *Health care systems: A global survey*: New Century Publications, 2011.

- [14] A. Waheed, M. Xiaoming, J. Karamat, and S. Waheed, "Comparison of Human Resource Planning and Job Analysis process in banking sector of Pakistan," 2016.
- [15] A. Sadri McCampbell, L. Moorhead Clare, and S. Howard Gitters, "Knowledge management: the new challenge for the 21st century," *Journal of knowledge management*, vol. 3, pp. 172–179, 1999.
- [16] T. H. Davenport, D. W. De Long, and M. C. Beers, "Successful knowledge management projects," *Sloan Management Review*, vol. 39, p. 43, 1998.
- [17] J. Papows, "Enterprise. com. Massachusetts: Perseus Publishing." 1999.
- [18] A. Jalal-Karim and W. Balachandran, "Interoperability Standards: the most requested element for the Electronic Healthcare Records significance," in *2nd International Conference–E-Medical Systems, E-Medisys2008*, 2008.
- [19] K. Häyrinen, K. Saranto, and P. Nykänen, "Definition, structure, content, use and impacts of electronic health records: a review of the research literature," *International journal of medical informatics*, vol. 77, pp. 291–304, 2008.
- [20] A. Razzaque and A. Jalal-Karim, "Conceptual Healthcare Knowledge Management model for adaptability and interoperability of EHR," in *European, Mediterranean & Middle Eastern Conference on Information Systems*, 2010, pp. 12–13.
- [21] F. Shaheen, N. Ahmad, M. Waqas, A. Waheed, and O. Farooq, "Structural Equation Modeling (SEM) in Social Sciences & Medical Research: A Guide for Improved Analysis," *International Journal of Academic Research in Business and Social Sciences*, vol. 7, 2017.

# The Applied Research of Image Dehazing Algorithms in License Plate Location

WANG RUI <sup>a,1</sup> and WANG GUOYU <sup>b</sup>

<sup>a</sup>*Binzhou Medical University;*

<sup>b</sup>*Ocean University of China*

**Abstract.** Urban transportation is developing rapidly, which leads to that intelligent traffic is the direction of future traffic. License plate recognition(LPR) technology is one of the key techniques. But in bad weather, images taken by cameras degrade dreadfully, which results in the poorer LPR effect. This paper introduces the propagating deconvolution, dark channel prior and color attenuation prior dehaze algorithms as the pretreatment of LPR, and proposes a license plate location method based on both color and edge information. In addition, this paper compares the influence of various dehaze algorithms on the license plates location.

**Keywords.** Propagating deconvolution, dehaze, License plate recognition, License plate location, intelligent transportation

## 1. Introduction

License plate recognition (LPR) technology is one of the core technologies in intelligent transportation, it can achieve the automation of the traffic system management, information technology. LPR technology is able to detect the monitored road vehicle and automatically extract the vehicle license information. It is based on digital image processing, pattern recognition and computer vision technology, and can obtain the license plate information from the vehicle images taken by the camera or video sequences. Since each car has only one license plate number, LPR technology can add the parking fee management, traffic control index measurement, the vehicle positioning, car alarm, automated highway speeding regulation, electronic police, highway toll station, and so on. Therefore, it has the real meaning in intelligent transportation to ensure traffic safety and urban security, prevent traffic jams and realize traffic automation management.

The complete LPR system includes image collection, image processing, and database management of three subsystems, as shown in figure 1. While the image processing sub system is the primary part which includes image pre-processing, License plate (LP) location, character segmentation and character recognition.

Image pre-processing refers to the operation that increases the target image, enhances the contrast of the target and background image, and facilitates the follow-up work of the LP location. LP location is to locate the license plate and extract the license plate image from a captured image. Thus it has the direct impact on character

---

<sup>1</sup> Corresponding Author.

segmentation and recognition and is an important prerequisite for license plate recognition in complex situations.

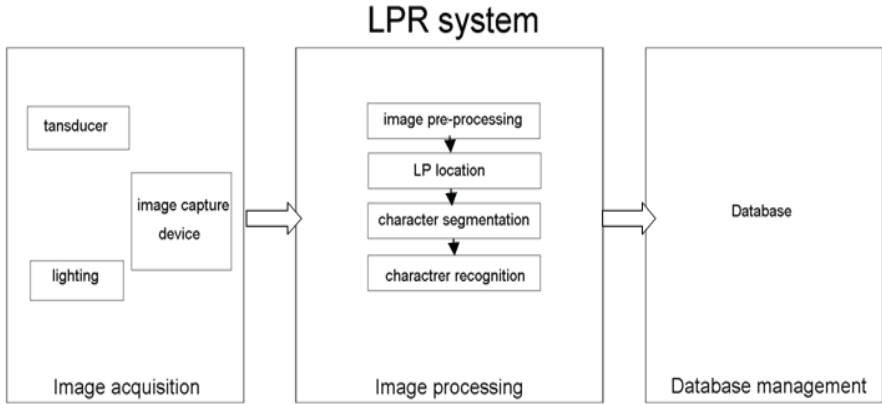


Figure 1. The composition of LPR system

However, in bad weather, the details of haze images are lost and the color are degraded. Researches [1-6] are applied in image dehazing,

In this paper we use the LP location method based on both color and edge characteristics for haze traffic images, and introduces three image dehazing algorithms [4-6] then compare the results by different dehazing algorithms as the pre-processing.

## 2. Image dehazing algorithms

### 2.1. Propagating deconvolution (PD) dehazing method

Propagating deconvolution is based on a layered model of the haze [4]. We can separate the haze into a series of very thin layers as show in Figure 2.

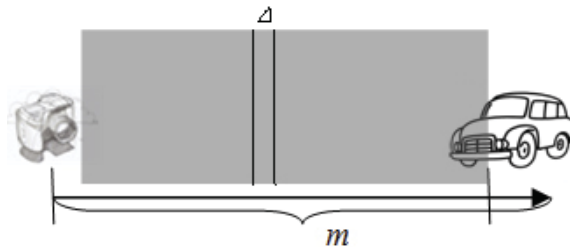


Figure 2. Layered decomposition of the haze

Assume that the depth between the target and the sensor is  $z$ , the sensor location was the 0 point.  $z$  can be divided into  $m$  layers; the thickness of each layer is  $\Delta$ , then  $z = m\Delta$ . The frequency domain expression of point spread function (PSF) of single layer can be described as (1),  $\rho$  is transmission parameter of unit thickness :



$$H_{\Delta}(u, v) = e^{(-\rho\Delta\pi(u^2+v^2))} \tag{1}$$

$\kappa(x, y)$  is represented the medium backscatter coefficient of unit thickness, whose Fourier transform is  $K(u, v)$ . At  $z$ , the thickness  $dz$  causes backscattering was:

$$dN_z(u, v) = K(u, v)AdzH_z(u, v) \tag{2}$$

To calculate the integral of equation (2) with  $z$ , then :

$$N_z(u, v) = AK(u, v)\int_0^z e^{-c+\rho\pi^2(u^2+v^2)z} dz = \frac{AK(u, v)}{c + \rho\pi^2(u^2 + v^2)}(1 - e^{-[c+\rho\pi^2(u^2+v^2)]z}) \tag{3}$$

Backscattering noise power spectrum expression is :

$$Pn_z(u, v) = A^2\sigma^2(1 - e^{-[c+\rho\pi^2(u^2+v^2)]z})^2/[c + \rho\pi^2(u^2 + v^2)]^2 \tag{4}$$

Let  $k(x, y) = k_0 + \gamma$ ,  $k_0$  is the DC component,  $\gamma$  is the random component and can be regarded as the Gaussian white noise with the variance of  $\sigma^2$ . the mean scattering intensity can be obtained by summing the backscatter noise of all layers:

$$S_d = N_z(0, 0) = (K(0, 0)A/c)(1 - e^{-cz}) = (k_0A/c)(1 - e^{-cz}) \tag{5}$$

Let  $\eta_1 = \rho/c$ ;  $\eta_2 = A\sigma/c$ ;  $\eta_3 = k_0A/c$ ; when  $z$  is infinite, and the area without any target can be regarded as the complete backscattering noise  $N_{\infty}(u, v)$ :

$$N_{\infty}(u, v) = AK(u, v)/[1 + \eta_1\pi^2(u^2 + v^2)] \tag{6}$$

The mean of infinity backscattering noise  $S_d^{\infty} = k_0A/c = \eta_3$ . The power spectrum of infinity backscattering noise is :

$$Pn_{\infty}(u, v) = \eta_2^2/[1 + \eta_1\pi^2(u^2 + v^2)]^2 \tag{7}$$

Defining the deconvolution depth parameter  $\alpha=cz$ , then the frequency domain expression of PSF is :

$$H_z(u, v) = e^{(-\alpha(\eta_1(u^2+v^2)+1))} \tag{8}$$

Intercepting the infinity area, which is a 2D signal, its power spectrum can be used to fitting the unknown parameters  $\eta_1, \eta_2$  by the least square method. Its mean value can be used to obtain  $\eta_3$ . The frequency domain expression of wiener filter is:

$$F(u, v) = \frac{H_z(u, v)}{|H_z(u, v)|^2 + \frac{P_g(u, v)}{P_f(u, v)}} G(u, v) \tag{9}$$

Where  $F(u, v)$ ,  $G(u, v)$  are the expression in frequency domain of the restored and the degraded image  $g(x, y)$  respectively.  $P_f$  is the power spectrum of  $g(x, y)$  getting rid of the scattering noise  $S_d$ . As long as the parameters  $\eta_1, \eta_2, \eta_3$  were obtained, the image degradation function will only relate to the transmittance  $e^{-\alpha}$ . The parameter  $\alpha$  is estimated as 0.5.

### 2.2. Dark channel prior(DCP) dehazing method

He [7] discovered that the dark channel of a clear image was almost dark on the statistics of a large amount of images. Almost in every local patch of a clear image there will exist very low gray intensity pixels which called dark pixels. The definition of the dark channel is consisted of the dark pixels.  $J^{dark}$  is described by the formula:

$$J^{dark} = \min_{\Omega}^c \in \{R, G, B\}(\min(J^c)) \tag{10}$$

$J^c$  is one color channel of  $J$  including RGB three channels, and  $\Omega$  is a local patch of  $J$ . First normalize the (11) which describes the haze degradation model in haze, and then put the minimum operators on both sides and calculate the dark channel on both sides, we can get (12):

$$I = tJ + (1 - t)A \tag{11}$$

$$\min_{R, G, B}(\min_{\Omega}(I^c)) = t \min_{R, G, B}(\min_{\Omega}(J^c)) + (1 - t)A = tJ^{dark} + (1 - t)A \tag{12}$$

The dark channel prior suggests that  $J^{dark}$  should tend to be zero. There is :

$$t = 1 - \min_{R, G, B}(\min_{\Omega}(\frac{I^c}{A})) \tag{13}$$

After the transmittance  $t$  is estimated, we can get the clear image  $J$ .

### 2.3. Color attenuation prior(CAP) dehazing method

Paper [6] proposed a new transcendental called color attenuation prior based on the relationship between scene depth and fog concentration: depth is related to the difference of luminance and saturation. It can be described as equation (14), where  $d(x)$  is the depth,  $v(x)$  is brightness,  $s(x)$  is saturation.  $\theta_0, \theta_1, \theta_2$  stand for the linear coefficients.  $\varepsilon(x)$  is a random variable of random errors.

$$d(x) = \theta_0 + \theta_1 v(x) + \theta_2 s(x) + \varepsilon(x) \tag{14}$$

Let  $\varepsilon(x) \sim N(0, \sigma^2)$ , then  $d(x) \sim N(\theta_0 + \theta_1 v + \theta_2 s, \sigma^2)$ ,  $v$  and  $s$  are known.

Since the depth  $d$  is unknown, Zhu combined clear image, random depth and random air light as the sample haze image as shown in Figure 3.

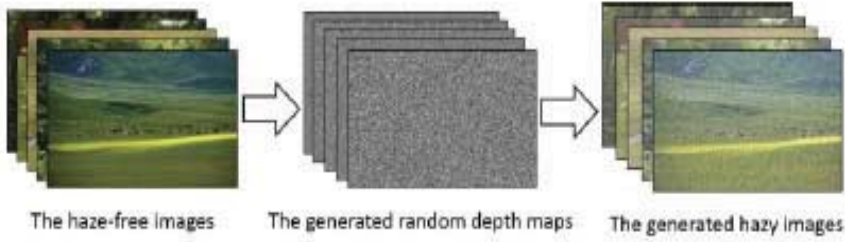


Figure 3. Layered decomposition of the haze

For the joint probability density function, assume that the probability of error per pixel is independent, then

$$L = \prod_{i=1}^n p(d(x_i) | x_i, \theta_0, \theta_1, \theta_2, \sigma^2) \tag{15}$$

Do Maximum likelihood estimation to (15), and operate Ln to both sides and find extremum :

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n dg_i - (\theta_0 + \theta_1 v(x_i) + \theta_2 s(x_i)) \tag{16}$$

$$\frac{\partial L}{\partial \theta_0} = \frac{1}{\sigma^2} \sum_{i=1}^n (dg_i - (\theta_0 + \theta_1 v(x_i) + \theta_2 s(x_i))) \tag{17}$$

$$\frac{\partial L}{\partial \theta_1} = \frac{1}{\sigma^2} \sum_{i=1}^n v(x_i) (dg_i - (\theta_0 + \theta_1 v(x_i) + \theta_2 s(x_i))) \tag{18}$$

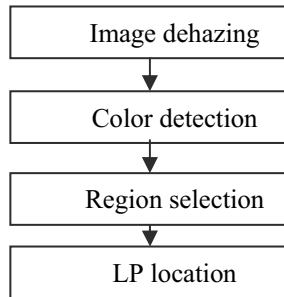
$$\frac{\partial L}{\partial \theta_2} = \frac{1}{\sigma^2} \sum_{i=1}^n s(x_i) (dg_i - (\theta_0 + \theta_1 v(x_i) + \theta_2 s(x_i))) \tag{19}$$

According to (16)-(19),  $\theta_0, \theta_1, \theta_2$  can be estimated by the gradient descent algorithm, where  $\theta_0=0.121779, \theta_1=0.959710, \theta_2=-0.780245, \sigma=0.041337$ . After determined these correlation coefficients, the depth  $d$  could be obtained. Since  $t=e^{-\beta d}$ ,  $\beta$  is the scattering coefficient an indicates the ability of dehazing. In the majority situation  $\beta=1.0$  will be suitable.

### 3. License plate location method in haze

License plate location methods are usually divided into the following two categories: one is based on the edge detection [7-9], the other is based on color feature [10,11]. In addition, LPR methods based on single feature are limited, [12-14] proposed methods based on both edge and color features.

The advantage of color detection is accuracy and make full use of the information. If we do edge detection directly, the image gray-scale transformation will lose some useful information. Therefore, the progress of the haze image LPR we use could be shown as Figure 4.



**Figure 4.** The progress of the haze image LPR

- Image dehazing: Input haze images, and output the haze free images.
- Color detection: Since the license plates in China are bluest and yellow, select these specific color areas can improve the detection speed. Use HSV color space to detect the color.
- Region selection: Weed out small and oversized areas, the threshold can be set up by the percents of the image size.
- LP location: Crop the region with obvious vertical edges.

### 4. Results

We use the three dehazing algorithms PD, DCP and CAP introduced in section 2, as the pre-processing method before license plate location. The experiments have been carried out on the output haze free images. Figure 5 shows the results of license plate location of two groups with original image, PD dehazed image, DCP dehazed image and CAP dehazed image respectively.

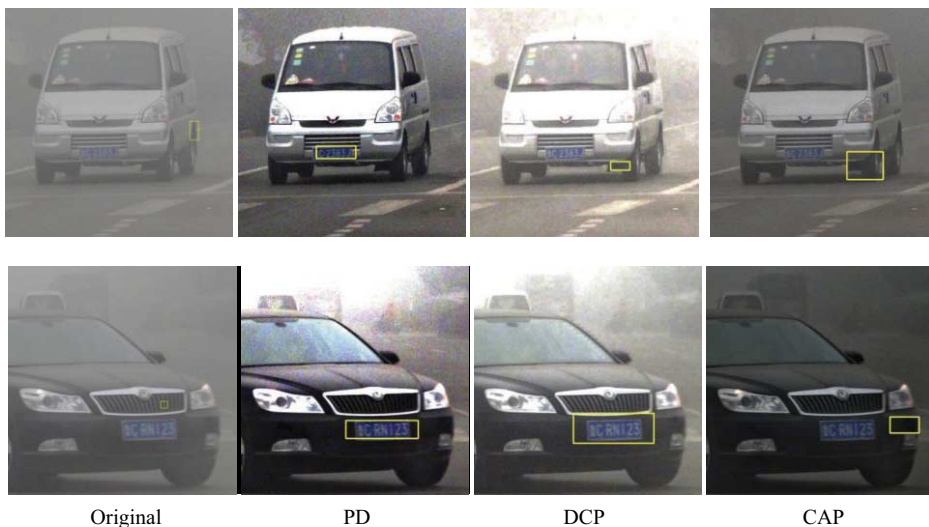


Figure 5. Haze image license plate location results

It can be seen that for the original haze images, the LP location algorithm can hardly locate on the right license plate region, because both the color and the edge information lost. The DCP and CAP algorithms belongs to image enhancement methods; image enhancement sometimes leads to color distortion when enhance the contrast. For DCP algorithm when the car is white, the color correction is biased because the dark channel is not suitable for white object. For CAP algorithm, the color restoration is not accurate for the HSV color detection. While PD algorithm is a kind of image restoration method, it uses the backscattering information in haze image to reduce color deviation. The results show that the PD restoration dehazing algorithm plays a effective performance in the pre-processing of LPR.

## 5. Conclusion

Haze image license plate recognition is an important technology in intelligent transportation. Since the accuracy of license plate localization algorithm associated with use of the characteristics detection method, multi-characteristics including both color and grayscale characteristics methods could improve the accuracy of license plate localization greatly. In this paper we combined the dehazing algorithm and the multi-characteristics based LP location algorithm as the haze image LP location method. All the dehazing methods play a good performance in strengthening the edge enhancing the contrast. The enhancement algorithms somewhat are lacking in color fidelity, while the restoration algorithm can effectively restore the true color information, and provides guarantee for color feature extraction in LP R system.

## References

- [1] R. Tan, "Visibility in bad weather from a single image" in Proceedings of IEEE Conference on CVPR. IEEE (2008), 1-8.
- [2] R. Fattal, "Single Image Dehazing." *ACM Siggraph* **8**(2008), 1-9.
- [3] J.P.Tarel and Nicolas Hautière, "Fast Visibility Restoration from a Single Color or Gray Level Image". Proceedings of IEEE International Conference on Computer Vision IEEE, (2009).2001-2008.
- [4] Wang Rui and Wang Guoyu, "Single image recovery in scattering medium by propagating deconvolution" *Optics Express*, **22**(7), (2014), 8114-8119.
- [5] Kaiming He, Jian Sun, and Xiaoou Tang, "Single image haze removal using dark channel prior," in Proceedings of IEEE Conference on CVPR. IEEE (2009) .1956-1963
- [6] Q. Zhu, J. Mai, and L. Shao, "Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior," *IEEE Trans On Image Processing*, **24**(11), (2015)3522-3533.
- [7] F. Martin, M. Garcia, L. Alba. New methods for automatic reding of VLP. In Proc. IAST Int. Conf. SPPRA.2002
- [8] Enyedi, B. Konyha, L. Szombathy, C. Fazekas. Strategies for fast license plate number localization, *Electronics in Marine*, (2004),579-584.
- [9] Zheng Danian, Zhao Yannan, Wang Jiabin, "An efficient method of license plate location," *Pattern Recognition Letters* **26**(2012), 2431-2438.
- [10] Zheng, Chengyong. A novel license plate location method on RGB color space. *Journal of Image & Graphics* (2010),1623-1628
- [11] CHEN Changtao et.al. Application of Lab spaces color segmentation in fast vehicle license plate location *Lab Application Research of Computers* **27**(2010):3191-3193.
- [12] Zhao, Bing. License plate location based on mixed characteristics. *Computer Engineering & Design* **28**(2007),5668-5670.
- [13] Jie, Guo, and S. P. Fei. Color and Texture Analysis Based Vehicle License Plate Location. *Journal of Image & Graphics* **7** (2002):472-476.
- [14] Pan, Wei, and N. Yang. A new method of vehicle license plate location under complex scenes. *International Conference on Advanced Computer Control IEEE Xplore*, (2010),134-138.

# Application of Grey Correlation Matching Model on Informatizational Assessment of Ports

Min ZHANG<sup>1</sup>

*Shipping Department of Economic Management, Tianjin Maritime College, Tianjin, China*

**Abstract.** Ports are quickening the pace of “intelligent port” with informatization exhibiting a strong competitive edge in the area of port operation. One of the challenges in the development is how to analyze the performance of informatizational applications in order to determine the main factors that affect the quality of them. According to the complex port informatizational condition and incomplete data collection, the Grey Correlation Matching Model is proposed to assess the informatization status of ports. This paper constructed the evaluation index, evaluated the index’s attributes combining the Analytic Hierarchy Process (AHP), and analyzed the performance of 11 ports to rank order by using the Grey Correlation Matching Model. Compared with the grey target model, the results show that the improved algorithm is more advantageous in dealing with the informatizational assessment. Furthermore, such results can be used in future resource planning stages, at the time of improve the intelligent ports.

**Keywords.** Intelligent port, Grey Correlation Matching Model, port informatization

## 1. Introduction

With the development of global economic integration, the new technology develops rapidly in the direction of scale, synergy and intelligence. As the node of a variety of modes of transport, each port might have multi-party information integration. The intelligent and information technology of port has become not only an important means to enhance the core competitiveness, but also the key to reduce logistics costs and improve logistics efficiency. Consequently, it is necessary to assess the informatization condition of ports. Therefore, the use of appropriate analytical methods to assess the port information, for the intelligent port of reasonable planning and thus further promote the development of the port is of great significance. Therefore, it is of great significance for the intelligent port to evaluate port informatization using suitable analysis method, in order to adjust and reform according to the needs of the development of port stability. In view of the incomplete information due to the imperfect index of port informatizational evaluation and instability of the relevant data, the informatizational assessment can be regarded as a grey system.

---

<sup>1</sup> Min Zhang, Shipping Department of Economic Management, Tianjin Maritime College, Tianjin, China; E-mail: tiantianlou@163.com.

In the port informatization research, the relevant research methods mainly include DEA Model, DEAHP Model and grey target model. These methods has achieved some success. Ma Dongmei [1] analyzed the influencing factors, established the performance evaluation system, defined the volume of the input-output, and discussed the level of port logistics informatization together with the DEA model. Zhangmin and Du Xiangran [2] constructed the evaluation index of the harmonious development between port economic growth and information industry, and uses an interactive DEAHP model for evaluating the port informatization performance, which would relatively comprehensively reflect the weights among all decision making units and among index's attributes. Zhangmin [3] introduced the Weighted Grey Target theory to assess the informatization status of ports, which concluded that Port of Shanghai had won the highest score belonging to the assessment Class I area, while Port of Lianyungang, Yantian, Jinzhou, Jiaxing and Wenzhou had got low scores belonging to the assessment Class IV area.

The gray correlation matching model (GCMM) developed by Zhu Qingxiang, Cao Qiufang, Liu Jing, Ji Haipeng, Wang Zhipeng and Lei Siyuan is a analysis method of identification, selection, classification and rank for commercial site location [4]. Based on the gray relational analysis method, Euclidean distance is introduced into the gray correlation matching model. The improved algorithm overcomes the shortcomings of gray relational analysis which is only analyzed from unilateral consideration. The example shows that the model clears the proximity between the data columns and improves the site selection efficiency, compared with the traditional method. However, together with the GCMM, the quantitative research on the level of port informatization development is still very rare.

## 2. Theory of Gray Correlation Matching Model

The gray correlation matching model is a grey relational analysis theory for pattern series processing, which can realize the quantization and ordering among the factors for grey system with uncertain information. The purpose of modal is to show the relative proximity of the selected points, helping the user to find the optimal choice in a short time.

The calculation steps of analysis method of GCMM are expressed as follows:

### 2.1. Establish the Effect of Sample Matrix

Suppose that there are  $n$  decision-making units,  $m$  evaluating indicators, and index set  $A_i = \{A_1, A_2, \dots, A_m\}$ ,  $i \in M = \{1, 2, \dots, m\}$ .

The raw data is standardized by the range-method. The effect of sample matrix is labeled as  $X = (x_{ij})_{m \times n}$ .

- If  $A_m$  is an efficiency indicator, then

$$x_{ij} = \frac{y_{ij} - \min \{y_{ij}\}}{\max \{y_{ij}\} - \min \{y_{ij}\}}, (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (1)$$

is referred to as the effect measure of efficiency indicator.



- If  $A_m$  is a cost indicator, then

$$x_{ij} = \frac{\max \{y_{ij}\} - y_{ij}}{\max \{y_{ij}\} - \min \{y_{ij}\}}, (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \tag{2}$$

is referred to as the effect measure of cost indicator.

### 2.2. Determine the Index Weight

The weight vector  $\omega = (\omega_1, \omega_2, \dots, \omega_m)^T$  was determined by AHP method, where

$$\sum_{j=1}^m \omega_j = 1 \quad (j = 1, 2, \dots, m).$$

### 2.3. Calculate the Weighted Normalized Decision Matrix

Set,  $Z = (z_{ij})_{nm}$ , then  $z_{ij} = x_{ij} \times \omega_j, (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$ . (3)

### 2.4. Determine the Positive Solutions and Negative Solutions of Weighted Normalized Matrix

Set,  $z^+ = (z_1^+, z_2^+, \dots, z_n^+)^T, z^- = (z_1^-, z_2^-, \dots, z_n^-)^T$ , then  $z_i^+ = \max_j \{z_{ij}\}, z_i^- = \min_j \{z_{ij}\}$ . (4)

### 2.5. Calculate the Gray Correlation Coefficient Matrix Between Evaluation Samples and Positive & Negative Ideal Solutions

The decision matrix is labeled as  $R = (r_{ij})_{m \times n}$ , where

$$r_{ij} \in [-1, 1], (i = 1, 2, \dots, n; j = 1, 2, \dots, m).$$

Set,  $R^+ = (r_{ij}^+)_{m \times n}, R^- = (r_{ij}^-)_{m \times n}$ , then

$$r_{ij}^+ = \frac{\min |z_i^+ - z_{ij}| + \xi \max |z_i^+ - z_{ij}|}{|z_i^+ - z_{ij}| + \xi \max |z_i^+ - z_{ij}|},$$

$$r_{ij}^- = \frac{\min |z_i^- - z_{ij}| + \xi \max |z_i^- - z_{ij}|}{|z_i^- - z_{ij}| + \xi \max |z_i^- - z_{ij}|}. \tag{5}$$

Where  $\xi$  is the resolution coefficient ( $0 < \xi < 1$ ), when  $\xi \leq 0.5463$ , the resolution is the best. Generally,  $\xi$  is 0.5.

2.6. Calculate the Gray Correlation Degree

$$\text{Set, } r_i^+ = \frac{1}{m} \sum_{j=1}^m r_{ij}^+, \text{ and } r_i^- = \frac{1}{m} \sum_{j=1}^m r_{ij}^-. \tag{6}$$

2.7. Calculate the Euclidean Distance from the Evaluation Samples to the Positive & Negative Ideal Solutions

$$\text{Set, } d_i^+ = \sqrt{\sum_{j=1}^m (z_{ij} - z_j^+)^2}, \text{ and } d_i^- = \sqrt{\sum_{j=1}^m (z_{ij} - z_j^-)^2}. \tag{7}$$

2.8. Correlation and Euclidean Distance Dimensionless Processing

$$\text{Set, } R_i^+ = \frac{r_i^+}{\max r_i^+}, R_i^- = \frac{r_i^-}{\max r_i^-}, D_i^+ = \frac{d_i^+}{\max d_i^+}, \text{ and } D_i^- = \frac{d_i^-}{\max d_i^-}. \tag{8}$$

2.9. Calculate the Index Comprehensive Correlation Degree

Set,  $S_i^+ = \alpha R_i^+ + \beta D_i^-$ ,  $S_i^- = \alpha R_i^- + \beta D_i^+$ , and  $i \in M$ .  $\alpha$  and  $\beta$  reflect the factors of preference degree, where  $\alpha + \beta = 1$ , and  $\alpha, \beta \in [0, 1]$ . Policymakers are allowed to determine the value according to the their own preferences. Generally,  $\alpha = \beta = 0.5$ .  $S_i^+$  and  $S_i^-$  reflect the close of the evaluation sample and the ideal solution. The larger the value of  $S_i^+$ , the better the sample; the larger the value of  $S_i^-$ , the worse the sample.

2.10. Calculate the Relative Proximity of the Sample

$C_i^+$  stands for the relative proximity of the sample.

$$C_i^+ = S_i^+ / (S_i^+ + S_i^-) \quad (i = 1, 2, \dots, n). \tag{9}$$

The larger the value, the better the sample it is.

**3. Application Examples**

In this paper, the data of 11 ports in China is used as the sample source, where some of the data were collected from the China's Ports-of-Entry Yearbook, and some were captured through research [1]. By analyzing the evaluation indexes of each sample point, the gray relational matching model is applied to research the informatization status of

**Table 1.** Data statistics of stylebooks.

Port	Information Technology Investment (Billion Yuan) <sup>a</sup>	Administrative Expenses (Billion Yuan) <sup>a</sup>	Net Profit (Billion Yuan) <sup>a</sup>	Operating Income (Billion Yuan) <sup>a</sup>	Import & Export Cargo Volume (Thousand Tons) <sup>b</sup>	Container Volume (Thousand Boxes) <sup>b</sup>
Tianjin	0.2	0.22752	1.007	12.7	159289	4629
Shanghai	0.46	0.64075	4.724	21.6	169046	22103
Ningbo	0.1	0.2289	2.468	6.98	130216	8921
Lianyungang	0.1	0.1598	0.149	1.5	57182	656
Qingdao	0.086	0.22005	0.98	4	148460	1904
Dalian	0.1	0.18797	0.845	3.95	69047	4546
Jiaxing	0.005	0.0211	0.1	0.7	5588	195
Wenzhou	0.004	0.02	0.084	0.7	4650	139
Rizhao	0.03	0.0486	0.504	3.45	111034	170
Yantian	0.02	0.09198	0.477	0.36	34175	7975
Jinzhou	0.01	0.04332	0.215	1.13	6328	9

<sup>a</sup> Data from [1]. <sup>b</sup> Data from China’s Ports-of-Entry Yearbook.

**Table 2.** Evaluating indexes system.

Target Layer	Decomposed Layer	Index Layer	Nature
Evaluating Indexes	Informationization Indicators	Information Technology Investment	+
		Administrative Expenses	-
		Net Profit	+
	Port Economic Indicators	Operating Income	+
		Import & Export Cargo Volume	+
		Container Volume	+

ports, in order to find some meaningful methods for the field of research. From the Table 1, the effect of sample matrix can be recognized.

Port information system includes the port public information service platform, data exchange platform and application system established by governments, ports, logistics and other relevant enterprises. The core is the data exchange platform which provides data sharing and exchange between the various application systems, to seamlessly connect the entire port-related business. The evaluation index should be a set of intrinsically linked indicators to evaluate the status and characteristics of Information Industry. Therefore, the paper select information technology investment, administrative expenses, net profit, operating income as informationization indicators.

Port economic development mainly refers to the current development of port. Port throughput and container throughput are unquestionably the most important and widely accepted indicators of port or terminal output. This paper selects indicators related to the level of port economic development, such as Import & Export cargo volume and container volume, to evaluate the port economic development.

So the following six objectives are considered: information technology investment, administrative expenses, net profit, operating income, import & export cargo volume and container volume, which are shown in Table 2.

**Table 3.** Judgment sheet of index.

	Information Technology Investment	Administrative Expenses	Net Profit	Operating Income	Import & Export Cargo Volume	Container Volume
Information Technology Investment	1	2	1/3	1/3	1/2	1/2
Administrative Expenses	2	1	1/3	1/2	1/2	1/2
Net Profit	3	3	1	2	3	3
Operating Income	3	2	1/2	1	2	2
Import & Export Cargo Volume	2	2	1/3	1/2	1	2
Container Volume	2	2	1/3	1/2	1/2	1

Among these objectives, information technology investment, net profit, operating income, import & export cargo volume and container volume belong to “efficiency indicator”, which means that the higher evaluation score, the better it is. Administrative expenses belongs to “cost indicator”, which means that the lower evaluation score, the better it is.

Based on the development of information technology of ports, experts of the port determined the index weights by pair wise comparison in accordance with the index of the scaling law [5].

The indicators judgment matrix is shown in Table 3.

The AHP method is used to determine the weighted normalized decision matrices. The detailed procedure is omitted here.

According to the improved gray relational analysis steps described above, the relative proximity is calculated as follows:

$$Z = \begin{bmatrix} 5.74 & 5.25 & 1.13 & 1.87 & 3.19 & 4.23 \\ 13.00 & 12.00 & 2.83 & 4.83 & 7.50 & 9.00 \\ 5.69 & 5.26 & 1.24 & 2.16 & 3.40 & 4.37 \\ 1.56 & 1.49 & 0.30 & 0.44 & 0.70 & 1.04 \\ 3.84 & 3.53 & 0.77 & 1.26 & 2.09 & 3.01 \\ 2.94 & 2.72 & 0.61 & 1.00 & 1.57 & 2.06 \\ 0.09 & 0.08 & 0.02 & 0.03 & 0.05 & 0.06 \\ 0.06 & 0.04 & 0.01 & 0.02 & 0.03 & 0.04 \\ 2.17 & 2.03 & 0.42 & 0.70 & 1.26 & 1.92 \\ 1.60 & 1.52 & 0.32 & 0.51 & 0.69 & 1.05 \\ 0.30 & 0.24 & 0.07 & 0.12 & 0.19 & 0.20 \end{bmatrix}$$

The gray correlation coefficient matrix between the evaluation samples and the positive & negative ideal solutions are calculated as follows:

$$R^+ = \begin{bmatrix} 0.47 & 0.47 & 0.45 & 0.45 & 0.46 & 0.48 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.47 & 0.47 & 0.47 & 0.47 & 0.48 & 0.49 \\ 0.36 & 0.36 & 0.36 & 0.35 & 0.35 & 0.36 \\ 0.41 & 0.41 & 0.41 & 0.40 & 0.41 & 0.43 \\ 0.39 & 0.39 & 0.39 & 0.39 & 0.39 & 0.39 \\ 0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.39 \\ 0.36 & 0.36 & 0.36 & 0.36 & 0.35 & 0.36 \\ 0.34 & 0.34 & 0.34 & 0.34 & 0.34 & 0.34 \end{bmatrix}, \text{ and } R^- = \begin{bmatrix} 0.53 & 0.53 & 0.56 & 0.57 & 0.54 & 0.52 \\ 0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.33 \\ 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 0.51 \\ 0.81 & 0.80 & 0.83 & 0.85 & 0.85 & 0.82 \\ 0.63 & 0.63 & 0.65 & 0.66 & 0.64 & 0.60 \\ 0.69 & 0.69 & 0.70 & 0.71 & 0.71 & 0.69 \\ 0.99 & 0.99 & 0.99 & 0.99 & 0.99 & 0.99 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.75 & 0.75 & 0.78 & 0.78 & 0.75 & 0.70 \\ 0.81 & 0.80 & 0.82 & 0.83 & 0.85 & 0.82 \\ 0.96 & 0.97 & 0.96 & 0.96 & 0.96 & 0.96 \end{bmatrix}.$$

The grey correlation degree and the Euclidean distance from the evaluation samples to the positive & negative ideal solutions can be calculated according to (6) and (7). Through the correlation and Euclidean distance dimensionless processing, the index comprehensive correlation degree can be calculated according to (8). The rank of assessment object depends on the size of the Relative Proximity of the Sample. The larger the value of  $C_i^+$ , and the better the port informatization performance. Put  $C_i^+$  into the order from large to small, we can get the sort results of port informatization evaluation, which are shown in Table 4. Therefore, the port informatization evaluations can be recognized, which are shown in Table 4. The maximum value is 0.857 calculated by the improved algorithm. This study illustrates that Shanghai port is the optimal for informatization environments.

#### 4. Effectiveness Analysis

It can be seen from the comparison results of Table 5 that the order has not changed, except Rank 7th and Rank 8th. The order of Lianyungang port and Yantian port has obviously changed compared with the weighted grey target theory.

According to the traditional method, the target distances of Lianyungang port is smaller, which means Lianyungang port is significantly better than Yantian port. While the improved method is used to calculate the relative proximity, the conclusion shows

**Table 4.** Results of calculation.

Port	$R_i^+$	$R_i^-$	$D_i^+$	$D_i^-$	$S_i^+$	$S_i^-$	$C_i^+$
Tianjin	0.465	0.542	6.921	0.438	0.452	3.731	0.108
Shanghai	1.000	0.333	0.000	1.000	1.000	0.167	0.857
Ningbo	0.475	0.528	6.714	0.446	0.461	3.621	0.113
Lianyungang	0.358	0.827	17.232	0.113	0.236	9.030	0.025
Qingdao	0.412	0.637	10.889	0.295	0.353	5.763	0.058
Dalian	0.389	0.699	13.285	0.220	0.305	6.992	0.042
Jiaying	0.334	0.995	21.737	0.003	0.168	11.366	0.015
Wenzhou	0.333	1.000	21.854	0.000	0.167	11.427	0.014
Rizhao	0.375	0.753	15.044	0.172	0.273	7.899	0.033
Yantian	0.359	0.822	17.131	0.115	0.237	8.976	0.026
Jinzhou	0.338	0.963	21.052	0.019	0.178	11.007	0.016

**Table 5.** Experimental results by comparison.

Port	Grey target theory		Grey correlation matching model	
	Results of target Distances <sup>a</sup>	Order	Results of relative proximity	Order
Tianjin	0.768	3	0.108	3
Shanghai	0.407	1	0.857	1
Ningbo	0.708	2	0.113	2
Lianyungang	1.167	7	0.025	8
Qingdao	0.935	4	0.058	4
Dalian	1.018	5	0.042	5
Jiaying	1.314	10	0.015	10
Wenzhou	1.319	11	0.014	11
Rizhao	1.041	6	0.033	6
Yantian	1.172	8	0.026	7
Jinzhou	1.297	9	0.016	9

<sup>a</sup> Data from [3].

that Yantian port is better than Lianyungang port. The improved algorithm is more advantageous, as follows in Table 5.

Considering the actual situation and the perspectives for development of the port informatizational condition, Yantian port has promoted the cloud computing industry through “container intelligent logistics platform”, which can achieve the convenience of container logistics. The platform can realize intelligent scheduling, track goods real-time through GPS positioning and deliver the goods on line, which is conducive to improving the efficiency of the operations and regulating the supervision and services of ports. So it is a relatively appropriate result that the Yantian port ranked seventh, which is also consistent with the improved calculation results. Therefore, the result shows that the GCMM is a simple and effective method, which has advantages in terms of confirming the informatizational level.

Considering the actual situation and the perspectives for development of the port informatizational condition, Yantian port has promoted the cloud computing industry through “container intelligent logistics platform”, which can achieve the convenience of container logistics. The platform can realize intelligent scheduling, track goods real-time through GPS positioning and deliver the goods on line, which is conducive to improving the efficiency of the operations and regulating the supervision and services of ports. So it is a relatively appropriate result that the Yantian port ranked seventh, which is also consistent with the improved calculation results. Therefore, the result shows that the GCMM is a simple and effective method, which has advantages in terms of confirming the informatizational level.

## 5. Conclusion

In this paper, the Grey Correlation Matching Model was applied to evaluate and assess the performance of informatizational port. The performance analysis results are very important because they help to detect the source of the degradation of the informatizational port. Good results were obtained for the classification rate and algorithm adaptability, which indicated that the GCMM is available to sort the port informatizational level. GCMM is a promising method for port informatizational status assessment, but

some indicators data are difficult to get. Therefore, specific quantitative standards remains to be further studied to solve the problem of evaluation design feasibility.

## References

- [1] Ma Dongmei, "The Evaluation and Research of Port Logistics Informatization Efficiency Based on DEA Modeling", *Journal of Dalian Maritime University*. vol. 6, (2012), 38–40.
- [2] Zhang Min, Du Xiangran, "Evaluation of Economic Coordination of Port Informatization", *Proceedings of 2013 International Conference of Information Technology and Industrial Engineering*, Wuhan, China (2013) August 7–8.
- [3] Zhangmin, "Application of Weighted Grey Target Theory on Informatizational Assessment of Ports", *Proceedings of International Conference on Cyberspace Technology*, CCT 2014, Beijing, China (2014) November 8–10.
- [4] Zhu Qingxiang, Cao Qiufang, Liu Jing, Ji Haipeng, Wang Zhipeng and Lei Siyuan, "Research on Optimal Location of Commercial Network Based on Improved Gray Relation", *Journal of Statistics and Decision*, vol. 8, (2016), 51–53.
- [5] Liu Cuilian, Liu Jianmei, Liu Nannan, and Wang Siting, "Application of DPSIR Model in Evaluation of Eco-port Group," *Journal of Shanghai Maritime University*, vol. 33, no. 6, (2012), 61–64.

# The Systematic Design of Road Traffic Signal Recognition Based on the Integration of ZigBee with Active FRID

Ziyi LIANG <sup>a,1</sup> Ping XU <sup>b</sup>

<sup>a</sup>*Changzhou College of Information Technology, Changzhou 213164, China*

<sup>b</sup>*Changzhou College of Information Technology, Changzhou 213164, China*

**Abstract.** In order to improve the safety of motorists and to realize driving automation and unmanned, a road traffic signal recognition system based on the integration of ZigBee network and FRID is designed. Firstly, the devices of Radio Frequency Identification (RFID) is installed on the original road signs, traffic lights and road markings. It can transmit traffic signals, which are consistent with the original encoded information. At the same time, the device is the terminal node of ZigBee network. Then, routers and coordinator nodes are installed in the necessary place. At last, the network of road condition information is formed in the way of ad-hoc network. When the cars equipped with traffic signal recognition system are driving in this network, they can receive encoded information of traffic signal and converts the received traffic signal encoded information into a voice signal, or converts it into a vehicle control signal to control the vehicle. The system has some good points, with a high recognition rate, long recognition distance, real-time, secure, reliable, easy to implement and so on.

**Keywords.** ZigBee, FRID, traffic signals, Intelligent Transportation, Information tunnel

## 1. Introduction

With the rapid development of the world economy and the increasing car popularizing, people are nagged by a series of problems caused, such as traffic jam, environment contamination, the increase of energy consumption, frequent occurrence of traffic accidents, lower transportation efficiency, etc. Because of limited land resources, it is unlikely to build new roads or expand old ones unboundedly, and measures as limiting the purchase of motor vehicles or odd-and-even license plate rule are just temporary measures. As a result, Intelligent Transportation System (ITS) has become the focus of research in some countries. The purpose is that vehicles can run freely and the traffic can be adjusted to optimal state.

Vehicle Control System is a major branch of Intelligent Transportation System, which is called Intelligent Car, because it helps drivers driving or realize unmanned operation. As visual information is most important road traffic scene information, Intelligent Car must have keen vision to recognize driving road, abnormal collision,

---

<sup>1</sup> Corresponding Author, Changzhou College of Information Technology, Changzhou, Jiangsu, China; email: laoliang\_99@163.com.



and traffic sign. For abnormal collision recognition, the common method is to install radar or infrared detector on a car, at the front and side, to decide the accurate distance between vehicle and obstacle. With emergency, onboard computer can sound an alarm or brake automatically. While traffic sign recognition mainly depends on image identification, deciding whether the sign exists. With color CCD camera, Videokarte, high speed data acquisition is achieved, and stored. The method of feature extraction is used to decide the existence of traffic sign among large-scale images. If the traffic sign does exist, further analysis will be done by pattern recognition. Due to factors, such as pollution, light, distance, angle, position, vibration, and occlusion, etc. the accuracy is around 80%. And because of the complicated computation, poor real-time, demand of high-speed CPU and comparatively large storage, the image recognition is not practical and popular. Furthermore, since traffic lights emit light actively, the computation of recognition is different from that of traffic signs.

In fact, in addition to keen vision, Intelligent Car can also have sharp auditory system. The author designed a Region Traffic Signal Control System (traffic signals includes traffic signal lamps, traffic signs, traffic line markings and the command of traffic policeman) based on the integration of Zigbee network and FRID, which provides Intelligent Car with auditory system. The basic idea is to install Rf transmitters on original Road Signs, traffic lights, and road markings, which can send coded information in line with original traffic signals. Rf transmitters being the terminal node of Zigbee network, router and coordinator node are installed to form the road conditions information network, or Information Tunnel. While passing the Tunnel, cars with the Region Traffic Signal Control System will transfer the received coded information to audio messages or signals to control the cars. In this traffic signals are not searched or found by Intelligent Car, but provided by traffic signal equipment instead. Therefore, the accuracy, real-time and reliability are greatly enhanced. The system can be applied to various city roads and expressways, which offers an effective way to help Intelligent Car with its utility and marketization.

## **2. Integration of ZigBee Network and RFID.**

### *2.1. Brief introduction of Zigbee*

Zigbee is a two-way wireless communications technology or wireless network technology (featured by close range, low complexity, and low power consumption, low data rate, and low cost), using the IEEE 802.15.4 protocol, similar to CDMA and GSM networks. Different from CDMA or GSM network, Zigbee network is established to realize automatic data transmission. Each Zigbee network node itself can be monitoring target, (for example, the data from its connected sensor can be collected and monitored), or automatically transfer data transmitted from other network nodes. Its data transmission module ranges from tens of meters to hundreds of meters, to several kilometers, and it supports infinitive expansion. Zigbee has the following features: low cost—all frequency bands are opened free, protocols simple and free; large capacity—a Zigbee network supports 255 equipments; safe-- use advanced AES — 128 encryption algorithm, and provides inspection and authentication of data integrity. There is no need to worry the conflict of signals because of its multi-access of carrier sense or

CSMA/CA. Zigbee has rapid self-organizing ability, which can conveniently network and rapidly establish wireless connections among all devices.

## *2.2. Brief introduction of RFID*

RFID can transform traditional tags electronically to communicate between reader and electronic tags by utilizing RF signals and space coupling. It is an automatic-recognition technology between reader and electronic tags. RFID system consists of electronic tags, reader, and application software system. RFID tags are often classified as passive ones and active ones, according to the origin of power. Passive RFID tags have no power, and its needed power is gained from the process of converting electromagnetic waves sent by reader, which requires the reader's transmitting power is great enough. Whereas in active RFID, tag and reader are in magnetic field, and signals are transmitted by coupling of electromagnetic waves. Active RFID tags have self-contained power pack, and readers don't have to send activated signals or instructions while reading the tag, which will actively send its stored U I D and other information to reader in the form of electromagnetic wave. Therefore, RFID with power is also called active RFID, its communication distance ranges from centimeters to hundreds of meters. The Region Traffic Signal Control System discussed here is active RFID.

## *2.3. Integration of ZigBee and RFID*

ZigBee can be utilized to network RFID reader or electronic tag, so as to combine their advantages. The way to integrate ZigBee and RFID can be categorized into the integration of electronic tag and ZigBee node, and the integration of RFID reader and ZigBee node. Before integration, electronic tag can only communicate with RFID reader. But after the integration, electronic tag can not only communicate with RFID reader, but also with other nodes. Via the multi-hop network of ZigBee, data can be transmitted further. One or more RFID readers can communicate with all electronic tags.

## **3. Systematic structure**

The systematic structure is shown by Figure 1, which consists of the following parts: traffic signal launcher, undertaken by RFID tag; on-board traffic signal receiver, undertaken by RFID reader; on-board media player and automatic driving controller; GPS and ZigBee network positioning module; ZigBee network, including terminal node, routing node, coordinator, among which terminal node is integrated with active RFID tag and installed on original traffic signal equipment. Routing nodes are installed where relay is needed. Terminal nodes join the coordinator network, reporting its node.

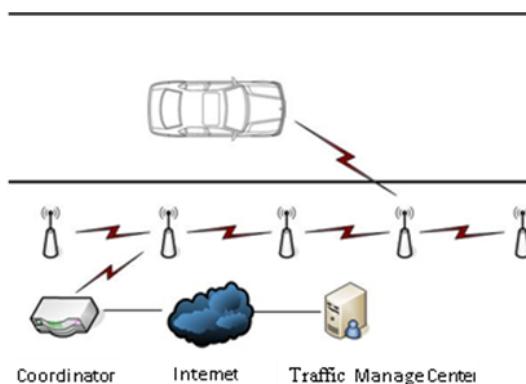


Figure 1. Systematic structure

operation to coordinator, receiving orders and data from coordinator, and terminal nodes can also be set as routing nodes when necessary; routing nodes forward data between terminal nodes and coordinator, which can extend network. Coordinator node, full function device, is the center of the whole network, responsible for the establishment and management of the network. Through the coordinator, customers configure and run the network. Traffic signals from all terminal nodes are set by coordinator, and all operation and fault information are collected by coordinator and later reported to control center. Coordinator can be connected with host computer via a serial port, or connected to the control center via internet.

Traffic control center can monitor the operation of traffic signal in real time via LED big screen and monitoring software. Since traffic signs and markings are usually stay still, data are only transmitted to or intervened by traffic control center when traffic signs are added or altered, or in case of new road marking due to reconstruction, or when traffic signal device is out of order. Usually traffic lights are subjected to coordination control of ZigBee, as all instructions are transmitted to the corresponding controller via ZigBee network, timing calibration remote controller, which remotely changes signal light phase and updates signal light timing strategy, coordinating with how long the lights should stay green, what the cycle time should be and when to turn on the green light, etc. Via ZigBee network, traffic control center releases the guidance information and current traffic conditions. Data about traffic flow or condition data of designated site can be collected by ZigBee as well.

#### 4. The design of system hardware

##### 4.1. The design of integration between ZigBee terminal node and RFID active tag

Figure 2 displays the structure of traffic signal transmitter, showing the layout of ZigBee network nodes and routing nodes. The transmitter has functions such as active RFID electronic tag, automatic networking, and wireless sensor networks nodes. The core of the node is chip CC2538SF53 from TI Company. In order to assure the reliability of the system and reduce the varieties of chips, transceiver chips of RFID also adopt CC2538SF53 from TI Company. CC2538SF53 is the ideal wireless microcontroller SoC suitable for the high-performance ZigBee. It has the powerful 32-

bit MCU based on ARM Cortex™ -M3, with the clock speed of 32MHz, 32KB On-Chip RAM and 512KB Flash memory. 32 general-purpose input and output (GPIO) and serial peripheral interface can be connected to other parts of the circuit board. Security hardware accelerator can realize the authentication and encryption fast and effectively while CPU dealing with applications. It can be adapted to run under the bad environment, as it is capable of temperatures ranging from -40°C to 125°C. For the sake of smooth development, CC2538SF53 includes a powerful debugging system and a Comprehensive drive library. The integration of CC2538 and the solid and comprehensive Z-Stack software solution freely offered by TI can provide the most powerful, most stable solution for ZigBee. CC2538 is integrated with CC2500 RF front. CC2500 is RF transceiver chip of Zigbee/IEEE 802.15.4 launched by TI Company, operating frequency being 2394-2507MHz, responsible for the receiving and transmitting of RF signals, as the physical layer on the protocol. The chip is integrated in the form of a kernel module CC2538 RF, communicating between SoC and CPU via SPI bus, receiving sensitivity as low as -98dBm, Programmable output power as high as 7dBm, with 768-byte memory space, 128-byte AES encryption scheme, 128-byte FIFO forward buffer, 128-byte FIFO accept buffer, which all can meet the requirement of the system. To increase the communication distance, RF power amplifier CC2591 is added, which is a high-performance RF front chip designed by TI Company for low-voltage and low-power wireless communication, communication range can be extended by 15 times. The terminal node's main task is to join coordinator network, report to coordinator whether there is an error or not at fixed period, accept traffic signals or guidance information set by control center, and periodically transmit encoded data or real-time traffic information (transmission period can be set by car speed and road conditions). Tags are always active and power consumption is rather large. As a road infrastructure can be powered by a regulated power supply.

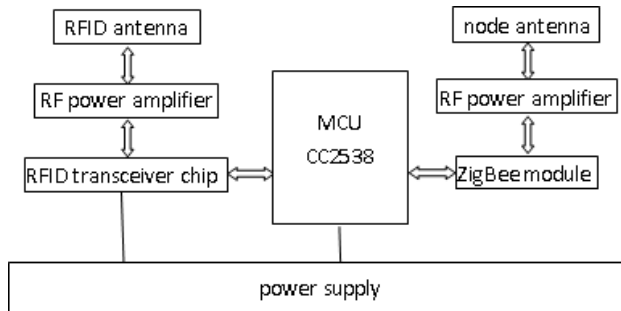


Figure 2 RFID tags and ZigBee nodes

#### 4.2. The Design of on-board RFID Reader and Control System

Figure 3 shows the structure of the on-board RFID Reader and Control System. The systematic design is based on modules, including RFID Reader RF, Reader main control, recording-voice integration & media player, automatic or unmanned driving control, GPS positioning, WSN positioning, and other auxiliary functions. The system works properly with the coordination of ARM controller, which adopts MCU 32-byte ARM Cortex™ -M integrated by CC2538SF53 chips.

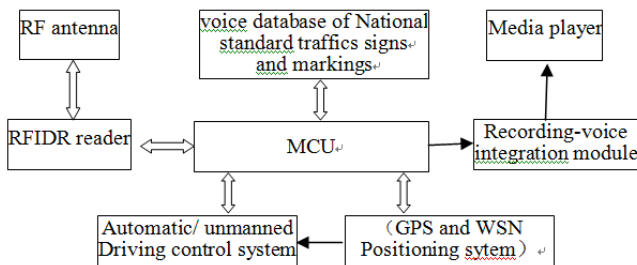


Figure 3. systematic structure of RFID reader and control.

- RFID RF Unit, Reader Main Control Unit**  
 The basic task for RFID Reader is to receive traffic signal code or guidance information sent by active RFID tags, to decode, to store & categorize data. Reader's RF Unit and Main Control Unit still adopt CC2538SF53 chips of TI Company, its hardware design almost the same as active electronic tags, and CC2591 is also added to Reader's front.
- Recording-voice integration & media player Unit Item**  
 While the car with reader entering the Information Tunnel, RF antenna sends traffic signals and guidance information to micro-processor via RFID transceiver chip. Software will decide on signal types. If it is standard traffic signs, or traffic markings, then visit the pre-recorded voice database of National standard traffic signs and markings, reading the corresponding voice data, and broadcasted by recording-voice integration & media player. If it is traffic light, then broadcast the color (red, green, yellow). If it is real-time traffic information, transmitting device sends voice data in MP3 code, which later is decoded by Recording-voice integration module and broadcast by media player. Therefore, drivers can get informed of traffic signal and road conditions without eyes off ground, so as to reduce the probability of the traffic accident.
- Automatic/ Unmanned driving control Unit**  
 Traffic signals and road conditions can assist drivers in driving, and control the car (including steering control, electronic air throttle, accelerator and auxiliary braking system, etc.) with the support of vision system, so as to realize automatic driving.
- GPS positioning Unit, WSN positioning Unit, and other auxiliary functions**  
 GPS and WSN positioning can be used for autonomous navigation and network navigation. Autonomous navigation means that in addition to aided positioning, vehicles can achieve navigation mission independently without other help. Network navigation can help communicate information with traffic information center via wireless communication network. Mobile devices, connected to GIS server via mobile communication network, perform functions as map storage and complicated calculation, etc. Customers can download map data from Server-side. The advantages of network navigation include more computing power, not bound by storage capacity.

## 5. The design of system software

System software includes ZigBee network software and onboard equipment software. ZigBee network software programs to meet customers' requirement in the form of protocol stack application, which mainly includes ZigBee coordinator's network construction, the network entry of router/ terminal node (active electronic tag), data exchange between router/terminal node and coordinator, information sending and receiving of router/ terminal, software management in control center, etc.

### 5.1. ZigBee node and active RFID tag software

ZigBee nodes can be divided into terminal node and router node. The embedded software of ZigBee node adopts Z-stack protocol stack of TI Company, and develops program in application layer. Terminal nodes are responsible for receiving data or instruction sent by coordinator, in two working modes: communication mode and sleep mode. For most of the time, nodes are in sleep state, and change into communication state while RFID module or coordinator is to send instruction. The node is integrated with RFID electronic tag, which is activated all the time and send almost the same traffic signal. Terminal nodes can work as router as well, transferring data from some nodes to other nodes when needed. This function can be found in protocol stack task.

### 5.2. Reader software

Reader's embedded programs are targeted for ARM Cortex™ -M32, which is used as a reader's controller, aiming to implement all operations of reader and other peripheral equipment. The operations mainly include: reader's initialization, control of reader chip's query tags, perform an anti collision algorithm , get the UID of the tag and the data sent by the tag . Reader will pack UID and data into data frames while reading an e-tag, and send to microprocessor of the main controller, which will broadcast or control vehicle according to these data.

There's not space to cover all the other software.

## 6. Conclusion

This system digitalizes the vision information in road traffic scene, and establishes an information tunnel for road conditions by means of the integration of ZigBee network with active RFID. Within this system, traffic signals are not sought or discovered by smart cars, but offered by traffic signal equipment. Therefore, the accuracy, real-time and reliability of traffic signal recognition have been greatly improved. The system can apply to different city roads and various express way, and finds a new way for smart cars' utility and marketization.

## References

- [1] Ziyi LIANG, Xinnan QU, Xianghui ZHAO "Exploration Based on ZigBee Traffic Signals Area Control System" *Applied Mechanics and Materials* Vols. 135-136 (2012) pp 535-539

- [2] LIANG Ziyi, Ma Zhenghua, Qu Xinnan “ Research on Wind Farm Remote Monitoring System Based on ZigBee and GPRS Technology”, *Computer & Digital Engineering* (2010,vol,38,No.11).
- [3] LIANG Long, WANG Chun-xue “A design of distributed attendance system based on RFID and Zigbee networks” *Manufacturing Automation* (2012,vol,34No.7)
- [4] LIU Yu,PENG Gang,WANG Tao,WANG Yan-qin “An Approach to Constructing the RFID System Network Based on the ZigBee Technology” *Journal of Guangxi Academy of Sciences*(2010,vol,26,No.4)
- [5] LI Rongkuan, TAN Fangming, ZHANG Lin “Design of RFID Reader Network Based on ZigBee Technology” *Programmable Controller & Factory Automation* 7(2010),99-93
- [6] JIA Yu lei, JIN Wu yin “Development of wireless control system of intelligentvehicle based on ZigBee technology” *Transducer and Microsystem Technologies*(2013,vol,32,NO.5),135-137
- [7] CC2538 SoC for 2.4-GHz IEEE 802.15.4 & ZigBee/ZigBee IP Apps User's Guide (Rev. C) <http://www.ti.com/product/cc2538>
- [8] Li ZOU, *Internet of Things and Intelligent Transportation*, Electronic Industry Press, Beijing China, 2012.

# Research on Path Planning of Automatic Handling Robot

Yuanbin HOU, Liu HE<sup>1</sup>, Chunfeng SONG, Yanyan ZHANG and Dong LI  
*School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China*

**Abstract.** Aiming at the problem that path planning for automatic handling robot in an environment with obstacles, the working environment model of the handling robot is analyzed. And then a path optimization algorithm based on fusing ant colony and particle swarm optimization algorithm is proposed. First of all, this algorithm uses the global search ability of particle swarm to go on a rough search and quickly plans the starting point to the end of the initial path. Then, the pheromone distribution is performed on the initial path. Finally, the ant colony algorithm is used to search the path carefully to get the optimal path. Experimental verification shows that compared with a single ant colony or particle swarm optimization algorithm, fused algorithm about ant colony and particle swarm optimization has a significant improvement in the number of iterations and path planning.

**Keywords.** Handling robot, path planning, particle swarm optimization algorithm, ant colony algorithm

## 1. Introduction

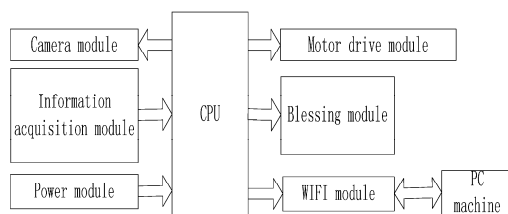
Automatic handling robot is an intelligent mobile handling device [1]. It can automatically complete the loading and unloading of goods, transportation and a series of work under the control of a computer. Because of its simple operation, flexible operation, it is widely used in the logistics storage system. Based on the given initial position and target location, path planning [2] can be used to plan a path which can not only avoid all obstacles but also meet certain criteria in an environment with obstacles.

Many experts and scholars at home and abroad have studied the shortest path length and put forward the solution. An improved ant colony algorithm based on artificial immune algorithm is proposed in literature [3], This algorithm introduces artificial immune system and processes the parameters of ant colony. Aiming at the search ability and convergence speed of particle swarm, the improved multi-objective particle swarm optimization algorithm is proposed in the literature [4] by Ji Zhuangzhuang. The use of hybrid genetic algorithm to solve the obstacle free path planning problem in the literature [5] was proposed by Umar Ali et al. They use adaptive weights to assign weights to each target. In [6], Qu Hong and Huang Liwei, according to the prior knowledge of the environment, improved the traditional ant colony algorithm to obtain

---

<sup>1</sup> Corresponding author. Professor, School of electrical and control engineering, Xi'an University of Science and Technology, Yanta Road No. 58, Xi'an 710054, Shaanxi Province, China; E-mail: 156500679@qq.com.





**Figure 1.** Automatic handling robot car system block diagram.

the global path, and combined with the rolling window theory to avoid collision with obstacles. Combined with the research status at home and abroad, we can find that because path planning plays an important role in automated logistics system. A path planning algorithm based on ant colony and particle swarm optimization is proposed in this paper, it can take advantage of the two algorithms to make the final path length shorter, and the efficiency of the robot higher.

## 2. System Description

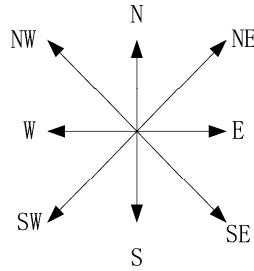
### 2.1. General Description

Automatic handling robot car is mainly composed of the main control chip, information collection module, blessing cargo module, camera, WIFI communication module and power module. The system framework is shown in Fig. 1. The information acquisition module consists of ultrasonic distance measurement sensor and photoelectric sensor, which can realize the distance detection and route detection of obstacles. According to the data collected by the data acquisition module, the main control chip (SCM) controls the motor drive module and the blessing module, and realizes the movement of the automatic handling robot and the grasping and placing of the goods. Then the camera module is used to transfer the image information around the robot to the PC through the WIFI module, and the upper computer is used to optimize the path.

### 2.2. Environmental Model Analysis of Handling Robot

Because the raster image is visual, simple structure, easy to operate and it is very easy to do simulation training on the computer [7], the environmental model of the handling robot is established by using the grid method in this paper. The grid method refers to the definition of the operating space of the handling robot as the same size grid two-dimensional space. And where the obstruction is marked as a shaded part and is represented by a black area, it is marked as a blank part, represented by a blank area [8]. Combined with the actual working environment of the transport robot car in the logistics system, assumptions as follows:

- The working area is a two-dimensional space, which is a finite static plane region.
- The size, location, starting point and destination location of static obstacles in the workspace are known.



**Figure 2.** The moving direction of the moving car.

- Automatic handling robot is regarded as a particle.
- Automatic handling robot can be arbitrary conversion direction, the direction of movement as shown in Fig. 2.
- Automatic handling robot can be arbitrary switching speed: acceleration, deceleration, uniform, stop, etc.

### 3. Path Optimization Algorithm Based on Ant Colony and Particle Swarm Optimization

#### 3.1. The Basic Principle of Particle Swarm Optimization (PSO) Algorithm

Particle swarm optimization (PSO) algorithm is an evolutionary algorithm to simulate the predation behavior of birds [9,10]. The algorithm is mainly based on the velocity and position search model. A flock of random search for food in space, Supposing the birds search for only one kind of food, and they know their current distance from food rather than where the food is. At this point, the best way is to search out the nearest food area.

Thus, PSO algorithm has some characteristics: the global searching ability is strong, the stability is strong, the convergence speed is fast, the search time is short, and it is easy to combine with other methods [11]. But in the later stage of the algorithm, the local search ability is poor and the feedback information is not fully used.

#### 3.2. The Basic Principle of Ant Colony (ACO) Algorithm

Ant colony algorithm (ACO), which is called the “ant algorithm”, is a bionic evolutionary algorithm proposed by Italy scholar Dorigo in the last century [12]. Ants communicate with each other through special substances, and pass messages to each other, they can choose the shortest path from the nest to the middle of the path according to the amount of pheromone left on the path, and change the path in real time according to the location of obstacles in the environment.

ACO algorithm is a swarm intelligence algorithm. After a long time of experts and scholars on the study of ACO algorithm, we can summarize its main features are the following:

- Each ant is independent of each other, respectively, search, which can improve the operational efficiency, and make the data became more reliable.
- At the beginning of the search process, the search for the optimal solution is chaotic. However, with the progress of the process, more and more ants begin to choose the path that is closer to the optimal solution. Therefore, the process is a self-organizing evolutionary process [13].
- Because ants are used to search the number of pheromone, therefore, with the increase in probability, which is the ants choose the path to be closer to the optimal solution, the number of pheromone is constantly convergence, and it is positive feedback growth to ensure the accuracy of the path selection.

According to the above characteristics, some important parameters which affect the path optimization of ACO algorithm are summarized, including the number of ants, the pheromone evaporation coefficient, the expected heuristic factor.

However, ACO also has some shortcomings. In the early stage of the search, it has a great blindness and it is easy to make the search time becomes larger, there may be a stagnation phenomenon, and it will miss the optimal solution.

### 3.3. Path Optimization Algorithm Based on Ant Colony and Particle Swarm Optimization

According to the characteristics of PSO and ACO algorithm, a fusion optimization algorithm for path planning of robot is obtained by combining ACO and PSO reasonably, the pros and cons of each other. Firstly, Suitable grid structure is established, and the starting point and the target point of the handling robot are determined, by using the global search ability of PSO algorithm to make a rough search [14] it will get the suboptimal solution to the path planning problem of the robot, and Using this sub optimal solution determine the distribution of pheromone in ACO algorithm. Then detailed search of the problem is implemented by using the positive feedback mechanism of ACO algorithm [15]. Finally, the optimal path of the robot is obtained.

The general steps of PSO and ACO algorithm fusion optimization are:

- Grid environment model is built, the start and end are determined, in the 2 dimensional plane, the initialization of each of the two algorithms is carried out with  $M$  ants, which can be iterated  $N$  times.
- PSO algorithm is used to adjust the pheromone in the ACO to realize the update of the speed and position by the global extreme value and the machine extreme value.

$$V_{i,j}(t+1) = \omega V_{i,j}(t) + C_1 \beta [P_{i,j}B - X_{i,j}(t)] + C_2 \beta [g_{i,j}B - X_{i,j}(t)] \quad (1)$$

$$X_{i,j}(t+1) = X_{i,j}(t) + V_{i,j}(t+1) \quad (2)$$

$$\tau_{m,n}(0) = C \quad (3)$$

Where,  $V(t)$ ,  $X(t)$  individually represent the velocity and position of the particle;  $PB$ ,  $gB$  represent the individual mechanism and global extremum;  $\beta$  represents

any random number between 0 and 1;  $C_1$ ,  $C_2$  are learning factor,  $C$  is constant,  $\tau_{m,n}(t)$  is the content of pheromone in the  $m$  node to the  $n$  node.

- Using ACO algorithm to search accurately.

At the  $t$  moment, the probability that the ant moves from the  $m$  node to the  $n$  node is:

$$P_{m,n}^k = \begin{cases} \frac{[\tau_{m,n}(t)]^\delta \left[ \frac{1}{d_{m,n}} \right]^\gamma}{\sum_{s \in A} [\tau_{m,n}(t)]^\delta \left[ \frac{1}{d_{m,n}} \right]^\gamma} & \text{if } j \in A \\ 0 & \text{else} \end{cases} \quad (4)$$

Where,  $A$  says the ant  $K$  will then select the node;  $\eta_{m,n}$  is a heuristic factor;  $\delta$ ,  $\gamma$  are information heuristic factor and expected heuristic factor, respectively;  $d_{m,n}$  represents the distance between  $m$  and  $n$  nodes.

The pheromone on the path is updated and adjusted as follows:

$$\tau_{m,n}(t+1) = (1-\rho)\tau_{m,n}(t) + \Delta\tau_{m,n}(t) \quad (5)$$

$$\Delta\tau_{m,n}(t) = \sum_{k=1}^m \Delta\tau_{m,n}^k(t) \quad (6)$$

Where,  $(1-\rho)$  is residual factor, in order to avoid the infinite superposition of pheromone,  $(1-\rho) \in (0,1]$ ,  $\Delta\tau$  represents the increment of pheromone, thus,  $\Delta\tau_{m,n}^k(t)$  represents the amount of pheromone left by ants in the inner path of the time. Normally, if the ants pass  $(m, n)$  node,  $\Delta\tau_{m,n}^k(t) = \frac{Q}{L}$ , Otherwise 0.  $L$  is the distance of an ant,  $Q$  is the strength of pheromone.

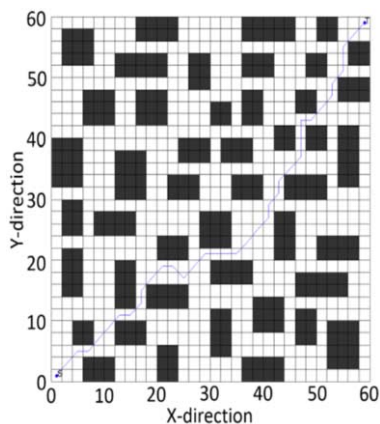
- Repeated iteration of the process, if the number is greater than or less than, you can continue to optimize the path, on the contrary, finished.
- The optimal path data is obtained and output, the algorithm is finished.

#### 4. Experiment and Analysis

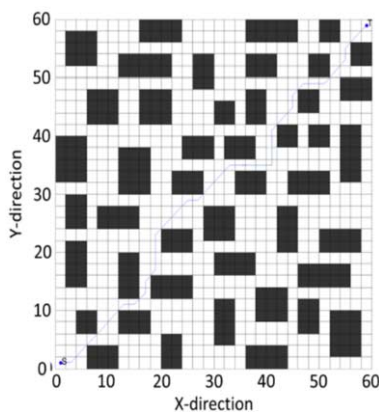
In order to compare the effectiveness of the proposed algorithm with a single particle swarm optimization algorithm and ant colony algorithm in the path planning of the automatic handling robot, the same environment was selected as the contrast experiment. According to the actual storage warehouse layout and MATLAB software, simulation diagram is obtained. Among them, the simulation of the grid size of  $60 \times 60$ , the length of each grid is set to 2 cm, the grid number is initialized to 0. According to X axis and Y axis direction, each grid is numbered sequentially, the location of the non target cargo storage area (i.e., the static obstacle) is marked as black, the rest is marked white and the goods are stored in 45, the starting point  $S$  coordinates is (1,1),

and the endpoint  $T$  coordinates is (59,59). Initialization value set to  $\alpha = 2.2$ ,  $\gamma = 9.0$ ,  $1 - \rho = 0.49$ , the maximum number of iterations was 100, the number of ants was about 100, and the pheromone intensity was  $Q = 10$ . The optimal path and convergence curves of the three algorithms in this environment are shown in Figs 3 and 4. Specific data analysis shows in Table 1.

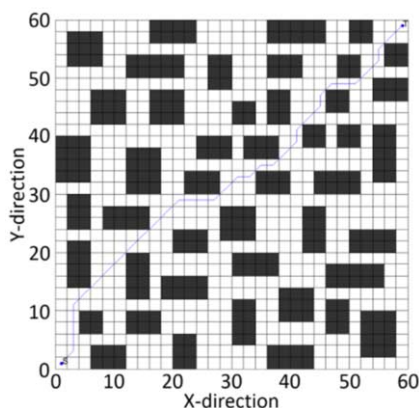
According to the chart and table, in the application of automatic handling robot car, the proposed PAAA is superior to PSO and ACO algorithm in terms of the optimal path length, the number of iterations and the running time. Among them, as can be seen from Fig. 3, the optimal path lengths of PSO and ACO were 95 cm and 93 cm, while PAAA was 90 cm, which was significantly better than the other two. As can be seen from Fig. 4, from the beginning of the algorithm to get the optimal path of the process PAAA the number of iterations required for 20 times, less than PSO 10 times, less than ACO 15 times. In addition, the optimal distance of PAAA search is 2.57 s, which is



(a) PSO searches the shortest path



(b) ACO searches the shortest path



(c) PAAA searches the shortest path

**Figure 3.** The optimal path search graph of the three algorithms in the same environment.

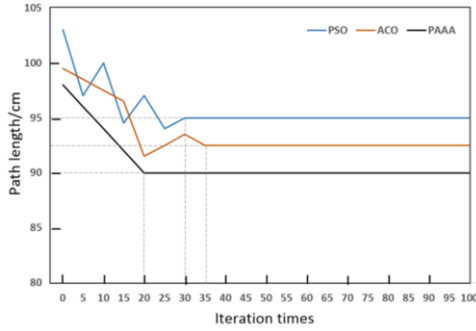


Figure 4. Convergence curves of the three algorithms in the same environment.

Table 1. Comparison of the optimal path parameters of the three algorithms.

Relevant parameters	PSO	ACO	PAAA
Optimal path length(cm)	95	93	90
Optimal run time(s)	5.38	4.76	2.57
Iteration 100 run time(s)	14.61	12.28	10.43
Minimum number of iterations	30	35	20

shorter than PSO by 2.81 s, shorter than ACO by 2.19 s. To sum up, the PAAA algorithm is superior to the single PSO and ACO algorithm in the performance of path planning.

### 5. Summary

Through the analysis of environment model of robot car manufacturing in this paper, the fusion algorithm PAAA is obtained and applied to path planning according to the advantages and disadvantages of PSO and ant colony algorithm. MATLAB simulation results verify that the number of iterations of the PAAA algorithm is significantly shorter, the path length of the planning is reduced and the convergence speed is accelerated by compared with the single PSO and ACO algorithm. The results show that the PAAA algorithm provides a good theoretical basis for the application of automatic handling robot path optimization.

### References

- [1] HuShuo. Design and production of intelligent transfer robot [J]. Coal Mine Machinery, 2015, (08):18–20.
- [2] Guangjie Han, Chenyu Zhang, Jinfang Jiang, ect. Mobile anchor nodes path planning algorithms using network-density-based clustering in wireless sensor networks [J]. Journal of Network and Computer Applications, 2017, 1(85):64–75.
- [3] ZhangYanYan, HouYuanBin, LiChen. Ant Path Planning for Handling Robot Based on Artificial Immune Algorithm. Computer Measurement and Control., 2015, 23(12):4124–4127.

- [4] WengLiGuo, JiZhuangZhuang, XiaMin, ect. Robot Path Planning Based on Improved Multi-objective Particle Swarm [J]. Journal of System Simulation, 2014, 25(12):2892–2898.
- [5] Umar Ali, Ariffin MKA, Ismail N, etc. Hybrid multiobjective genetic algorithms for integrated dynamic scheduling and routing of jobs and automated-guided vehicle (AGV) in flexible manufacturing systems (FMS) environment [J]. International Journal of Advanced Manufacturing Technology, 2015, 81(9–12):2123–2141.
- [6] QuHong, HuangLiWei, KeXing. Research of Improved Ant Colony Based Robot Path Planning Under Dynamic Environment [J]. Journal of University of Electronic Science and Technology of China, 2015, 44(2):260–265.
- [7] LiuXiaoLei, JiangLin, JinZuFei, GuoChen. Mobile robot path planning based on environment modeling of grid method in unstructured environment [J]. Machine Tool and Hydraulics, 2016, (17):1–7.
- [8] LiangJiaJun, ZengBi, HeYuanLie. Research on path planning algorithm for cleaning robot based on improved potential field grid method [J]. Journal of Guangdong University of Technology, 2016, (04):30–36+43.
- [9] Cao Jianfang, Chen Junjie, Zhao Qingshan. An optimized scheduling algorithm on a cloud workflow using a discrete particle swarm [J]. Cybernetics and Information Technologies, 2014, 14(1):25–39.
- [10] LiuXianZheng, Wen JiaLiang, PanYan, Wu Pengfei, LiJinYuan. OPF Control of DC-Grid Using Improved PSO Algorithm [J]. Power System Technology, 2017, (03):715–720.
- [11] Pawan Kumar Tiwari, Deo Prakash Vidyarthi. Improved auto control ant colony optimization using lazy ant approach for grid scheduling problem [J]. Future Generation Computer Systems, 2016, 78–89.
- [12] LvLin, YuYongQing, RenMin, PanXiao. Agglomerative hierarchical clustering based on ant colony optimization algorithm [J]. Application Research of Computer, 2017, (01):114–117.
- [13] LiYan, XuXiangRong, XuHao, LiShuang. Three-dimensional Trajectory Planning of Aerial Robot Based on Ant Colony Algorithm [J]. Journal of Anhui University of Technology (Natural Science), 2015, (04):360–365.
- [14] YangJingMing, MuXiaoWei, CheHaijun, HuZiYu, HouYuhao. Improved multi-objective particle swarm optimization algorithm based on multiple strategies [J]. Control and Decision, 2017, (03):435–442.
- [15] HouMengTing, ZhaoZuoPeng, GaoMeng, ZhangNaNa. Ant colony optimization multipath routing algorithm adopted angle factor [J]. Computer Engineering and Applications, 2017, 53(1):107–112.

# Metro Intelligent Riding System Based on Dynamic Hybrid People Identification

Yuanbin HOU, Liu HE<sup>1</sup>, Yun BAI and Chen LI

*School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China*

**Abstract.** For the problem of uneven distribution of passengers and poor flow of passenger in subway, a intelligent riding system based on dynamic hybrid people identification is established by the analysis tools of the dynamic hybrid algorithm. First of all, the collection of the car image is used to have pretreatment that includes gray and binarization, the degree of congestion within a single compartment is defined as four levels though the proportion of passengers in the image. And then, image recognition is done by the dynamic hybrid algorithm, the level of congestion of the compartment is determined. Ultimately, the level is displayed to the next station in the form of warning lights and text. Experiment and comparison results show that the background elimination method is proposed to improve the recognizable identity of the population in this paper, and the effective identification time is 6.28 s, it is much less than the running time of the subway station, this system can improve the capacity of the subway and the environment of the subway.

**Keywords.** Dynamic hybrid, people identification, intelligent riding, early warning broadcast

## 1. Introduction

With the development of economy and the increasing of urban population, the pressure of urban transportation is increasing. As a city fast track, Subway plays a very important role in the public transport system, and eases the pressure on road traffic for people to travel to bring great convenience to a certain extent. However, due to the sharp increase in the number of population, the subway crowded situation still can not be ignored [1]. There are uneven distribution of passenger flow in crowded subway, it is very meaningful that using image processing technology to detect the distribution of passengers in crowded subway.

At present in the field of image processing, many experts are studying all over the world. For example Beijing Automation Research Institute of Chinese Academy of Sciences researched on visual surveillance of traffic scene and human visual surveillance [2]; University of Science and Technology China considered high resolution SAR image target recognition model and high efficient algorithm. Due to the image processing technology to the number of personnel is now used for large shopping malls

---

<sup>1</sup> Corresponding author. Professor, School of electrical and control engineering, Xi'an University of Science and Technology, Yanta Road No. 58, Xi'an 710054, Shaanxi Province, China; E-mail: 156500679@qq.com.



and other traffic flow detection, in metro areas it is mostly confined to emergency evacuation of personnel and rarely involved in the distribution of the subway train passengers. Thus, metro intelligent riding system based on dynamic hybrid people identification is established. The system uses the background subtraction and the background subtraction of the dynamic mixing algorithm and the definition of the degree of congestion for image processing and identifying the number of passengers at each entrance, meanwhile, this result will be transmitted to the next stop of the subway traffic routes and inform the passengers on the subway. This system can help passenger to make a reasonable choice of subway entrance and to meet the needs of passengers to the greatest extent.

## 2. System Structure

Metro intelligent bus system device based on dynamic crowd identification includes: subway camera module, computer, WIFI module, microcontroller, warning indicator light, text broadcast module, power module, etc. Subway camera uses OV7670 (PAL) module to complete the vehicle real-time monitoring and video capture, it can send the collected image signal to the terminal computer, then the terminal computer can complete the image dynamic crowd recognition process and send the identification signal to the target microcontroller through the WIFI module. Microcontroller receives the corresponding commands to achieve the warning function of the warning lights and text. The structure of the system is shown in Fig. 1.

The function of the system mainly includes image processing, signal analysis, display and so on. Image processing refers to the input of the video or continuous image, the system analysis and comparison of the various features of the input image, and finally gets the results to be identified. Through the sequential image acquisition, image preprocessing, target recognition algorithm, dynamic analysis of multiple processing steps, the system examines subway passenger compartment distribution, then show the congestion of all the entrances and exits of the subway. The flow chart of system identification is shown in Fig. 2.

According to the investigation of the maximum capacity and load of the electric bus, congestion can be divided into four levels: loose, general, more crowded, especially crowded; the corresponding colors are green, yellow, orange and red.

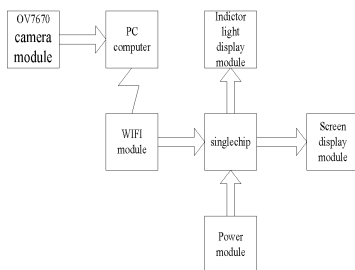


Figure 1. Subway system structure diagram.

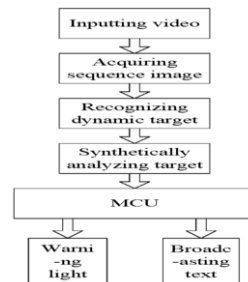


Figure 2. System identification flow diagram.

### 3. Image Pre-Processing

#### 3.1. Graying and Binarization

Image gray is a common method used in image processing, it transforms the RGB diagram of the 3 channel to the gray scale of 0–255. There are a lot of image processing methods, such as the maximum between class difference method, etc. [3–5]. Because this system is not complicated to deal with the gray scale on population distribution. The R, G, B channel pixel value as the mean to get the gray value.

The two value of the image is that the gray image change into a pure black and white image, it means the gray value is converted to 0 or 255 [6,7]. The completed method is to select a threshold from 0 to 255, the gray value of all pixels is compared with the value of this value, and we can finally get a whole and part of the two values of the characteristic image, available 1 representation. When the image element value is greater than the threshold, then it can be taken 255 and recorded as foreground objects, contrast to, when the threshold value is less than the threshold value, the gray value is 0, and the other area is recognized as the other region [8,9].  $T_k$  is the threshold.

$$I(x, y) = \begin{cases} 255 & D_k(x, y) > T_k \\ 0 & \text{else} \end{cases} \quad (1)$$

According to the design of the intelligent subway system, the image of the image of the specific gray and two processing results in Figs 4(b) and (c).

#### 3.2. The Definition of Congestion Level in Image Processing

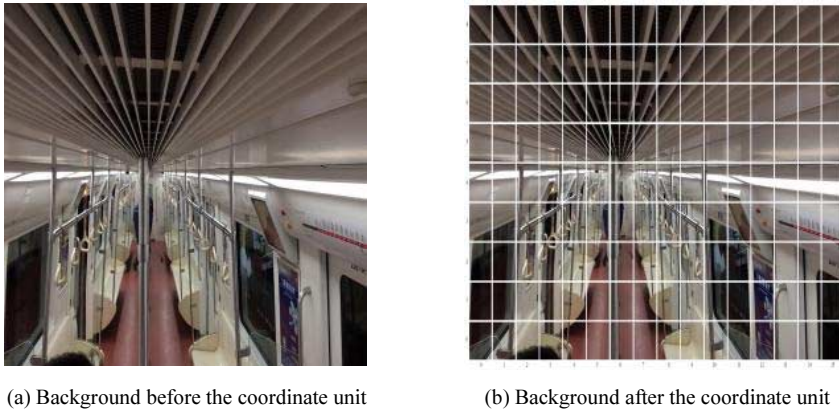
In the image processing, there is no specific analysis method to define the crowd congestion degree, so it is necessary to re define the image of the background of the subway. Figure 3 for subway car background processing, Fig. 3(a) is the background image and the background of the car after the erase map, which is the background of the unit coordinate, Fig. 3(b) is the picture that is to split the compartment background into 144 cells and after the coordinate of the map.

**Definition 1.** It is defined the subway background which is generally not covered by the distribution of the crowd to change as a hard background, otherwise defined as the soft background.

Hard background can be ignored in the background of the application of the background of the distribution of the subway staff, such as the top of the subway; Soft background is easily covered by the floor, seats, billboards, billboards and compartments.

**Definition 2.** The background image, which is composed of 144 cells, is removed from the background image, and the 80 unit of the hard background is used as the 64 unit of the subway train.

In particular, for the analysis of the level of congestion of the subway cars should be reasonable to reduce the number of car to get off the target. According to the



**Figure 3.** Subway car background processing.

long-term investigation and the analysis of the behavior and habits of urban population: passengers will generally reach the target mobile station in advance to the door, waiting to get off. If the background difference method is adopted to identify the area where the number of the car is off, the 80 cell of the signal monitoring and collecting part need to be eliminated.

It should be stated that the cell is covered more than 50% that is counted as 1, which can eliminate the error caused by other interference.

#### 4. Dynamic Mixing Crowd Recognition Algorithm

According to the variation of human behavior and environment in the carriage of the subway, a dynamic mixed crowd recognition algorithm was proposed in this paper. And the background elimination method and mixed Gauss model of hybrid parallel processing, subway car crowd recognition. First of all, we should do static processing of background elimination method, and the process of background subtraction is to select a reference frame as the reference image, using the current frame and the reference image do the differential operation [10], eliminate the background, and show the dynamic target. Now we only do static processing, that is: only empty compartments to eliminate the background, a frame of car image to the station within a specified time is taken as the target image.

The static processing can identify all the people in the car after eliminating the background. Secondly, the mixed Gauss model is used to deal with the dynamic time between the departure time of the subway and the time of a specific vehicle image. According to the behavior of the crowd to get off the subway, we can identify the crowd who want to get off the subway. Finally, using the first two steps of the static processing image and the mixed Gauss model to Subtraction operation aimed at getting the image all of the passengers inside the car. According to the degree of congestion to determine the definition of interior congestion level and make timely warning and reporting.

### 4.1. Background Elimination

Background subtraction method is a common method for moving object extraction, and it can get more accurate target image. Background subtraction is the difference between the current frame and the reference image which is the firstly selected reference frame, in this way, the moving background is subtracted and the moving target is shown. If the reference image is selected properly, the moving object can be segmented accurately [11].

As a common method of image processing, background subtraction is often used in the identification of many large occasions. It has a good effect on the environment such as shopping malls, outdoor squares, etc. This paper will link in the experiment, the recognition results which are separate the result of the proposed algorithm and background elimination compared and analyzed.

### 4.2. Mixed Gauss Model

The mixed Gauss model is one of the background difference methods, and it is a kind of modeling method in complex background [12]. For the monitoring system, over a period of time the background image is almost constant, or there are some regular perturbations in the background compared to foreground objects. It can use the appropriate reference frame as the background difference, that is, we can extract the foreground image of the video image efficiently by the way of difference [13].

In this paper, the adaptive Gauss mixture model is used to detect the movement of the mobile population in the subway car, it can realize the real-time identification of the effect of the vehicle personnel. Formula (2) is the concrete expression of the background difference method.

$$D_K = f_K - B_K \tag{2}$$

Thereinto,  $f_K$  represents the current frame,  $B_K$  represents the background image.

Set  $t$  time's a pixel  $(x, y)$  of the color vector is the three-dimensional column vector which consists of the three color composition of R, G, B.

$$\mathbf{X}_{xy,t} = [\mathbf{R}_{xy,t}, \mathbf{G}_{xy,t}, \mathbf{B}_{xy,t}]^T \tag{3}$$

The  $(X, Y)$  is viewed as a color vector  $(X_{xy,1}, \dots, X_{xy,t})$  in time, it can be viewed as a stochastic process which is independent of other pixel vectors, it also is a probability density function. For a pixel in an image frame, if it can be described by the probability density function [14], the pixel is a background point, otherwise it is the former attractions.

Due to the complexity of the pixel distribution of the image frame, people usually use a number of probability density function to simulate the pixel color history. If you use the  $K$  Gauss distribution to describe the distribution of color vectors, the probability [15] of defining the current pixel  $\mathbf{X}_{xy,t}$  is:

$$P(\mathbf{X}_{xy,t}) = \sum_{k=1}^K w_{xy,t} \eta(\mathbf{X}_{xy,t}, \boldsymbol{\mu}_{xy,t}, \sum_{xy,k,t}) \tag{4}$$

Thereinto,  $w_{xy,k,t}$  represents the weight of the  $K$  Gauss distribution, the mean vector and covariance matrix are respectively represented by  $\mu_{xy,t}$  and  $\sum_{xy,k,t}$ . With the increase of  $K$  values, the difficulty of the calculation will improve,  $\eta$  is the Gauss probability density function, and its specific definition is:

$$\eta(X_{xy,t}, \mu_{xy,k,t}, \sum_{xy,k,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} \left| \sum_{xy,k,t} \right|^{\frac{1}{2}}} \bullet e^{-\frac{1}{2}(X_{xy,t} - \mu_{xy,k,t})^T \sum_{xy,k,t}^{-1} (X_{xy,t} - \mu_{xy,k,t})} \tag{5}$$

According to the definition of the formula, it can be seen that the value of  $w$  is larger, the actual value is more consistent with the corresponding Gauss distribution, however, the standard deviation is smaller, the distribution is more stable. Thus,  $w/\delta$  can be used to describe the  $K$  Gauss distribution, they can be divided into  $B$  subsets and  $K-B$  foreground subsets.

According to the order of Gauss distribution, we can respectively calculate the distance from  $X_{xy,t}$  to the mean vector  $\mu_{xy,t}$ . The distance which is the first of the  $K$  results to satisfy the following formula matches with observed value  $X_{xy,t}$ , if they are matching,  $M_{xy,k,t}$  will be set to 1, otherwise, it is 0. Calculating formula which is used for match is as follows.

$$M_{xy,k,t} = \begin{cases} 1, & \text{if } |X_{xy,t} - \mu_{xy,k,t}| < \lambda\sigma_{xy,k,t} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

$\lambda$  generally take 2.5–3.

When the dynamic recognition algorithm is used for identification, the monitoring field information collected by the metro vehicle camera is transformed into a digital image sequence inputting computer through the image acquisition card.

The steps of the computer image processing are as follows. Firstly, the computer performs background elimination to get the total number of soft background  $X$ ; then, the Gauss mixture model is used to detect the moving target, get the total number of moving objects in the area  $Y$ , and calculate the total number of people who are get off the subway; finally, the crowded level of the subway can be got.

## 5. Test and Analysis

### 5.1. Processing Results

In order to study the effectiveness of the system, a video of the passenger flow in a section of Xi'an subway line is selected. We can make the following definition based on the investigation of the congestion degree of passengers in different stations of Xi'an Metro Line 1, and the load requirement in the Xi'an Metro Co., Ltd. Branch operating standards:

**Definition 3.** Assuming that the number of passengers between the two adjacent doors is  $X$ , the number of passengers who are getting off is  $Y$ , and the number of remaining passengers is  $Z = X - Y$ .  $Z \leq 15$ , the congestion degree is loose;  $15 \leq Z \leq 30$ , the

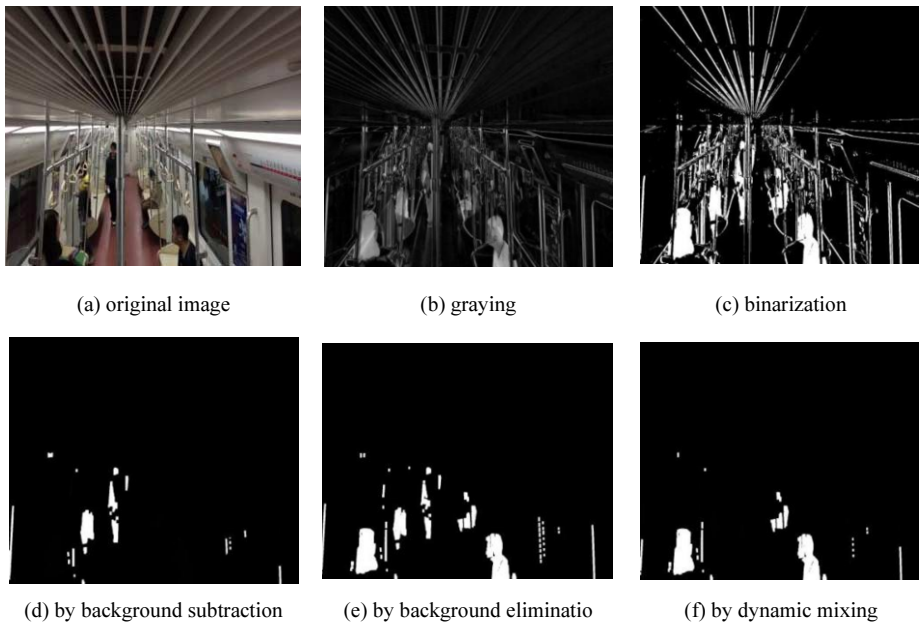
congestion degree is common;  $30 \leq Z \leq 45$ , it is more crowded;  $45 \leq Z$ , it is crowded. Each level allows 1–2 human error tolerance because of the uncertain factors of passenger at the entrance.

We can image contour image processing in OpenCV produced by the Intel Software. Figure 4 is the process of image processing for dynamic mixing crowd recognition, Fig. 4(e) is the effect chart by eliminating Fig. 3(a) which is the background, Fig. 4(f) is the result of the subtraction of Figs 4(e) and 4(d) by using dynamic hybrid algorithm.

By comparing Figs 4(e) and 4(f): The results of this method are significantly different from the usual background elimination methods. The classical background subtraction method can only eliminate the background of the actual image of the crowd, but it cannot identify the number of the specific crowd, there is a large recognition error in practical application. In this paper, the hybrid recognition algorithm can eliminate the number of passengers, and congestion degree definition can be used to further eliminate the interference factors in the image, and the reliability is significantly improved compared with the background elimination method.

Through the description of the scene in the subway car and the dynamic mixed crowd recognition process shown in Fig. 4, we can see that there are 61 frames from the crowd to successfully extract the crowd, video capture rate is 25 frames/s.

By the definition of 3 and the total number of people which is treatment of the Fig. 4(f), we can determine the time of the crowded compartment for green loose, the results are arranged on the PC and sent to the controller to carry on the corresponding processing, then the text and the indicating lamp are displayed.



**Figure 4.** Dynamic mixing population identification.

**Table 1.** Implementation time of each algorithm (unit: ms).

Time algorithm	Background elimination	Graying	Binarization	Background subtraction	Dynamic mixing	Total time
1	13.36	19.21	0.70	17.46	15.22	65.95
2	13.27	15.33	0.67	16.24	14.24	59.75
3	14.25	17.43	0.73	14.34	14.15	60.90
4	14.38	17.56	0.61	15.24	13.22	61.01
5	13.62	24.63	0.75	14.78	13.69	67.47
average time	13.78	18.83	0.69	15.61	14.10	63.01

## 5.2. Processing Time

Selecting the algorithm of the background elimination, graying, binarization, background subtraction and dynamic mixing. We can write programs on the PC machine and process a frame image, its operating environment is a 64 bit operating system, memory RAM is 4.00 GB, processor is AMD A6-3400M APU with Radeon(tm) HD Graphics 1.40 GHZ. The implementation time of each algorithm is shown in Table 1 (two decimal places).

It can be seen from the table that the average total length of the dynamic hybrid algorithm proposed in this paper is 63.01 ms/frame, for crowd recognition image which the acquisition rate is 25 frames/s and experiences 61 frames, its acquisition time is 2.44 s, the image processing time is about 3.84 s, So the total length of the process is about 6.28 s, time is far less than the subway station in advance and the arrival of the real arrival of the total length of 80 s (adjacent stations). Combined with image processing results, this algorithm can eliminate the influence of the scene of the car facilities, light and other factors in a very short time to complete image processing functions and have a strong practical value.

## 6. Conclusion

Aiming at the problem of bad order of passengers in the subway operation, intelligent vehicle system based on dynamic mixing crowd identification is proposed. Through the image background of the subway, the distribution of the compartment can be obtained; processing erased image by unit coordinates, using dynamic mixing method analysis the number of passengers to get new personnel distribution. It can complete the subway car congestion degree of real-time warning and text broadcast through the four levels of compartment congestion degree and the image eliminating about the number of people who are getting off the subway. By comparing the test results, it showed that the proposed algorithm improve the recognition of the image compared with the background subtraction method, obtain the congestion degree in a short period of time (6.28 s) between adjacent stations (80 s) and provide a reference to improve the subway rules.

## References

- [1] Wang Jing, Peng Jin shuan. Identification and release system for the number of passengers in Urban Rail Transit [J]. *Intelligence City*, 2016, 03:146.

- [2] Jia Shuanhang. Analysis of large passenger flow in subway station [J]. Modern Enterprise, 2013.
- [3] Fu Yaoqing, Liu Huanlai. .NET of cutting image based on Apple platform Method research and implementation [J]. Silicon Valley, 2013.
- [4] Wen Zhuqing, Luo Wei, Du Huaying. Research on the License Plate Recognition Based on BP Neural Network Algorithm [J]. Modern Computer, 2015, 35(4):65.
- [5] Liu Yan, Wang Tong, Ding Hui, etc. Research on Risk Assessment DEA Model of Crowd Crushing and Trampling Accidents in Subway Stations [J]. China Safety Science Journal, 2013, 23(10):100.
- [6] Liu yuhong, Wang zhiyang, Yang jiayi, etc. A New Color Image Binarization [J]. Chinese Journal of Medical Physics, 2013, 1(30):3873.
- [7] Gong Yishan, Wang Peng. Identification algorithm for invoice number based on template matching [J]. Journal of Shenyang University of Technology, 2015, 37(6):673.
- [8] Hu Binbin, Tu Zhengzheng, Zheng Ai hua, etc. Anomaly Detection in Crowded Scenes Based on Cell Motion Feature [J]. Computer Applications and Software, 2014, 31(8):157.
- [9] Ko B C, Cheong K H, and Nam J Y. Fire detection based on vision sensor and support vector machines [J]. Fire Safety Journal, 2009, 44(3):322.
- [10] Lin yongzheng, Wu xiaoyu, Li hengjian, etc. Cheating behavior detection in examination room based on back- ground subtraction [J]. Journal of University of Jinan, 2015, 29(6):406.
- [11] Yasen Aizezi, Aishan Wumaier. Background Subtraction Method Based on Gray Feature Model in Moving Target Detection [J]. Computer Engineering, 2015, 41(6):240.
- [12] Fan Wenchao, Li Xiaoyu, Wei Kai, etc. Moving Target Detection Based on Improved Gaussian Mixture Model [J]. Computer Science, 2015, 42(5):286.
- [13] Wang Guoqiang, Gai QiLin, Yu Huai yong, etc. Video Target Detection algorithm based on Background Subtraction [J]. Journal of Engineering of Heilongjiang University, 2014, 5(4):64.
- [14] Jiang Ming, Pan Jiaoli. An improved method of mixture Gassian model for background estimation [J]. Microcomputer and Its Applications, 2011, 30(11):31.
- [15] Li Ju, Li Keqing, Su Yonggang. Moving target detection algorithm combined with Markov random walk and Gauss mixed model [J]. Journal of electronic measurement and instrumentation, 2014, 28(5): 533.



# Energy Management Strategy for a Hybrid Vehicular Power System Based on Haar Wavelet Transform

Bo LI<sup>1</sup>, Shaoyi BEI, Jingbo ZHAO, Ting ZHOU, Zhang NI and Qing LIU  
*Vehicle and Traffic Engineering School, Jiangsu University of Technology,  
Changzhou, 213001, China*

**Abstract.** Hybrid vehicular power system is a complicated and nonlinear system, and its energy management strategy is one key factor for vehicle performance. The hybrid powertrain consisting of engine, Ni-MH battery and ultra-capacitor is designed, and the energy management strategy for a hybrid vehicular power system is proposed by Haar wavelet which makes vehicle power demand into high frequency and low frequency parts, then distributed among engine, Ni-MH battery and ultra-capacitor by genetic algorithm with the aim of reducing fuel consumption, therefore, improves the system performance and lengthens the service life of components. System modeling and simulation are conducted in Matlab, and the results show that this method of power distribution can meet the design requirements, and have better fuel economy compared to fuzzy control strategy.

**Keywords.** Hybrid vehicular, energy management, wavelet transform, fuel economy

## 1. Introduction

Hybrid power system uses a variety of power sources. It allocates reasonably various parts of the energy demands to obtain the optimal control objectives (higher fuel economy, lower Emissions, increase in the service life of parts, etc.) which the traditional diesel locomotive cannot achieve by using combination of traditional automotive internal combustion engine drive and battery electric drive technology. The energy management strategy is the core algorithm of the hybrid system and it is also a key technology of today's hybrid vehicle.

The hybrid electric vehicle (HEV) power system [1–3] is an extremely complex time-varying nonlinear system and the road conditions are diverse, which result in dramatic changes in vehicle power demand. Therefore, how to make use of and allocate the power demand of engine, Ni MH battery and super capacitor is a difficult point. Currently, hybrid energy management strategy mainly includes: rule-based steady-state control, fuzzy logic control [4–7], instantaneous optimal control and global optimal control. However, these control strategies do not consider the effects of the dramatic changes in the power demands on the internal combustion engine and battery performance. Besides, they do not give full play to the characteristics of super-capacitive dynamic adaptability and recycle the braking energy.

---

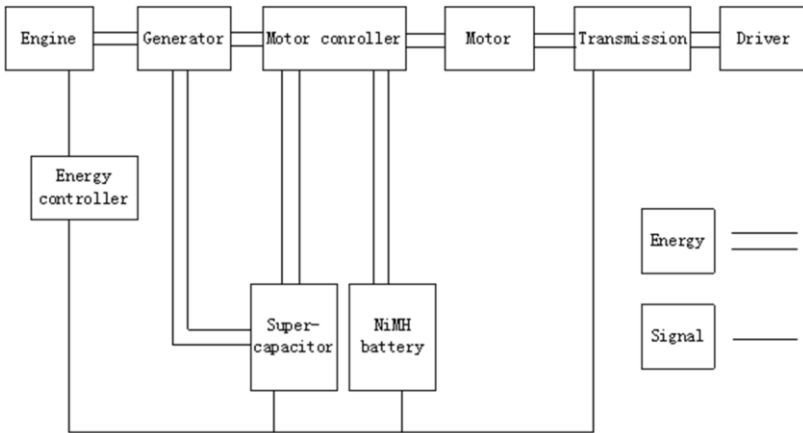
<sup>1</sup> Corresponding author. Li Bo, Vehicle and Traffic Engineering School, Jiangsu University of Technology, Changzhou, 213001, China; E-mail: czbolifly@sina.com.

In this paper, multi-energy management strategy is established based on Haar wavelet aimed at engine, Ni-MH battery and super-capacitor hybrid system. In a certain cycle conditions, multilayer wavelet is decomposed for the dynamic system power demands. The low frequency power is allocated to the engine and the NiMH battery while the high frequency power is allocated to the super-capacitor by using genetic algorithm with the minimum fuel consumption. Energy can be recycled during braking by NiMH battery and super-capacitor to improve power system performance, fuel economy and extend service life of parts compared with fuzzy control strategy.

**2. Structural Model of Hybrid Power System**

Figure 1 shows the tandem hybrid power-driven system. The engine drives generator to generate electricity, together with the electric energy from the NiMH battery and the super-capacitor, are transmitted to the motor via the motor controller and the vehicle is driven by the transmission drive. If the SOC (state of charge) of NiMH battery or super-capacitor is low, it can be charged by engine driven generator to maintain normal work. The overall structure of this kind of dynamic system is simple because there is no structure of engine-driven wheel.

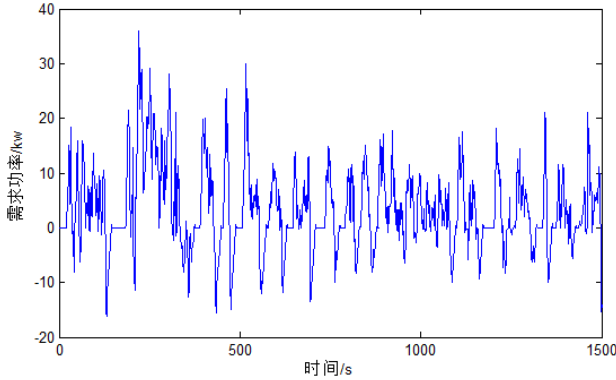
Regardless of driving conditions, because the engine only provides power for the generator, it has been working in a stable economic zone to reduce fuel consumption. This advantage makes tandem hybrid models more suitable for frequent start-up and low-speed urban areas.



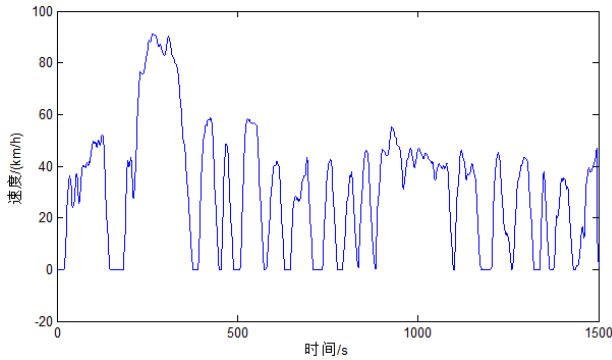
**Figure 1.** Hybrid power system structure.

**Table 1.** Hybrid power system power parameters.

Engine	NiMH battery	Super-capacitor
Rated power	Rated power	Rated power
53 kw	20 kw	19 kw
Maximum power	Monomer capacity	Monomer capacity
64 kw	40 A.h	29.3 F



**Figure 2.** FTP-75 power curve.



**Figure 3.** FTP-75 speed curve.

### 3. Cycle Conditions

FTP-75 (Federal Test Procedure) cycle condition is selected from AVL CRUISE software as the simulation condition of the hybrid power system, and the specific power curve is shown in Fig. 2. The maximum power demand is 35.5616 kw and the minimum power demand is  $-15.3595$  kw. The cycle time is 1500 s and the driving distance is 11.9938 km. Figure 3 is the FTP-75 cycle speed change curve.

### 4. Power Allocation Algorithm Based on Haar Wavelet Transform and Genetic Algorithm

Wavelet Transform is a method for time-frequency representation. The basic idea is to assume that the high and low frequency components of the signal have different time-varying characteristics [8,9]. Generally, the low frequency part of the spectrum changes slowly with time, and the high frequency changes rapidly. Therefore, you can obtain the ideal frequency and time resolution in different time-frequency domain according to such characteristics of the frequency and time axis. The overall calculation

step of wavelet transform (WT) is the convolution of a function (signal) and the corresponding wavelet basis function. The function is further analyzed and processed after decomposed into various components of different time and frequency.

The wavelet transform concrete steps: the wavelet transform of the finite energy function  $f(t)$  is that the mother wavelet function  $\varphi(t)$  performs  $b$  displacement and do the inner product with the function  $f(t)$  to be analyzed under scale factor  $a$ .

$$\begin{aligned} WT_f(a, b) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \varphi^* \left( \frac{t-b}{a} \right) dt \\ &= \langle f(t), \varphi_{a,b}(t) \rangle \end{aligned} \quad (1)$$

Where:  $a$  is the scale factor ( $a > 0$ ), and  $b$  is a positive and negative time-shift factor.

$\varphi^* \left( \frac{t-b}{a} \right)$  is the conjugate function for  $\varphi(t)$ . The symbol  $\langle x, y \rangle$  represents the inner product and its specific meaning is

$$\langle x(t), y(t) \rangle = \int x(t) y^*(t) dt \quad (2)$$

$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \varphi \left( \frac{t-b}{a} \right)$  is the scaling and displacement of the wavelet function.

$T, a, b$  are continuous variables, so this wavelet transform is Continuous Wavelet Transform (CWT). The corresponding frequency domain is expressed as

$$WT_f(a, b) = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{+\infty} F(w) \psi^*(aw) e^{+jwb} dw \quad (3)$$

Where:  $F(w), \psi(w)$  are  $f(t), \varphi(t)$  Fourier transform respectively.

Different wavelet functions show different properties in smoothness, symmetry, orthogonality and compact support, and it is often difficult to establish wavelet functions with the above characteristics at the same time. In the practical application, the appropriate wavelet function is selected by compromising these characteristics according to the characteristics of the signal function and the specific decomposition of the processing requirements. Haar wavelet is a kind of mother wavelet which is widely used to check and locate the transient characteristics of the signal. Compared with other wavelet bases, Haar has the shortest filter length in the time domain. So it is more suitable for analyzing and processing HEV power demand signal. It can be shown from Fig. 2 that this power curve contains a large number of sharp power mutation points. Haar wavelet is defined as follows.

$$\varphi(t) = \begin{cases} 1, & \text{if } 0 < t < 0.5 \\ -1, & \text{if } 0.5 < t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Wavelet reconstruction is the inverse wavelet transform, and the detail coefficients and approximate coefficients obtained by wavelet transform are constructed back to the original signal. The inverse transform of continuous and discrete wavelet is as follows.

$$\begin{aligned}
 f(t) &= \frac{1}{C_\varphi} \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} WT_f(a,b) \varphi_{a,b}(t) db \\
 &= \frac{1}{c_\varphi} \int_0^{+\infty} \frac{da}{a^2} \int_0^{+\infty} WT_f(a,b) \frac{1}{\sqrt{a}} \varphi\left(\frac{t-b}{a}\right) db
 \end{aligned}
 \tag{5}$$

$$f(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} W(j,k) \varphi_{j,k}(t)
 \tag{6}$$

As shown in Fig. 5, the power demand of the hybrid system under FTP-75 is decomposed and reconstructed three times as the original signal  $f(n)$  to obtain the high frequency and low frequency parts needed for the power distribution of the hybrid system.  $A(n)$  ( $A3(n)$ ) is the approximate signal (low frequency signal), and  $D(n)$  ( $D3(n) + D2(n) + D1(n)$ ) is the detail signal (high frequency signal). The relationship between them and the original signal  $f(n)$  is shown in Eq. (7).

$$f(n) = A3(n) + D1(n) + D2(n) + D3(n)
 \tag{7}$$

The high-frequency signals can be assigned to nickel-metal hydride batteries and super-capacitors. The low frequency signal can be assigned to the generator and the NiMH battery. The specific distribution value is solved according to the multi-objective optimization genetic algorithm.

The standard genetic algorithm [10] consists of four basic steps of initialization, selection, crossover, and mutation. The concrete structure is shown in Fig. 4.

#### 4.1. Initial population Generation

Population initialization, also known as coding, is to use a number of code string to optimize the solution parameters. When the binary coding is used, this correspondence is the relationship between decimal and binary conversion. According to the difference between the highest power and the lowest power, the choice of  $x$  value changes between  $-20-40$ . So there are 6 binary code string can fully express all genotypes of chromosomes, that is, all genotypes between 000000–111111.

#### 4.2. Fitness Calculation

The objective function (fuel consumption rate) is selected as the original fitness function, respectively, with different weights to represent the different distribution coefficient.

$$V(x) = K_{1r} f_1 + K_{2r} f_2 + K_{3r} f_3$$

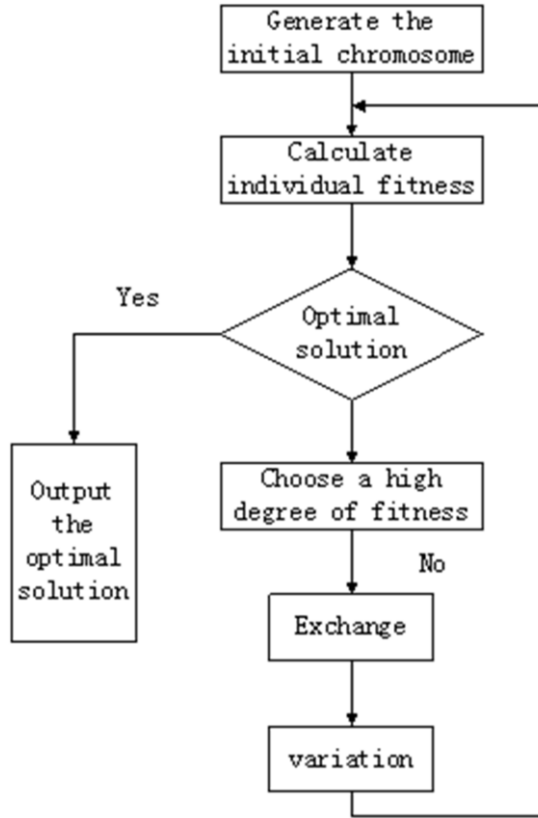


Figure 4. Basic steps of genetic algorithm.

Where:  $V(x)$  is the target function.  $K_{1t}$ ,  $K_{2t}$ ,  $K_{3t}$  are the energy distribution coefficients of the engine, NiMH battery and super-capacitor at time  $t$ .  $f_1$ ,  $f_2$ ,  $f_3$  are fuel consumption rates of the engine, NiMH battery and super-capacitor.

#### 4.3. Constraints

In order to allow the three energy sources of the hybrid system to operate in the normal power range, they are constrained within the rated power (as shown in Table 1). At the same time, the SOC of NiMH battery and super-capacitor is required between 20% and 80%.

### 5. Simulation Analysis

In order to observe and verify the performance of the hybrid system under specific driving conditions, the control structure shown in Fig. 6 is built. FTP-75 power demands decompose into low-frequency power and high-frequency power through wavelet transform.

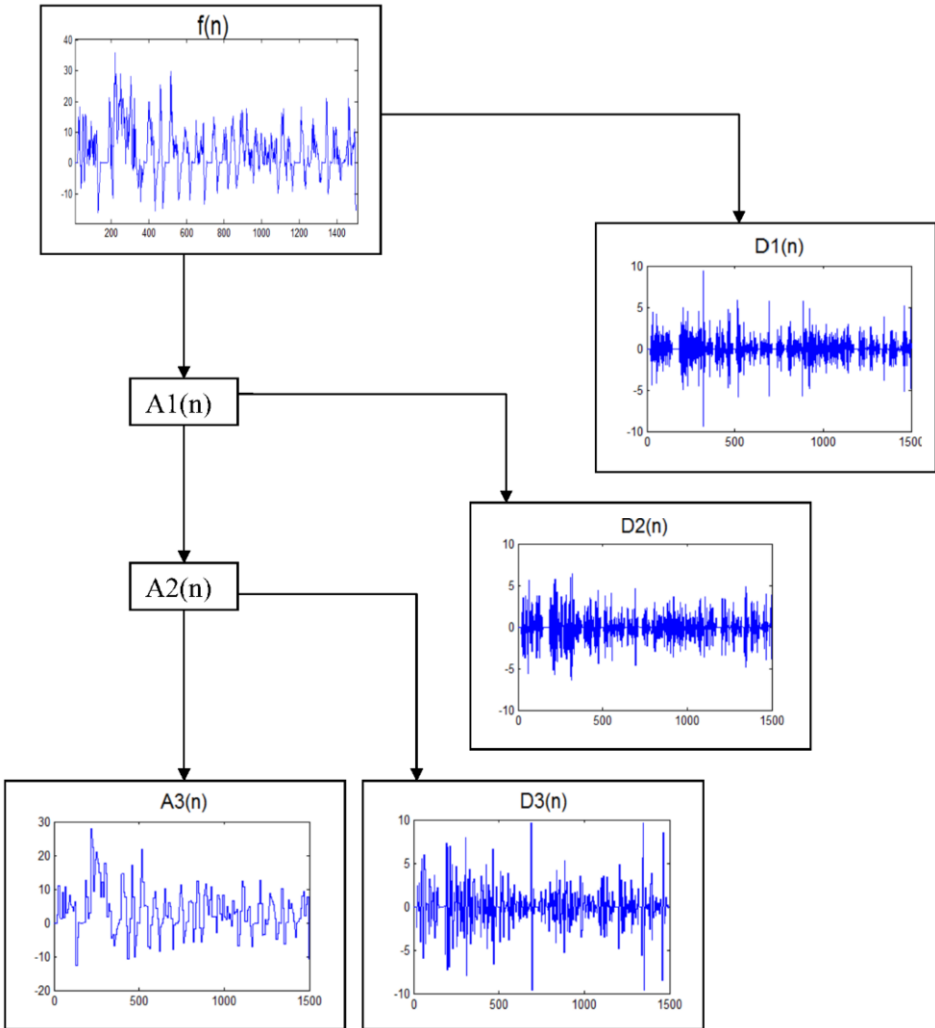


Figure 5. FTP-75 power demand decomposed by wavelet transform.

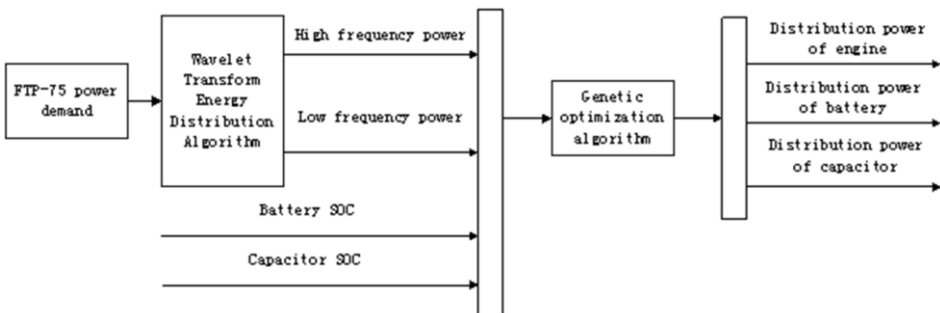
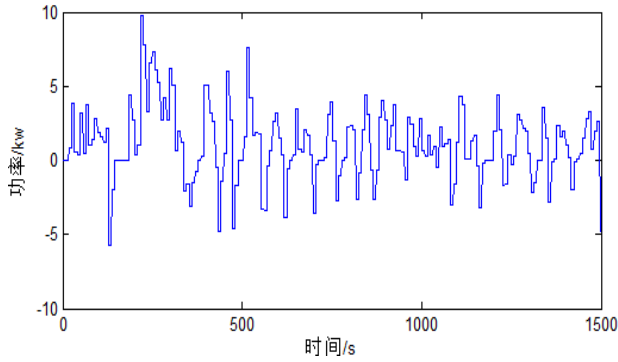
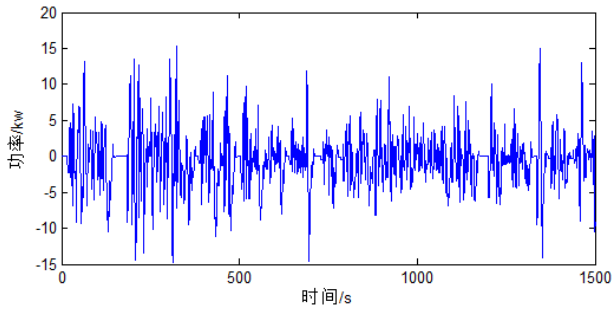


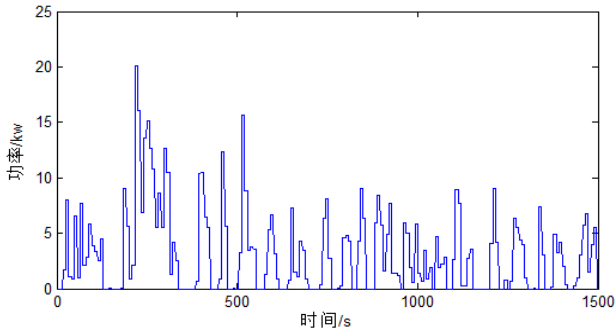
Figure 6. Hybrid system simulation analysis process.



**Figure 7.** NiMH battery distribution power.



**Figure 8.** Super capacitor group allocation power.



**Figure 9.** Engine output power.

Then, the power characteristics of engine, NiMH battery and super-capacitor are combined with and the SOC value of battery and capacitor (the SOC is controlled as much as possible between 20% and 80%. If the SOC is less than 20%, the generator charges it).

The demand power of each of the three energy sources is determined by using the genetic optimization algorithm. Respectively, Figs 7, 8, and 9 show the NiMH batteries, super capacitors and engine power allocation.



The following Figs 10 and 11 are the SOC curves of NiMH battery and super-capacitor under the condition of FTP-75 cycle. It can be seen that the small changes are between the expected 20% and 80%. Because the power demands of NiMH battery are all from low-frequency power, the SOC of NiMH battery changes smoothly. On the contrary, the power demands of super-capacitor are all from high-frequency power, so the SOC of super-capacitor changes intensely.

The fuel characteristic surface of the engine is shown in Fig. 12. Figure 13 shows the distribution of the operating point of the engine based on the wavelet transform energy control strategy and the fuzzy control strategy. The blue \* represents the engine operating point based on the wavelet transform, and the red + represents the operating

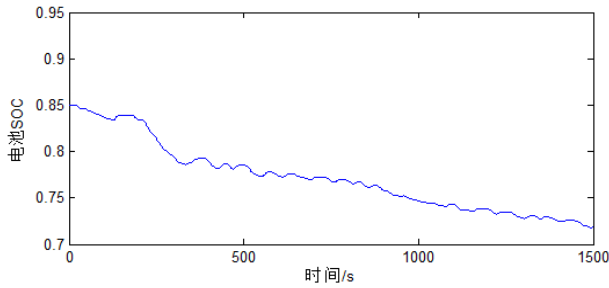


Figure 10. SOC change of NiMH battery.

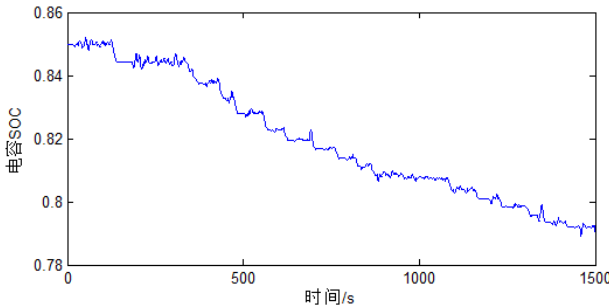


Figure 11. SOC change of super-capacitor.

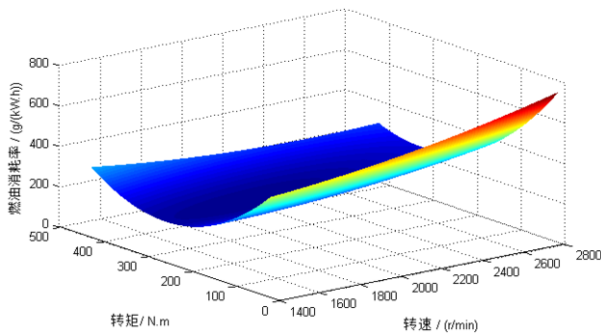


Figure 12. Engine Fuel Characteristics Surface.

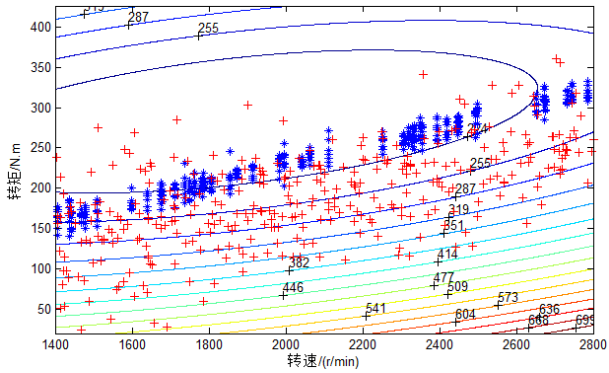


Figure 13. Fuel consumption of the engine under the two control strategies.

point of the engine based on the fuzzy strategy. It is obvious that the blue \* engine working points are mostly between 224–255 (g/(kw.h)) and the distribution is concentrated; the red + engine operating points are generally distributed between 446–224 (g/(kw.h)), and the distribution is dispersed. Therefore, the energy control strategy based on wavelet-genetic algorithm is more stable than the fuzzy energy control strategy and has better fuel economy.

## Acknowledgments

The research described in this paper was financially supported by National Natural Science Foundation of Youth Fund Project (61503163), Changzhou Science and Technology Support Plan Project (CE20150084), National Spark Program (2015GA105003) and National Torch Program (2014GH590518).

## References

- [1] J. Song, X. Zhang, Y. Tian. Research on control strategy of mild parallel hybrid electric vehicle [C]. International Conference on Electric Information and Control Engineering. IEEE, 2011:2512–2516.
- [2] Y. Ates, O. Erdinc, M. Uzunoglu, B. Vural. Energy management of an FC/Uchybrid vehicular power system using a combined neural network-wavelet transform based strategy [J]. International Journal of Hydrogen Energy, 2010, 35:774–783.
- [3] C.-Y. Li, G.-P. Liu. Optimal fuzzy power control and management of fuel cell/battery hybrid vehicles [J]. Journal of Power Sources, 2009, 192:525–533.
- [4] Y. Eren, O. Erdinc, H. Gorgun, M. Uzunoglu, B. Vural. A fuzzy logic based supervisory controller for an FC/UC hybrid vehicular power system [J]. International Journal of Hydrogen Energy, 2009, 34: 8681–8694.
- [5] H. Zhong, F. Wang, G.Q. Ao, et al. An optimal torque distribution strategy for an integrated start–generator parallel hybrid electric vehicle based on fuzzy logic control [J]. Proceedings of the Institution of Mechanical Engineers Part D Journal of Automobile Engineering, 2008, 222(1):79–92.
- [6] M. Amirabadi, S. Farhangi. Fuzzy Control of a Hybrid Power Source for Fuel Cell Electric Vehicle using Regenerative Braking Ultracapacitor [C]. Power Electronics and Motion Control Conference, 2006. Epe-Pemc 2006. International. IEEE, 2006:1389–1394.
- [7] X. Wu, X. Wang, L. Mao. A research on the fuzzy control of energy management for an ISG hybrid electric vehicle [J]. Automotive Engineering, 2011, 33(7):558–557.
- [8] C. Capilla. Application of the Haar wavelet transform to detect microseismic signal arrivals [J]. 2006, 59(1):34–46.

- [9] L. Joseph, T. Minh-Nghi. A wavelet-based approach for the identification of damping in nonlinear oscillators [J]. *Int. J. Mech. Sci.*, 2005, 47(8):1262–1281.
- [10] Y. Liu, N. Lu, X. Yin. A Hybrid Method for Structural System Reliability-Based Design Optimization and its Application to Trusses [J]. *Quality & Reliability Engineering*, 2015, 32(2):595–608.

# Design and Development of Chongqing Traffic Weather Service System Based on WebGIS

Shigang Han <sup>a,1</sup>, Baicheng xia <sup>a</sup> and Fanhua Min <sup>a</sup>

<sup>a</sup>*Chongqing Meteorological Bureau*

**Abstract.** To satisfy the weather service requirements of Chongqing traffic departments, this paper designed the traffic weather service system of Chongqing (TWSSC) based on WebGIS by using B/S technology and taking the advantage of the client technology. TWSSC integrates highway information, highway meteorological observing data and forecasting and early warning service information, and apply such technology as tile map, Skyline 3D map and road weather inversion technique. TWSSC has functions of querying Chongqing highway information, monitoring weather conditions in real time, issuing early warning information, and visualizing meteorological service data and products.

**Keywords.** WebGIS, traffic weather, system development

## 1. Introduction

Highway transport and communication is the basis of national economic development, which is the prerequisite for the normal operation and coordinated development of production, circulation, distribution and consumption. As for the research of highway safety under adverse weather, some developed countries and regions have already developed the meteorological observing and forecasting system to make highway transport and communication safer under some adverse weather conditions <sup>[1-3]</sup>. Some advanced provincial meteorological agency in china have carried out related research, for example, Jiangsu Meteorological Bureau has make remarkable progress on the fog monitoring and forecasting, Shanghai Meteorological Bureau has taken the lead in developing a system for monitoring and forecasting the visibility of urban roads, which has become an integral part of the Shanghai urban road information system <sup>[4]</sup>. Guangdong Meteorological Bureau has carried out the research on the development and application of mountainous highway visibility forecasting system, which have produced acceptable forecast accuracy <sup>[5]</sup>. The highway mileage in Chongqing creeps up to 1000 kilometres. It forms the skeleton of the “three ring, ten shoot, triple highway”, covering almost all the cities of Chongqing. But the highway safety in Chongqing is threatened by such serious meteorological disasters as dense fog, heavy rain, thunderstorm, high wind, snow, freezing, high temperature, etc. To mitigate the loss of lives and money caused by adverse weather conditions, Chongqing municipal government has urged meteorology and transportation departments to cooperate better, actively develop meteorological monitoring and early warning software, and manage

---

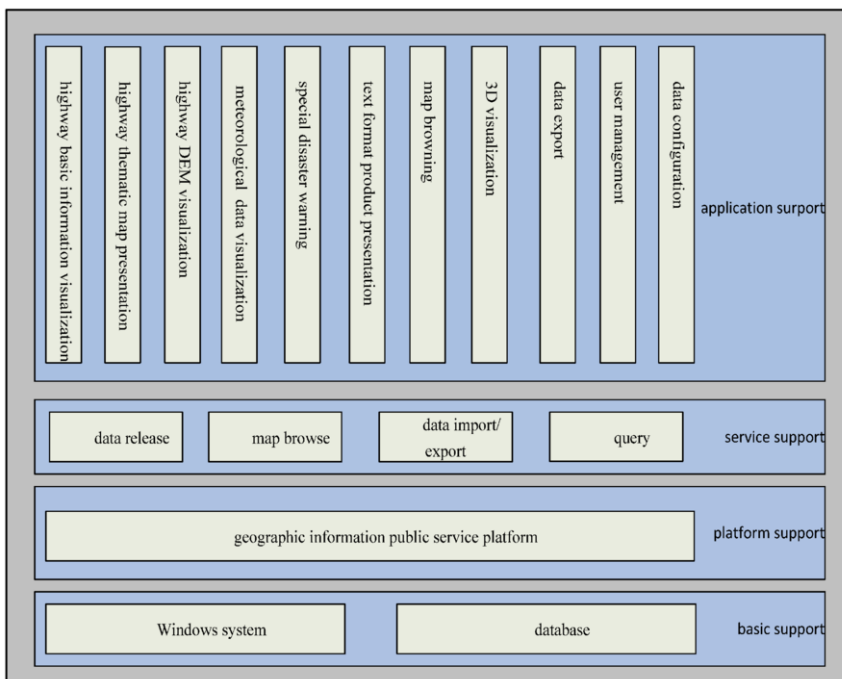
<sup>1</sup> Corresponding Author.

highway traffic more scientifically. TWSSC will help Chongqing transportation department enhance its ability of disaster prevention and mitigation, and actively deal with the adverse effects of disastrous weather.

## 2. System design

Based on the B/S architecture, TWSSC is developed to provide a variety of map applications using ArcGIS software and some API interfaces, which is made by Chongqing public service system of geographic information office. The foundation layer provides necessary functionality. API interface provides basic map services and subject services, and Skyline software is used to provide function of 3D.

As shown in Figure 1, TWSSC provides the necessary functional applications by support of the foundation support layer. TWSSC is based on the perfect physical layer. Physical equipment mainly includes servers and network equipment. The software support layer provides development and deployment environment and necessary database software support. The data layer is composed of map data and transaction data. Base map data is accessed and issued by way of tile data. Thematic map data is prepared and released inside TWSSC. The geographical information data includes base map data, highway thematic data, 3D terrain data etc. Transaction data, such as basic information data, thematic information data, service data, system management data, covers a wide range of data. They are all necessary support of the system. The platform layer supports the operation of the system and the data flow. The map platform layer provides the data. Interface platform layer provides corresponding interfaces. The



**Figure 1.** System architecture

application layer has functions of system management and displaying basic information, service product and 3D data. The user layer is composed of various types of users, such as professional subscribers, monitor, administrator and public visitor, etc.

### 3. System development and application

TWSSC is expected to be able to visualize Chongqing highway information, meteorological observing data and related products, provide weather warning and timely effective meteorological serve for the highway management department.

#### 3.1. Data preparation

The related data is firstly prepared. Spatial data includes base map data, highway data, 3D terrain data. Traffic data includes basic data, thematic data, products and service data.

#### 3.2. Software development

As shown in Figure 2, TWSSC is composed by four modules, including basic information visualization module, service product visualization module, three-dimensional visualization module, and system management module. The basic geographic information functions of all modules as positioning, browsing, map query and other related services are provided by the public service platform of geographic information office. The basic thematic application data of all modules is provided by the highway and weather database.

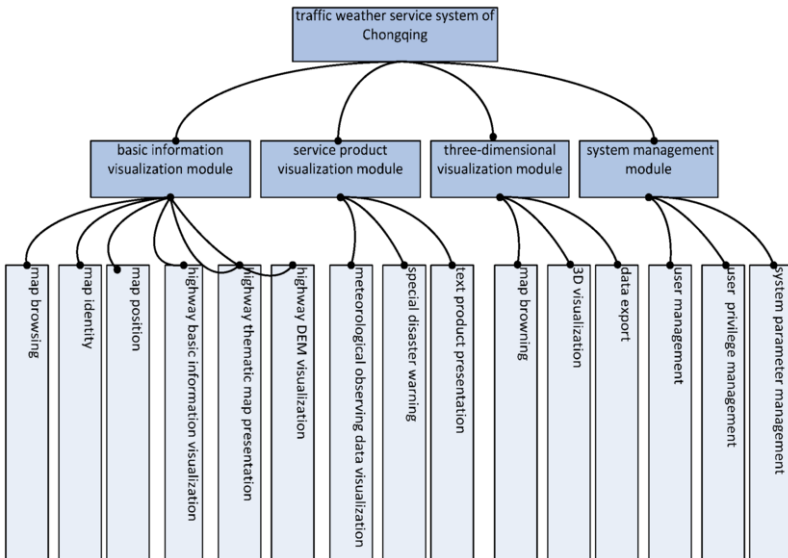


Figure 2. System function diagram

Basic information visualization module can visualize on basic map highway route ID, the key intersection, cities, towns, information billboard, toll gates and service stations along the highway. The module can get and make curve graph of elevation data along specific highway routes too.

Service product visualization module can visualize on basic map observing data from meteorological stations along high way routes, which covers such element as visibility, road conditions, road surface temperature, wind, rainfall, air temperature and relative humidity. If an element value exceeds the pre-set threshold, it will be marked with the corresponding icon the map, and the value will be marked in red to indicate an alarm. Some products, such as disaster warning and forecasting products for fog, heat wave, gale, heavy rainfall, freezing, are prepared and stored in Microsoft word or text format documents. They can be listed and browsed inside TWSSC. Other products, as daily report, weekly review, monthly report, important report, forecast abstracts and loss assessment report can be listed and browsed too.

Three-dimensional visualization module is used to visualize highway routes overlaying three-dimensional map data, and browse relevant meteorological observing data and forecasting and early warning data in flight mode along specific route.

### 3.3. Key technology

**Tile Map:** TWSSC provides a variety of map applications with ArcGIS. Its base map is provided in form of map service by the public service platform and issued by way of tile-map technology. The map service mainly consists of the tile map database and Ajax technology. Server-side tile map library deals with hierarchical tile cutting of the spatial data in the database, generates tiles of different levels, and establishes map tile pyramid model. It manages tile map data storage using the way of database or directory, and creates linear four-binary-tree tile index. Each level of pyramids corresponds to certain scale of the data set. Scale of the bottom level is largest, and the top level has the smallest. Each scale data set corresponds to one or more levels of pyramid map tiles. Ajax technology is applied to provide a variety of rich client operations, so that displaying details of map tiles and geographic information can be perfectly controlled so as to meet the complex needs of the map operation. When an extent of maps is requested, the server only returns the corresponding map tile data to the user which is not generated in real time, so the burden of the server is greatly reduced, and the response time of the system is shortened too.

**Three-dimensional Skyline development:** Skyline software can create an interactive 3D map environment by using aerial image, satellite data, digital elevation model and other data sources. It not only allows users to quickly integrate data, update the database, and support large database and real-time information flow communication technology, but also can display 3D geospatial images almost in real time<sup>[6-7]</sup>. Calling skyline API, TWSSC can access Chongqing 3D terrain data, and overlay highway data layer above it, and display three dimensional terrain map along the specific route at specified velocity, observing data and forecasting data can also be displayed as kind of reference information when necessary.

**Road weather data inversion technology:** by overlaying weather data, forecasting and early warning data and highway data, we can figure out the extent and degree Chongqing highway is affected by adverse weather, and produce products in

geographic or text format. Geographic products include all kinds of road thematic maps, which render weather condition maps with different colours. Text product describes in detail the type of weather and the degree of influence.

#### 4. Conclusion

TWSSC is developed by integrating traffic weather database and WebGIS technology. The software can display highway data, and access meteorological observing data in real time early warning information. It can display meteorological products that made for highway managing departments. TWSSC can be used by relevant departments to penetrate into highway information and real-time meteorological data, and by professional subscriber to browse meteorological service products. TWSSC can issue and deliver meteorological observing, forecasting and warning data and products, and has powerful human-computer interface and functions. Its degree of automation is high, which ensures TWSSC timely issue and orderly store forecasting products. At present, TWSSC is able to satisfy the needs of the highway management department, and provide meteorological services for the higher authorities and professional subscribers.

#### References

- [1] A. Sass. Numerical forecasting system for the prediction of slipper roads. *Journal of Applied Meteorology*.1997,36(1):801-817.
- [2] J. Thornes. A comparison of UK road ice prediction models. *Meteorological magazine*.1991,61(3):51-57.
- [3] A. Keskinen. Road Weather 30 Years- Hindsight, Insight, Foresight. *Proceedings of the 12th SIRWEC*. German Press, 2004.
- [4] F.Qian. Climatic Characteristics analysis and prediction of heavy fog in Nanjing-Shanghai highway route. *Research and Application of atmospheric science*.2001,32(2):32-37.
- [5] Y X.Yang. An envisage of developing a real-time meteorological disaster monitoring and decision-making system.*highway*.2003,50(7):90-93.
- [6] B Z.Chen. Development of three-dimensional piping systems based on the Skyline. *Modern Surveying and Mapping* : Proceedings of the 14th surveying and mapping meeting for the six provinces and one city Institute.2012, pp. 174-175.
- [7] L J.Zhao. Highway 3D GIS system development based on skyline. *Proceedings Symposium of Computer branch of China Highway and Transportation Society*.2010, pp. 9-10.



# PCA Based Improved Face Recognition System

Fayaz Ali DHAREJO<sup>a</sup>, Munsif Ali JATOI<sup>b</sup>, Zongbo HAO<sup>a</sup> and Majid Ali TUNIO<sup>c</sup>

<sup>a</sup>University of Electronic Science and Technology, China

<sup>b</sup>Indus University of Pakistan

<sup>c</sup>Mehran University of Engineering and Technology, Pakistan

**Abstract.** Due to the rapid evolution of modern image processing and pattern recognition techniques, there exists a variety of biometric techniques like fingerprints, iris (retina) scans, and speech recognition etc. nowadays. However, among them, face recognition is still the most common technique which is in use due to the fact that it is easy to install and has less complexity. It has been a prominent research field for security applications such as video surveillance, fraud detection, person tracking and crowd recognition. This research work discusses and implements the facial recognition by using MATLAB environment in real time. In this work, the face recognition system is implemented using Principal Component Analysis (PCA) and Eigenface approach by dealing with large dataset. A methodology is proposed to produce more accuracy and efficiency by removing the unrelated space from the image. Also, a graphical Interface System (GUI) is developed in order to make our system clearer and to measure the training time. Furthermore, large databases (ORL and private Database named face-100) are tested through PCA and Eigenfaces approach and then the person identification is carried out. The system successfully recognized the human faces and worked better in different conditions like illumination and blur conditions of the face. The rate of male and female accuracy is also calculated.

**Keywords.** Recognition, PCA, GUI, facial images, cropped function, large database, gender recognition

## 1. Introduction

Face recognition is a part of bio-metrics. In bio-metrics methods, basic traits of human are matched to the existing data and result depends on matching the identification of a human being is traced. Facial features are extracted and implemented through algorithms which are efficient and some modification are done to improve the existing algorithms as described early techniques (PCA, ICA, LAD) [1]. The detection and recognition of faces can be done by computers and can be applied to a variety of practical applications such as criminal identification, security systems, identity verification etc. Face detection and recognition are used in many places nowadays, in social networking sites and websites hosting images [2–10]. Face recognition is the most common technique among other biometric techniques due to its easy installation anywhere, in buildings, airports, hospitals, streets, superstores and other places to recognize individuals among the crowd. Therefore, this system is unique than others due to these particular applications. Face recognition is not perfect and suffers from various conditions such background variance, Scale variance, Orientation variance, illumination variance,

Emotion variance, Noise variance, etc. and still these problems are under the observations [3]. Due to these challenges researchers are engaged to increase the rate of accuracy by working on different algorithm [1,2].

Since 1960s the automatic recognition carried out and many improvements are made through modeling techniques advancement in algorithms but by automatic recognition many problems creates in effect of face recognition so it is remaining a challenging issue [11]. Automatic recognizing is of the face can be divided into two groups based on the face representation [9].

1. Appearance-based, is holistic texture features and is work as to either whole-face or specific regions in a face image.
2. Feature-based, which uses geometric facial features (mouth, eyebrows, cheeks etc.), and geometric relationships between them [8–10].

### *1.1. Overview of the System*

There are many advanced techniques for face recognition such as SVM, LDA, Cluster technique, PCA etc. Support Vector Machine (SVM) is useful technique for classification but it will not work when we have some missing entries in feature vectors [11]. PCA and LDA both are redundancy and featured extraction methods while PCA is very old but it still dominate in some cases [11]. Many people worked on PCA before to produce the maximum possible results. In this experiment, we have been tried to increase the performance by applying few parameters i.e. retina filter, cropped function, GUI. We have used the face based approach which has been proved effective technique. This is a Real-Time face recognition system based on Principal Component Analysis (PCA) algorithm that takes the images from digital camera, processes it and identifies the person from the database. Features extracted from a face are processed and compared with most close face present in the database. If a face is recognized it is known and if not recognized or not existing already in the database it is unknown. The system block diagram is shown in Fig. 1.

## **2. Methodology**

### *2.1. Face Recognition Process*

Face Recognition process can be divided into two parts one is image processing and other is face recognition. Image processing part includes image acquisition, detection while the face recognition process part includes an artificial intelligent part that is composed of Principal Component Analysis. PCA is used for feature extraction and dimension reduction [1,6]. In this experiment the Images are taken using a HD web camera LOGITEC C920 which is used as sample prototype then only face of image is detected and trained in real time during training the face image is proceeded by using geometric normalization, the features of face image are extracted using PCA technique then the data is stored in users database. The recognition process deals with the face of image which is acquired to identify the person. The identification compares the acquired biometric information against corresponding all persons in the database while verification deals with the only those templates to corresponding the claimed identity.

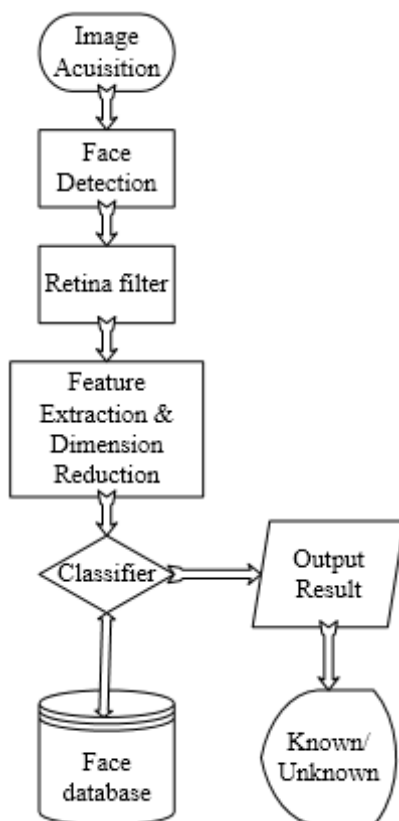


Figure 1. System Block diagram.



Figure 2. Effect of Retina filter.

*Image acquisition:* In our system the image is taken and then extracts a region of face then face detection is applied through PCA to detect the face and extracts the information according to features matched from the database [2].

*Use of Retina filter:* In this research work, our recognition process after detection, we also use retina filter to overcome the illumination variation problems because at different lighting the recognition performance altered. After detection if the face appears to be dark then retina filter helps to overcome the lighting problem and due this our recognition performance is improved [11].



Figure 3. Sample Images ORL database.



Figure 4. Sample Images Face-100 database.

Some detected images do not imply so clear result so after applying retina filter all images can be

*Face recognition/verification:* During face recognition module feature extraction and classification process, the input to this process is face image from two sources camera and database during the feature extraction the normalized image is considered as feature vector [4]. The purpose of classification is matching the captured image with the image stored in the database.

In our experiments, we tested two types Databases. One is well known ORL database that is taken at Olivetti Research Laboratory of Cambridge university of United Kingdom from 1992 to 1994 which are consisting of 400 images of 40 people in different 10 poses. These images were collected by in collaboratively. Shown in Fig. 3.

We created the second database (Face-100) in order to compare our system how accurately respond in different conditions. In this database we increased the number of poses, 100 images of 5 people in different 20 poses, there are some Images from our database as shown Fig. 4.

While improving the performance of the system under the problems of blur and illumination, the input images are given to the system then these images are normalized by preprocessing step. Then the database is used as input images from ORL database for initial evolution. The system is tested by Using Principal Component Analysis in the feature extraction step.

## 2.2. PCA Algorithm/Eigenface Approach

PCA is very old and well-known method called principal component analysis which is used for reducing the dimensions of an image also called redundancy method. It simply tells how to reduce the data to lower dimension from complex data sets which produce

the optimal linear least-square (LS) decomposition of a training set. Kirby and Sirovich applied PCA for representing faces and Turk and Pentland extended PCA for identifying faces [3,5,10]. In many applications such as face recognition and image compression a helpful statistical technique called PCA, nowadays it's widely used especially for face recognition. PCA commonly referred to as the use of Eigenfaces [10]. The Eigenfaces method took a holistic approach to recognizing the face [7]. A facial image is a point from a high dimensional image space and a lower dimensional image is found, where classification becomes easy. It is the adequate and efficient method to be used in face recognition due to its simplicity, speed, and learning capability. The major advantage of using PCA is using Eigenface approach which helps in reducing the size of the database for recognition of test images.

Steps:

Let be the face image contains  $\Gamma(x, y) \in \mathfrak{R}^{M \times N}$  in our experiment. The image matrix  $I \in \mathfrak{R}^{112 \times 92}$  is taken into consideration.

1. Prepared the training faces where each face is represented by  $T_1, T_2, \dots, T_m$  and each face of the image is represented as  $\Gamma_i$ .
2. Compute the average mean of M training images

$$\psi = \sum_{i=1}^M \Gamma_i \tag{1}$$

3. Subtract the average face vector such that:

$$\Phi_i = \Gamma_i - \psi \tag{2}$$

Each face is subtracted from the average which is called mean centered image.

4. Calculate the co variance matrix which is constructed as:

$$C = \frac{1}{M} \sum_{i=1}^m \Phi_n \Phi_n^T \tag{3}$$

5. Calculate Eigenvalues and Eigenvectors of co variance matrix. Consider the matrix  $L \in \mathfrak{R}^{M \times M}$ . Then,

$L = A^T A$ . Then compute Eigenvalues  $v_i$  of  $L = A^T A$  such that:

$$A^T A v_i = u_i v_i \tag{4}$$

The relationship between  $u_i$  and  $v_i$  are provided below:

$$A^T A v_i = u_i v_i \quad \text{and} \quad A A^T A v_i = u_i A v_i \tag{5}$$

The Eigenvalues and their corresponding Eigenvectors are related as follows:

$$u_i = Av_i \quad (6)$$

Project the training samples on the into the Eigenfaces a.

The test image,  $\Gamma$  is projected into the face space to obtain a vector [13].

$$\omega_i = u_i^T (\Gamma_i - \psi) \quad (7)$$

where  $u_i$  is the  $i$ th Eigenfaces and  $i = 1, 2, 3, \dots, K$ . Finally, the weight is calculated as

$$\omega_i = u_i^T (\Gamma_i - \psi).$$

### 3. Experiment Results

#### 3.1. Testing Parameters

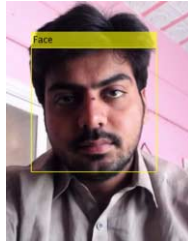


Figure (a)  
(112 × 92)  
(A colored face image)



Figure (b)  
(112 × 92)  
(Grey scale face image)

This system is developed on MATLAB R2013a environment. A colored face image is converted to grayscale image as gray scale images are easier for applying computational techniques in image processing. A gray scale face image is scaled for a particular pixel size of (112 × 92) because many input images are of different size whenever face is taken for recognition.

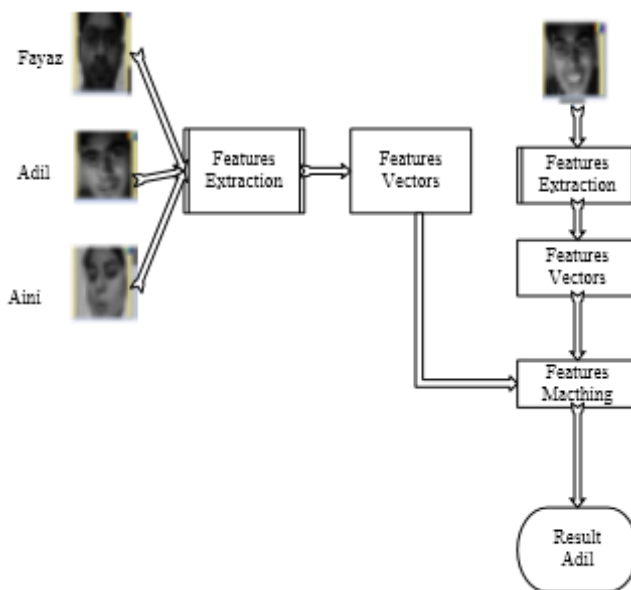
#### 3.2. Training Set

Database for the different set of conditions is maintained.

Hundred different expressions for forty different people and hence, creating a 100 × 40 that is equal to 4000 different set of face images. Rotated images in the left and right direction and by making the training set different illumination conditions are needed to be considered, size variations in an input face image can also change the output therefore, input images by varying their size are also taken for recognition.

The recognition process can be more understood from this diagram in which we have divided it into two phases i.e. One is Training phase and other is testing phase. These phases are elaborated below:

*Training Phase Testing Phase*



3.3. Graphical User Interface(GUI)

Nowadays, GUIs are widely used in many applications such as office mobile phones household devices, gaming device etc. The good accessibility, very high productivity and easy to use are the main advantages of a graphical user interface (GUI). Electronic GUI devices help users to finish their tasks at the higher rate as compare to the older interfaces. In comparison with text-based interfaces also, users do not need to remember and write many text commands, since GUI consists of interactive graphical elements with the same functions because to memorize the things also not a good factor to do always So GUI is very helpful to understand our recognition system.

There are two mainly types of text boxes, an edit text box and a static text box. To controls all groups together by using the panel button we can insert a panel around them. First, select the static text and then insert a static text box into the GUI. The property inspector can be used to change the text which is displayed by editing the String option. It appears to " $u'(x) + r(x)u(x)$ ". We have used an edit box to enter the functions  $r(x)$ ,  $f(x)$  and the boundary values  $u(0)$  and  $u(1)$ , our GUI is saved as differential equation as shown in Figure (a). It can change tags as well that is very useful in this project.

*Programming the GUI*

Clicking on the edit m-file button, it will open the automatically generated file which controls the actions of the GUI.

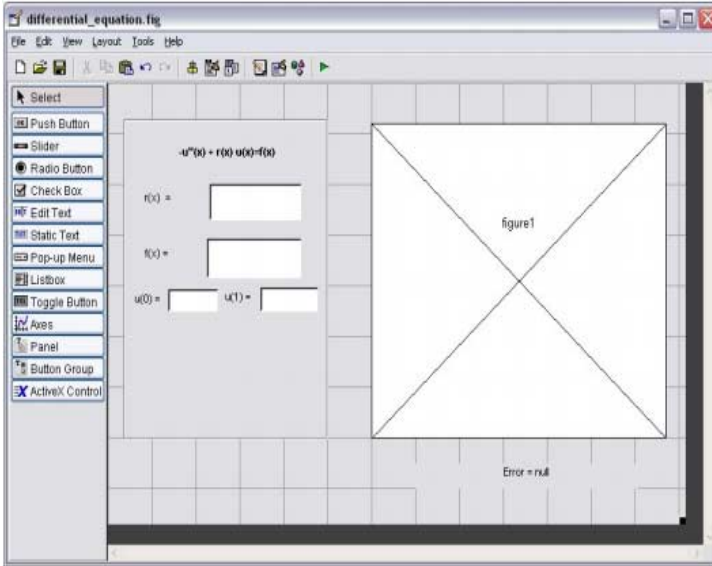


Figure 5(a). GUI system.



Figure 5(b). GUI system.





Figure 5(c). GUI system.

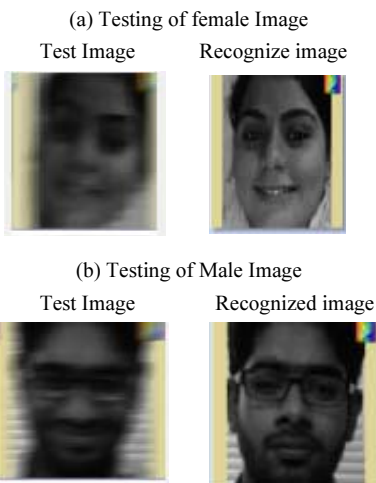


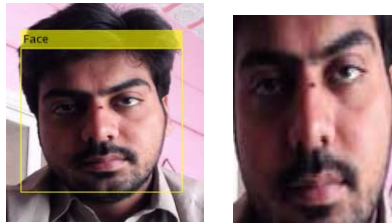
Figure 6. Male/female Recognition.

Table 1. Gender recognition rate variation according to number of training images.

Difference of male/female Recognition Rates	Number of training images			Total Average (%)
	300 (%)	200 (%)	100 (%)	
Female Recognition Rate	79	74.3	70.1	74.4
Male Recognition Rate	92	91.2	89.9	91.03

After training the images following recognition rate under observation, shown in Table 1.

This experiment is even tested under the blur images as well and it successfully recognized the persons from the database, in following table, we calculated the male and female recognition rate. For this whole experiment, we have images of five stu-



**Figure 7.** Cropped/Without cropped.

**Table 2.** Recognition Rate using PCA.

No. Testing Images	No. Training Images	Recognition rate (Not Cropped ORL database)	Recognition rate (Cropped Face-100 database)
1	2	90%	100%
1	10	100%	100%

**Table 3.** Recognition rates of real time images.

Technique	Number of testing images		Total Average (%)
	70(%)	90(%)	
PCA	92	88	90

dents with each having 20 poses collectively 100 images there, the input images are given to the system with the use of PCA technique.

### 3.4. Crop Function

Additionally, we have cropped the detected face for recognition to optimize the accuracy. ROI (Region of Interest) is extracted from frames or faces after detection process applying cropped function in our database (Face-100). A small difference is observed in rate of recognition in comparison with not cropped data base (ORL). The background variance problem can also be addressed through the usage of this function.

Finally, in Table 3, we have calculated the recognition rate of our system after applying all techniques.

## 4. Conclusion

In this experiment, we implemented the face recognition system using Principal Component Analysis (PCA) and Eigenface approach. The System recognizes male and female faces and provides 91% & 74% accuracy in results respectively. For this, various ways are adopted to optimize the performance in terms of accuracy. To address the issue of illumination, retina filter is used. However, by cropping face the unrelated background space from the face image is removed. A Graphical user Interference (GUI) is developed and tested to observe the recognition time and analyze the results in statistical way. The difference between recognition system for male and female faces was also observed in this research work and thus, the rate of accuracy for developed

system was estimated. While good results have been observed by implementing these techniques on specific database (ORL & Face-100), Improvement can be made further to deal with problem like big database, face orientation and different poses.

## Acknowledgment

I take this opportunity to thanks My supervisor Professor Zongbo Hao who provided me all resources and HIS precious time.

## References

- [1] L. C. Paul and A. Al Sumam, "Face recognition using principal component analysis method," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 1, pp. 135–139, 2012.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, pp. 711–720, 1997.
- [3] P. Shemi and M. Ali, "A principal component analysis method for recognition of human faces: Eigenfaces approach," *International Journal of Electronics Communication and Computer Technology (IJECCCT)*, vol. 2, pp. 139–144, 2012.
- [4] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91, IEEE Computer Society Conference on*, 1991, pp. 586–591.
- [5] Y. V. Lata, C. K. B. Tungathurthi, H. R. M. Rao, A. Govardhan, and L. Reddy, "Facial recognition using eigenfaces by PCA," *International Journal of Recent Trends in Engineering*, vol. 1, pp. 587–590, 2009.
- [6] P. S. Sandhu, I. Kaur, A. Verma, S. Jindal, I. Kaur, and S. Kumari, "Face Recognition Using Eigen face Coefficients and Principal Component Analysis," *International Journal of Electrical and Electronics Engineering*, vol. 3, pp. 498–502, 2009.
- [7] G.-C. Feng, P. C. Yuen, and D.-Q. Dai, "Human face recognition using PCA on wavelet subband," *J. Electronic Imaging*, vol. 9, pp. 226–233, 2000.
- [8] A. K. Bansal and P. Chawla, "Performance evaluation of face recognition using PCA and N-PCA," *International Journal of Computer Applications*, vol. 76, 2013.
- [9] N.-S. Vu, S. Schwab, P. Bouges, X. Naturel, C. Blanc, T. Chateau, et al., "Face Recognition for Video Security Applications," in *Workshop Interdisciplinaire sur la Sécurité Globale. Troyes, France (cit e p. 119)*, 2013.
- [10] N. H. Barnouti, "Face Recognition using PCA-BPNN with DCT Implemented on Face94 and Grimace Databases," *Methodology, International Journal of Computer Applications (0975-8887)*, vol. 142, no. 6, May 2016.
- [11] H. Jia and A. M. Martinez, "Support vector machines in face recognition with occlusions," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 136–141.

## Biography of Authors

**1. Fayaz Dharejo** was born in Sindh province, Pakistan, in 1993. He received the B.E. degree from the Quaid E Awam University of Engineering Science and Technology of Pakistan (QUEST) in 2016. He is currently pursuing the Master degree with the Department of Information and Software Engineering, UESTC. His research interests include Image Processing, Deep learning, Machine Learning.

**2. Munsif Jatoi:** Dr. Munsif Ali Jatoi has completed his Ph.D. (Electrical and Electronic Engineering) from Universiti Teknologi PETRONAS, Malaysia in 2016. Prior to this, he has completed M.Sc. (Advanced Photonics and Communications) from The University of Warwick, UK and B.E. (Electronics) from Mehran University of Engineering and Technology, Jamshoro, Pakistan in 2009 and 2007 respectively. Dr. Jatoi has more than 7-years of teaching experience locally and internationally in the capacity of Assistant Professor,

Lecturer and Graduate Assistant respectively. He has total 19 journals and conferences at his credit. He has presented his research work in various international exhibitions and won two silver medals for his performance in ITEX and SEDEX in Malaysia. Dr. Jatoi has filed 5 patents in the field of EEG source localization and has co-authored a book chapter with CRC Taylor in Francis, US. His research interests are brain signal processing, EEG inverse problem, Epilepsy prediction, brain connectivity and applied mathematics for neuroscience.

**3. Zongbo Hao** is an associate professor at the School of Information and Software Engineering, UESTC. He received his Ph.D. degree from Harbin Institute of Technology in 2007. He worked as an academic visitor closely with Prof. Ebroul Izquierdo in Queen Mary, University of London from Jan 2013 to Jan 2014. His research areas are Deep Learning, Image Processing, Surveillance based on Videos.

**4. Majid Ali Tunio:** He received the B.E. degree from the Quaid E Awam University of Engineering Science and Technology of Pakistan (QUEST) in 2016. He is currently pursuing the Master degree with the Department of Information and Software Engineering, UESTC. His research interests include Image Processing, Deep learning, Machine Learning.

# 60-GHz $15 \times 20$ Flat Array Antenna

Jiangmei TANG, Bin YAO and Qinhong ZHENG<sup>1</sup>

*College of Physical Science and Electronic Information, Yunnan Normal University*

**Abstract.** The 60 GHz band has gained great interest as an enabler for multi-Gb/s wireless links. Due to the high 60 GHz path loss (about 82 dB per 5 m), a high gain antenna is required to maintain budget requirements. Basically, phased-array antennas that provide higher gain and beam scanning capabilities are attractive for 60 GHz non-line-of-sight (NLOS) communication scenarios. This paper presents an array antenna consisted of  $15 \times 20$  array elements for working in 60 GHz. The substrate of the antenna is FR4 epoxy with relative dielectric constant 4.4 and dielectric loss tangent 0.02. The size of the array antenna is  $30 \times 20$ . The interval of antenna elements is 13.14 mm. The HFSS's simulation results show that the peak gain, efficiency, and impedance bandwidth of the flat array antenna can achieve 28.76 dBi, 95.978%, and 20 GHz, respectively. At the same time, the array antenna also has a good directivity.

**Keywords.** 60 GHz, array antenna, gain, wireless communication

## Introduction

Modern communication and sensor systems require a wider bandwidth, higher resolution and compactness resulting in operating frequency up to the millimeter wave (mmWave,  $f > 30$  GHz) or even sub-mmWave regime. 60 GHz band has a rich spectrum of resources, high wireless transmission rate, the advantages of high security. The multiple Gigabit/s wireless communication technology development at 60-GHz enables a new range of high throughput wireless applications such as high quality multimedia services, high definition video distribution, wireless HDMI and more. In the next generation mobile communication network 60 GHz band is considered as small cells, both indoor and outdoor. 60 GHz technology has attracted significant benefits and demonstrated great potential. Undoubtedly, it has become one of the major working frequency bands for future wireless communication systems. Therefore, the study of 60 GHz millimeter-band antenna design has attracted more and more interest in related fields.

Including the quasi-eight-chip patch antenna and inverted F-shaped antenna structure [1], the gain reached 19 dBi and 12.5 dBi, respectively, measuring bandwidth of 12.5 GHz and 10 GHz; There are single, double-serial cascaded microstrip rectangular patch array structure [2], in which a single serial cascade structure antenna gain is 13.4 dBi, and dual-serial cascade structure of the antenna Gain up to 14.6 dBi; And the use of grid array structure in the form of microstrip antenna [3], the gain of antenna is 13.5 dBi. The impedance bandwidth can achieve 8.7 GHz. The overall grid length is 11.085 mm, width is 4.85 mm. On the basis of this structure, the four grid arrays are arranged again to form a circularly polarized microstrip array antenna [4], Axial band-

---

<sup>1</sup> Corresponding author. E-mail: zheng\_qh62@aliyun.com.

width is 7 GHz. The antenna gain is 15.5 dBi at 60-GHz; There are also microstrip antennas based on textile material substrates [5] with gain up to 9.3 dBi; and a V-shaped dipole microstrip antenna [6] with a gain of only 4.4 dBi.

The high gain antennas shaped like comb-line can realize gain of 20.3 dBi [7]. Additionally, the tri-band antenna integrated array antenna and dual-band unipolar gain can reach 11.3 dBi at 64.5 GHz, the gain bandwidth can reach 13.5 GHz [8]. Also, a 4 × 4 microstrip patch antenna array based on the FR-4 substrate can achieve a high gain of 15.79 dBi and a measured bandwidth of 3.02 GHz [9–12]. In recent years, communication systems work more and more popular in the v-band because of their high speed of data transmission speed, stronger anti-jamming capability and higher security.

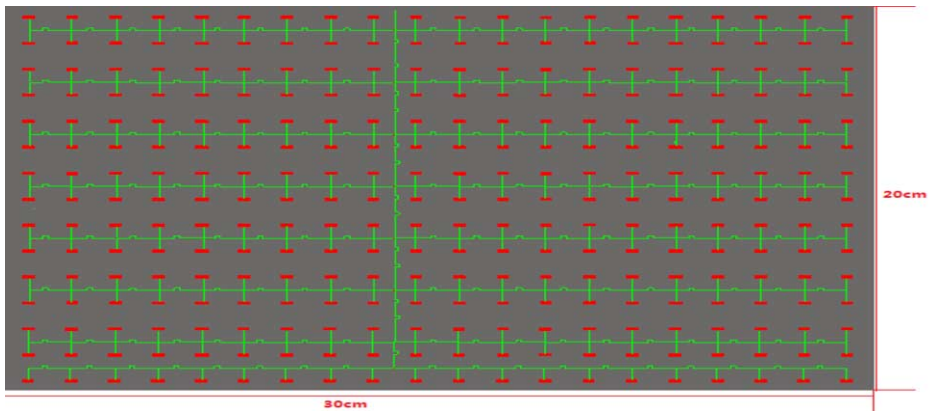
In particular, the 60 GHz millimeter wave communication system. Because 57 GHz to 66 GHz is an unauthorized band. Most proposed antenna arrays use complex structures in v-band, Such as FR-4 PCB / Epoxy Resin / Polypropylene / Epoxy Composite (PEPEC) Substrate [13], Low Temperature Ceramic High Gas Substrate [14–17], Substrate Integrated Waveguide (SIW) Feed Network. The IEEE 802.11ad standard supports PHY rates up to 6.7 Gb/s [21]. A 60 GHz system offers higher throughput than existing 802.11ac solutions.

Due to the high path loss at 60 GHz (about 82 dB for 5 m link), high gain antennas are needed to maintain budget requirements. Phased array waveguide antenna elements are preferred. This paper design a 60-GHz 15 × 20 waveguide array antenna. The simulation knows that the peak gain achieve 28.76 dBi, Which is the previous work did not materialize .This paper is organized as the following. Section 2 introduces the antenna structures including its principles. Describes the detailed of antenna designing procedures. The simulation (based on HFSS soldering via finite element method) and the measurement of antenna performance are placed in Section 3.

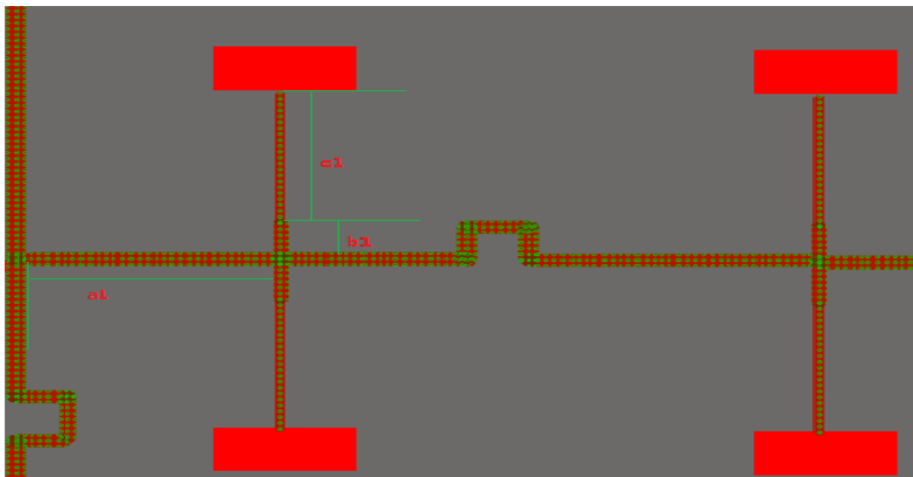
## 1. Array Antenna Design

The proposed antenna design is based on [24] when the interval of antenna elements is 13.14 mm. This proposed array antenna has an outstanding directivity.

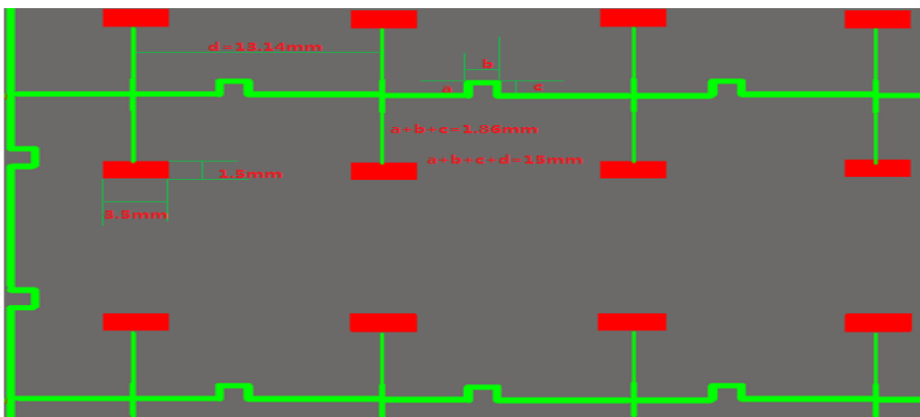
Figure 1 is drawn using software AD10. The array antenna design is shown in Fig. 1(a). The length of the entire array antenna is 30 cm and the width is 20 cm. The antenna feed point is designed to be the top position in the middle of the antenna and contains 300 elements. As shown in Fig. 1(c), the interval of antenna elements is 13.14 mm. The wavelength of the frequency for 60 GHz is 5 mm. According to the principle of the antenna unit length is generally 0.7 times the wavelength, width is 0.5 times the wavelength. everyone waveguide element length is 3.5 mm, width is 1.5 mm, height is 1 mm. The material used for this array antenna is FR4 \_epoxy material. The relative dielectric constant of this material is 4.4, and the dielectric loss tangent is 0.02. In order to achieve two elements phase equality  $a + b + c + d = 15 \text{ mm}$  . Which is three times the wavelength of 5 mm. As shown in Fig. 1(b), In order to achieve the impedance matching [25] using three types microstrips to connect two waveguide elements. The impedance of the antenna array is 50 ohm. The impedance of a1 is 40 ohm. The impedance of b1 is 44.7 ohm. The impedance of c1 is 50 ohm. The microstrip line



(a)



(b)



(c)

**Figure 1.** Antenna design (all units are in mm). (a) entire antenna model, (b) microstrip line design, (c) antenna fine structure.

bandwidth is designed according to [25]. The bandwidth of the microstrip line a1 is 0.481 mm. The bandwidth of the microstrip line b1 is 0.346 mm. The length of the microstrip line b1 is one quarter of the wavelength [26]. Because the quarter-wavelength microstrip line has an impedance conversion function. When the length of the microstrip line is equal to the quarter wavelength, the impedance of b1 is equal to the impedance of the a1 multiplied by the c1 impedance and then the square root. The bandwidth of the microstrips c1 is 0.241 mm. In order to realize the phase of the entire antenna equal the spacing of the previous line and the next line is also 15 mm.

## 2. Antenna Simulation and Analysis

The HFSS software simulation was used to verify the performance of the waveguide array antenna [27,28]. For the problem of plane periodic structure of antenna array, when creat a model based on HFSS simulation software, it is not necessary to create an entire antenna array model. Only can we need to create single array element model, then combining with the master-slave boundary conditions and port stimulus settings can construct and calculate the performance of any number of units formed array antenna. Figure 2 is a model of an element. Array element model consists of two parts, waveguide structure and free space. The ground rectangular waveguide surface is set to wave port excitation. In HFSS, since the surface of the contact background is automatically to set as the ideal conductor surface, the surface around the waveguide does not need to set the boundary conditions. The upper surface of the free space model needs to be set to floquet port incentivig, Used to simulate periodic surface radiation problems. The around of the free space model need to be set the master-slave boundary conditions. Fou surfaces of free space model has two master-slave boundary conditions. left and right sides are one master-slave boundary before and after sides are one master-slave boundary. The master-slave boundaries same team. The U,V axes direction must be

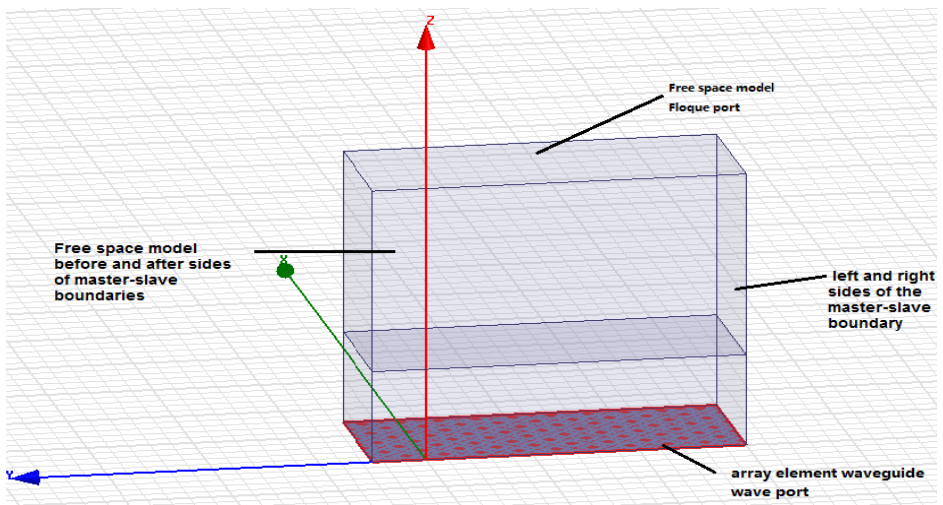


Figure 2. Single array element model.



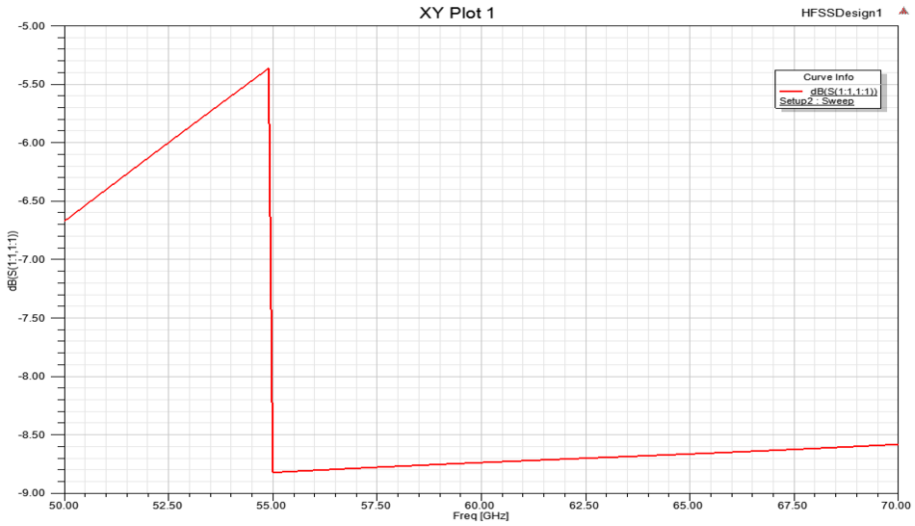


Figure 3. The change of  $S_{11}$  parameter along with frequency.

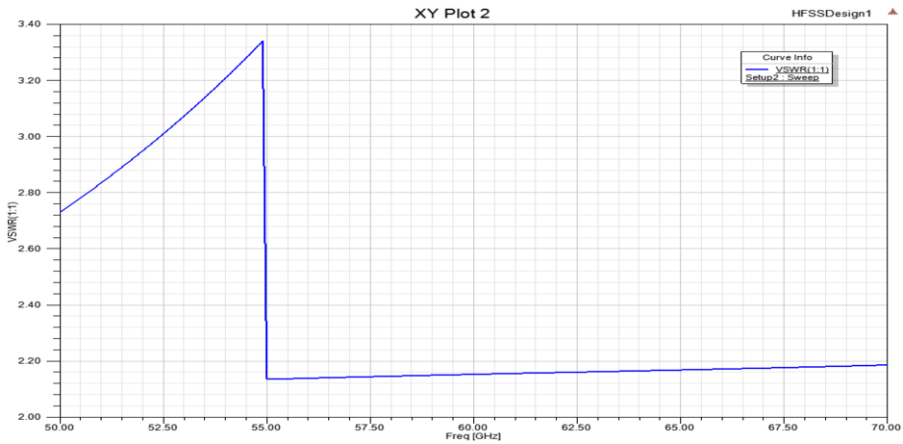


Figure 4. The change of VSWR parameter with frequency.

completely consistent. The scan angle or phase delay must be consistent with the sweep angle or phase delay of floquet incentivings. In this article, use scan angle to define the phase error between the two elements. Froming Fig. 3 we can see clear that the frequency froming 50 GHz to 70 GHz  $S$  parameter is less than  $-5$  dBi. The frequency froming 55 GHz to 70 GHz  $S$  parameter is less than  $-8.5$  dBi. Because the worst possible  $S$  parameter values allowed by the project are less than  $-5$  dBi [12]. So this antenna meet the requirements of engineering.

In the Fig. 4, Frequency from 50 GHz to 70 GHz the VSWR less than 3.4. Especially, when the frequency changes from 55 GHz to 70 GHz the VSWR less than 2.2. When the frequency is 50 GHz, the voltage standing wave ratio is the smallest. The minimum value of the VSWR is less than 2.16.

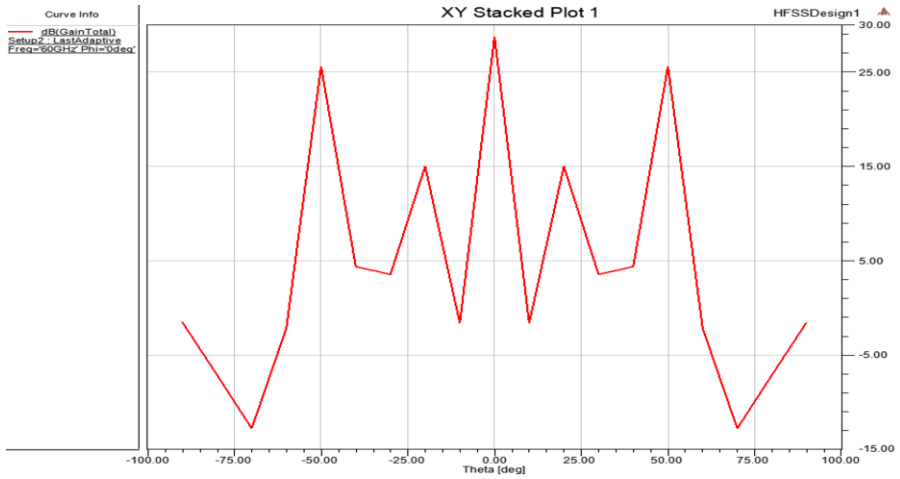


Figure 5. The change of gain with theta.

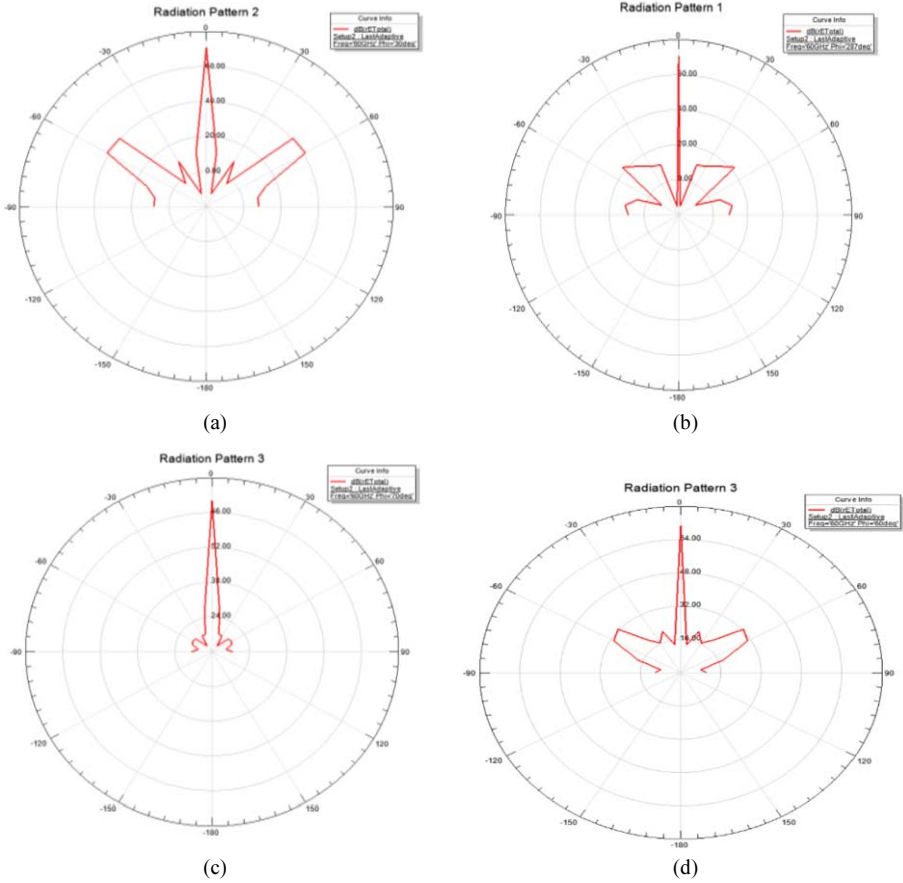
Froming the Fig. 5, we knowing the gain is given about  $\theta = 0$  symmetrical distribution when deg, the maximum gain of the waveguide array antennas is 28.76 dBi. When  $\theta = 70, -70$  deg, array antennas gain is 25.4 dBi. When  $\theta = 20, -20$  deg, array antennas gain is 15 dBi.

Figure 6 shows the simulated radiation pattern of rETotal. In the Fig. 6(a) freq = 60 GH, Phi = 287 deg, in the Fig. 6(b) freq = 60 GH, Phi = 30 deg, in the Fig. 6(c) freq = 60 GH, Phi = 70 deg, in the Fig. 6(d) freq = 60 GH, Phi = 60 deg. From  $-90$  deg to  $90$  deg, we can know the array antenna's main lobe extremely thin, the maximum value of rETotal is 70.05 dB in four pictures. Compared with the main lobe, the side lobe values is very small, very few. As are shown in the four figures, we knowing Gain is about  $\theta = 0$  symmetrical distribution. In the Fig. 7 the maximum of directivity is 29 dBi. When  $\theta = -50, 0, 50$  deg this array antenna has good directivity. Directivity is also about  $\theta = 0$  symmetrical distribution.

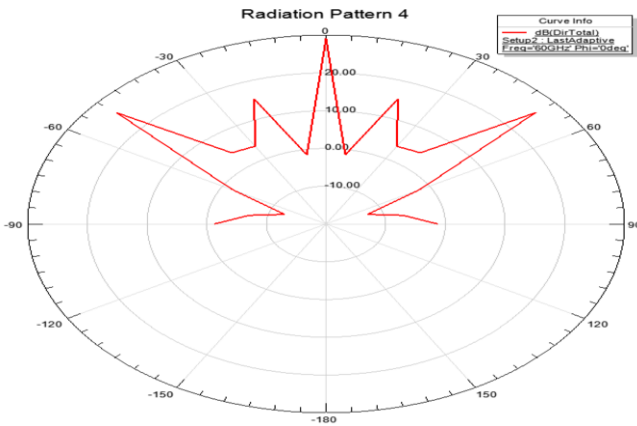
Figure 8 shows a three-dimensional radiation pattern simulation of rETotal at 60 GHz. As shown in Fig. 7, The 3-D radiation pattern like a durian uneven outside surface. This maximum value is 70.05 dBi. The entire distribution is symmetrica about XOZ and YOZ plane.

### 3. Conclusion

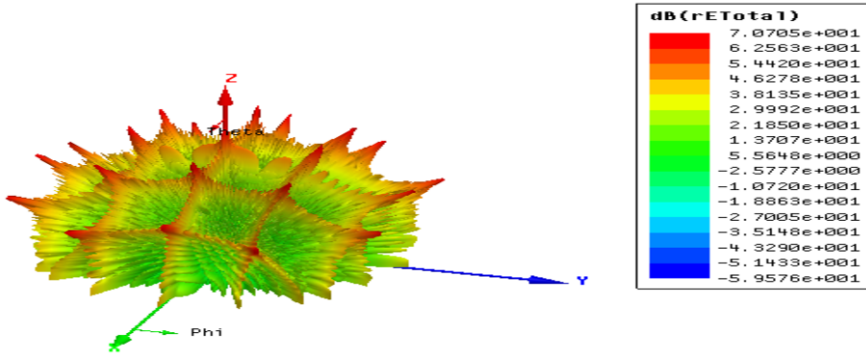
A 60-GHz  $15 \times 20$  waveguide array antenna designing are presented for mmwave commmunication. Compared to other 60 GHz array antenna designs, this proposed array antenna has an outstanding gain. This peak gain up to 28.76 dBi. The array antenna has a wider bandwidth. The impedance bandwidth of array antenna can achieve 20 GHz. At the same time, this array antenna also has a good direction. The proposed mm-wave antenna array is designed using flexible HDI PCB technology.



**Figure 6.** The simulated radiation pattern of rETotal. (a) Freq = 60 GHz, phi = 287 deg to simulate the radiation pattern of rETotal. (b) Freq = 60 GHz, phi = 30 deg to simulate the radiation pattern of rETotal. (c) Freq = 60 GHz, phi = 70 deg to simulate the radiation pattern of rETotal. (d) Freq = 60 GHz, phi = 60 deg to simulate the radiation pattern of rETotal.



**Figure 7.** The simulated radiation pattern of directivity.



**Figure 8.** Three-dimensional radiation pattern simulation of rETotal at 60 GHz.

## References

- [1] Kärnfelt, C., Hallbjörner, P., Zirath, H. (2006). High Gain Active Microstrip Antenna for 60-GHz WLAN/WPAN Applications. *IEEE Transaction on Microwave Theory and Techniques*, 54(6), 2593–2603.
- [2] Zhang, B., Zhang, Y.P. (2011). Analysis and Synthesis of Millimeter-Wave Microstrip Grid-Array Antennas. *IEEE Antennas and Propagation Magazine*, 53(6), 42–55.
- [3] Zhang, B., Zhang, Y.P. (2010). A Circularly-Polarized Microstrip Grid Array Antenna for 60 GHz-Radios, *Microwave Conference Proceedings (APMC)* (pp. 2194–2197).
- [4] Chahat, N., Zhadobov, M., Anwar, S. (2013). 60-GHz Textile Antenna Array for Body-Centric Communications. *IEEE Transactions on Antennas and Propagation*, 61(4), 1816–1824.
- [5] Chen, I.S., Chiou, H.K., Chen, N.W. (2009). V-band on-chip Dipole Based Antenna. *IEEE Transactions on Antennas and Propagation*, 57(10), 2853–2861.
- [6] Hayashi, Y., Sakakibara, K., Nanjo, M., Sugawa, S., Kikuma, N., Hirayama, H. (2011). Millimeter-wave microstrip comb-line antenna using reflection-canceling slit structure. *IEEE Transaction on Antennas and Propagation*, 59 (2), 398–406.
- [7] Tu, Z.H., Zhang, Y.P., Chu, Q.X. (2012). Tri-band antenna in package for single-chip WLAN/60 GHz radio. 2012 IEEE Asia-Pacific Conference on Antennas and Propagation (pp. 281–282).
- [8] Lu, B., Luo, J., Zhang, L., Wang, Y., Yue, R.F. (2012). A 60 GHz microstrip antenna array based on PCB/Polypropylene composite substrate. 2012 IEEE 11th International Conference on Solid-State and Integrated Circuit Technology (pp. 1–3).
- [9] Masood, Ur-Rehman, Qammer, Hussain Abbasi. (2015). Design of band-notched ultra wideband antenna for indoor and wearable wireless communications. *IET Microw. Antennas Propag.*, 9(3), 243–251.
- [10] Daniels, R.C., Heath, R.W. (2007). 60 GHz wireless communications: emerging requirements and design recommendations. *IEEE Veh. Technol. Mag.*, 2(3), 41–50.
- [11] Xue, Z.H., Li, W.M. (2011). *Antenna array analysis and synthesis*. Beijing: Beijing Aerospace University Publisher.
- [12] Hucheng, Sun., Ziliang, Wang. (2012). 60-GHz Circularly Polarized U-Slot Patch Antenna Array on LTCC. *IEEE Antennas and Propagation Society*, 8(10), 430–435.
- [13] Yang, T.Y., Hong, W.Y. (2014). Wideband millimeter-wave substrate integrated waveguide cavity-backed rectangular patch antenna. *IEEE Antennas and Wireless Propagation Letters*, 13(2), 205–208.
- [14] Gu, X., Kam, D.G., Liu, D., Piz, M. (2013). Enhanced multilayer organic packages with embedded phased-array antennas for 60-GHz wireless communications. *IEEE Electronic Components and Technology Conference (ECTC)* (pp. 12–16).
- [15] Bisharat, D.J., Liao, S., and Xue, Q. (2016). High gain and low cost differentially fed circularly polarized planar aperture antenna for broadband millimeter-wave applications. *IEEE Trans. Antennas Propag.*, 1(64), 121–126.

- [16] Wang, L., Sheng, W.-X. (2013). Wideband high-gain 60-GHz LTCC L-probe patch antenna array with a soft surface. *IEEE Trans. Antennas Propag.*, 4(61), 1802–1809.
- [17] Chin, K.-S., Jiang, W., Che, W., Chang, C.-C., Jin, H. (2014). Wideband LTCC 60-GHz antenna array with a dual-resonant slot and patch structure. *IEEE Trans. Antennas Propag.*, 1(62), 174–182.
- [18] Tang, J.M., Zheng, Q.H., Yao, B. Using the pattern function of angular momentum improve the directivity of binary antenna array, *Artificial Intelligence and Industrial Engineering Conference Proceedings (AIIE) 2016* (pp. 291–294). Beijing, China.
- [19] Li, X.Y. (2007). *Microwave technology and the microwave circuit*. Guangzhou; South China University of Technology Publisher.
- [20] Bozzil, M.A., Georgiadis, Wu, K. (2011). Review of substrate-integrated waveguide circuits and antennas. *IET Microw. Antennas Propag.*, 5(8), 909–920.
- [21] Li, M.Y., Liu, M. (2013). *Electromagnetic simulation design*. Nanjin: People's Posts and Telecommunications Publisher.
- [22] Collado, A., Via, S., Georgiadis, A., Bozzi, M. (2010). Optimized design of substrate integrated waveguide cavity based oscillators. *Applied Computational Electromagnetics Conf. (ACES)*. Tampere.

# Digital Image Watermarking Tools: State-of-the-Art

Surekha BORRA<sup>a,1</sup>, Lakshmi H R<sup>a</sup>, Nilanjan DEY<sup>b</sup>, Amira S. ASHOUR<sup>c</sup> and Fuqian SHI<sup>d</sup>

<sup>a</sup>*Department of ECE, K.S. Institute of Technology, Bangalore, India*

<sup>b</sup>*Department of Information Technology, Techno India College of Technology, West Bengal, 740000, India*

<sup>c</sup>*Department of Electronics and Electrical Communications Engineering, Faculty of Engineering, Tanta University, Tanta, 31527, Egypt*

<sup>d</sup>*College of Information and Engineering, Wenzhou Medical University, Wenzhou, 325035, P.R. China*

**Abstract.** The concept of watermarks started with inserting logos into paper currency and legal documents. However, drastic improvements in digitization process, devices and broad band communication have made digital transmission of data inexpensive, fast and eased the way for data storage, access, distribution and retrieval. The net effect is that data is easily copied, edited and redistributed without proper acknowledgements to the content owners without loss of information. The present work addressed these concerns by providing an overview of existing image watermarking tools, associated features, functional limitation and scope of future research and development. It also suggested the suitability of tools towards class of users and briefed possible attacks on image watermarking.

**Keywords.** Image watermarking, Visible watermark, Attacks, Steganography, Cryptography

## 1. Introduction

Sharing digital contents in the form of audio, images, and video increases with the rapid growth in the internet access. This surge in data sharing has led to a serious concern called Intellectual Property (IP) violation, which has various forms including counterfeiting, modifying users' data without authorization and claiming false ownership. This lead to huge loss monetarily for copyrights owner of such data [1]. Several software tools for steganography, cryptography and watermarking [2,3] along with various data copyrights management policies have been proposed in the literature to address the piracy related issues [4,5]. In addition, cryptographic tools for data protection allow data understanding by authorized users only. However, there is no prohibiting provision of the unauthorized users from copying and data distributing. On the contrary, steganography tools protect the data existence in some other information, but are limited to secure point-to-point data exchange.

---

<sup>1</sup> Surekha Borra, Department of ECE, K.S.Institute of Technology, Bangalore, India

Digital watermarking has been actually popularized in 1990 with rapid growth in internet technology and entertainment industry. The watermarks have been embedded in different types of media with various degrees of visibility to suit a variety of applications including ownership authentication and copyright protection. In digital watermarking, a watermark data is embedded into a multimedia entity, which can be detected to make a declaration about the entity. This watermark can be visible, such as visible logo or invisible to be hidden in the object. These watermarks are used to detect unauthorized copying and to suit the author authentication. Digital watermarking has a variety of applications including medical imaging, digital cameras, and image databases that are based on either spatial domain techniques or frequency domain techniques. These methods are carried out in watermark embedding/ extracting algorithms. Furthermore, the digital watermarking industry has gained its popularity after the appearance of music files free distribution by Napster file sharing services. Selecting proper watermarking tool should assure the quality of Intellectual Property. However, there is lack of availability about of current watermarking tools. Also, lot of new tools have been introduced recently, while many existing tools are being upgraded. This paper overviews some of the available online watermarking tools for embedding watermarks in an image. The organization of the remaining sections is as follows. Section 2 briefs the requirements of watermarking tools. Section 3 provides overview of the available online watermarking tool. Section 4 describes various types of attacks that may affect the efficiency of watermarked image. Section 5 lists the latest applications of watermarking and provides an overall discussion about the gaps and future directions. Lastly, Section 6 concludes the paper.

## **2. Digital Watermarking Requirements**

Digital watermarking tools [4] are widely preferred for data security in case of multiple distribution or publication of data. Watermarking process involves embedding copyright data (watermark) in Intellectual Property image either visibly or invisibly for the purpose of authentication of digital media, copyright protection, copy control or tracking digital distribution. The watermark is usually associated with the origin of cover image, such as ownership details or signature or logo. Visible watermark embedding creates visible trace of watermark on the cover image, while invisible watermarks hides the watermark imperceptibly in either spatial or transformed domain. Watermarking has an additional requirement of robustness towards the possibility of unintentional and intentional attacks on the cover image compared to steganography. There are several watermarking tools available in the market to cater for ownership authentication. The general watermarking terminologies include host, watermark, watermarked, attack and test images. A host image is a raw digital image/video/audio image that is to be copyrighted and in which the watermark is to be inserted. Watermark (W) is either a random sequence of bits or meaningful information, which has some significance with the host image (HS). It can be a logo, company, secret, message or a copyright notice. The watermarking phases are namely embedding, attacking, and extracting. During embedding, the watermarks are embedded either visibly or invisibly into host image with a variety of spatial/transform/hybrid embedding algorithms, predefined parameters and secret key. The embedding algorithm hides single/multiple watermarks in significant components of the host image and are

designed to be robust and secure even after corrupting/attacking the host image to some extent so that the watermark is retained in host image. Once the host image is watermarked, it becomes a watermarked image that can be stored or published, wherein the general image processing operations or attacker plays a vital role. In the insecure channel, attackers may try to remove the inserted watermarks subjects the watermarked image to various attacks, with a view of detecting/removing/distorting the watermark for various purposes. The embedded watermarks are extracted when desired by the owners at extracting phase using secret keys. The essential criteria used to analyze, to compare and to assess watermarking systems and tools for its suitability to any application are the imperceptibility, robustness, unambiguous, interoperability, security and false positive.

### 3. Digital Image Watermarking Tools

A variety of watermarking tools are available on the internet based on the type of host data, such as image watermarking tools, audio watermarking tools and video watermarking tools. Moreover, along with watermark creation, embedding and extraction some tools also allow modification of the host data as well as the watermarked data. There are some tools that may restrict the watermarks to only those available in their data base, whereas some tools may allow for creation of one's own watermark. The Table 1 provides brief descriptions of 26 image watermarking tools presently available online.

**Table 1.** Summary of 26 Image Watermarking Tools

Sl. No	Tool Name	Description
01	Umark Lite [6]	The tool has 4 distinct features arranged in separate tabs. It provides the option of either logo or series of texts to be set as watermark. The user can modify the text content, its color, shadow, font style, size and position. The opacity of logo can also be adjusted. An option to save the watermark texts or logos so that the user can use them consistently is also available.
02	WTM [7]	This is a powerful online tool that can insert visible watermark to users' images. It allows the user to personalize various aspects of watermarks, such as color, position and text size. The tool runs on Windows XP OS and higher versions. One can get access to Bernstein Compatible database of watermarks.
03	WMT PLUS [8]	This tool has a special feature that lets the user to watermark multiple photos simultaneously. User can insert text, symbols or image as watermark. Resizing, renaming multiple photos after watermarking make it an intuitive and user friendly tool. It supports various image formats including JPEG, PNG and TIFF. There are 17 in-built watermark profiles.
04	Water marquee [9]	This is an online tool for quick watermarking. It is a fully online tool with no downloadable version. Either text or an uploaded logo can be used as watermark. The font, color, style and position of the watermark can be adjusted as per the users' requirement. One can watermark up to 5 images simultaneously.
05	Lunapic [10]	This is an online watermarking tool with many options apart from watermarking. One can watermark the image with text along with Copyright or TM symbols. The users can apply color filters, add frames, and change the color effects to the image. One can adjust the position of the watermark. The tool also allows direct uploading of images using one's instagram or Facebook accounts.
06	Visual water mark [11]	This tool enables batch watermarking. The images can be resized after being watermarked. The tool offers several watermark templates and appearance options. The tool can be run on both Windows as well as Mac OS platforms. It is available for download in India, Australia and United Kingdom.



07	Mass water mark [12]	This tool allows simultaneous watermarking of multiple images. It supports text and image watermarks. It has Flickr and Picasa connectivity. Hence users can easily share the watermarked images. The users can also optimize the images by changing the color styles, brightness levels, sizes and contrast. Blur, smoothen, sharpen and mean removal features are also incorporated.
08	Water mark lib [13]	This free quick watermarking tool helps users protect their images using text or image watermarks. One can customize the font style, size, position, opacity and color of the watermark. The tool supports various image formats like JPEG, JPG, BMP and PNG. The tool offers simultaneous watermarking of multiple images. It allows simultaneous uploading of multiple images in a go from a specified folder.
09	Alamoon [14]	This watermark tool enables users to add text watermarks. It supports various image formats like JPEG, BMP, TIFF, GIF, VDA, ICB, PBM, PNG, PSP, PSD and PIX. It also provides some image enhancement features.
10	TSR water mark [15]	This is an easy- to- use tool. Even people who are not familiar with watermark tools can easily use this. Batch watermarking feature up to 100 photos is supported.
11	1-More water marker [16]	This watermark tool can insert text or emblem as watermarks to protect images against misuse. It allows batch watermarking of many images at a time. There is a "Hidden Text" feature included which allows imprinting of notes into the graphic file.
12	Photo Watermark Professional [17]	This is a trial version tool that can run on Windows platform. It requires less storage i.e. it is a light downloadable tool for watermarking images. It is widely used in India, Turkey and Pakistan.
13	Fast water mark [18]	This tool allows fast watermarking of images. It doesn't require installation. One can simply unzip the file contents to any folder and use it. The tool only allows text to be embedded as watermarks. It doesn't support images or graphics to be added as watermark. There are basic options like editing the text, size, color, etc. But more advanced options like changing shadow effects, opacity of text are not supported.
14	Watermark Master [19]	It is a video and image watermarking tool. It contains an easy interface, various filters, six different re-sampling algorithms, color adjustment, rotate and image sharpening. It supports text, GIF, image and blur area watermarks. There is a preview area where one can view the watermarked image before finalizing it.
15	Watermark.w s [20]	It is an online watermarking tool. It supports image and text watermarks. One can do batch watermarking of multiple image files. Batch image resizing and compressing is an added feature. There are both free and paid versions. The free version limits a maximum of 0.5 MB per photo. It doesn't let the user save preferences of watermarks for future use.
16	Cooltweak [21]	This watermarking tool allows users to resize the images after watermarking. The users can share the images after watermarking them. It is a freely downloadable light weight version. It guarantees adware – free experience.
17	Picture stamper [22]	This watermarking tool presents an easy to use interface. It supports text, image or logo watermarks. It can watermark multiple images at once. The tool saves users watermark designs for later use.
18	Water mark passion [23]	Watermark Passion is a small utility that provides user with a simple method of safeguarding pictures that user wants to use with blogs, articles or other content that the user posts online.
19	Easy watermark studio lite [24]	This watermarking tool adds text or image watermarks. It can watermark one or multiple images simultaneously. The tool offers preconfigured set of watermarks for easy copyright solution. It supports various image formats like TIFF, BMP and TIFF. It comes in freeware and paid versions. The paid Pro version includes various filters, image modifying options and customer support. Pro version offers multiple layers watermarking for images. It also allows for color conversion and image reflection options. The users can set the path and choose which folder to save the watermarked image on the system.
20	Bytescout watermarking [25]	This watermarking tool offers single and batch image watermarking. Only text watermark can be used. User can orient the watermarks either horizontally, diagonally or vertically. There are additional options like blur and texture of the image. The tool also allows changing the format of the watermarked images.
21	Snagit [26]	This tool offers batch watermarking. It is a screen capture and editing tool. It only allows image watermarks. Text watermarks are not supported. A text can be

		converted to an image and used as watermark in case it is to be used.
22	Jaco Watermark [27]	This is a free watermarking tool which allows text watermarks. This is an open source Java project. Users can change the font style, color, transparency and size. Batch processing is supported. The Pro version also supports image watermarks.
23	Star Watermark [28]	Star Watermark supports text and image watermarks. It supports batch watermarking. The tool is useful for artists who wish to protect their artwork and professional photographers. Also, company logos can be added as watermark for brand building.
24	Arclab Watermark [29]	Arclab Watermark Studio adds visible multilayered watermarks to photos. It supports batch watermarking. Both text and image watermarks can be inserted. The tool also provides options to resize, convert formats and create thumbnails for images.
25	BiggerBids Image Watermarking Tool [30]	BiggerBids Watermarking Tool is a robust watermarking tool for adding text and logo as watermarks. It provides various options like setting margins, aligning, and resizing the images.
26	Batch Watermark Creator [31]	Batch Watermark Creator adds image and text watermarks to photos, pictures and images. As batch processing is supported, thousands of files can be processed in few seconds. One can also edit the images, resize them, and create borders. Many image formats like JPEG, BMP, PIX, TIFF, TG, PBM, and GIF, are supported.

The tool mentioned in [6] includes various processing and memory management optimization techniques for speeding up the watermarking process. It provides batch watermarking option. During the batch process, the user can pause or cancel the operation. Apart from enabling users to protect their photos, this tool is also suitable for amateurs who want to experiment with watermarking. One can use the tool to tag names of the people in a photo, add description, such as date or time or place where the photo has been taken, add text descriptions about the photo and create memes or jokes by placing appropriate text as watermark on the image. Since the tool allows use of images as watermark, one can merge two images, wherein one is the host image and other is the watermark and creates a whole new watermarked image, which looks like a photo mashup. Hence, the users can try out various things with their photos as per their amusement and interest. The watermarking tool in [7] allows the users to insert visible watermarks. The tool in [8] supports text and image watermarks, wherein professional photographers can benefit from this tool as it supports batch watermarking. Also, a photographer can create a personalized watermark with the user name or institution name and add it to the whole batch. However, the tool doesn't support editing of some of the built in stickers, which is not very flexible. Water Marquee [9] is an online tool that requires no downloading. One can add multiple watermarks and it supports both text and image or logo watermark in the same picture. Many amateur wedding photographers have given good feedback about the tool as it allows them to add logo and text to a group of photos. However, the maximum size limit of the host image is 1 MB. Hence, it may be an issue for professional photographers with high resolution images in which image needs to be compressed such that it lowers its quality. Furthermore, one can perform batch watermarking of only 5 images at a time which is a small number for professional photography. Lunapic [10] is an online watermarking tool. It is more than just a watermarking tool in the sense that it allows users to add filters and animations along with watermarks. It is quite popular among active social media users as it allows the users to add and to edit photos directly through facebook and picassa accounts. The observation by various reviews, first time users who have never used watermark may find the tool little difficult to use due to its limited explanation, but it has many options. The Visual Watermark [11] is a downloadable

tool for watermarking. It has a special feature of adding metadata, which allows the photographers to add their contact details or details about the image that can be viewed in the file details. This is useful when the details are too big to be put as watermark. This includes automatic resizing and watermark placing based on features in the image. The Mass watermark [12] is a downloadable watermarking tool, which works on Windows and Mac platforms. This tool is useful for users who want to watermark multiple images in multiple folders quickly and there is no online version. Moreover, the tool is linked to the OS, wherein tool license will not work if the OS crashes and is reinstalled or updated. WatermarkLib [13] is a free watermarking tool that supports text and image watermarks. Alamoan [14] watermark tool is downloadable, which comes in free and pro versions. The free version has a disadvantage that only one image at a time can be watermarked. It is a useful tool that helps amateur photographers to add time stamp and descriptions to the images, but it contains no additional features like filters or animations. The TSR [15] is a watermarking tool, which offers text watermarks. The images or logos are not supported. Also, drag and drop option is not available. The 1-more water-marker [16] is a downloadable tool that supports text and emblems as watermark. It allows users to adjust intensity and transparency of watermarks. The main drawback in this tool is that it doesn't support many image formats. This makes it less lucrative for professional photographers. The Photo watermark tool [17] is a downloadable tool and it supports text watermark. The general user review is that this watermarking tool has excellent customer support system, which is very prompt in responding to user queries.

Fast watermark [18] is a downloadable tool that requires no installation, which is an added advantage. It only supports text watermarks. Also, the tool allows watermark to be placed in one of the 9 preset locations. This is not very useful in professional photography as the locations may end up being prominent positions in the image. Watermark master [19] is a video and image watermarking tool as it supports text, image, audio and video watermarks. The tool offers filters of various types. It is a tool which can be used for video and image editing and watermarking in the same platform. Moreover, the interface is simple and easy to use. Watermark.ws [20] is an online tool that requires setting up an account first. It provides various features for customizing the watermark as per user requirements. It is helpful for organizations or companies to add their logos to the image. Furthermore, the compress feature in the watermark is not very good as it degrades the quality of image considerably. Cooltweak [21] has a special feature, which creates a context menu in the Windows Explorer. There were few bugs in the menu bar options, but they have been fixed in the new versions. The tool offers fast watermarking solution for image copyrights protection. Picture Stamper [22] has a minimalistic yet user friendly interface. It allows adding shadow box around the watermark and supports various image formats.

Watermark Passion [23] is a downloadable tool that requires installation. The interface is very simple with few options. The tool doesn't provide much flexibility to place the watermark wherever needed and not included many customizable options. Hence, it is unsuitable for professional use. However, those looking for a simple watermarking tool to learn the basics may benefit from this. Easy watermark [24] supports both text and image watermarks. It allows drag and drop of watermark to desired position and has various features useful for professional use. In furtherance, there are certain restrictions imposed, such as, one can only import JPG, BMP or TIF

format images only. Also, the batch processing feature is limited to only 30 images per batch. Tool [25] is a screen capture and image watermarking tool. It has various image editing features that come in handy, but the main drawback is that it supports only image watermarks. Tool [26] supports only text watermarks and the interface is not user friendly and many reviews suggest that the service is not satisfactory. Tool [27] supports both text and image watermarks. Hence, it can be used for branding as one can set customized preset watermarks. However, the main drawback with this tool watermarks are stored as large thumbnails so that the user takes more time in browsing the list. Hence, the interface and layout can be made more users friendly.

The tool [28] allows text and symbol watermarks. The tool allows watermarking individual files or bunch of folders. However, if an unwanted folder is selected, there is no option to remove it, wherein the user has to start over again. The tool [29] is a quick watermarking tool with a simple interface in which the help file is very useful with FAQs and tutorial. In addition, the tool [30] provides various features like alignment, resizing, adding margins to the image along with watermarking. However, the tool has several bugs and upload hangs sometimes, wherein the user is forced to restart. Lastly, tool [31] allows logos and text watermarks, which supports batch watermarking. The demo version doesn't support customized watermarks.

#### 4. Types of Attacks on Watermarks

Although, there are plenty of watermarking algorithms and tools that are available online, watermarking is still a challenging task when dealing with attacks. The attack in watermarking terminology aims to remove or traces of watermark to thwart the authentication goals. Knowing the attacks types and their analysis assists the identification of the best tool for a particular application in competing environment. Tools such as, Checkmark [32], Optimark [33] and StirMark [34] are the few software that assess the robustness of the developed watermarking tools. The attacks in general are categorized into removal, geometrical, cryptographic, oracle, protocol, and security attacks.

- **Removal Attacks:** These attacks aims at removing watermarks from the watermarked image by intentionally performing image processing operations, such as quantization, lossy compression, averaging, re-modulation, demodulation, collusion attacks, block replacement attacks, de-noising and filtering. The Lossy compression of images which includes JPG and VQ compressions, which restricts the size of watermark. It has been observed that to perform re-modulation attack, the attacker first forecasts watermark using variety of filters (median, high pass and wiener), then subtracts it from the watermarked image and finally adds the Gaussian noise to it. This kind of attacks is commonly referred to as collusion attacks. However, on the other hand, as a mosaic attack, the attacker splits the watermarked image into small portions and tries to reassemble it using HTML table, with an intention of removing the inserted watermark.
- **Geometrical Attacks:** This aims at distorting the watermark rather than removing it by creating non-synchronization among extracted and original watermarks. The attackers try cropping the image from sides, delete/edit/shift some rows or columns of pixels randomly or with an intension of distorting the watermark. This creates synchronization problems, while extracting the watermarks. It is common to apply a combination of attacks than a single attack. The common geometrical attacks are

rotation, shearing, translation, affine transformation, scaling, aspect ratio changes, cropping, column/line removals, jitter, and random bending. It is to be noted that these attacks can be applied either locally or globally and in systematic or random way.

- **Protocol Attacks:** It aims at creating ambiguity of ownership and by attacking the watermarking application itself, using the concept of invertible watermarks. The protocol attacks sets another requirement for design of watermarking tools i.e., watermark extraction must be impossible from any images, which are in fact not watermarked. In other way, the false positive rate must be at minimum. Also, protocol attacks take the advantage of loopholes in the management or implementation process of watermarking.
- **Copy Attacks:** It aims at prediction of watermark and replicates it on to other data without the knowledge of secret keys involved. Also, development of image-dependent watermarks is the best solution to thwart these kinds of attacks.
- **Legal Attacks:** It attempts in doubting the technical evidences on watermarks and watermarking schemes, while proving the ownership in the courts.
- **Cryptographic Attacks:** This envisages in finding the lengthy secret keys by exhaustive searches.
- **Oracle Attacks:** It aims at generating the original host image from watermarked image using a watermark detector algorithm.
- **Disable Detection Attacks:** It aims at breaking the relation between the watermark and host carrying it without affecting the existence of watermark.
- **Ambiguity Attacks:** The attacker embeds multiple fake watermarks in order to mislead the detector. The ambiguity attacks occur in systems with multiple watermarks. In such cases, sometimes the order in which the watermarks are inserted is ambiguous.

## 5. Discussion

The study of piracy and security issues is an evergreen interest due to drastic improvements in the digitization process, devices and broad band communication that makes the digital transmission of data inexpensive, fast and easy day-to-day. Copyright protection methods have evolved over two decades by now. The developments of new types of watermarking techniques, algorithms and tools have paved the way for a wide variety of applications [36-39], such as copyright protection/ownership assertion: healthcare, transaction tracking, copy control, annotation and privacy control, tamper detection, invisible marking on paper and ID card security. After analyzing the case studies of customers using these aforementioned watermarking tools, the most challenging issue that is existent in designing watermarking tools has been its robustness and invisibility of the watermark. These two characteristics trade off against each other because one affects behavior of other. For instance, if a watermark is placed at the region of interest of image, they would less probably be attacked, but the quality degrades. On the other hand, if watermarks are placed in non-region of interest, they can be cropped easily. With invisible watermarks, the watermarks purpose is served only as if they robustly survived in the host image irrespective of a variety of defined and undefined attacks.

The case studies indicated that in spite of availability of plenty of watermarking tools in the market; employing invisible watermarking of sensitive data such as military

maps, and medical images is not much practical and quite accepted mainly due to trade off among several properties of audio watermarking such as inaudibility, payload and robustness. It is noticed that still most professional artists/photographers are employing visible watermarks to protect their artworks/photos at multiple scales. Furthermore, the visible watermarking tools are popularly employed for displaying advertising messages. In furtherance, the extended requirements of watermarking tools may include real-time processing, mathematical formulations and support to multiple resolutions, multiple formats and multi-media. Moreover, the automated techniques and benchmarks for proper and simple assessment of watermarking tools for a particular application need to be developed. The development of efficient and intelligent watermarking techniques and tools to resolve the existing problems is the need for the hour. In addition to that, combination of machine intelligence techniques such as Fuzzy Logic, Artificial Neural Network (ANN) and genetic algorithms may resolve this trade-off issue [40-43]. Watermarking has several applications in the medical domain applications as reported by several researchers [44- 46].

## 6. Conclusion

The piracy has been a serious concern as illegally distributed copies are leading to huge loss to Intellectual Property owners due to rapid uploading of multimedia files over the Internet. Furthermore, the digital watermarking of multimedia content has become a very active research area in the recent years, which necessitates the understanding of watermarking tools. Watermarking tools are employed as a way to protect Intellectual Property rights and to prevent illegal forgery and piracy. The recent researches on watermarking tools have been concerned with proposing robust algorithms for watermark embedding and extraction in multimedia data host. The present research has analysed and presented an overview of more than 26 different image watermarking tools that are currently available online. This study summarized various features of watermarking tools and suggests the suitability of its usage in accordance with users. Also, briefed various possible attacks on image watermarking tools, which needs to be addressed at large in future towards copyrights for data protection in terms of ownership security and intellectual property.

## References

- [1] N. S. Gavini, S. Borra, Lossless watermarking technique for copyright protection of high resolution images, *2014 IEEE Region 10 Symposium*, (2014, April), 73-78.
- [2] H. R. Lakshmi, B. Surekha, S.V. Raju, Real-time Implementation of Reversible Watermarking, *Intelligent Techniques in Signal Processing for Multimedia Security* (2017), 113-132.
- [3] N.Dey, P. Das, S. Biswas, A. Das, S.S. Chaudhuri, Feature analysis for the reversible watermarked electrooculography signal using low distortion prediction-error expansion, *2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)* (2012),624-627.
- [4] N.Dey, S.S. Ahmed, S. Chakraborty, P. Maji, A. Das, A., S.S.Chaudhuri, Effect of trigonometric functions-based watermarking on blood vessel extraction: an application in ophthalmology imaging, *International Journal of Embedded Systems*, **9(1)** (2017) 90-100.
- [5] S. Acharjee, S.Chakraborty, S.Samanta, A.T.Azar, A.E.Hassanien, N.Dey, Highly secured multilayered motion vector watermarking, *International Conference on Advanced Machine Learning Technologies and Applications*, (2014) 121-134.
- [6] <https://www.uconomix.com/>
- [7] <http://www.pearlmountainsoft.com/watermark/>
- [8] <http://www.pearlmountainsoft.com/watermarkplus/>

- [9] <https://www.watermarquee.com/>
- [10] <http://www.196.lunapic.com/editor/>
- [11] <https://www.visualwatermark.com/>
- [12] <http://www.masswatermark.com/>
- [13] [http://download.cnet.com/WatermarkLib/3000-2192\\_4-10963794.html](http://download.cnet.com/WatermarkLib/3000-2192_4-10963794.html)
- [14] <http://alamoon.com/>
- [15] <https://www.watermark-image.com/>
- [16] <https://1-more-watermarker.en.softonic.com/>
- [17] <http://www.photowatermark.com/>
- [18] <https://fast-watermark.en.softonic.com/>
- [19] <http://www.videocharge.com/Products/wm/main.php>
- [20] <https://www.watermark.ws/>
- [21] <http://www.cooltweak.com/>
- [22] <http://amin-ahmadi.com/picture-stamper/>
- [23] [http://www.majorgeeks.com/mg/getmirror/watermark\\_passion.1.html](http://www.majorgeeks.com/mg/getmirror/watermark_passion.1.html)
- [24] <http://www.easy-watermark-studio.com/easy-watermark-studio/lite-version.html>
- [25] <https://bytescout.com/products/enduser/watermarking/watermarking.html>
- [26] <http://snagit.1.software.informer.com/download-snagit-watermarks/>
- [27] <http://jaco-watermark.sourceforge.net/>
- [28] <http://www.star-watermark.com/>
- [29] <https://www.arclab.com/en/watermarkstudio/>
- [30] <http://www.biggerbids.com/watermark/>
- [31] <http://batch-watermark-creator.software.informer.com/6.0/>
- [32] <http://watermarking.unige.ch/checkmark/>
- [33] <http://poseidon.csd.auth.gr/optimark/>
- [34] <http://www.petitcolas.net/fabien/watermarking/stirmark/index.html>
- [35] S. Borra, H.R. Lakshmi, Visual Cryptography Based Lossless Watermarking for Sensitive Images. *International Conference on Swarm, Evolutionary, and Memetic Computing*, (2015) 29-39.
- [36] A. Rashid, Digital Watermarking Applications and Techniques: A Brief Review. in *International Journal of Computer Applications Technology and Research*, **5 (3)** (2016), 147-150.
- [37] A.K. Pal, P.Das P, N.Dey, Odd-Even embedding scheme based modified reversible watermarking technique using Blueprint, (2013).
- [38] N.Dey, S. Samanta, X.S.Yang, A.Das, S.S.Chaudhuri, Optimisation of scaling factors in electrocardiogram signal watermarking using cuckoo search, *International Journal of Bio-Inspired Computation*, **5** (2013), 315-26.
- [39] B.Chen, G.W. Wornell, Quantization index modulation: A class of provably good methods for digital watermarking and information embedding, *IEEE Transactions on Information Theory*, **47** (2001), 1423-1443.
- [40] D. Wang, Z.Li, N.Dey N, A.S.Ashour, R.S. Sherratt, F.Shi, Case-based Reasoning for Product Style Construction and Fuzzy Analytic Hierarchy Process Evaluation Modeling using Consumers Linguistic Variables. *IEEE Access*, (2017).
- [41] T. He, L.Cao, V.E. Balas, P.McCauley, F. Shi. Curvature manipulation of the spectrum of Valence-Arousal-related fMRI dataset using Gaussian-shaped Fast Fourier Transform and its application to fuzzy KANSEI adjectives modeling. *Neurocomputing*, **174** (2016), 1049-59.
- [42] Z. Li, T. He, L.Cao, T. Wu, P. McCauley, V.E.Balas, F.Shi, Multi-source information fusion model in rule-based Gaussian-shaped fuzzy control inference system incorporating Gaussian density function. *Journal of Intelligent & Fuzzy Systems*, **29(6)** 2(015) 2335-44.
- [43] Y.B. Amar, I.Trabelsi, N. Dey, M.S. Bouhlel, Euclidean Distance Distortion Based Robust and Blind Mesh Watermarking, *International Journal of Interactive Multimedia and Artificial Inteligence*, **4** (2016).
- [44] N.Dey, S.Samanta, X.S. Yang, A. Das, S.S. Chaudhuri, Optimisation of scaling factors in electrocardiogram signal watermarking using cuckoo search, *International Journal of Bio-Inspired Computation*, **5(5)** (2013), 315-326.
- [45] N.Dey, S. Bose, A. Das, S.S. Chaudhuri, L. Saba, S. Shafique, et al., Effect of watermarking on diagnostic preservation of atherosclerotic ultrasound video in stroke telemedicine, *Journal of medical systems*, **40(4)** (2016) 1-14.
- [46] D.Wang, T. He, Z. Li, L. Cao, N. Dey, A.S. Ashour,..., F. Shi, Image feature-based affective retrieval employing improved parameter and structure identification of adaptive neuro-fuzzy inference system, *Neural Computing and Applications* (2016), 1-16.

# Fault Diagnosis of Mine Main Fan Based on Improved Particle Swarm Wavelet Neural Network

Yong Yang<sup>a</sup>, Xiaodan Yuan<sup>a,1</sup>, Chenchen Cui<sup>a</sup>

<sup>a</sup>School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China

**Abstract.** In this paper, due to no clear match relationship between the fault symptom and fault type of the main fan in coal mine, an improved particle swarm wavelet neural network model is used to diagnose the fault of the main fan in coal mine. Firstly, the vibration signal characteristics of 5 kinds of common faults of ventilators are introduced, and the feature vectors of faults are extracted by multi-scale wavelet transform. Then wavelet neural network is used for fault diagnosis, and the selection of initial parameters of neural network is optimized by the improved particle swarm optimization algorithm. The training speed of wavelet neural network is increased and the training precision is improved to get better fault diagnosis results by adding weight factor, steering operator and random disturbance. The simulation results show that the improved particle swarm wavelet neural network model has better stability, and the accuracy is improved by 2% by comparing with the traditional wavelet neural network. This method can be better applied to the diagnosis of ventilator.

**Key word.** Main fan, Vibration signal, Improved particle swarm optimization, Wavelet neural network, Fault diagnosis

## 1. Introduction

Mine main fan is the key equipment in coal mine production, it is mainly responsible for sending the fresh air to the coal mine and diluting the harmful gases in coal mine to provide a good working environment for underground workers. If the mine fan have failure or even serious things that the wind stopped, it will threaten the personal safety of underground workers, and also be a serious blow to coal production<sup>[1]</sup>. Therefore, it is necessary for the mine main fan to have a fault diagnosis, identify and solve problems to ensure the safety of coal mine production.

In recent years, a variety of ventilator fault diagnosis model was established by scholars at home and abroad, such as the BP neural network<sup>[2]</sup>, fuzzy neural network, support vector machine, wavelet neural network, in these methods, the wavelet neural network has strong robustness, and it is more suitable for the fault diagnosis of ventilator, wavelet neural network BP algorithm is used to train the network, and the BP algorithm is easy to fall into local minimum, train many times, and affect the

---

1. Corresponding Author, School of electrical and control engineering, Xi'an University of Science and Technology, Yanta Road No.58, Xi'an 710054, Shaanxi Province, China. E-mail: yxd0820yuan@163.com



accuracy of network training.

In order to solve these problems, taking into account the particle swarm algorithm has strong global search ability, applied to the neural network initial parameters of the neural network can effectively avoid falling into local optimum, this paper based on the standard particle swarm algorithm is added to the absorption operator and random disturbance of thought, it will put forward the model of improved particle swarm wavelet neural network diagnosis, verify the rationality and validity the model through simulation analysis.

## 2. Main fault types and feature extraction of ventilator

### 2.1. Main fault types and fault characteristics of ventilators

The main fault of mine main fan is divided into several types, such as the installation is not in place, failure occurs during debugging or in operation and so on, this paper mainly analyzes the fault in the running process. Common types of fan faults include rotor imbalance, rotor misalignment, mechanical foundation looseness, rubbing fault and blade failure.

The ventilator is composed of two parts, the fan and the motor, and the two parts are connected by shaft or coupled device. There will be vibration when the fan is working<sup>[3]</sup>. The vibration signal contains a rich variety of laws and characteristics. It is a common diagnostic method to analyze the vibration signals to diagnose the faults. When the fan is out of order, the vibration signal is different from the normal vibration signal. In one same frequency band, the frequency component will be very different, that is, the energy of the signal is different. The following table for vibration characteristics of fault types.

**Table 1.** The vibration characteristics of fault types

Fault type	Fault signal characteristic
Rotor unbalance	The energy is mainly concentrated on the fundamental frequency (fan operating frequency) The amplitude increases with the increase of rotating speed
Misalignment of rotor	The spindle frequency is mainly 2fr, it accompanied by fundamental frequency and higher harmonics
Pedestal Looseness	The frequency of vibration is mainly base frequency, fractional harmonic and sub harmonic, often accompanied by 2fr, 3fr
Rubbing fault	In addition to the fundamental frequency, it also contains some higher harmonics, such as 2fr, 3fr, etc
Blade fault	The vibration frequency is mainly $N * fr$ , and around the frequency, a large number of side bands are modulated by fr

### 2.2. Fault feature extraction of ventilators

The principle of Multi-scale wavelet analysis is that for any signal, can be decomposed into the high-frequency part of the  $D_1$  and low-frequency part of the  $A_1$ , and then the low-frequency part  $A_1$  is decomposed again,  $A_1$  is divided into high frequency part  $D_2$  and low frequency  $A_2$ <sup>[4]</sup>. So repeatedly the signal  $x(t)$  can be decomposed into any space.

$$x(t) = D_1 + D_2 + \dots + D_n + A_n (n = 1, 2, 3 \dots) \quad (1)$$

The Mallat algorithm can be used to calculate the wavelet coefficients in different

space<sup>[5]</sup>.

$$\begin{cases} c_k(n+1) = g_{l-2k}c_l(n) \\ d_k(n+1) = h_{l-2k}d_l(n) \end{cases} \quad (n=0,1,2,\dots) \quad (2)$$

Where  $g$  is low pass filter coefficient,  $h$  is high pass filter coefficient,  $c_0$  is the original time series, it can be decomposed into high frequency coefficient vector  $d_1, d_2, \dots, d_n$  and low frequency coefficient vector  $c_n$ .

The decomposed wavelet coefficients are reconstructed to obtain the energy of each frequency band, then the main energy of the frequency band is extracted. That is, the characteristic value of the signal in the frequency band.

### 3. The model of improved particle swarm wavelet neural network

#### 3.1. Wavelet neural network

Because the BP neural network learning efficiency is low, this paper improve the learning performance of neural network structure based on Neural Network by adjusting the network weights and using wavelet function instead of neural network hidden layer input function to optimize the neural network .

The compact wavelet neural network is used in this paper. Based on the topology of the three layer BP neural network, the wavelet function is used to replace the Sigmoid function in the hidden layer of the standard BP neural network<sup>[6]</sup>. The network structure of the neural network is shown in Figure 1. The number of network input nodes is  $n$ , the number of hidden nodes is  $m$ , and the number of network output nodes is  $l$ .

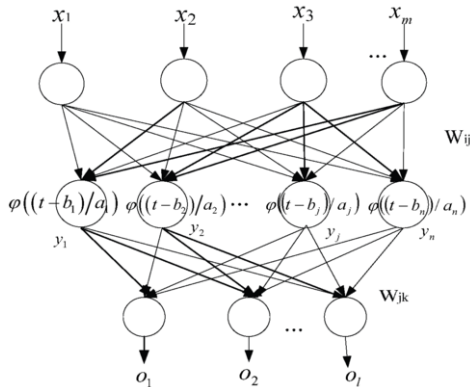


Figure 1. The structure of wavelet neural network

Among them, the wavelet function selects the representative Morlet function:

$$\varphi(t') = \cos(1.75t') \exp\left(-\frac{t'^2}{2}\right), \quad t' = \frac{t - b_k}{a_k} \quad (3)$$

Assume that the input of the wavelet neural network is  $x_i$ , and the connection weights between  $i$  node of the input layer and  $j$  node of the hidden layer is  $\omega_{ij}$ .

Therefore, the input of the hidden layer  $j$  node is  $net_j = \sum_{i=1}^n \omega_{ij}x_i$ , which is excited by the wavelet function Morlet to get the following formula.

$$y_j = \varphi\left(\frac{net_j - b_j}{a_j}\right) = \varphi\left(\frac{\sum_{i=1}^n \omega_{ij}x_i - b_j}{a_j}\right) \tag{4}$$

The connection weights between node  $j$  of the hidden layer and node  $k$  of the output layer is  $\omega_{jk}$ , Therefore, the input of the output layer is  $net_k = \sum_{j=1}^m \omega_{jk}y_j$ .

After the excitation of Sigmoid function, the output of the node  $k$  in the output layer of the wavelet neural network is expressed as the following formula<sup>[7]</sup>.

$$o_k = f(net_k) = f\left(\sum_{j=1}^m \omega_{jk} \varphi\left(\frac{\sum_{i=1}^n \omega_{ij}x_i - b_j}{a_j}\right)\right) \tag{5}$$

The error function of network output is calculated as follows.

$$E = \frac{1}{l} \sum_{k=1}^l (d_k - o_k)^2 = \frac{1}{l} \sum_{k=1}^l \left( d_k - f\left(\sum_{j=1}^m \omega_{jk} \varphi\left(\frac{\sum_{i=1}^n \omega_{ij}x_i - b_j}{a_j}\right)\right) \right)^2 \tag{6}$$

Although the learning efficiency of wavelet neural network is relatively high, there are still some problems in its application<sup>[8]</sup>, for example, when the training sample is large, the convergence rate of the network is reduced and even the convergence is not present, and the initial value of the network is difficult to be set as the initial value that we expect.

### 3.2. The research on Improved Particle Swarm Optimization

The initial value of the wavelet neural network is related to the training algorithm in the network, so it is necessary to improve the algorithm or re-select the appropriate training algorithm. In this paper, an improved particle swarm optimization algorithm instead of the traditional BP training algorithm is used to optimize the parameters of wavelet neural network.

In the standard particle swarm optimization algorithm, a single particle is called an individual, when each individual optimizes the track in the search space, the optimal trajectory is determined by the speed and position of the individual<sup>[9]</sup>. In the process of repeated optimization, the individual velocity and position are updated based on comparison of the optimal location of the individual and the best position of the swarm search. The velocity and position of the standard particle swarm update are as follows.

$$v_{id}^{k+1} = v_{id}^k + c_1r_1(p_{id} - z_{id}^k) + c_2r_2(p_{gd} - z_{id}^k) \tag{7}$$

$$z_{id}^{k+1} = z_{id}^k + z_{id}^{k+1} \tag{8}$$

Where  $v_{id}^k$  is the  $d$  dimensional velocity component of number  $i$  of individual after the  $k$  times optimization,  $z_{id}^k$  is the  $d$  dimensional location component of number

$i$  of individual after the  $k$  times optimization,  $P_{id}$  is currently the best location for individuals,  $P_{gd}$  represents the optimal location of the current population,  $k$  is the number of individual optimization.  $r_1$  and  $r_2$  take any number between 0 and 1.  $c_1$  and  $c_2$  are particle search factors, it takes any number between 0 and 2.

The standard particle swarm optimization algorithm is easy to fall into local optimization, three improvements are made to the improved particle swarm optimization algorithm, the specific performance is as follows:

- Weight factor  $\omega$

The standard particle swarm with weight factor can improve the global searching ability of the particle and extend the search space of the individual<sup>[10]</sup>, The change range of particle inertia weight can change the search speed. The speed and position of the particle are updated as follows:

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id} - z_{id}^k) + c_2 r_2 (p_{gd} - z_{id}^k) \tag{9}$$

Among them,  $\omega = 0.9 - \frac{0.5k}{k_{max}}$ ,  $k$  said the number of optimization for particles,

$k_{max}$  is the maximum number of optimization for particles.

- Steering operator  $\lambda$

The steering operator  $\lambda$  is added to the standard particle swarm optimization algorithm to adjust the direction and speed of individual optimization. The direction of individual optimization is determined by the positive and negative values of the steering operator, when the operator is positive value, the individual is optimized to the global optimal solution, and vice versa. The numerical value of the steering operator determines the speed of individual optimization, the greater the value, the faster the speed. The general values of the operator are (-1, 1), which is determined by the difference of individual fitness and population fitness, the smaller the difference indicates that the individual is closer to the global optimal solution. The velocity update formula after adding the steering operator is shown in formula (10).

$$v_{id}^{k+1} = \omega v_{id}^k + \lambda_i (c_1 r_1 (p_{id} - z_{id}^k) + c_2 r_2 (p_{gd} - z_{id}^k)) \tag{10}$$

Where  $\lambda_i$  is a guide operator.

- Excitation factor  $\delta$

In the process of optimization, the standard particle swarm is easy to fall into the local optimal<sup>[11]</sup>. When the individual falls into the local optimum, the change of the individual fitness will be very small, then we can add the random disturbance to make the individual jump out of the local optimum range and re-find the global optimal solution. Given a fitness difference  $\mathcal{E}$  ( $\mathcal{E}$  is relatively small constant), and then the difference  $\Delta f_i$  between the fitness values of the particles separated by N generations is calculated. if  $\Delta f_i < \mathcal{E}$ , excitation factor is activated, the particle velocity update formula is

$$v_{id}^{k+1} = \delta v_{max}, \delta \in [0,1] \tag{11}$$

Where  $V_{\max}$  is the maximum speed set by the algorithm.

To sum up, adding direction operator and random disturbance based particle swarm algorithm to solve the easy to fall into local optimum, slow convergence and poor local search ability, it provides support for further optimization of wavelet neural network parameters.

### 3.3. Optimization of wavelet neural network based on Improved Particle Swarm Optimization

The particle swarm algorithm with disturbance to broaden the search area of particle swarm<sup>[12]</sup>, the learning steps of the model of the improved particle swarm wavelet neural network proposed in this paper are as follows.

Step 1: The structure of wavelet neural network is initialized, the input layer is set as  $n$  layer, the hidden layer is set to  $m$  layer, and the output layer is set to  $k$  layer.

Step 2: Initialize the particle swarm. The number of individuals, the initial velocity and initial position of the individual, and the individual oriented operator are set randomly in the population. The initial value of the steering operator is generally 1.

Step 3: Evaluation of individual fitness. The fitness function of PSO algorithm is used to get the error function of the network.

Step 4: Evaluation of particle optimization. According to individual fitness value judgment particle optimization effect, if the fitness value is better than the individual extreme  $p_i$ ,  $p_i$  is updated to this position, and the optimal position of each individual is stored in the optimal set of population.

Step 5: Individual velocity and position update. If  $\Delta f_i < \varepsilon$  ( $\varepsilon$  is set to a very small number), the speed is adjusted according to the formula (11), otherwise according to the formula (10). The position of individual update according to the formula (8).

Step 6: Whether the algorithm meets the end condition was judged, if the conditions are met, the initial parameters of the wavelet neural network are the output of the optimized parameters.

Step 7: Using wavelet neural network for simulation and prediction.

## 4. The simulation of fault diagnosis of ventilator

The fault diagnosis of mine axial-flow fan is analyzed in this paper<sup>[13]</sup>. Firstly, the db10 wavelet function is used to analyze the fault vibration signal by 8 layer wavelet analysis. The decomposition coefficients of each layer are obtained and reconstruct it. Then, the total energy of each frequency band is extracted, and the feature vector is constructed ( $d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8$ ). The training samples selected 15 sets of data, respectively, for the five common faults corresponding to the eigenvector, each fault to take three samples as a training sample.

The extracted feature vectors as wavelet neural network input, so the input node number is 8. The output for the fault type, with 5 nodes, respectively (1, 0, 0, 0, 0), (0, 1, 0, 0, 0), (0, 0, 1, 0, 0), (0, 0, 0, 1, 0), (0, 0, 0, 0, 1) said the unbalance, misalignment

and looseness, rubbing fault, blade fault. The number of nodes in the hidden layer after repeated simulation test is set to 12.

The population size of the improved particle swarm is set to 30. The dimension of particle is the sum of the number of weights and the number of the threshold, the number of weights is 156, the number of the threshold is 17, so the particle dimension is 173. The maximum algebra is 200, the learning factor is set to 2, the initial value of the individual oriented operator is set to 1, the maximum speed of the particle is 1. The fitness function is the error function of the network, and the individual fitness difference is  $\varepsilon$  0.01.

In the experiment, the population size of the improved particle swarm is set to 25, the evolution algebra is about 200, the learning factor is set to be 2, the optimal fitness is set to be 2, the initial value of the individual oriented operator is set to be 1, and the fitness difference of the given individual is 0.01. The input of wavelet neural network is 8 vectors (d1,d2,d3,d4,d5,d6,d7,d8), the output is a fault type with 5 nodes, respectively (1, 0, 0, 0, 0), (0, 1, 0, 0, 0), (0, 0, 1, 0, 0), (0, 0, 0, 1, 0), (0, 0, 0, 0, 1) said the Rotor unbalance, misalignment and pedestal looseness, rubbing fault, blade fault. The number of nodes in the hidden layer after repeated simulation test is set to 12. The initial weights of the wavelet neural network is the output of the improved particle swarm algorithm, the network learning rate is 0.01.

The simulation results show that using the improved particle swarm optimization algorithm to optimize the neural network parameters of the network fitness as shown in Figure 2.

Diagnose the following 4 test samples,  $T1=\{0.043, 0.412, 0.039, 0.848, 0.218, 0.037, 0.065, 0.043\}$ ,  $T2=\{0.034, 0.960, 0.128, 0.142, 0.156, 0.061,0.045, 0.057\}$ ,  $T3=\{0.028, 0.361, 0.055, 0.029, 0.017, 0.081, 0.926, 0.049\}$ ,  $T4=\{0.468, 0.502, 0.063, 0.390, 0.616, 0.071, 0.046, 0.032\}$ . Respectively using wavelet neural network (WNN) and the improved particle swarm wavelet neural network (Improved PSO-WNN) and the results of the two algorithms are compared.

You can see from Figure 2, the wavelet neural network training speed started quickly, but the convergence speed is slow in the latter part. About 33 times after iteration, the error is about 0.03, the error is larger, and easy to fall into the local advantages. Compared with the wavelet neural network, the convergence rate of the improved PSO-WNN is fast. After 18 iterations, the error reaches 0.008, and the error precision is relatively small, and it has been greatly improved.

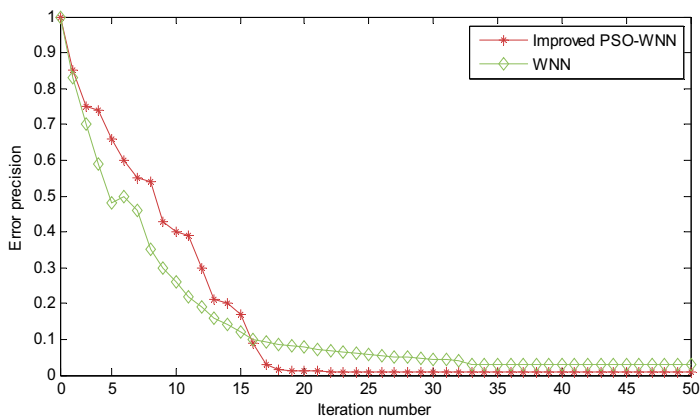


Figure 2. Error precision of two algorithms

Table 2 shows two different algorithms, simulation results and error precision analysis results. The results of wavelet neural network and improved particle swarm wavelet network are correct, but the error of PSO-WNN is lower than that of wavelet neural network, the accuracy of the system is increased by 2%. The simulation results show that the improved PSO algorithm optimize the parameters of wavelet neural network, which can improve the efficiency of the system.

**Table 2.** Comparison of the results of the two algorithms

Test sample	Fault type	Algorithm	Diagnosis result					Diagnostic error
T <sub>1</sub>	0 1 0 0 0	WNN	0.134	0.897	0.301	0.118	0.157	0.0315
		Improved PSO-WNN	0.035	1.078	0.125	0.108	0.085	0.0084
T <sub>2</sub>	1 0 0 0 0	WNN	0.875	0.171	0.145	0.100	0.202	0.0233
		Improved PSO-WNN	0.932	0.092	0.069	0.109	0.096	0.0072
T <sub>3</sub>	0 0 0 0 1	WNN	0.129	0.221	0.123	0.174	0.802	0.0300
		Improved PSO-WNN	0.112	0.064	0.082	0.113	1.092	0.0089
T <sub>4</sub>	0 0 0 1 0	WNN	0.218	0.069	0.206	0.899	0.098	0.0229
		Improved PSO-WNN	0.084	0.032	0.057	1.099	0.131	0.0077

## 5. Summary

According to the fault diagnosis of mine main fan of real-time and accuracy requirements are relatively high, and proposed an improved particle swarm optimization algorithm of wavelet neural network model to solve the traditional wavelet neural network learning efficiency is low, easy to fall into local optima<sup>[14]</sup>. Fault diagnosis, first decomposition and reconstruction of vibration signal of the fault, get a different energy band, and form a feature vector for the wavelet neural network input, optimal network parameters of the solutions obtained by adding a weighting factor, guide operator and improved particle swarm algorithm with random perturbation idea, the wavelet neural network has the characteristics of high efficiency training, high accuracy and fast convergence of learning, can quickly diagnose the faults. The simulation results show that compared with the traditional wavelet neural network, the training accuracy of the improved particle swarm wavelet neural network is improved by about 2%.

## References

- [1] Wang Haisheng, Statistical analysis of coal mine production safety accidents in China in 2013[J], *Chinese Zhongzhou coal*, 2014 (9): 77-79.
- [2] Wang Yibing, Fault diagnosis of mine fan based on neural network[D], Shanghai: *Shanghai Jiao Tong University*, 2012.
- [3] Qu Zhongyang, Fan performance evaluation and testing based on vibration signal analysis [D], Shanghai: *Shanghai Jiao Tong University*, 2014.
- [4] Hou Yuanbin, Du Jingyi, Wang Mei, *Neural network* [M], Xi'an: Xi'an Electronic and Science University press.2007.
- [5] Jing Shuangxi, Huawei, coal mine ventilator fault diagnosis of wavelet based on support vector machine[J], *Journal of coal industry*, 2007 (1): 98-102.
- [6] Song Haiqiang. Development of mine ventilator monitoring and diagnosis system based on wireless sensor network[D], Beijing: *Beijing University of Technology*.2012.
- [7] Li Dongchen, The Application of wavelet neural network based on LabVIEW in fault diagnosis of coal mine ventilator[D], Anhui: *Anhui University Of Science and Technology*, 2012.
- [8] MA Wen-jing, Power transformer fault diagnosis of genetic algorithm based on wavelet neural

- network[J], *Techniques of Automation and Applications*, 2008,27(12): 17-30.
- [9] Cheng Shengfeng, cheng Xiaohua, Yang Lu. Application of wavelet neural network with improved particle swarm optimization in transformer fault diagnosis [J], *Power System Protection and Control*. 2014 (10): 37-42.
- [10] Jun Shi, Xiaoping Liu. Multiresolution analysis and orthogonal wavelets associated with fractional wavelet transform[J]. *Image and Video Processing*, 2015, 19(9): 228-230.
- [11] Chen Caixia, Based on modified particle swarm optimization and wavelet neural network prediction model and its application [D], Wuhan: *Huazhong Normal University*, 2011.
- [12] LI Xin-bin, ZHU Qing-jun. An improved particle swarm optimization algorithm for multi-objective optimization of reactive power[J]. *Transactions of China Electrotechnical Society*, 2010, 25(7): 137-143.
- [13] Jing Shuangxi, Leng Junfa. The fault diagnosis of mine ventilator based on wavelet neural network[J]. *Coal Journal*, 2004 (6): 736-739.
- [14] Ji Yanjie, Chen Xiaoshi, Hu Bo, A prediction of effective parking lots based on wavelet transform and particle swarm wavelet neural network[J], *Journal of Jilin University*.2016 (3): 399-405.



# Statistical Image Analysis Based Automated Leaves Classification

Manar Bati AL-OTAIBI<sup>a</sup>, Amira S. ASHOUR<sup>a,b,1</sup>, Nilanjan DEY<sup>c</sup>, Rahaf Abdullah AL QUTHAMI<sup>a</sup>, Asrar Abdullah AL-NUFAEI<sup>a</sup>, and Fuqian SHI<sup>d</sup>

<sup>a</sup>Computer Science Department, CIT College, Taif University, KSA

<sup>b</sup>Department of Electronics and Electrical Communications Engineering, Faculty of Engineering, Tanta University, Egypt

<sup>c</sup>Department of Information Technology, Techno India College of Technology, Kolkata, India

<sup>d</sup>College of Information and Engineering, Wenzhou Medical University, Wenzhou, PR China

**Abstract.** Plants recognition and classification is a challenging process due to the high variability in the plants' features and shapes. Numerous methodologies incorporating image processing were improved to tackle this process for early stage recognition of diseases. Leaf recognition has popular/wide range of agriculture practical applications. Consequently, the current work is interested in the recognition and classification of parsley and basil leaves along with the recognition of their infected parts. An image analysis is used to extract different statistical features from the leaves' dataset. From such statistical features a recognition/classification processes are performed to classify the fresh and infected leaves in each leaf type as well as to classify the two-leave species. The classification process was performed using neural network. The experimental results depicted that the classification accuracies for the three tested cases, namely fresh/infected basil, fresh/infected parsley, and fresh basil/parsley were 80%, 80.0%, and 100.0%; respectively.

**Keywords.** Image processing, image analysis, de-noising, leaf recognition, classification, neural network

## 1. Introduction

Plants have a vital role in the environment for nature balance and conservation of the earth's biosphere. In addition, several countries mainly depend on the agriculture for economic issues. Moreover, some plants are entering in many pharmaceutical agents' due to their medical properties, while others can be used as sources of alternative energy (bio-fuel). However, various types of diseases and pests can affect the harvests; which lead to remarkable decrease in the harvest quality as well as affect the production. Typically, the diseases/pests' effects appear on the stems/leaves of the plant. The infection can appear in the form of incomplete (damaged) parts or spots on the stems/leaves [1]. Plant infection and diseases influence the optical characteristics of

the plants that lead to signatures of the infected plants. In order to increase the harvest (plants) productivity, cultivators have to contact experts for their advising concerning the treatment and control suggestions based on the infected plant shape and other characteristics. Furthermore, some situations require distinguishing between the different plants' species. Consequently, leaves' recognition/classification is obliged to mark the diseased/ infected parts or the percentage of the infected regions to the whole leaf for successful fostering of crops.

Protection, recognition and classification of diverse plant types require a plant database to extract information from their images. Thus, efficient and quick automated methods motivate the active research domains. Recognition/classification of plants has several approaches based on the plant part under concern (root, leaf, flower, fruit) [2-4]. Leaf recognition and classification are complex processes due to the irregular leaves' shapes. Recent decades have witnessed a revolution in digital image processing, analysis and machine vision in several industrial, medical, and agriculture applications. Generally, digital images have high impact in the modern society. They considered to be a significant component in several application domains, such as computer vision, pattern recognition, agriculture and industrial automation. Computer vision and pattern recognition systems have been developed toward automated procedures for plant recognition and/or classification. Image processing includes the use of several algorithms for enhancement, compression, features extraction, segmentation, and classification. Such algorithms are deployed in various application' domains. Thus, digital image processing and analysis are required to extract features from the plant's images for plants' characteristics/quality measurements with less time consuming and accuracy compared to the traditional methods.

The main contribution of the current work is to recognize and classify the infected plants. An automated technique to analyze the leave's images using ImageJ as well as to recognize the different leaves and the fresh/infected ones is proposed. The proposed system consists of three modules, namely i) image analysis module for the leaves using ImageJ software, ii) leaf recognition/classification module based on the extracted feature vectors using Matlab software, and iii) neural network based leaves classification. The first module is deployed to perform leaves' image analysis to extract the statistical features such as the area, perimeter, and the standard deviation that characterize each leaf type as well as to determine the feature vector that used in the next module. The second module is concerned with leaves recognition and classification using the Euclidean distance (ED) to recognize the leaf type and its status (fresh or infected). The third module classifies the leaves into fresh/ infected for the species under concern as well as classifies the fresh leaves' species using neural network, i.e. classify the fresh/infected basil, fresh/infected parsley, and fresh basil/parsley cases.

## **2. Related Work**

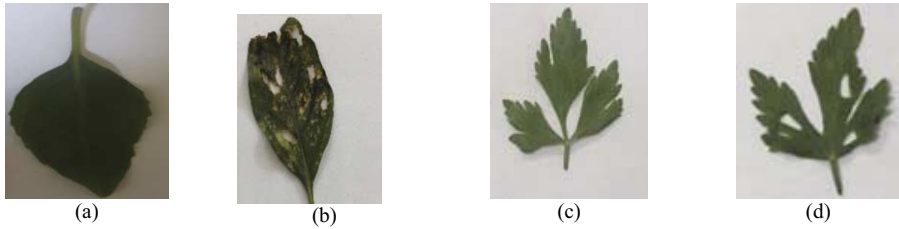
The impact of insects and diseases on the plants has about 40% potential annual loss in the world food production [1]. Detecting the plant's disease is supposed to be measurable and can be distinguished from healthy (fresh) tissues. The traditional agricultural managing systems are inaccurate and time consuming. Thus, the innovative technologies become a promising scheme to optimize the agricultural production based on image processing to automate such applications including images analysis and processing phases. Several researchers were interested with leaf recognition as a mark

for diseased plants in their first stage. Gwo and Wei [2] suggested a feature extraction approach for leaf contours identification. The extracted features were the invariant to scale and rotation. Kulkarni *et al.* [3] applied a framework for recognizing plants using vein, shape, color, and texture features. The experimental results have been carried out on the Flavia leaf dataset. The results established 93.82% accuracy rate. Using the leaf vein and shape, Lee and Hong [4] implemented a leaf recognition system using Fast Fourier Transform (FFT) with the distance between centroid and contours in the detected leaf image. For the leaf recognition, twenty-one leaf features were extracted. Priya *et al.* [5] proposed a leaf recognition method for plant classification using support vector machine (SVM). A dataset for tea leaf images were captured using digital cameras or scanners in the system. The tea leaf images can be identified accurately in the preprocessing phase by fuzzy denoising using Dual Tree Discrete Wavelet Transform (DT-DWT) in order to remove the noisy features and boundary enhancement to obtain the shape of leaf accurately. In [6], an automated plant recognition system based on leaf images was proposed. A hybrid features set involving these models has been generated to determine the feature vector that lead to better analysis. Afterward, the neural networks were employed for different species the classification process. Wijesingha and Marikar [7] designed an automatic endemic Herbarium leaves identification system. The extracted features have been then used as input to a Probabilistic neural network (PNN). Sulc and Matas [8] introduced a novel scheme for visual leaf identification using a pair of local feature histograms. The local features were extracted from the leaf interior and the border. In order to attain an invariance desirable level, the leaves were described with multi-scale histograms of rotationally invariant features, where local binary patterns (LBP) was used. In [9], four modules have been presented, namely image acquisition, preprocessing, recognition and results display for automated recognition and classification of plants based on their leaves. The extracted features were deployed for the recognition module, which were finally displayed. The experiments were executed on 12 types of leaves that achieved 97.9% classification accuracy. Du *et al.* [10] used a shape matching technique for plant species identification system based on the leaf images using the Douglas-Peucker approximation algorithm. The leaves' shape has been implemented to outline the invariant attributes sequence. A recognition accuracy of 92% was achieved for recognizing the intact leaves. Based on the plant leaf's shape, Sosa *et al.* [11] proposed a plant leaf recognition system using Neural Networks (LeafRApp). A hybrid features extraction technique using the Moment-Invariant method which extracted the first four moments of the image has been involved.

From the preceding literatures, it is established that leaf recognition is a wide domain that has several approaches. It is considered the milestone for further image processing phases such as the classification process. Consequently, the proposed work is concerned with leaves recognition/classification.

### 3. Methodology

The proposed system was implemented using ImageJ as an image processing tool for leaves' images analysis. In order to test the proposed technique, basil and parsley samples are obtained by capturing images for fresh and infected leaves as illustrated in Figure 1.



**Figure 1.** Dataset captured image samples: (a) fresh basil, (b) infected basil, (c) fresh parsley, and (d) infected parsley.

Afterward, Matlab software (R2011a) is employed for recognition/classification using the Euclidean distance approach and neural network. The dataset is obtained by capturing images using digital camera for two types of leaves, namely Basil and Parsley. It contains both fresh and infected samples to be tested. Each category contains 15 sample images. The image analysis step is initialized by pre-processing for de-noising using the median filter.

### 3.1 Median filter

Typically, the captured images are often corrupted by random variations in illumination, intensity, have poor contrast or noise due to the use of the digital camera. Therefore, it is necessary to remove such noise before any further image analysis/processing [12]. There are different types of filters such as mean filter, median filter, and Gaussian filter. Since, the proposed system is concerned with the recognition of the leaves boundaries as well as the fresh/infected regions. Thus, in the proposed work median filter is used to enhance the dataset images as it preserves the image edges. The median filter is mainly based on the idea of replacing the intensity values of each pixel with another new value over a neighborhood of fixed size. It is a nonlinear digital filter approach used for noise reduction and edges' preservation while noise removal. Moreover, it has the following advantages: no contrast reduction during its steps, does not shift boundaries, less sensitive to the outliers compared to the mean filter [13].

### 3.2 Histogram based Threshold

For image analysis, the threshold method is required. The thresholding process is based on the replacement of each pixel in the image by another black or white pixel if the image intensity is less than, or greater than certain constant (threshold); respectively. Thus, there are several techniques that used to determine the threshold value, such as i) histogram shape –based approach, ii) clustering-approach, iii) entropy-based approach v) spatial approach [14]. In the proposed system ImageJ is supported by the histogram approach for threshold selection. Whereas, thresholding creates binary images from gray-level images through turning all pixels based on the specified threshold. The histogram represents the numerical data distribution using graphical representation, as it provides a continuous variable estimation of the probability distribution [15]. The main idea of histogram is to divide the entire values' range into a series of non-overlapping intervals. Afterward, count the values within each interval [16].

### 3.3 Image analysis and Feature extraction

Image processing, machine learning and pattern recognition are based on feature extraction which represents a set of informative and non-redundant measured data that used further in learning/classification phases. Feature extraction reduces the required resources amount which describes a large set of data/image [17]. Since, image analysis provides a meaningful information from images by using automatic/semiautomatic techniques including image description, pattern recognition and computer vision. The main difference between image analysis and image processing is that the final result of image analysis technique has the form of numerical output rather than an image. Therefore, prior to the feature extraction step, image analysis is required. Generally, feature extraction can be defined as the transformation of the input data into a set of features. In proposed system, statistical feature are extracted using image analysis for the dataset under concern. The extracted statistical features using ImageJ as an output from the image analysis are such as area, Perimeter, Mean gray value, Standard deviation (StdDev), Centroid, Circularity (Circ.), Feret's diameter (Feret), integrated density (RawIntDen), and Median.

### 3.4 Recognition and Classification using the ED

Plants species recognition can be simply identified by the investigation of their leaves, flowers, and fruits. However, once the plant is damaged or infected, the recognition becomes tremendously complicated. Typically, recognition in image processing is a branch of machine learning that interested with the recognition of regularities in data or patterns [19]. It can be definite as a labeling problem based on models of the object under concern. All recognition systems include models for the object either implicitly or explicitly and utilize feature detector according to this model. Recognition systems can use supervised learning, where a training "labeled" data exist, or unsupervised learning, when no labeled data are available [20]. In the present work, the ED is used to identify the groups under concern and thus classify the different groups. The ED is defined as the distance between two points in the Euclidean space using the following formula [21]:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \quad (1)$$

Where,  $p = (p_1, p_2)$  and  $q = (q_1, q_2)$  are the points which required to find the distance between them. Thus, in order to identify the fresh plant from the infected one for any species. The feature vectors that identify a significant difference between the two categories are used for each category. Then, the based on the  $d(p, q)$  between the input feature vector and the different categories' feature vectors, the category of this input can be identified based on the least distance. Therefore, in the proposed work recognition was performed on different cases: fresh/infected basil, fresh/infected parsley, and basil/parsley; respectively.

### 3.5 Classification Using Neural Network

Basically, the NN system consists of three layers, namely the input layer, hidden layer and output layer as shown in Figure 2.

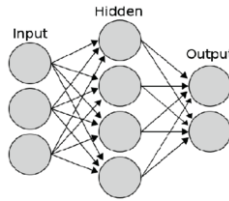


Figure 2. Architecture of the NN

Figure 2 depicted that the input neurons of the input layer transfer the data to the hidden layer. Similarly, the hidden layer transfers this data to the output layer. The weights help the manipulation between the input and output to various layers [22-24]. In the present proposed system, the NN approach was used for the classification process. It consists of three main phases, namely: i) training phase, where the training dataset (part of the dataset that consists of attribute values) is involved to learn its entity into the corresponding class label to construct the classification model. During the training step, the NN model tries to obtain adequate knowledge to realize how the entities are classified into the giving classes. During this phase, the dataset is used to adjust the neural network's weights. ii) The validation phase set is used to establish the neural network performance, where the dataset is used to minimize over-fitting. iii) The Evaluation (testing) phase, where the trained model's accuracy is tested with another portion of the dataset (test data) [25].

#### 4. Proposed System

The preceding methodologies are used in the proposed system for basil/parsley as well as their fresh/infected images recognition and classification. Figure 3 illustrated the block diagram of the proposed system.

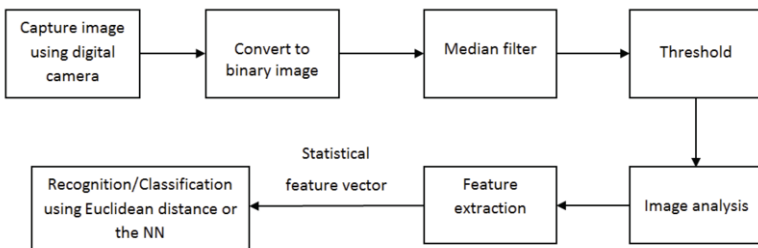
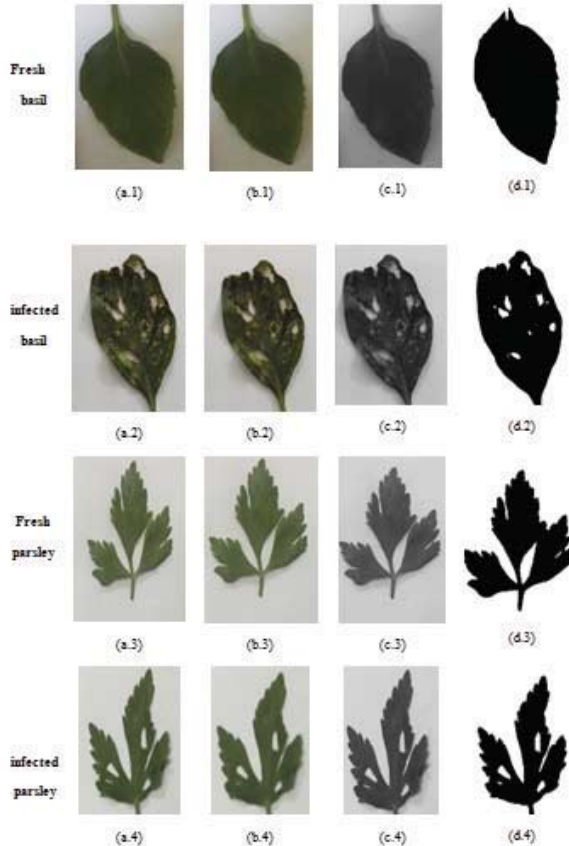


Figure 3. Block diagram of the proposed system

Figure 3 demonstrated the steps of the proposed system as follows: i) capture the leaves' images from the two species for both fresh and infected cases using digital camera. These captured images forms the dataset that used as an input to the proposed system, ii) convert the image to binary as it simple in process, iii) apply median filter to remove noise from the image, iv) apply the threshold method to the binary image, v) analysis the image using ImageJ tools, vi) extract the statistical features from the analysis result, vii) employ the Euclidean distance method for the recognition step, and finally viii) the classification step is performed for three cases. The classification cases are a) classify fresh basil leaves and the infected ones, b) classify fresh and infected parsley leaves, and c) classify basil and parsley leaves.

## 5. Results and Discussion

The experimental results of the proposed system are applied on the previously mentioned dataset as follows. The image analysis steps is used for the preceding describe datasets to extract statistical feature for further recognition and classification steps. Figure 4 demonstrates the applied procedure for samples from the datasets.

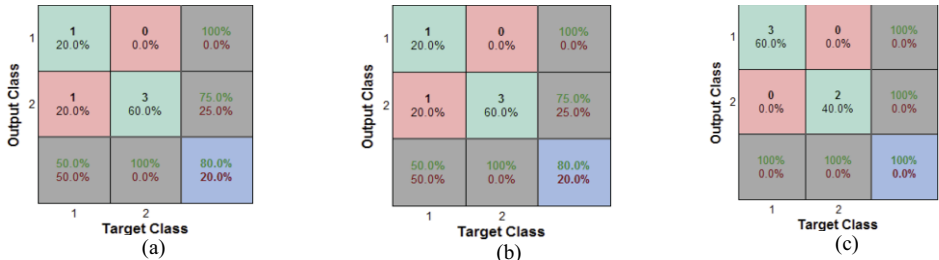


**Figure 4.** leaves' samples for basil and Parsley:(a.1, a.2, a.3, and a.4) original images of leaves, (b.1, b.2, b.3, and b.4) filtered images, (c.1, c.2, c.3, and c.4) images in 8-bit format, and(d.1, d.2, d.3, and d.4) threshold images.

In figure 4 (a), the original captured images using digital camera are displayed. The images become smoother with clear edges by using median filter in figure 4(b).8-bit images to convert into grayscale is shown in figure 4(c). The binary threshold image showing black leaves on white background threshold in is used to create binary images as in figure 4(d). In order to attain the features vector, the previous image analysis statistical results are used where various measured features are extracted. In order to perform recognition/classification, certain significant feature vector is used based on each case according to ten samples from each group (fresh basil, infected basil, fresh parsley, and infected parsley). The tested cases are recognized/classified into: i) fresh and infected basil, ii) fresh and infected parsley, and iii) basil and parsley. The selected feature vector for each case was based on the average value of the features that provides a significant difference between the classes which we want to classify and

distinguish from each other. Afterward, in order to perform the recognition phase, the Euclidean distance is employed to recognize the previously mentioned cases where the selected feature vectors of ten images are used as a training set and the remaining five images from each group are used as testing set. The results of tested sets for the recognition of the different cases: fresh/infected basil, fresh/infected parsley, and basil/parsley; respectively, established 100% recognition in all cases.

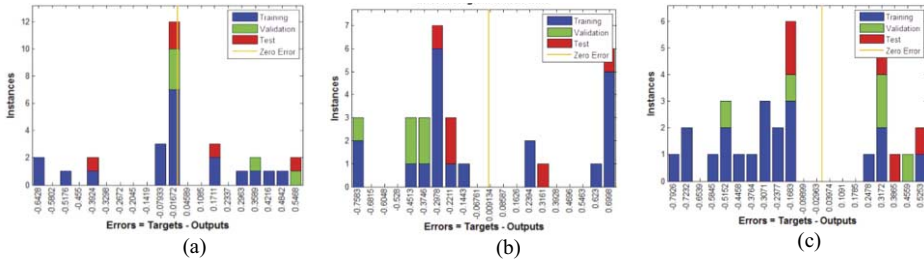
Furthermore, the neural networks in the proposed work for classification consists of input layer, hidden layer (ten neurons), and output layer. For the classification purpose, each image set is divided into 70% of the images are used for feature extraction and training (20 images), 30% for validation and test (10 images) in each image set. Since, the performance evaluation is essential in all automated classification systems to evaluate the ability to classify the dataset into proper categories [26]. The accuracy and the confusion matrix are calculated. The confusion matrix contains information about actual and predicted classifications. Generally, the confusion matrix entries can be defined as follows: i) True positive (Tp) is the number of positive instances categorized as positive (Correctly identified ‘positive’), ii) False positive (Fp) is the number of negative instances categorized as positive (Incorrectly identified ‘positive’), iii) False negative (Fn) is the number of positive instances categorized as negative (Incorrectly rejected ‘negative’), and iv) True negative (Tn) is the number of negative instances categorized as negative (Correctly rejected ‘negative’). Figure 5 illustrates the results of confusion matrix of the different cases: fresh/infected basil, fresh/infected parsley, and basil/parsley; respectively for the test phase.



**Figure 5** The test confusion matrix for the three cases: (a) fresh/infected basil, (b)fresh/infected parsley, and (c)basil/parsley.

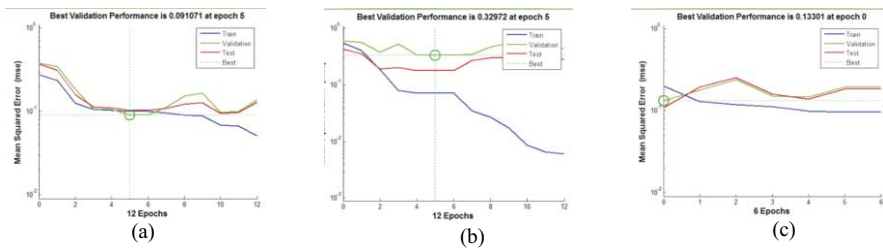
Figure 5 depicts the correct classification percentage for the three cases are 80.0%, 80.0%, and 100.0% respectively. Whereas the incorrect classification percentage for the three cases are 20.0%, 20.0%, and 0.0% respectively. The error histogram plot for the three cases is illustrated in Figure 6 to obtain supplementary verification of the network performance.





**Figure 6** Error histogram for the three cases: (a) fresh/infected basil, (b)fresh/infected parsley, and (c) basil/parsley.

In Figure 6, the blue bars stand for the training data, the green bars stands for the validation data, and the red bars stands for the testing data. The error histogram plot provides an indication of the outliers. The outliers are the data points where the fit is extensively poorer than the majority of data. In addition, the classification performance is measured in terms of the mean squared error which shown in log scales in Figure 7.



**Figure 7** cross-entropy versus epochs plot depicting the performance of the model for the three cases: (a) fresh/infected classification, (b)fresh/infected parsley, and (c)basil/parsley.

Figure 7 illustrates that in all cases the MSE versus the number of iterations is decreased as the network was trained. The performance is shown for each of the training, validation and test sets for each classification case. From the preceding results, it is established that the statistical features can be used efficiently for plants recognition and classification using Matlab software. Selected codes are given in the appendix. Consequently, the proposed system can be used to recognize and classify more different species as well as infected ones. As a future work, it is recommended to use largest dataset to obtain improved classification results. Furthermore, other classifiers can be used and compared with the NN for classifying different types of the plants as well as the fresh and infected ones. Other feature vector can extract from different types of leaves and the process of choose feature vectors make it more automated. Also, features selection algorithms can be applied to select the proper features form the extracted feature vectors. A comparative study with previously published works [27-31] can be performed in the future.

**6. Conclusion**

The main aim of the present work is to recognize/classify leaves, where plants have innumerable advantages including medicinal properties. They are for therapy due to their low cost. Plants recognition is hence a very essential task for plant enthusiasts,

herbal medicine researchers and other several industries and applications. This work employed the NN approach for the classification process and performance evaluation to evaluate the ability to classify the dataset into proper categories. As a conclusion the automated proposed system using image processing achieved correct classification percentage for the three cases fresh/infected basil, fresh/infected parsley, and basil/parsley of 80.0%, 80.0%, and 100.0% values; respectively. Additionally, it achieved incorrect classification percentage for the three cases of 20.0%, 20.0%, and 0.0% values; respectively. The obtained recognition rate with all the cases was 100%.

## References

- [1] E. C. Oerke, H. W. Dehne, Safeguarding production—losses in major crops and the role of crop protection, *Crop protection*, 23(4) (2004), 275-285.
- [2] C. Y. Gwo, C. H. Wei, Plant identification through images: Using feature extraction of key points on leaf contours, *Applications in plant sciences*, 1(11) (2013).
- [3] A. H. Kulkarni, H. M. Rai, K. A. Jahagirdar, P. S. Upparamani, A leaf recognition technique for plant classification using RBPNN and Zernike moments. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(1) (2013), 984-988.
- [4] K. B. Lee, K. S. Hong, An implementation of leaf recognition system using leaf vein and shape. *International Journal of Bio-Science and Bio-Technology*, 5(2) (2013), 57-66.
- [5] C. ArunPriya, T. Balasaravanan, A.S.Thanamani, An efficient leaf recognition algorithm for plant classification using support vector machine, *IEEE 2012 International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME)*, (2012, March), 428-432.
- [6] J. Chaki, R. Parekh, Plant leaf recognition using shape based features and neural network classifiers. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2(10) (2011).
- [7] D.Wijesingha, F. M. M. T. Marikar, Automatic detection system for the identification of plants using herbarium specimen images, *Tropical Agricultural Research*, 23(1) (2012).
- [8] M. Sulc, J.Matas, Texture-Based Leaf Identification. *In Computer Vision-ECCV 2014 Workshops* (2014, September),185-200.
- [9] G.Kaur, G. Kaur, Classification of biological species based on leaf architecture, *International Journal of Engineering Research and Development*, 1(6) (2012), 35-42.
- [10] J. X. Du, D. S. Huang, X. F. Wang, X. Gu, Computer-aided plant species identification (CAPSI) based on leaf shape matching technique. *Transactions of the Institute of Measurement and Control*, 28(3) (2006), 275-285.
- [11] J. P. M. Sosa, G.A.Solano, J.A.Obico, An application using canny edge detection and multilayer perceptron for recognizing leaves of tropical plants, *Mathematics and Computers in Biology and Biomedical Informatics*, (2013), 53-61.
- [12] A. Fabijańska, Two-pass median filter for impulse noise removal. *Automatyka/AkademiaGórnictwo-Hutniczaim. StanisławaStaszica w Krakowie*, 13(2009), 807-820.
- [13] E. Arias-Castro, D.L. Donoho, Does median filtering truly preserve edges better than linear filtering?, *Annals of Statistics*, 37(3) (2009).
- [14] Y.Zhang, Optimal multi-level Thresholding based on Maximum Tsallis Entropy via an Artificial Bee Colony Approach, *Entropy*, 13 (4) (2011) 841–859.
- [15] K. Pearson, Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 186 (1895), 343–414.
- [16] D. Howitt, D. Cramer, *Statistics in Psychology*, Prentice Hall (2008).
- [17] V. P. Rathi, S.Palani, Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis, (1895).
- [18] T. Ferreira, W. Rasb, *ImageJ user guide*, (2012).
- [19] C. M. Bishop, Pattern recognition. *Machine Learning*, 128, (2006), 1-58.
- [20] L. G. Lim, Automatic Classification Of Colonic Histopathological Images Using Adaptive Neuro-Fuzzy Networks And Genetic Algorithms. PhD Thesis. Coventry: Coventry University (2011).
- [21]E. Deza, M.M. Deza, *Encyclopedia of Distances*, Springer, (2009), 94.
- [22] S.Hore, S. Chatterjee, S.Sarkar, N. Dey, A. S. Ashour, D.Bālas-Timar, V. E. Balas, Neural-based prediction of structural failure of multistoried RC buildings, *Structural Engineering and Mechanics*, 58(3) (2016), 459-473.

- [23] A.M. Elazouni, I.A. Nosair, Y.A. Mohieldin, A.G. Mohamed, Estimating resource requirements at conceptual stage using neural networks, *J. Comput. Civil Eng.*, 11(4) (1997), 217-223.
- [24] H. Erdem, Prediction of moment capacity of reinforced concrete slabs in fire using artificial neural networks, *Adv. Eng. Softw.*, 41(2) (2010), 270-276.
- [25] S.J.S. Hakim, H. Abdul Razak, Modal parameters based structural damage detection using artificial neural networks - a review, *Smart Struct. Syst.*, 14(2) (2014), 159-189.
- [26] S. Beagum, A.S. Ashour, N. Dey, Bag-of-Features in Microscopic Images Classification, *Classification and Clustering in Biomedical Signal Processing*, 1(2016).
- [27] A. S. Ashour, S. Beagum, N. Dey, A.S. Ashour, D.S. Pistolla, G. Nguyen, , ..., F. Shi, Light microscopy image de-noising using optimized LPA-ICI filter, *Neural Computing and Applications*, (2016)1-17.
- [28] D. Wang, Z. Li, L. Cao, V. Balas, N. Dey, A.S. Ashour, ..., F. Shi, Image fusion incorporating parameter estimation optimized Gaussian mixture model and fuzzy weighted evaluation system: a case study in time-series plantar pressure dataset, *IEEE Sensors Journal*, (2016).
- [29] D. Wang, Z. Li, N. Dey, A.S. Ashour, R.S. Sherratt, F. Shi, Case-Based Reasoning for Product Style Construction and Fuzzy Analytic Hierarchy Process Evaluation Modeling Using Consumers Linguistic Variables, *IEEE Access*, 5 (2017), 4900-4912.
- [30] Z. Tian, N. Dey, A.S. Ashour, P. McCauley, F. Shi, Morphological segmenting and neighborhood pixel-based locality preserving projection on brain fMRI dataset for semantic feature extraction: an affective computing study, *Neural Computing and Applications*, (2017), 1-16.
- [31] S. Kamal, N. Dey, S.F. Nimmy, S.H. Ripon, N.Y. Ali, A.S. Ashour, ..., F. Shi, Evolutionary framework for coding area selection from cancer data, *Neural Computing and Applications*, (2017)1-23.

# Fractional Dimension Parameters Modelling and Optimization Based on Pulverized Coal Image

Zheng WANG <sup>a,1</sup> and Mei WANG <sup>a</sup>

<sup>a</sup>*School of Electric and Control Engineering, Xi'an University of Science and  
Technology, China*

**Abstract.** Because of the complex environment in coal mines, it is extremely hard to process pulverized coal images with large sum of noise. Fractional dimension characterization and model in bounded variation space are presented to optimize pulverized coal image. This paper puts on fractional calculus theory in bounded variation space and discusses the characterization model to enhance image optimization stability. Applying the regional characteristics of the pulverized coal image, the fractional dimension characteristic parameter  $u$  and regularized parameter  $\lambda$  of each image pixels are modified by adaptive algorithm. To evaluate image optimization effect, Numeric experimentations show that the PSNR (peak signal-to-noise ratio) and ERI (edge-preserving index). The two quantitative indicators have better performance considerably with the improved algorithm than the traditional algorithm. Because of the sound smoothing effect in the "non-texture region" and the excellent texture retention capacity in the "texture region", fractional dimension characterization in bounded variation space optimization algorithm is effective and expeditious image filtering approach. The advanced algorithm is employed to observe pulverized coal in underground and succeeds in achieving satisfactory outcomes.

**Keywords.** Fractional dimension; Characteristic Parameter; Bounded Variation Space; Pulverized Coal Optimization; PSNR; EPI

## 1. Introduction

Pulverized coal will explode in the following three conditions: the pulverized coal has explosive property; there exist certain concentration of coal particles in the air and there is hot reservoir around the pulverized coal [1]. The first condition is inherent property and the third one is inevitable in mechanized coal mining operation, so we must restrict pulverized coal concentration to prevent safety accidents. Pulverized coal detection is very important and it has great significance on classification measurement and exact detection of particle concentration on line. The physical structure of pulverized coal is one of the factors that affect the safety performance. The image features of single particle can be obtained by pulverized coal image processing, which can not be obtained by other methods [2].

The algorithms research for particular environment such as coal mine are rare, nevertheless this paper analyses the detection and identification of pulverized coal in

---

<sup>1</sup> Corresponding Author.

depth. Image filtering can get rid of the image background disturbance which would affect pulverized dust detection while it can preserve the effective image information.

Analyzing with the mathematic perspective, the texture detail is weak derivative information that is fractional dimension derivative characteristics [3]. So fraction characterization optimal algorithm is fit for dealing with these problems.

## 2. Fractional dimension characteristic parameter evaluation

We can differentiate the "texture area" from the "non-texture area" based on local variance of image and propose selection methods of regularization parameters  $\lambda$  and fractional order  $u$ , such based on these parameter adaptive algorithms are given.

### 2.1. regularization parameter modeling

In order to study how to choose adaptively regularization parameter  $\lambda$ , we can give the parametric adaptive fractional variational PDE model:

$$\partial_{u(x,y)} J(s) - \lambda(x,y) \sum_{j \geq 0} 2^{-2j\beta} [\lambda(x,y)(f-s)]_j = 0 \tag{1}$$

Where  $\beta$  is the scale space factor.

To the constant  $s(x,y)$ , The above formula can be converted into

$$\sum_{j \geq 0} 2^{2j\beta} [\partial_{u(x,y)} J(u)]_j = Q(x,y)(f-s) \tag{2}$$

Where  $Q(x,y) = \lambda^2(x,y)$ .

Such the estimation of  $\lambda$  is transformed into the estimation of  $Q(x,y)$ .

In order to relate  $Q(x,y)$  to the regional structural characteristics in the image, the neighborhood information in the image each point determines the value of  $Q(x,y)$ .

Assuming

$$Q(x,y) = \int_{\Omega} \tilde{Q}(\tilde{x}, \tilde{y}) w_{x,y}(\tilde{x}, \tilde{y}) d\tilde{x}d\tilde{y} \tag{3}$$

Where  $\tilde{Q}$  is undetermined,  $w_{x,y}(\tilde{x}, \tilde{y}) = w(|x-\tilde{x}|, |y-\tilde{y}|)$  is a standardized radially symmetric smooth window function[4], satisfying

$$\int_{\Omega} w_{x,y}(\tilde{x}, \tilde{y}) d\tilde{x}d\tilde{y} = 1. \tag{4}$$

Both sides of formula (2) are multiplied by  $(f-s)$  at the same time, and then are integrated of  $x, y$  at the same time.

$$\int_{\Omega} (f-s) \sum_{j \geq 0} 2^{2j\beta} [\partial_{u(x,y)} J(s)]_j dx dy = \int_{\Omega} Q(x,y)(f-s)^2 dx dy \tag{5}$$

Formula(2) is substituted into (5) :

$$\int_{\Omega} (f-s) \sum_{j \geq 0} 2^{2j\beta} [\partial_{u(x,y)} J(s)]_j dx dy = \int_{\Omega} \left[ \int_{\Omega} w_{x,y}(\tilde{x}, \tilde{y})(f-s)^2 dx dy \right] \tilde{Q}(x,y) dx dy \tag{6}$$

Supposing

$$P_S(x, y) = \int_{\Omega} w_{x,y}(\tilde{x}, \tilde{y}) \cdot [f(\tilde{x}, \tilde{y}) - s(\tilde{x}, \tilde{y})]^2 d\tilde{x}d\tilde{y} \tag{7}$$

Then the sufficient condition of formula (6) is:

$$\tilde{Q}(x, y) = \frac{(f - s) \sum_{j \geq 0} 2^{2j\beta} [\partial J_u(s)]_j}{P_S(x, y)} \tag{8}$$

To estimate  $\lambda(x, y)$  is equal to estimate  $P_S(x, y)$ . In fact, if the average of  $(f - s)$  in  $P_S(x, y)$  is equal to zero, then  $P_S(x, y)$  is the local variance of the residual error[5]  $v = (f - s)$  actually. Considering the quality of  $w_{x,y}(\tilde{x}, \tilde{y})$ ,  $P_S(x, y)$  satisfies the following formula:

$$\frac{1}{|\Omega|} \int_{\Omega} P_S(x, y) dx dy = \frac{1}{|\Omega|} \int_{\Omega} (f - s)^2 dx dy = k_s \sigma^2 \tag{9}$$

where  $k_s \geq 1$ ,  $\sigma^2$  is the noise variance[6]. Obviously  $k_s = \frac{Var(f - s)}{\sigma^2}$ ,

Where  $Var(\cdot)$  represents variance.

After the image noise is processing with the fixed parameters, we can calculate the local variance of the residual image[7] and represent as  $P_{sROF}(x, y)$ , and then calculate

$$k_{sROF} = \frac{Var(f - s_{ROF})}{\sigma^2}. \text{Assuming}$$

$$\tilde{Q}(x, y) = \frac{(f - s) \sum_{j \geq 0} 2^{2j\beta} [\partial J_u(s)]_j}{k_{sROF} \cdot \sigma^2} \cdot \frac{P_{sROF}(x, y)}{k_{sROF} \cdot \sigma^2} \tag{10}$$

Where  $\sigma^2$  is the noise variance.

According to formula(2), the signs of  $(f - s)$  and  $\sum_{j \geq 0} 2^{2j\beta} [\partial J_u(s)]_j$  are identical, so  $\tilde{Q}(x, y) \geq 0$ . In the image non-texture area[8]  $P_{sROF}(x, y) \approx \sigma^2 \ll k_{sROF} \sigma^2$ , so  $\tilde{Q}(x, y)$  is small relatively.  $Q(x, y)$  and  $\lambda(x, y) = \sqrt{Q(x, y)}$  derived are small in the non-texture area, so the noise is processed effectively. In the image texture area[9],  $P_{sROF}(x, y) \gg k_{sROF} \sigma^2$ , so  $\tilde{Q}(x, y)$  is large relatively.  $Q(x, y)$  and  $\lambda(x, y) = \sqrt{Q(x, y)}$  derived from formula(3) are large in the texture area, so the texture detail is kept well.

### 2.2. Fractional order evaluating

As for the selection of  $u$ , when  $u > 1.0$  the algorithm may suppress the “staircase effect”[10]. But when  $u$  is too large, the denoising effect[11] is not good and the image visual effect is bad in non-texture area. So based on comprehensive consideration of the above two respects,  $u$  should be taken  $1.0 \leq u \leq 1.7$  in the image non-textured areas. While Considering denoising and the texture keeping in the image

texture areas,  $1.7 \leq u \leq 2.2$ .  $P_{sROF}(x, y) \approx \sigma^2 \ll k_{sROF}\sigma^2$  in the non-texture area and  $P_{sROF}(x, y) \gg k_{sROF}\sigma^2$  in the texture area, so  $u$  can be represented as:

$$u = \begin{cases} 1 + \frac{2}{5\pi} \arctan(P_{sROF} - \min(P_{sROF})), & P_{sROF} < k_{sROF} \cdot \sigma^2 \\ 1.3 + \frac{4}{5\pi} \arctan(P_{sROF} - k_{sROF} \cdot \sigma^2), & P_{sROF} \geq k_{sROF} \cdot \sigma^2 \end{cases} \tag{11}$$

In the non-texture area that is  $P_{sROF} < k_{sROF} \cdot \sigma^2$ , the value of  $u$  is near to 1.7, while in the texture area that is  $P_{sROF} \geq k_{sROF} \cdot \sigma^2$ , the value of  $u$  is near to 2.2. The range of 1.7 and 2.2 is determined by the practice experience, and the specific value of each point is determined adaptively in accordance with the local information  $P_{sROF}$  of the point [12].

### 3. Improved Fractional Dimension characterization model Optimization

The essence of fraction characterization [13] is that the integer order is extended to the fraction order. The fractional prediction variational model based on ROF [14] is applied in the paper. The fractional model is as follow:

$$\min \left\{ E(s) = \int_{\Omega} |D^u s| dx dy + \lambda \int_{\Omega} (\log s + f/s) dx dy \right\} \tag{12}$$

Where  $f$  represents the detected image,  $s$  represents unknown approximate image.  $\lambda$  represents the regularization parameter of each image pixel.  $|D^u s| = \sqrt{(D_x^u s)^2 + (D_y^u s)^2}$ ,  $D_x^u s$  refers to the  $u$  order partial derivatives about  $x$ ,  $D_y^u s$  refers to the  $u$  order partial derivatives about  $y$  [15].

By using variational calculus [16], formula (12) can be transformed as follow:

$$s \cdot H^u(s) - \lambda \cdot \frac{f - s}{s} = 0 \tag{13}$$

Where  $H^u(s) = \overline{(-1)^u} \left[ \overline{D_x^u} (D_x^u s / |D^u s|) + \overline{D_y^u} (D_y^u s / |D^u s|) \right]$ ,  $\overline{D^u}$  is the conjugate operator of  $D^u$ .

The size of the image  $s = [s(i, j)]_{i,j=1}^N$  is  $N \times N$ , supposing  $i, j < 1$  or  $i, j > N$ ,  $s_{i,j} = 0$ . Then discretization operator  $\Delta_x^u s$  of  $D_x^u s$  and  $\Delta_y^u s$  of  $D_y^u s$  are as follows respectively:

$$\left. \begin{aligned} (\Delta_x^u s)_{i,j} &= \sum_{k=0}^{K-1} (-1)^k \binom{u}{k} u(i-k, j) \\ (\Delta_y^u s)_{i,j} &= \sum_{k=0}^{K-1} (-1)^k \binom{u}{k} u(i, j-k) \end{aligned} \right\} \tag{14}$$

Then

$$\left[ H^u(s) \right]_{i,j} = \sum_{k=0}^{K-1} (-1)^k \binom{u}{k} p_1(i+k, j) + \sum_{k=0}^{K-1} (-1)^k \binom{u}{k} p_2(i, j+k) \tag{15}$$

Where

$$p_1(i, j) = \frac{(\Delta_{x^u s}^u)_{i,j}}{\sqrt{(\Delta_{x^u s}^u)_{i,j}^2 + (\Delta_{y^u s}^u)_{i,j}^2 + \varepsilon^2}}, \tag{16}$$

$$p_2(i, j) = \frac{(\Delta_{y^u s}^u)_{i,j}}{\sqrt{(\Delta_{x^u s}^u)_{i,j}^2 + (\Delta_{y^u s}^u)_{i,j}^2 + \varepsilon^2}}, \tag{17}$$

Where  $\varepsilon^2$  is a tiny constant so that the denominator is not zero [17]. K need not be too large in real calculations, here K=20.

So s represents as follow:

$$s^{(n+1)} = s^{(n)} + \Delta t \cdot \left[ -s^{(n)} \cdot H^u(s^{(n)}) + \sum_{j=0}^L 2^{-2jm_j} \left( \lambda \cdot \frac{f - s^{(n)}}{s^{(n)} + \varepsilon^2} \right)_j \right] \tag{18}$$

Where L represents the number of iterations, u and  $\lambda$  will update adaptively in the iterative operation [18]. The numeric experimentation indicates that the model parameters would be selected adaptively based on the image region characteristics [19]. Such can obtain more effective filtering in non-textured area.

Supposed that these parameters are constant and image size is M×M, the algorithm computational complexity is O(N<sup>2</sup>) for an iterative [20]. The computation of algorithm with which processing adaptive parameters was more complex than the computation of algorithm with which processing constant parameters [21].

#### 4. Experimental Analysis of coal particle image optimization

Experiment 1: when  $\lambda = 0.001$ , u take different values and the times of iteration are 1000.using fractional optimization algorithm to process international standard test image cameraman.tif which were added White Gaussian noise ,where the standard deviation  $\sigma = 20$ . Comparing the noise suppression effect and staircase effect when u were different values. As shown in Figure1 and Figure 2.

When the algorithm was iterated 1000 times with different u values, Figure 2 gives the errors between the denoising image and the original image with the iterative times

changed. Here the error  $e = mean(|u^{(n)} - f_0|)$ , where  $f_0$  is original image,  $mean(\cdot)$  is the mean value.

From the error curves in Figure 2 it was shown that the error curves between the denoising image and the original image decreased at first and then increased for different u values. Besides  $u=0.6$ , the difference of the error minimums was not large. When the errors reached the minimum values, the error would increase with the iteration and the increasing speed was more and more slow. In fact, when the numbers of iterations were large enough, the error would reach an equilibrium state which meant



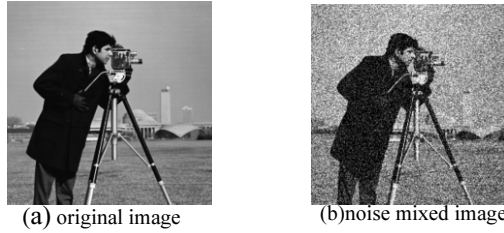


Figure 1 international standard test image cameraman.tif and thenoise mixed image

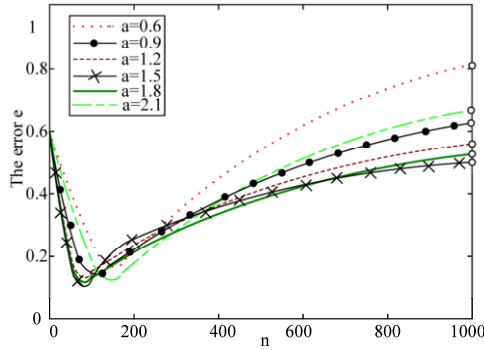


Figure 2 the error curve between the denoising image and original image with u values changing

the iterative convergence. From the error curve it could be seen that in the range [1.0,1.8] the larger  $u$  was, the smaller the error of achieving balance was finally. This meant that with the iterative convergence the greater  $u$  could make the smaller error between the denoising image and the original image. It was shown that the original image information keeping better.

Numerical experiments show that the more  $u$  is the better the capacity of "staircase effect" suppressing and textures keeping are, while the noise reduction capacity will decline. Particularly in the image non-textured areas remaining a lot of noise makes visual effect being unsatisfactory. To resolve this problem, the simplest method is that the image is processed subregionally. The characteristic parameters adjust adaptively based on the image structure. Fractional derivative Variation model and algorithm have good performances in suppressing the "staircase effect", improving the PSNR as well as keeping texture detail.

Experiment 2: we can process the noise of figure 1.(a) image with improved algorithm and compare the model denoising effects when  $u$  take different values. PSNR is a measure index of filtering effect.

$$PSNR = 10 \times \log \left( \frac{255^2}{MSE} \right) \tag{19}$$

$$MSE = \frac{\sum_{n=1}^{Framesize} (I^n - P^n)}{Framesize} \tag{20}$$

Where 255 is peak value, that is the maximum of 8 bits notation, MSE refers to Mean Square Error,  $I$  (superscript  $n$ ) represents the  $n$ th pixel value of the detected image,  $P$  (superscript  $n$ ) represents the  $n$ th pixel value of the worked image. The unit of PSNR is dB. The more the PSNR value is, the less the distortion is.

For a given  $u$ , when the PSNR reaches the maximum iteration terminates. Considering  $u$  take different values respectively, the algorithm is applied for optimization. Table 1 lists the different PSNRs when  $u$  takes different values in the image, correspondingly.

**Table 1.** The comparison of the image denoising effect PSNR

number	$u$	Iteration	PSNR
1	0.3	122	24.55
2	0.6	140	29.87
3	0.8	155	28.97
4	1.1	145	29.55
5	1.3	160	31.65
6	1.5	165	33.78
7	1.7	170	34.56
8	2.0	185	34.45
9	2.2	180	33.26
10	2.5	185	30.27

From the different PSNRs comparison, it is shown that the PSNR increased at first and then decreased with the increase of  $u$ . When  $u$  is small, the texture maintaining is not good, so the PSNR is not high. With the increase of  $u$ , texture maintaining effect enhanced, thus the PSNR improved. When  $u$  was too large, the texture maintaining effect was still good. But the denoising capability decreased and too much noise was residual in the image, so the PSNR decreased. Thus, from image denoising and texture maintaining two aspects, the larger  $u$  was does not mean that the better  $u$  was. When  $u < 1$ , there would be serious "staircase effect", while  $u$  was too large, the model denoising effect would not be good, so the general recommendation  $1.3 < u < 2.2$ . Here we selected  $u = 1.7$  through numerous tests.

Experiment 3: The experimental subjects were the gray images of  $256 \times 256$  pixels (all pictures were shrunk to show conveniently ).The test platform of algorithm was: Pentium Dual-Core CPU G3420, 4GRAM.the experiment was implemented with Matlab7.1. Firstly added the original image with the Gaussian white noise , its mean was 0 and the standard deviation  $\sigma = 20$  , The size of the square window template was  $M = 10$  .The window template was implemented to analyze the image filtered effect.

The comparative results of four methods (the Wiener filter, Wavelet filtering, integer-order variational algorithm and fractional-order variational algorithm) were as follows Figure 3, The PSNR and EPI of above-mentioned methods were as Table 2. Figure 3 shows the comparative result:

The comparative result reveals that the first three methods had the good filtering effect. In the "non-texture areas" of the image adaptive algorithm had the denoising effect and the "staircase effect" suppressing capability, and it had a certain texture maintaining capacity in the "texture areas". The improved algorithm is an efficient image optimization method. By fractional dimension modeling, the image processing model is more sophisticated. What's more, the fractional dimension optimal algorithm was more effective in texture retention capacity and image edges than other three algorithms.

The fractional-order variational algorithm remained good results in the PSNR and EPI. EPI is expressed as the ratio of the difference between two grey values at the edge of image and the original image, its value showed the ability to keep the edge. It was assigned usually in the range  $[0,1]$ . The comparison results in PSNR and EPI were shown in table 2.

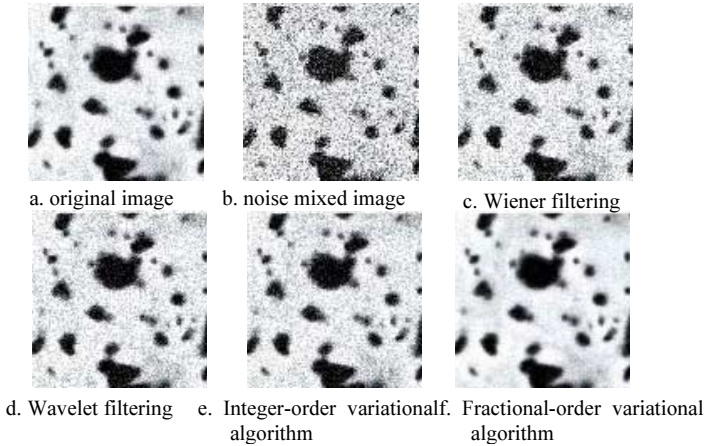


Figure3 the comparison result by different algorithms for coal dust image

**Table 2.** The comparative result of PSNRs and EPIs with three algorithms

Iteration		Original image	Wiener	Wavelet	Integer-order variational	Fractional-order variational
0	PSNR	19.565	-	-	-	-
	EPI	1.000	-	-	-	-
80	PSNR	-	22.476	23.587	26.478	33.546
	EPI	-	0.435	0.453	0.654	0.734
120	PSNR	-	23.576	24.243	27.221	34.659
	EPI	-	0.486	0.513	0.676	0.766
180	PSNR	-	24.547	25.438	28.887	36.589(*)
	EPI	-	0.612	0.642	0.764	0.886(*)
240	PSNR	-	25.520	27.547	29.587(*)	35.488
	EPI	-	0.687	0.665	0.876(*)	0.842
280	PSNR	-	26.456	28.519(*)	28.685	34.211
	EPI	-	0.765	0.786(*)	0.823	0.848
320	PSNR	-	27.142(*)	27.543	27.664	33.443
	EPI	-	0.763(*)	0.654	0.789	0.835

(\*), represented the optimum value for the specific algorithm.

Table 2 showed that to achieve the PSNR optimal value, the number of iterations needed in the improved fractional dimension optimization algorithm were less than other three algorithms. In addition, the PSNR in improved algorithm was the largest. When the number of iterations reached to 180, image texture details and edge were preserved well. Through the comparative analysis of EPI, it was showed that

fractional dimension optimization algorithm was more effective than conventional filtering algorithms.

## 5. Conclusions

The following conclusions can be acquired from this paper:

- Based on pulverized coal image filtering, a fractional dimension optimization model was proposed and adaptive algorithm was employed to select model parameters.
- The fractional dimension optimization model was coupled with adaptive algorithm. Compared with the conventional filtering methods, the simulations were carried out to demonstrate that the resulting coupled model combined the two model advantages. So the PSNR improvements, texture maintaining and "staircase effect" suppressing had better results than the two models without coupling. The fractional dimension optimization algorithm could filter and preserve the image texture information more effectively. It could suppress the image disturbances well. The improved algorithm had better convergent performance and was numerically stable.
- The improved algorithm model can be broadened further to other image process fields such as image compression, image segmentation and so forth. It has a wide application prospect and far-reaching theoretical significance. It is expected that more research will be continued in the future.

## References

- [1] Sugata Banerji, Atreyee Sinha, Chengjun Liu, New image descriptors based on color, texture, shape, and wavelets for object and scene image classification, *Neurocomputing*, **25**(2013),256-264.
- [2] Tianzhu Zhang, Si Liu, Changsheng Xu, Hanqing Lu, M 4 L: Maximum margin Multi-instance Multi-cluster Learning for scene modeling[J], *Pattern Recognition* **25**(2013),325-331.
- [3] Mandelbrot B B and Vainness J W, Fractional Brownian Motions, *Fractional Noise and Applications. SIAM Review***10**(2010),422-437.
- [4] Yao Kui, Zhang Xia, Research Announcements on the Fractional, *Calculus of A Type of weiersrtass Functions* **31** (2010),324-331.
- [5] Raoul R. Nigmatullin. Alain Le Mehuate, Is there geometrical/Physical Meaning of the fractional integral with complex exponent, *Journal of Non-Crystalline Solids*, **15**, (2009), 2888-2899.
- [6] Lei Yang, Nanning Zheng, Mei Chen, Yang Yang, Jie Yang, Categorization of Multiple Objects in a Scene Using a Biased Sampling Strategy, *International Journal of Computer Vision*, **16**, (2013 ),135-142.
- [7] M. Fazel Zarandi, R. Gamasae, Type-2 fuzzy hybrid expert system for prediction of tardiness in scheduling of steel continuous casting process, *Soft Computing* **43** (2012),365-372.
- [8] Galic I., Weickert J., Welk M., Bruhn A., Belyaev A., Seidel H.P. .Image compression with Anisotropic diffusion, *Journal of Mathematical Imaging and Vision*, **31**, (2014),255-269
- [9] Gao G., Zhao L., Zhang J., Zhou D., Huang J, A segmentation algorithm for SAR images based on the anisotropic heat diffusion equation, *Pattern Recognition*, **41**, (2013),3035-3043.
- [10] Osher S., Burger M., Goldfarb D., Xu J., Yin W, An iterative regularization method for total variation based image restoration, *Multiscale Modeling and Simulation*, **4**, (2015),460-489.
- [11] Burger M., Osher S-, Xu J., Gilboa G., Nonlinear inverse scale space methods for image restoration, *Lecture Notes in Computer Science***52**(2015),25-36.
- [12] Burger M., Gilboa G., Osher S., Xu J., Nonlinear inverse scale space methods, *Communications in Mathematical Sciences***4**(2016),175-208.
- [13] Barbu T., Barbu V., Biga V, Coca D., A PDE variational approach to image denoising and restoration, *Nonlinear Analysis*, **10**, (2012), 1351-1361.

- [14] Chen B W, Wang W W, Qin Q Q., Infrared dim target Detection Based on fractional integral operator, *Control and Decision* **27**(2012), 147-151.
- [15] Liao Z, Zhu Z T, Liang S, et al., Subspace identification for fractional order hammerstein systems based on instrumental variables, *Int J of Control, Automation and Systems*, **10**, (2012), 947-953.
- [16] Liao Z, Peng C, Li W, et al., Robust stability analysis for a class of fractional order systems with uncertain parameters, *J of the Franklin Institute*, **348** (2011), 1101-1113.
- [17] Kaslik E, Sivasundaram S., Analytical and numerical methods for the stability analysis of linear fractional delay differential equations, *J of Computational and Applied Mathematics* **236**(2012), 4027-4041.
- [18] Fioravanti A R, Bonnet C, Ozbay H, et al., A numerical method for stability windows and unstable root-locus calculation for linear fractional time-delay systems, *Automatica*, **48**, (2012), 2824-2830.
- [19] Trigeassou J C, Maamri N, Sabatier J, et al., State variables and transients of fractional order differential systems, *Computers and Mathematics with Applications*, **64**, (2012), 3117-3140.
- [20] Wang Y, Wei Y H, Zhu M, et al., A move LMI approach for robust stabilization of uncertain fractional order systems, *Proc of Chinese Control Conf. Xi'an*, **27**, (2013), 1475-1479.
- [21] Yu J., Further results on delay-distribution-dependent robust stability criteria for delayed systems, *Int J of Automation and Computing*, **8**, (2011), 23-28.

# Modified Cuckoo Search Based Neural Networks for Forest Types Classification

Sankhadeep CHATTERJEE <sup>a</sup>, Nilanjan DEY <sup>b</sup>, Soumya SEN <sup>c</sup>, Amira S. ASHOUR <sup>d,1</sup>,  
Simon James FONG <sup>e</sup>, Fuqian SHI <sup>f</sup>

<sup>a</sup>Department of Computer Science and Engineering, University of Calcutta, Kolkata, India, E-mail: chatterjeesankhadeep.cu@gmail.com

<sup>b</sup>Department of Information Technology, Techno India College of Technology, Kolkata, India, E-mail: neelanjan.dey@gmail.com

<sup>c</sup>A. K. Choudhury School of Information Technology, University of Calcutta, Kolkata, West Bengal, India, Email: soumyasen1@acm.org

<sup>d</sup>Department of Electronics and Electrical Communications Engineering, Faculty of Engineering, Tanta University, Egypt, E-mail: amirasashour@yahoo.com

<sup>e</sup>Department of Computer and Information Science Data Analytics and Collaborative Computing Laboratory University of Macau, Taipa, Macau SAR, E-mail: ccfong@umac.mo

<sup>f</sup>College of Information and Engineering, Wenzhou Medical University, Wenzhou, PR China, E-mail: sfq@wmu.edu.cn

**Abstract.** Pixel classification in land scape images is a challenging process especially in forest images due to the similar spectral features of pixels situated close to each other. Previously, meta-heuristic coupled artificial neural network (ANN) models have been used to classify the two-different species, namely Japanese Cedar, Japanese Cypress and one mixed forest class. Previous attempts have shown reasonable improvement in the classification process using genetic algorithm (GA) supported neural network over other traditional approaches. Consequently, in the current work, a modified Cuckoo Search (CS) supported Neural Network (NN-MCS) classifier is proposed. The lévy flight associated with cuckoo search has been modified using McCulloch's method of generating stable random numbers. The proposed approach is compared with GA-NN using single objective function and CS-NN (ANN trained with CS) classifiers in terms of confusion matrix based performance metrics. The results depicted the dominance of the suggested NN-MCS model compared to the CS-NN model to a greater extent.

**Keywords.** Artificial Neural Network, Cuckoo Search, McCulloch's Method, Forest Type

---

<sup>1</sup> Corresponding Author Email: amirasashour@yahoo.com

## 1. Introduction

In the soft computing and machine learning framework, pixel classification has a significant potential in remote sensing applications. Typically, several natural, environmental and other effects reduce the classification accuracy in remote sensing. Thus, several studies have been conducted to improve the classification accuracy [1–8]. The classification problem can be formulated to classify pixels in to different groups corresponding to a particular land cover type. Afterward, the trained classifier can be conducted to find out the pixel membership whose class is unknown. In satellite images of forests, the classification process complicated due to the increased number of pixels as one pixel may have membership to multiple classes. This results in reasonable uncertainty in the classification process. Numerous unsupervised learning supported approaches have been reported to tackle the same efficiently [9-12]. Such unsupervised methods include split and merge [14], Fuzzy c-means [13], and NN [15, 17 – 22]. During classification the problem of uncertainty poses a great challenge, especially in case of forest images, as they contain higher level of pixel variation in a relatively small geographic area. Johnson *et al.* [16] proposed a support vector machine based geographically weighted variable supported method for classification. Slight improvement has been shown over the methods not supported by geographically weighted variables.

Recent development in machine learning has pointed out ANN or simply NN [38 - 45] as one of the most potential classifier for solving the classification tasks. In order to achieve maximum accuracy, the ANN weight vectors should be optimized, which is realized by optimizing an objective function (usually root-mean-square error (RMSE)). However, this traditional learning algorithm may cause an untimely convergence to local optima. In order to overcome such drawback, meta-heuristic optimization procedures are involved to train the NN in order to achieve the expected accuracy [17-27, 29]. Meta-heuristic supported ANN, such as GA trained ANN has been employed in classifying forest types. Nevertheless, the classification accuracy can be enhanced by using recent meta-heuristic algorithms, such as the CS. The current study reported forest classification using CS trained ANN. Studies have revealed that traditional CS can be improved by using McCulloch's random numbers in levy flights [28]. Thus, a modified CS algorithm has been employed to train ANN in the current work. This hybrid NN-MCS model is compared with the traditional NN-CS and NN-GA models in terms of confusion matrix based performance measuring criterions.

The remaining sections structure is arranged as follows. Section 2 defines the proposed method with introducing the variation of CS used and other experimental details. Section 3 introduces the ANN training method by modified CS. Section 4 described the dataset used in current study in detail. Finally, section 5 reported the experimental results and discussion.

## 2. Proposed Modified Cuckoo Search

In the current work, a modified CS algorithm is employed to enhance the convergence rate of the CS procedure so that it can work efficiently in time constraint situations. A new method is proposed for modeling the Lévy flight in the CS algorithm to achieve less expensive method that produces stable random variables. This method can be applied for modeling the Lévy flight. There are various methods proposed by many

researchers to generate stable random numbers. Most of the procedures depend on the inverse distribution of a set of pseudo-random numbers, which should be consistently distributed within (0, 1) range [28]. Large number of simulations can be efficiently and reasonably performed by applying such procedures, however, these methods are expensive and may introduce inaccuracies in each that propagate to the Monte Carlo results. Therefore, in the current work, a faster method is proposed that can be used in some real world problems by using the Lévy flight production in CS algorithm for training the ANNs.

---

**Algorithm 1: Modified Cuckoo Search with McCulloch algorithm**

---

*Set* the population  $x_{i,j}$   
*Calculate* the fitness value for a defined objective function:  $f(x)$ ;  $x = [x_1, x_2, \dots, x_n]^T$   
**If** (Iteration < Max(Iteration)) **then**  
     *Preserve* the current finest solution to create new solution space  
     *Calculate* fitness value  
     *Store* the finest nest  
     **if**  $k < \text{prob}_a$  **then**  
         *Use* McCulloch’s method to generate the step size ( $\alpha$ )  
         *Replace* the worst nest by Lévy flight  
         *Compute* the fitness value  
         *Store* the best nest  
         *Increase* the Counter  
         *Search* for the best fitness that has been computed till this stage  
     **else**  
         *Store* those nests  
     **end**  
**Compute** the optimum solutions

---

In the preceding algorithm, the value of  $\alpha$  should to be kept within the range 0 to 2, otherwise overflow can arise. Therefore, in case of the real world experiments, the value of  $\alpha$  is restricted to the range  $0 \leq \alpha \leq 2$ . The step size is given by:

$$s_z = \gamma \frac{1}{\alpha} Z_{cn} \tag{1}$$

here,  $\gamma$  represents the scale of the problem, Z is the generated random variable, which converges very fast (except for the large values) to Lévy distribution as follows:

$$Z_{cn} = \frac{1}{n^\alpha} \sum_{i=1}^n w_i \tag{2}$$

where, w is the nonlinear transformation to achieve rapid distribution convergence. The solution set is updated by using the generated step size. The solution set represents the eggs in the nests to produce new collection of values for the next iteration.



### 3. Modified Cuckoo Search in training Artificial Neural Networks

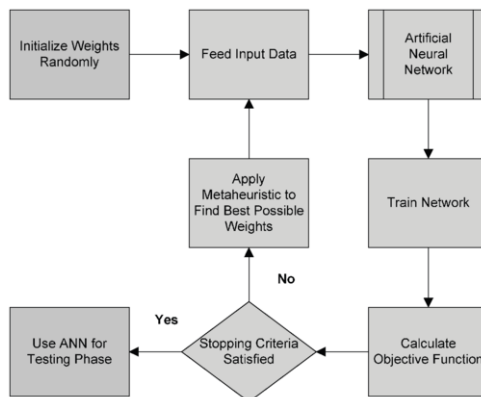
The ANN is one of the most used modeling approaches [30, 31]. It achieves accurate classification even with very small dataset. It can handle imprecise relationships during its training stage. The ANN structure consists of interconnected computational neurons that involved in the mathematical mapping through the learning process to adjust the weight value. Initially, the training phase is started by a part of the dataset to classify its inputs accompanied by its class label to create the classification. Afterward, a validation stage is performed to confirm the efficiency of the trained model using extra dataset. Finally, the evaluation phase is used to test the classification model accuracy using another set of test data. In general, the artificial neuron uses the input signal ( $x$ ) and their equivalent weights ( $w$ ) to form the input ( $N_j$ ). This input is surpassed to a linear threshold until it exceeds the output signal ( $y$ ) to other neuron. If  $N_j$  exceeds the threshold of that neuron, the neuron is inspired. The net input ( $N_j$ ) is calculated by the following equation:

$$N_j = \sum_{i=1}^n w_{ij}x_i \quad (3)$$

where,  $n$  is the number of input signals,  $w$  is the weight and  $x$  is the signal strength. Consequently, the output ( $y$ ) is computed as follows:

$$y = \begin{cases} 1, & \text{if } N_j \geq \theta_j \\ 0, & \text{if } N_j < \theta_j \end{cases} \quad (4)$$

here, ( $\theta_j$ ) is the bias. The sigmoid and logistic functions can be used as an activation functions. The perceptron learning rule assures optimal weight vector for ANN [32]. For the MLP-FFN experiments, two-layer perceptron feed-forward network can be conducted. The flow of current work engages different meta-heuristics, namely the CS and modified CS in the ANN training phase to find the optimal weight vectors of the network as depicted in Figure 1.



**Figure 1.** Flow of training of a typical ANN by using meta-heuristic optimization

#### 4. Experimental Methods

The current work utilized dataset [16] information of forest area in the province of Ibaraki Prefecture of Japan (Latitude 36° 57 N, Longitude 140° 38 E), which approximately covers area of 156 square km. The landscape contains mainly two species, namely Japanese Cypress planted forest and Japanese Cedar and mixed natural forest [16] that have been mapped by using the spectral features near visible infrared wavelengths obtained from satellite images captured by ASTER satellite. The dataset has 27 attributes, namely spectral information in the red, green, and near infrared wavelengths.

Several studies revealed that performance comparison in terms of accuracy is not sufficient to claim the ingenuity of a model [34, 36, 37]. Thus, the proposed NN-MCS model is compared with the NN-GA and NN-CS models in terms of accuracy, precision, recall, and F-Measure. These performance metrics are designed from the confusion matrix [35] of testing phase for different models under consideration. Performance measures are as follow:

$$Accuracy = \frac{tp + tn}{tp + fp + fn + tn} \quad (5)$$

$$Precision = \frac{tp}{tp + fp} \quad (6)$$

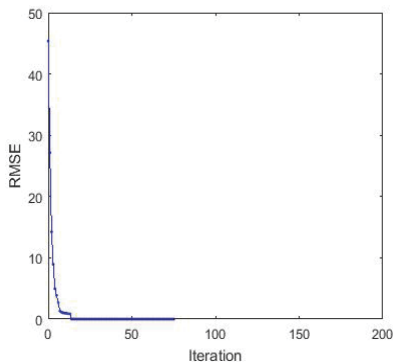
$$Recall = \frac{tp}{tp + fn} \quad (7)$$

$$F - Measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

Here, true positive, false positive, true negative, and false negative are signified by 'tp', 'fp', 'tn', and 'fn'; respectively.

#### 5. Results and Discussion

Simulation based experiments is tested using the dataset defined in section 3. The experimental procedure has been designated in Figure 1. The GA objective is selected to minimize the RMSE (in section 2). Figure 2 illustrates the RMSE versus iterations for MCS during training phase of ANN indicating the convergence of MCS in the training phase of ANN.



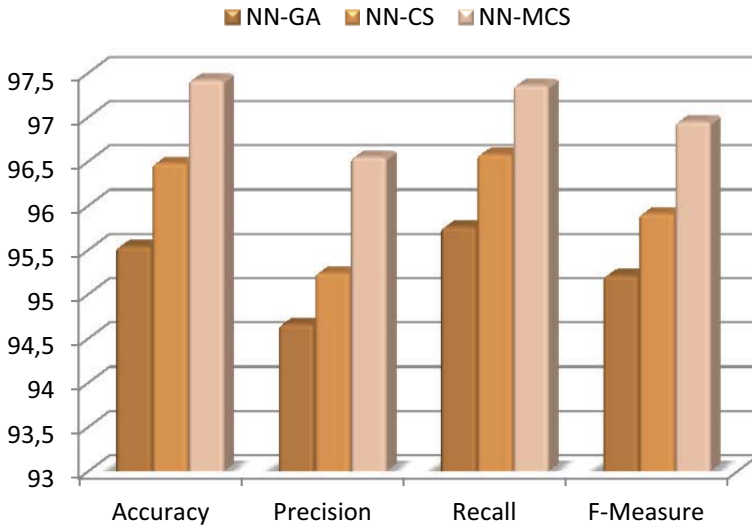
**Figure 2.** Plot of RMSE versus iterations for MCS during training phase of ANN

Table 1 reports the experimental results for NN-GA, NN-CS and NN-MCS classifiers.

**Table 1.** Performance measures of different algorithms

	<b>NN-GA</b>	<b>NN-CS</b>	<b>NN-MCS</b>
<b>Accuracy</b>	95.54	96.48	97.42
<b>Precision</b>	94.66	95.24	96.55
<b>Recall</b>	95.76	96.58	97.36
<b>F-Measure</b>	95.21	95.91	96.95

Table 1 illustrates that the NN-GA model has performed significantly well with an accuracy, precision, recall and F-Measure of 95.54%, 94.66%, 95.67%, 95.21%; respectively, while the proposed method achieved 97.42%, 96.55%, 97.36% and 96.95% values for the same metrics; respectively. Figure 3 illustrates the classification performance measures.



**Figure 3.** Comparative study of different classifiers in terms of different performance metrics

The comparative analysis in Figure 3 reveals that the proposed NN-MCS is higher in terms of all the measured metrics. The results indicate that NN-MCS is at least superior to the models considered in the current study for image classification task under present work.

## Conclusion

The current work proposed a MCS trained NN (NN-MCS) classification procedure to classify two forest types with one mixed class both species. The results proved the dominance of NN-MCS compared to the NN-CS (NN trained with simple CS) for classification of pixels in forest images. Results have indicated that MCS supported ANN is better in classifying different forest types. The proposed model has been found to be effective in classifying the forest types than simple version (NN-CS). Extended analysis of different models included performance analysis in terms of different performance metrics measurements.

## References

- [1] C. Kontoes, G.G. Wilkinson, A. Burrill, S. Goffredo, J Megier, An Experimental System for the Integration of GIS Data in Knowledge-Based Image Analysis For Remote Sensing of Agriculture, *International journal of geographical information systems*, 7 (1993), 247–262.
- [2] P. Gong, P.J. Howarth, Frequency-based contextual classification and gray-level vector reduction for land-use identification, *Photogrammetric engineering and remote sensing*, 58 (1992), 423–437.
- [3] J. Stuckens, P.R. Coppin, and M.E. Bauer, Integrating contextual information with per-pixel classification for improved land cover classification, *Remote sensing of environment*, 71 (2000), 282–296.
- [4] J. San miguel-ayanz, G.S. Biging, Comparison of single-stage and multi-stage classification approaches for cover type mapping with tm and spot data, *Remote sensing of environment*, 59 (1997), 92–104.

- [5] G.M.Foody, Approaches for the production and evaluation of fuzzy land cover classification from remotely-sensed data. *International journal of remote sensing*, 17 (1996), pp. 1317–1340.
- [6] F.J. Gallego, Remote sensing and land cover area estimation, *International journal of remote sensing*, 25 (2004), 3019–3047.
- [7] M. Pal, P.M. Mather, An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment*, 86 (1996), 554–565.
- [8] S.E Franklin, D.R. Peddle, J.A. Dechka, and G.B. Stenhouse, Evidential reasoning with Landsat TM, DEM and GIS data for land cover classification in support of grizzly bear habitat mapping, *International Journal of Remote Sensing*, 23 (2002), 4633–4652.
- [9] S.Bandyopadhyay, S.K.Pal, Pixel Classification Using Variable String Genetic Algorithms with Chromosome Differentiation, *IEEE Transactions on Geoscience and Remote Sensing*, 39(2) (2001), 303–308.
- [10] U.Maulik, S. Bandyopadhyay, Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification, *IEEE Transactions on Geoscience and Remote Sensing*, 41(5) (2003), 1075–1081.
- [11] S. Bandyopadhyay, U. Maulik, Genetic Clustering for Automatic Evolution of Clusters and Application to Image Classification, *Pattern Recognition*, 35(2) (2002), 1197–1208.
- [12] S. Bandyopadhyay, U. Maulik, A. Mukhopadhyay, Multiobjective Genetic Clustering for Pixel Classification in Remote Sensing Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 45(5) (2007), 1506–1511.
- [13] R. L. Cannon, R.Dave, J.C. Bezdek, M. Trivedi, Segmentation of a Thematic Mapper Image using Fuzzy c-means Clustering Algorithm, *IEEE Transactions on Geoscience and Remote Sensing*, 24 (1986), 400–408.
- [14] R. H. Laprade, Split-and-merge Segmentation of Aerial Photographs, *Computer Vision Graphics and Image Processing*, 48(1988), 77–86.
- [15] A.Baraldi, F. Parmiggiani, A Neural Network for Unsupervised Categorization of Multivalued Input Pattern: An Application to Satellite Image Clustering, *IEEE Transactions on Geoscience and Remote Sensing*, 33(1995), 305–316.
- [16] B.Johnson, R. Tateishi, Z. Xie, Using geographically weighted variables for image classification. *Remote Sensing Letters*, 3(6) (2012), 491–499.
- [17] S. Dehuri, S.B. Cho, A hybrid genetic based functional link artificial neural network with a statistical comparison of classifiers over multiple datasets, *Neural Comput Appl* , 19(2) (2010), 317–328
- [18] K.Socha C. Blum, An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training, *Neural Comput Appl* 16(3) (2007), 235–247.
- [19] C. Ciancio, G.Ambrogio, F. Gagliardi, R. Musmanno, Heuristic techniques to optimize neural network architecture in manufacturing applications. *Neural Comput Appl*. (2015).
- [20] S. Hore, S. Chatterjee, S. Sarkar, N. Dey, A. S. Ashour, D. Bălas-Timar and V. E. Balas, Neural-based Prediction of Structural Failure of Multi Storied RC Buildings, *Structural Engineering and Mechanics*, 58(3) (2016).
- [21] M. SZ, S. Saremi, S.M. Mirjalili, Designing evolutionary feedforward neural networks using social spider optimization algorithm, *Neural Comput Appl*, 26(8) (2015), 1919–1928.
- [22] J.F. Chen, Q.H.Do, H.N.Hsieh, Training artificial neural networks by a hybrid PSO-CS algorithm, *Algorithms*, 8 (2015)292–308
- [23] S. Chatterjee, S.Sarkar, S. Hore, N. Dey, A.S.Ashour, V.E.Balas, Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings, *Neural Computing and Applications*, (2016).
- [24] S. Hore, S. Chatterjee, R. Kr. Shaw, N. Dey, J. Virmani, “Detection of Chronic Kidney Disease: A NN-GA based Approach”, CSI - 2015; 50th Golden Jubilee Annual Convention, At Delhi, Volume: Springer - AISC, November 2015. (in press)
- [25] S. Hore, S. Chatterjee, V. Santhi, N. Dey, A. S. Ashour, V. E. Balas, F. Shi, Indian Sign Language Recognition using Optimized Neural Networks, 2015 International Conference on Information Technology and Intelligent Transportation Systems (ITITS 2015), At Xi’an, China, Volume: Springer - AISC, December 2015.
- [26] S. Chatterjee, S. Ghosh, S. Dawn, S. Hore, N. Dey, Forest Type Classification: A Hybrid NN-GA Model Based Approach, Third International Conference on Information System Design and Intelligent Applications, At Vishakhapatnam, Volume: Springer - AISC, January 2016.
- [27] S. Chatterjee, S. Hore, N. Dey, S. Chakraborty, A. S. Ashour, Dengue Fever Classification using Gene Expression Data: A PSO based Artificial Neural Network Approach, 5<sup>th</sup> International Conference on Frontiers in Intelligent Computing: Theory and Applications; Springer, June 2016.

- [28] S. Chakraborty, S. Chatterjee, N. Dey, A. S. Ashour, A. Ashour, F. Shi, K. Mali, Modified Cuckoo Search Algorithm in Microscopic Image Segmentation of Hippocampus, *Microscopy Research and Technique*, May 2017.
- [29] S. Chatterjee, S. Sarkar, S. Hore, N. Dey, A. S. Ashour, F. Shi, D. Le, Structural Failure Classification for Reinforced Concrete Buildings Using Trained Neural Network based Multi Objective Genetic Algorithm, *Structural Engineering and Mechanics*, Techno Press (2017).
- [30] A.J. Maren, C.T. Harston, R.M. Pap, Handbook of neural computing applications. Academic Press, San Diego (2014).
- [31] D.R. Baughman, Y.A. Liu, Neural networks in bioprocessing and chemical engineering. Academic press, San Diego (2014).
- [32] R. Rojas, Neural networks: a systematic introduction. Springer Science & Business Media, Berlin (2013).
- [33] N. Karayiannis, A.N. Venetsanopoulos, Artificial neural networks: learning algorithms, performance evaluation, and applications. Springer Science & Business Media, New York (2013).
- [34] C. Drummond, C. Robert, C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling, In Workshop on learning from imbalanced datasets II, vol. 11. 2003
- [35] D.M.W. Powers, Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation, *Journal of Machine Learning Technologies*, 2 (1) (2011) 37–63.
- [36] N. Japkowicz and S. Stephen, The class imbalance problem: A systematic study, *Intelligent Data Analysis*, 6(5) (2002) 429–450.
- [37] M. Gao, X. Hong, S. Chen and C. J. Harris, A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems, *Neurocomputing*, 74 (2011), 3456–3466.
- [38] S. Chatterjee, R. Chakraborty, N. Dey, and S. Hore, A quality prediction method for weight lifting activity, (2015), 95-8.
- [39] Z. Tian, N. Dey, A.S. Ashour, P. McCauley, F. Shi, Morphological segmenting and neighborhood pixel-based locality preserving projection on brain fMRI dataset for semantic feature extraction: an affective computing study, *Neural Computing and Applications*, (2017), 1-16.
- [40] S. Chatterjee, S. Paladhi, S. Hore, and N. Dey, Counting all possible simple paths using Artificial Cell division mechanism for directed acyclic graphs, *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1874-1879. IEEE, 2015.
- [41] D. Wang, Z. Li, N. Dey, A.S. Ashour, R.S. Sherratt, F. Shi, Case-Based Reasoning for Product Style Construction and Fuzzy Analytic Hierarchy Process Evaluation Modeling Using Consumers Linguistic Variables, *IEEE Access*, 5 (2017), 4900-4912.
- [42] S. Hore, S. Chatterjee, V. Santhi, N. Dey, A.S. Ashour, V. E. Balas, and F. Shi, Indian Sign Language Recognition Using Optimized Neural Networks, In *Information Technology and Intelligent Transportation Systems*, (2017) pp. 553-563.
- [43] Z. Li, K. Shi, N. Dey, A.S. Ashour, D. Wang, V.E. Balas, V. E., F. Shi, Rule-based back propagation neural networks for various precision rough set presented KANSEI knowledge prediction: a case study on shoe product form features extraction. *Neural Computing and Applications*, (2016) 1-18.
- [44] S. Chatterjee, N. Dey, A. S. Ashour, C. V. A. Drugarin, Electrical Energy Output prediction using Cuckoo Search supported Artificial Neural Network". World Conference on Smart Trends in Systems, Security and Sustainability (WS4 2017) At London, Volume: Springer LNNS series, February 2017.
- [45] S. Chatterjee, R. Nag, N. Dey, A. S. Ashour, Efficient Economic Profit Maximization: Genetic Algorithm based Approach, World Conference on Smart Trends in Systems, Security and Sustainability (WS4 2017) At London, UK, Volume: Springer LNNS series, February 2017.

# Monocular Vision Based Object Recognition and Tracking for Intelligent Robot

Qian Gao<sup>a,1</sup>, Lilian Zhang<sup>a</sup>, Xiaoping Hu<sup>a</sup>, Xiaofeng Wang<sup>a</sup>

<sup>a</sup>College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha, China

**Abstract.** The intelligent robot is a system that can detect the surroundings and make corresponding actions according to environment changes. It can perform the corresponding functions according to the pre-set instructions. Navigation is one of the important directions for robot research. The robot can move from the starting point to the destination under the control of corresponding module.

The main work of this paper is based on the visual tracking of markers and general targets. For tracking of markers, the EPnP algorithm is used to solve the camera's position and orientation relative to marker so that the robot can track it. For general target tracking, the AdaBoost algorithm is employed to recognize the target. Then we can get the relative position and realize tracking by using the RGBD information.

**Keywords.** Object recognition, Object tracking, EPnP, AdaBoost

## 1. Introduction

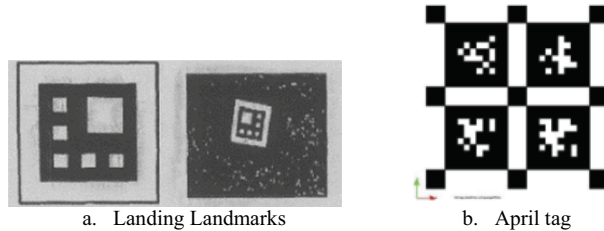
With the development of society, people put more demands on the robot both in social production and people's lives. Compared with other sensors, visual sensors have a range of advantages. Using visual sensors can perceive rich environmental information and the cost is relatively low. Visual sensors are passive measurement sensors that don't emit energy, so visual sensors don't interfere with each other. Using of visual sensors for navigation has high accuracy and good safety performance, what's more, it saves the robot's electrical energy. Therefore, the study of robot vision navigation and motion control is of great importance to promote the industrialization of robots, and has a strong practical significance.

At present, tracking marker is relatively mature in robot monocular vision tracking system. As the special landing sign in figure a, it is used in an autonomous landing system<sup>[1]</sup> for an unmanned helicopter. The marker is pre-manually designed. Based on the principle of computer vision, the position and orientation information of robot relative to the marker is obtained from image. This paper uses the marker (April tag) provided by ethz-asl laboratory<sup>[2-5]</sup>. April tag has rotation invariance. After the camera calibration, it can accurately calculate the relative pose of the camera with respect to the marker and provide all the tracking control parameters. Marker based tracking technology is usually applied to the autonomous landing system of UAVs<sup>[6]</sup> and the

---

<sup>1</sup> Corresponding Author, College of Mechatronics Engineering and Automation, National University of Defense Technology, Deya Road No.109, Kaifu District, Changsha, Hunan Province, China; E-mail: 492570232@qq.com.

terminal of robots navigation system<sup>[7]</sup>. In this work the image information is collected by the Kinect camera mounted on the Turtlebot, a ground mobile robot. Then estimate the positioning to the fixed marker through the image processing. During the movement of the marker, the robot is controlled to track the target stably and maintain a certain distance.



**Figure 1.**

The recognition and tracking of general targets based on computer vision has been developing rapidly in recent decades. But there are still many difficult problems to be solved. For example, target recognition and tracking of sequence images is a typical and challenging task. First, we need to identify automatically the target before tracking. At present, automatic target recognition is based on image segmentation<sup>[8]</sup>, feature extraction and intelligent learning. In this paper, machine learning is used to recognize the general target. Then controlling the movement of the robot to achieve the tracking of general targets.

## 2. The Tracking of Marker

According to the camera geometry model, for the rule target like April tag, the external parameters can be obtained according to the relation between the corners of the April tag in the image and in the real physical world. This external parameters defining the position of the camera relative to April tag at this time. We can get the relative position through the external parameters and then track the marker.

### 2.1. Camera calibration

In the pinhole camera model, like eyes, the camera needs to collect more light use the lens. However, the disadvantage of using this lens to obtain more light is not only causes us to deviate from the simple pinhole geometry, but also introduces distortions from the lens. Therefore, camera calibration is needed to correct the main deviation from the pinhole model due to the use of lenses. In this paper, Kalibr calibration toolbox is used to calibrate the camera to obtain the camera's internal parameters and distortion coefficients.

### 2.2. Camera positioning relative to marker

After calibrating the internal parameters and the distortion coefficient of the camera, the external parameters can be obtained according to the correspondence between the corner of the marker in the image and in the real physical world.



Positioning based on icon feature is also known as PnP (Perspective-n-Point) problem. It determines the relative position and pose of the camera according to the  $N$  feature points on the object. In this experiment, we use EPnP<sup>[9]</sup> to solve the pose. Let the reference points, that is, the  $n$  points whose 3D coordinates are known in the world coordinate system, be  $p_i, i = 1, \dots, n$ . Similarly, let the 4 control points we use to express their world coordinates be  $c_j, j = 1, \dots, 4$ .

When necessary, we will specify that the point coordinates are expressed in the world coordinate system by using the  $w$  superscript, and in the camera coordinate system by using the  $c$  superscript. We express each reference point as a weighted sum of the control points  $p_i^w = \sum_{j=1}^4 \alpha_{ij} c_j^w$  with  $\sum_{j=1}^4 \alpha_{ij} = 1$ . Where the  $\alpha_{ij}$  are homogeneous barycentric coordinates. They are uniquely defined and can easily be estimated. The same relation holds in the camera coordinate system and we can also write  $p_i^c = \sum_{j=1}^4 \alpha_{ij} c_j^c$ .

In theory the control points can be chosen arbitrarily. However, in practice, we have found that the stability of our method is increased by taking the centroid of the reference points as one, and to select the rest in such a way that they form a basis aligned with the principal directions of the data.

We now derive the matrix  $\mathbf{M}$  in whose kernel the solution must lie given that the 2D projections of the reference points are known. Let  $\mathbf{A}$  be the camera internal calibration matrix and  $\{u_i\}_{i=1,\dots,n}$  the 2D projections of the  $\{p\}_{i=1,\dots,n}$  reference points. We have

$$\forall i, \quad w_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_c \\ 0 & f_v & v_c \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^4 \alpha_{ij} \begin{bmatrix} x_j^c \\ y_j^c \\ z_j^c \end{bmatrix} \quad (1)$$

Substituting this expression in the first two rows yields two linear equations for each reference point:

$$\sum_{j=1}^4 \alpha_{ij} f_u x_j^c + \alpha_{ij} (u_c - u_i) z_j^c = 0 \quad (2)$$

$$\sum_{j=1}^4 \alpha_{ij} f_v y_j^c + \alpha_{ij} (v_c - v_i) z_j^c = 0 \quad (3)$$

Hence, by concatenating them for all  $n$  reference points, we generate a linear system of the form  $\mathbf{M}\mathbf{x} = 0$ , where  $\mathbf{x} = [c_1^c{}^T, c_2^c{}^T, c_3^c{}^T, c_4^c{}^T]^T$  is a 12-vector made of the unknowns, and  $\mathbf{M}$  is a  $2n \times 12$  matrix. The solution therefore belongs to the null space, or kernel, of  $\mathbf{M}$ , and can be expressed as

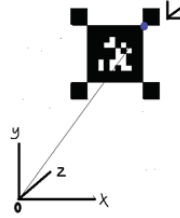
$$\mathbf{x} = \sum_{i=1}^N \beta_i v_i \quad (4)$$

Where the set  $v_i$  are the columns of the right-singular vectors of  $\mathbf{M}$  corresponding to the  $N$  null singular values of  $\mathbf{M}$ .

### 2.3. Tracking marker

After obtaining the pose relation between the camera and the marker, the pose relation is represented by the rotation vector  $\mathbf{R}$  and the shift vector  $\mathbf{t}$ . The next step is to use the Turtlebot to track the marker and keep it at the target distance. Since the camera is

fixed on the Turtlebot, we only consider the translation vector, regardless of the rotation vector.

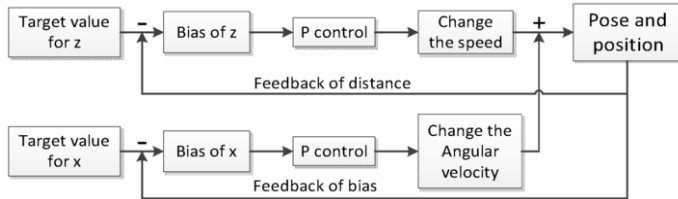


**Figure 2.** The translation of the camera relative to the first corner of the marker

As shown in the figure 2, point O is the origin of the camera coordinate system. The blue color point is the first corner detected in the marker. Regardless of the rotation vector, the origin of the camera coordinate system and the blue color point of the world coordinate system marker can be overlapped by translation. If the translation vector is obtained as  $t = [x \ y \ z]^T$ , we can overlap the two point by move camera a vector t.

The basic objective of the experiment is to keep the Turtlebot and marker distance at 75 cm (+ - 5 cm) on the Z axis and the first corner of the marker on the X axis is located on the center axis of the camera image plane (+ - 2 cm).

The control loop is shown in figure 3:



**Figure 3.** Control loop

We need to set the minimum speed and maximum upper speed for Turtlebot. Otherwise, Turtlebot cannot move because of the friction force or it may collide because of too much speed. This experiment only needs proportional control to meet the control requirements. The selection of the P parameter cannot be too large, otherwise the Turtlebot will wobble back and forth.

### 3. The Tracking of general targets

For the identification and tracking of general targets, the target recognition tracking system should be used to track the target intuitively. Firstly, the feature is extracted from the target, and then the classifier is trained. The target is identified by the trained classifier and the RGBD image of Kinect is used to track the pose.

#### 3.1. Object recognition

In the object detection, we need to have a sub-window continuous shift sliding in the image to be detected. Each point of the sub-window will calculate the characteristics of the region. Then we use a trained cascade classifier to filter the feature. Once the

feature passes through the screening of all strong classifiers, it is determined that the region is the object to be identified. For the representation and selection of features, as well as the construction of classifiers, the following schemes are adopted:

Haar classifier=Haar-like features+ Integral diagram method+AdaBoost+ Cascade

- Test with Haar-like features

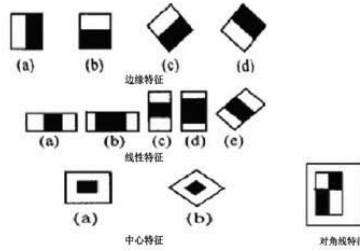


Figure 4. Harr-like features

- Accelerating Haar-like feature evaluation using integral graphs
- Training strong classifiers using the AdaBoost algorithm<sup>[10][11]</sup>

The algorithm takes as input a training set  $\{(x_1, y_1) \dots (x_N, y_N)\}$ , where  $x \in \chi, \chi \in \mathbb{R}^N$  and each label  $y_i$  is in some label set  $\{-1, +1\}$ . AdaBoost calls a given weak or base learning algorithm repeatedly in a series of rounds.

**Step1.** Initialize the weight distribution of training data. All weights are set equally:  $1/N$ .

$$D_1 = (w_{11}, w_{12}, \dots, w_{1i}, \dots, w_{1N}), w_{1i} = \frac{1}{N}, i = 1, 2, \dots, N \quad (5)$$

**Step2.** Perform multiple rounds of iterations, Use  $m = 1, 2, \dots, M$  to represent the number of iterations. Using the training data set with the weight distribution  $D_m$ , the basic classifier is obtained:

$$G_m(x): \chi \rightarrow \{-1, +1\} \quad (6)$$

The goodness of a weak hypothesis is measured by its error:

$$e_m = P(G_m(x_i) \neq y_i) = \sum_{i=1}^N w_{mi} I(G_m(x_i) \neq y_i) \quad (7)$$

Calculate the coefficient of  $G_m(x)$ , where  $\alpha_m$  represents the importance of  $G_m(x)$  in the final classifier (Get the weight of the basic classifier in the final classifier):

$$\alpha_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} \quad (8)$$

Update the weight distribution of the training data set (Get a new weight distribution for the sample):

$$D_{m+1} = (w_{m+1,1}, w_{m+1,2}, \dots, w_{m+1,i}, \dots, w_{m+1,N}) \quad (9)$$

$$w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp(-\alpha_m y_i G_m(x_i)), i = 1, 2, \dots, N \quad (10)$$

Where  $Z_m$  is a normalization factor (chosen so that  $D_{m+1}$  will be a distribution).

$$Z_m = \sum_{i=1}^N w_{mi} \exp(-\alpha_m y_i G_m(x_i)) \quad (11)$$

**Step3.** Combine weak classifiers:

$$G(x) = \text{sign}(f(x)) = \text{sign}\left(\sum_{m=1}^M \alpha_m G_m(x)\right) \quad (12)$$

- Using filter cascades to cascade strong classifiers together to improve the recognition accuracy

### 3.2. Object tracking.

Since the monocular camera can't obtain the depth information, the RGBD camera is used to obtain the depth information to control the distance between the robot and the target.

The control flow chart is the same as the tracking control flow chart of mark. However, compared with the marker tracking, the P parameter setting of this experiment is more difficult. The biggest difficulty in the control part of this test is the adjustment of the two-axis P parameter.

## 4. Test Results and Analysis

Firstly, the motion information is acquired by the Kinect camera mounted on the Turtlebot of the ground mobile robot. Then, the positioning function of the relative marker is realized by the processing of the image information. And in the process of moving marker, the robot is controlled to track the target stably and maintain a certain distance. What's more, we identify the general targets, and then control the movement of the robot to achieve the tracking of general targets.



**Figure 5.** Turtlebot equipped with Kinect camera

#### 4.1. EPnP algorithm verification

In the tracking of the marker, the EPnP is used to solve the position. When the distance is too small, the image tag shape is not complete so that can't solve. And when the distance is longer 3 meters, the distance value of the solution jumps sharply. Therefore, the following four distances are selected. By calculating the distance of the marker at different distances, 100 sets of data are calculated at each position to test the accuracy of the algorithm.

**Table 1.** EPnP test results for distance

Real distance(cm)	Min value(cm)	Max value(cm)	Average value (cm)	error(cm)
65.55	64.95	66.11	65.54	0.01
123.26	122.59	123.31	123.09	0.17
185.11	183.67	186.10	184.83	0.28
303.52	279.29	331.21	306.75	3.23

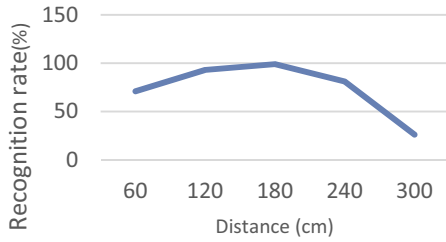
As shown in Table 1, in the distance that can detect the complete tag, the distance of the error is increasing with the distance increases. This is because with the increase of distance, the collected images are more and more blurred. The extraction of corner coordinates is not accurate enough, so the error of the calculated distance gradually increased.

From the above experiments we can see that in the marker tracking, calibration plate rotation angle should not change too fast. It will result in insufficient number of tag collected in the camera so that we can't solve the position. In the video, the robot can't move smoothly several times because of the rapid change of the moving angle of the calibration board. When the robot movement pauses, we move the calibration board slightly into the camera's range of vision, and then the robot continues to follow smoothly.

#### 4.2. AdaBoost algorithm verification

In the experiment of tracking a general target, a book is selected as the target to be tracked. Similar to the EPnP experiment, this experiment collects one hundred pictures with the book at different distances for identification. The results are shown in Table 2 and Figure 6:

distance(cm)	Recognition rate (%)
60	71
120	93
180	99
240	81
300	26



**Figure 6.** Recognition rate line chart

From the result we can see that when the target distance is shorter than 180cm, the recognition rate increases with the increase of distance. When the target distance is longer than 180cm, the recognition rate decreases gradually as the distance increases. From 240cm to 300cm reduction is particularly evident. This is due to the fact that the pixels used in the training classifier are not high. But when the distance is small, the details of the characteristics of obvious, so the recognition rate is not high.

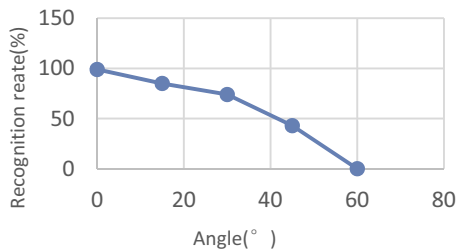
Next, as shown in Figure 7, the target is rotated at an angle at the highest recognition rate point. The effect of target deformation on the recognition rate is observed, as shown in Table 3 and Figure 8:



**Figure 7.** Turn the target around a certain angle

**Table 3.** Recognition rate on different angle

Angle(°)	Recognition rate(%)
0	99
15	85
30	74
45	43
60	00



**Figure 8.** Recognition rate line chart

From the result we can see that with the increase of rotation, the recognition rate decreases gradually. The recognition rate is decreased from 0° ninety-nine percent to zero percent in 60°. This is due to the greater the rotation angle the greater the deformation. The greater the difference between the pictures used in training classifiers, the lower the recognition rate.

Finally, we carry out an interference experiments using a similar object. The results are shown in Figure 9:



Figure 9. Recognition rate line chart

From the recognition results, AdaBoost algorithm has a certain anti-interference ability.

#### 4.3. Experimental video link

Marker-based tracking:

[http://v.youku.com/v\\_show/id\\_XMjc3MjkzNjg4OA==.html?spm=a2hzp.8244740.userfeed.512~5~5~513~5~A](http://v.youku.com/v_show/id_XMjc3MjkzNjg4OA==.html?spm=a2hzp.8244740.userfeed.512~5~5~513~5~A)

Tracking based on general object:

[http://v.youku.com/v\\_show/id\\_XMjc3MzAwNjAwNA==.html?spm=a2h3j.8428770.3416059.1](http://v.youku.com/v_show/id_XMjc3MzAwNjAwNA==.html?spm=a2h3j.8428770.3416059.1)

## 5. Conclusion

The main work of this paper is based on the visual control which realizes the robot tracking of marker and general targets. In the tracking of marker, the EPnP algorithm is used to solve the camera's position and orientation relative to marker so that the robot can track the marker. For general target tracking, the AdaBoost algorithm is used to recognize the target. Then we can get the relative position and realizing tracking through the RGBD information.

As shown in the experiment, within a certain range of space, the accuracy of EPnP algorithm for tracking markers is high. In the experiment of tracking general objects, the classifier trained by AdaBoost has strong anti-interference ability in target recognition. Moreover, within the appropriate distance range, the recognition rate is also guaranteed when the deformation of the object is acceptable.

## References

- [1] Sharp C S, Shakerna O, et al. A vision system for landing an unmanned aerial vehicle. IEEE international conference on robotics and automation, 2001, 2: 1720:1727

- [2] Paul Furgale, Joern Rehder, Roland Siegwart (2013). Unified Temporal and Spatial Calibration for Multi-Sensor Systems. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan
- [3] Paul Furgale, T D Barfoot, G Sibley (2012). Continuous-Time Batch Estimation Using Temporal Basis Functions. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 2088-2095, St. Paul, MN
- [4] J. Maye, P. Furgale, R. Siegwart (2013). Self-supervised Calibration for Robotic Systems, In Proc. of the IEEE Intelligent Vehicles Symposium (IVS)
- [5] L. Oth, P. Furgale, L. Kneip, R. Siegwart (2013). Rolling Shutter Camera Calibration, In Proc. of the IEEE Computer Vision and Pattern Recognition (CVPR)
- [6] Valenti, Roberto G., I. Dryanovski, and J. Xiao, Keeping a Good Attitude: A Quaternion-Based Orientation Filter for IMUs and MARGs, *Sensors* 15.8(2015), 19302-19330
- [7] Sharp C S, Shakerna O, et al. A vision system for landing an unmanned aerial vehicle. IEEE international conference on robotics and automation, 2001, 2: 1720-1727
- [8] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619
- [9] Lepetit V, Moreno-Noguer F, Fua P. EP n P: An Accurate O ( n ) Solution to the P n P Problem[J]. International Journal of Computer Vision, 2009, 81(2): 155-166
- [10] Yoav Freund, Robert E. Schapire. A Introduction to Boosting[J]. Journal of Japanese Society for Artificial Intelligence, 14(5): 771-780, September, 1999.
- [11] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1): 119-139, August 1997.
- [12] Peter Harrington. Machine Learning in Action[M]. Manning Publications, 2012.



# Mice Liver Cirrhosis Microscopic Image Analysis Using Gray Level Co-Occurrence Matrix and Support Vector Machines

Yu Wang<sup>1</sup>, Luying Cao<sup>1</sup>, Nilanjan Dey<sup>2</sup>, Amira S. Ashour<sup>3</sup>, and Fuqian Shi<sup>1,\*</sup>

<sup>1</sup>College of Information and Engineering, Wenzhou Medical University, Wenzhou 325035, People Republic of China

<sup>2</sup>Department of Information Technology, Techno India College of Technology, West Bengal, 740000, India

<sup>3</sup>Department of Electronics and Electrical Communications Engineering, Faculty of Engineering, Tanta University, Tanta 31111, Egypt

\*Corresponding author: sfq@wmu.edu.cn, Tel. +86-577-86689913

**Abstract.** Liver cirrhosis is a commonly chronic disease that often requires checking liver pathological microscopic images. To reduce the intensity of the work of doctors, pre-classification work needs to be issued and in this paper, related liver microscopic image classification analysis was proposed. Firstly, the extraction methods of mice liver images were taken using different features for different image features, and secondly, image features were obtained to form the original data-set; finally, the image classification experiments were conducted using gray level co-occurrence matrix (GLCM) and a developed support vector machine (SVM). The best classification model derived from the established characteristics is that GLCM performed the highest accuracy of classification; the training model using 11 features was accurately that only trained by 8 GLCMs. The experimental results preliminarily verified the feasibility of the method used in the experiment of classifying liver microscopic images of mice, and laid the foundation for further construction of computer aided diagnosis system.

**Keywords.** Liver cirrhosis, microscopic image, gray level co-occurrence matrix, support vector machine, computer aided diagnosis

## 1. Introduction

To date, the incidence of hepatocellular carcinoma (HCC) ranks sixth in all kinds of cancer, and the mortality rate ranks the third. The degree of malignancy of liver cancer is high and the rate of development of the disease is fast. When symptoms appear, the patient is usually in the middle or advanced stage. Therefore, the early diagnosis of liver disease is of vital importance for the treatment of liver disease. With the continuous development of artificial intelligence, information technology and image processing technology, computer aided diagnosis system is much more used in computer-aided diagnosis (CAD) in clinical diagnosis work. The intensity of work of clinicians is relatively large [1-3]. It is inevitable that a certain degree of missed

diagnosis and misdiagnosis is inevitable, and the application of CAD can help to solve the problem. At present, most of the field of liver assistance diagnosis is to use B-mode ultrasound or computer tomography (CT) and the scanning data can be the imaging basis for their diagnosis. Li used the scanning data obtained during liver CT image based on texture feature extraction using a variety of methods to train the artificial neural network classifier, in normal and tumor tissue, cyst and other lesions, three diagnostic sub problems of hepatocellular carcinoma and liver hemangioma, the diagnostic accuracy rate reached 0.9797 and 0.9851 respectively. And 0.9753 [4]. Deepalakshmo et al. also used gray level co-occurrence matrix, Lloyd's spectral feature measure and other methods to classify the liver ultrasound data of normal tissue, cyst, benign and malignant cancer, and obtained satisfactory results [5].

Cirrhosis (liver) is a common chronic liver disease. It may be caused by a variety of reasons, such as viral hepatitis and chronic alcoholism. The pathogenesis is progressive fibrosis of the liver. When the disease did not receive timely treatment, the course continues to progress, and eventually lead to liver hepatic lobule; its structure was destroyed and rebuilt by the blood circulation and the formation of liver cirrhosis, and then developed into a fatal disease caused by liver cancer. In this paper, the application of texture feature extraction gray level co-occurrence matrix -GLCM and support vector machine-SVM for image processing; the mice liver micro image and extract the texture features of microscopic images were addressed for two classification SVM, more accurately distinguish normal tissue image and image characteristics of liver cirrhosis laid the foundation for the realization of the next with CAD system.

## 2. Methods and Materials

### 2.1. Texture features extraction

Each image has its own features that distinguish it from other types of images, such as brightness, edges, and textures. Some of these features need to be obtained by transformation or processing, such as histograms and principal components. The feature extraction methods include statistical methods, spectral features and structural features. The corresponding methods include gray level co-occurrence matrix method and spectrum based analysis method [6] [7].

#### 1) Gray co-occurrence matrix method

The texture is formed by the gray distribution in the alternately changes space position, there will be a gray relation between two pixels so it has a certain relationship in spatial location in the image, known as the spatial correlation of image intensity; describe the texture using gray level spatial correlation is the ideological foundation the gray level co-occurrence matrix. GLCM is the classical and most commonly used method of texture statistical analysis. Let  $P_\delta$  be GLCM, gray level be  $L$ , then  $P_\delta$  is a  $L \times L$  matrix and element  $P_\delta(i, j)$ ,  $(i, j = 0, 1, 2, \dots, L-1)$  is defined as probability of space relation  $\delta = (D_x, D_y)$  and frequency of pixels  $i$  and  $j$ . Space position relations have four types in vertical, horizon and  $\pm 45^\circ$ ,  $P_\delta$  can be generated following the  $\delta$ .

$$P_\delta = \begin{bmatrix} P_\delta(0,0) & P_\delta(0,1) & \dots & P_\delta(0,j) & \dots & P_\delta(0,L-1) \\ P_\delta(1,0) & P_\delta(1,1) & \dots & P_\delta(1,j) & \dots & P_\delta(1,L-1) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P_\delta(i,0) & P_\delta(i,1) & \dots & P_\delta(i,j) & \dots & P_\delta(i,L-1) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P_\delta(L-1,0) & P_\delta(L-1,1) & \dots & P_\delta(L-1,j) & \dots & P_\delta(L-1,L-1) \end{bmatrix}$$

GLCM describes the two-order statistical feature of image texture and is suitable for many species; the advantages of GLCM is easily description on the spatial relationship between pixels in color, and it is an invariant of gray level transformation monotonous. Commonly used feature parameters by GLCM are texture, two-rank angle, contrast, correlation, entropy and so on.

### 2.2. Gabor wavelet transform

Supposed that  $f(x,y)$  is size of  $M \times N$ , 2-dimensional discrete transform of Gabor is,

$$I_{p,q}(x,y) = \sum_s \sum_t f(x-s,y-t) \varphi_{p,q}^*(s,t) \tag{1}$$

where,  $s$  and  $t$  are variables of size of filter mask,  $x$  and  $y$  are coordinates of pixels;  $p$  and  $q$  present scale and direction of wavelet,  $p=0,\dots,P-1$ , and  $q=0,\dots,Q-1$ .  $P$  and  $Q$  present all scales and directions of wavelet.  $\varphi_{p,q}^*$  denotes

complex conjugate of  $\varphi_{p,q}$  and  $\varphi_{p,q}$  also is a wavelet transform function of Gabor, that can be presented as

$$\varphi_{p,q}(x,y) = a^{-p} \varphi(x',y') \tag{2}$$

where,  $a > 1$  is a factor of scale. And we have that,

$$x' = a^{-p}(x \cos \theta + y \sin \theta), \quad y' = a^{-p}(-x \sin \theta + y \cos \theta) \tag{3}$$

where,  $\theta = q\pi / Q$ , if let Gabo function as mother wavelet, then it can be presented as,

$$\varphi_{p,q} = \left( \frac{1}{2\pi\sigma_x\sigma_y} \right) \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right) \exp(2\pi jwx) \tag{4}$$

Where,  $w$  presents bandwidth of Gabor,  $\sigma_x$  and  $\sigma_y$  present constants of Gaussian envelope following directions  $x$  and  $y$ .

According to the results of two-dimensional Gabor wavelet transform, the image means and standard variance, contrast and entropy can be calculated as texture features.

### 2.3. Support Vector Machine

#### 2.3.1. Recognition methods using SVM

Support vector machines (SVM) is a new generation learning algorithm developed based on statistical learning theory. For linearly separable training samples  $(x, y)$ , a classifier surface can be separated without error. The greater the spacing between categories, the better the generalization. The largest classification surface is the optimal classification surface. By applying the two-programming method, we can find that the optimal classification function is:

$$f(x) = \text{sgn}\left(\sum_{i=1}^n T_i y_i (x_i \cdot x) + b^*\right) \quad (5)$$

where,  $T_i$  is operator of Lagrange,  $b^*$  is threshold of classification.

For nonlinear un-separable, nonnegative relax term  $\alpha_i$  and  $C$  can be inducted to get optimal classification as same as to Eqn. (5).

For nonlinear problems, it can be transformed into a linear problem in high-dimensional space by kernel functions, and then the optimal classification plane is obtained in the transformation space, Eqn.5 can be transformed as

$$f(x) = \text{sgn}\left(\sum_{i=1}^n T_i y_i K(x_i, x_j) + b^*\right) \quad (6)$$

Kernel function include:

Linear:

$$K(x, y) = x \cdot y \quad (7)$$

Polynomial

$$K(x, y) = (x \cdot y + 1)^d, d = 1, 2, \dots \quad (8)$$

Radial basis function

$$K(x, y) = \exp(-r\|x-y\|^2) \quad (9)$$

Sigmoid

$$K(x, y) = \tanh(b(x \cdot y) - c) \quad (10)$$

#### 2.3.2. Parameters optimization of SVM

When using SVM to do the classification prediction, we need to adjust the related parameters (mainly the penalty parameter  $C$  and the kernel function parameter  $g$ ), and the ideal parameters need to be selected for getting better prediction accuracy. This study used cross validation (K-Cross Validation K-CV) theory, that is, the original data were divided into  $K$  group, and each data subset respectively a validation set, the  $K - 1$  group and the rest of the subset of data as the training set; it will get  $K$  models. Using  $K$  models, the average number of classification accuracy of the final set is used as the performance index of the classifier under this K-CV. This method can effectively avoid the occurrence of over learning and under learning, and finally achieve better accuracy for the prediction of the test set.

### 3. Results and Discussion

#### 3.1. Features selection

This study acquired 16 normal mice liver microscopic images and 16 cirrhosis images. The normal image is labeled as a positive sample, and the liver cirrhosis image is labeled as a negative sample. These images have been consulted by the pathologist and confirmed from the same mice. Fig. 1 shows part of those 32 images.

Mice liver cirrhosis has the key features of pseudolobule, and with the emergence of tissue fibrosis, the characterization of the image pixels to pixels can be found as uneven distribution and large amplitude of changes. Therefore, in the selection of feature dimensions, the emphasis is placed on the detection of spatial positions and directions. We use gray level co-occurrence matrix method and selected 8 feature parameters for all images. Similarly, 3 characteristic parameters are selected by Gabor wavelet transform. The characteristic dimensions of the selection are shown in Table 1.

#### 3.2. Classification result using SVM

In this study, all the feature dimensions are combined to form a  $32 \times 11$  data set, and then 10 images are randomly extracted from two sets of images, and a total of 20 images are formed. Then form the test set with the remaining 12 image data. Then, the original data are normalized. Finally, we establish a model of support vector machine, first use the training set to train the model, and then the test set prediction, the kernel function used in the model for polynomial kernel function. In parameter optimization, 2-fold cross validation, 5-fold cross validation, and test set optimization methods are adopted respectively, and the results are shown in Table 2.

From Table 2, the best classification model derived from the established features, we used GLCM method and obtained the highest accuracy of classification; the training model using 11 features of the model can be accurately, that only trained by 8 GLCMs to get the best gain the accuracy rate of classification. Although it can be noted that the accuracy of classification has reached better results, there is a great deal of contingency in the process of random data extraction, and the classification accuracy is very volatile. This shows that the construction of experimental model is not stable enough, the reason is the amount of data is too small and secondly, the feature extraction of images is not rich enough that resulted not enough in the available resolution feature representation

### 4. Conclusion

The experiments in this paper, the mice liver microscopic images were processed using different methods of feature extraction for multidimensional image features; image classification experiments were conducted using SVM and GLCM texture based features. We got the ideal results and verified the feasibility of the methods proposed in this paper, then lay the foundation for the application of computer aided diagnosis system in microscopic image of liver pathological tissue of mice. In future work, the next step is to extract more features of the resolution of the significance; in addition, much more training data need to be collected.

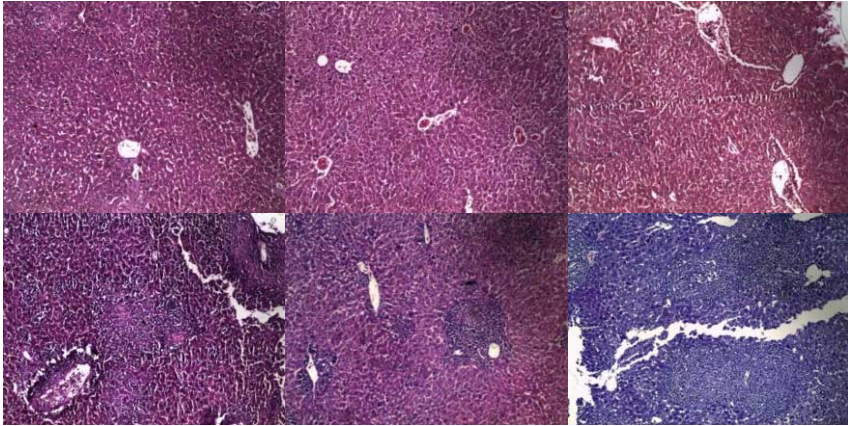


Figure 1. part mice liver microscopic images

Table 1 Features selection for GLCM and Gabor wavelet

GLCM	Gabor Wavelet Transform
2-rank moment	Binarization uniformity
Inertia moment	Binarization variance
Entropy	Binarization entropy
Correlation	
Uniformity	
Variance	
Non-similarity	
Inverse difference moment	

Table 2 The parameters of the best classification accuracy in the experiment of image classification

Features	Methods	Penalty function -c	Kernel-g	Accuracy (%)
3	2-fold cross validation.	2.2974	4.0000	60.0
	5-fold cross validation.	0.2500	1.8661	60.0
	test set optimization	1.2311	1.8661	66.7
8	2-fold cross validation.	0.2500	2.4623	90.0
	5-fold cross validation.	0.2500	0.2500	85.0
	test set optimization	0.2500	3.2490	91.7
11	2-fold cross validation.	0.2500	0.2500	80.0
	5-fold cross validation.	0.2500	0.2500	80.0
	test set optimization	0.2500	0.2500	91.7

References

[1] Dey, N., Ashour, A. S., Ashour, A. S., & Singh, A., Digital analysis of microscopic images in medicine. Journal of Advanced Microscopy Research, 10, 1-13, 2015.

[2] Amira S. Ashour, Samsad Beagum, Nilanjan DeyAhmed S. Ashour, Dimitra Sifaki Pistolla, Gia Nhu Nguyen, Dac-Nhuong Le, Fuqian Shi, Light microscopy image de-noising using optimized LPA-ICI filter, Neural Comput & Applic. doi:10.1007/s00521-016-2678-9

- [3] Sarwar Kamal, Nilanjan Dey, Sonia Farhana Nimmy, Shamim H. Ripon, Nawab Yousuf Ali, Amira S. Ashour, Wahiba Ben Abdesslem Karaa, Gia Nhu Nguyen, Fuqian Shi, Evolutionary framework for coding area selection from cancer data, *Neural Comput & Applic.* DOI 10.1007/s00521-016-2513-3
- [4] Li Wei, Design and implementation of CAD system based on multiphase liver images, PhD Thesis, Shanghai Jiaotong University, 2010
- [5] Sergios Theodoridis, Konstantinos Koutroumbas, *Pattern Recognition*, Publishing House of Electronics Industry, 2000.
- [6] Fuqian Shi, Shape Preserving Fitting Model for Affective Curves Extraction: An Affective Computing Method on fMRI Dataset, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(5): 117-128, 2013.
- [7] Fangfang. Chen, Xingfang Jiang, and Zhongyi Jiang, The Research of Recognition on Oceanic Internal Waves Based on Gray Gradient Co-occurrence Matrix and BP Neural Network, in *Proc. Photonics and Optoelectronics (SOPO)*, 2011, pp. 1-4

# The Overview of Chinese Cooperative Intelligent Transportation System Vehicular Communication Application Layer Specification and Data Exchange Standard<sup>1</sup>

Yijia Feng<sup>a</sup>, Dazhi He<sup>a</sup>, Lei Niu<sup>b</sup>, Ming Yang<sup>c</sup>, and Yunfeng Guan<sup>a</sup>

<sup>a</sup>Shanghai Jiao Tong University, China

<sup>b</sup>Chongqing Changan Automobile Co., Ltd, China

<sup>c</sup>Neusoft Corporation, China

**Abstract.** Vehicle-to-everything (V2X) communication is considered worldwide to be an appropriate technology to enable a large number of the intelligent transportation system (ITS) applications for traffic safety and efficiency. However, since there is no completed underlying V2X communication standard in China, it is difficult for Chinese automotive enterprises and research institutions to develop ITS applications in large scale. Therefore, the Society of Automotive Engineers of China (C-SAE) together with China ITS Industry Alliance (C-ITS) decided to develop Application Layer Specification and Data Exchange Standard first to motivate the development of the Chinese cooperative intelligent transport system industry. This paper introduces the details of the Chinese Cooperative Intelligent Transportation System Vehicular Communication Application Layer Specification and Data Exchange Standard. Included in the discussion are the description of the application scenarios and their technique requirements, the summary of the message set in the data exchange standard, and the design of the application layer data interactive interfaces. It is noteworthy that the message set in this standard is complete for the included application scenarios and the designed application layer interfaces are compatible with the underlying V2X communication protocols like DSRC and LTE-V. Finally, several directions worth further consideration are suggested for future standardization.

**Keywords.** V2X, cooperative intelligent transport system, application layer specification, data exchange standard

## 1. Introduction

In recent years, with the enormous economic development and urban expansion, the Chinese vehicle population is increasing rapidly. However, the road accidents and congestion, as well as the air pollution caused by the traffic, have become severe social problems for modern life [1]. In China, thousands of people died from car crashes

---

<sup>1</sup>This paper is supported in part by National Natural Science Foundation of China (61420106008, 61671297 and 61601286), National High Technology Research and Development Program (2015AA015802), National Science and Technology Major Project (2017ZX03001027).

The authors: Yijia Feng, Dazhi He, and Yunfeng Guan are from Shanghai Jiao Tong University. Lei Niu is from Chongqing Changan Automobile Co., Ltd. Ming Yang is from Neusoft Corporation. The corresponding author is Dazhi He (e-mail: hedazhi@sjtu.edu.cn).



every year, and the traffic congestion in metropolis causes tremendous economic loss every day. Therefore, it is urgent to motivate Chinese intelligent transportation industry to develop new technology and set up specific standards to mitigate the aforementioned situations.

Vehicle-to-everything (V2X) communication is considered worldwide to be an appropriate technology to solve the above problems. It means in the vehicle network, a vehicle can be connected to everything, including other vehicles (V2V), the roadside infrastructures (V2I), pedestrians (V2P), and the network (V2N) [2]. Hence, vehicles are no longer separated elements driving on the road, and they can sense the road traffic around themselves and exchange information data with others. Thus, V2X technology is able to enhance the traffic safety and efficiency by raising a variety of hazard warnings and road traffic notifications to the drivers.

There are two mainstream V2X communication technologies in the world at present: one is DSRC, the other is LTE-V.

Dedicated short-range communication (DSRC) standards are designed to support V2X technology in the United States and other countries. In 1999, the Federal Communications Commission (FCC) in the U.S. allocated 75MHz (from 5.850GHz to 5.925GHz) bandwidth for DSRC in vehicle environment. The U.S. Department of Transportation (DOT) has estimated that vehicle-to-vehicle (V2V) communication based on DSRC can address up to 82% of all crashes in the U.S., potentially saving thousands of lives and billions of dollars[3]. In December 2016, the U.S. DOT issued a proposed rule that would introduce V2V communication technology into all new light-duty vehicles in the U.S., which would prevent "hundreds of thousands of crashes every year"[4]. Furthermore, European Union, Japan and some other countries all launched their own V2X communication standards and started a lot of related academic research and implementation work[5][6].

LTE-V, short for LTE-V2X, is an alternative technology for V2X. LTE-V is proposed in 3GPP mainly by CATT, Huawei, and Qualcomm. It is based on the cellular network architecture and can reuse the existing cellular infrastructures and spectrum. Additionally, it supports operations outside of the infrastructure network coverage with technologies such as enhanced device-to-device (D2D) communications[7][8]. Chinese communication associations and enterprises are active to promote this standard's academic research and industrial realization.

In China, since it has not been decided yet whether the V2X lower layer communication standards adopt DSRC or LTE-V, a large number of obstacles hinder the development of Chinese intelligent transportation system (ITS) industry. Without the definition of the basic applications, it is difficult to develop the standards focusing on vehicular communication and thus it is not able to support the large scale V2X application implementation. Therefore, the Society of Automotive Engineers of China (C-SAE) together with China ITS Industry Alliance (C-ITS) decided to develop Application Layer Specification and Data Exchange Standard first to motivate the development of the Chinese cooperative intelligent transport system industry.

This paper introduces the details of the Chinese Cooperative Intelligent Transportation System Vehicular Communication Application Layer Specification and Data Exchange Standard. The remainder of the paper is organized as follows: Section II describes the compilation process of the standard. Section III elaborates the scope of the standard. Then, Section IV describes the application scenarios in the system and analyzes the technique requirements with specified classifications. And Section V introduces the message set supporting the application requirements. Furthermore, in

Section VI, the application layer data interactive interfaces are raised for flexible application development and compatibility with the lower layer communication protocols. Finally, Section VII summarizes the paper and proposes several new directions for further standardization.

## **2. The Compilation Process of the Standard**

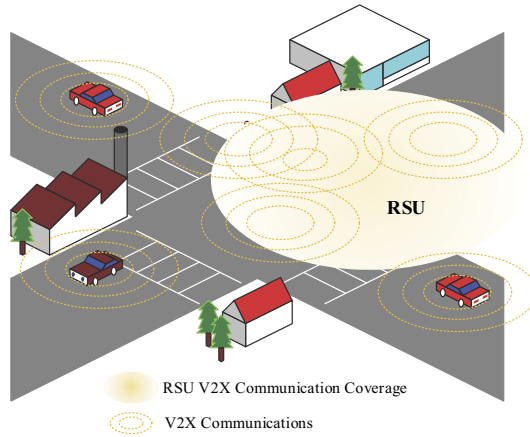
This standard is called Cooperative Intelligent Transportation System Vehicular Communication Application Layer Specification and Data Exchange Standard. It specifies the terminologies, definitions, application scenarios, data sets, and data-exchange interfaces in the cooperative intelligent transportation system, aiming at supporting application scenarios with vehicular communication of any type.

Launched by C-SAE and C-ITS, the standard committee involves automobile industry, communication enterprises, equipment provider, universities, research institutions, and etc. The standard's chief sponsors are Changan Automobile, GM China, and Tsinghua University. Additionally, 43 organization units including SAICMOTOR, JAC MOTORS, Neosoft, PATEO, Huawei, and Shanghai Jiao Tong University, participate in the standard compilation session. The standard committee opens its membership application to all the technicians of any company or institution who are concerned about this standard. The committee members have their rights to draft the standard, to participate in the technical seminar, to vote on the candidate standard, to share other standard information, and etc.

The compilation process of this standard started in Feb. 2016 and ended in Apr. 2017. At the beginning of the compilation process, the committee collected 40 basic application scenarios, then voted 17 typical application scenarios for compilation. Afterwards, from Jul. 2016 to Dec. 2016, the committee had a seminar every month to discuss the application scenarios and their technique requirements, the data structure of the message set, the application layer data interfaces, and etc. Especially, in Oct. 2016, the Changan Automobile, GM China, Tsinghua University, and Yanfeng Visteon successfully completed the vehicle-interconnection tests based on this standard in Shanghai. During Jan. 2017 to Apr. 2017, the standard was finalized with the first formal edition. Finally, the standard will be officially enacted in Jun. 2017.

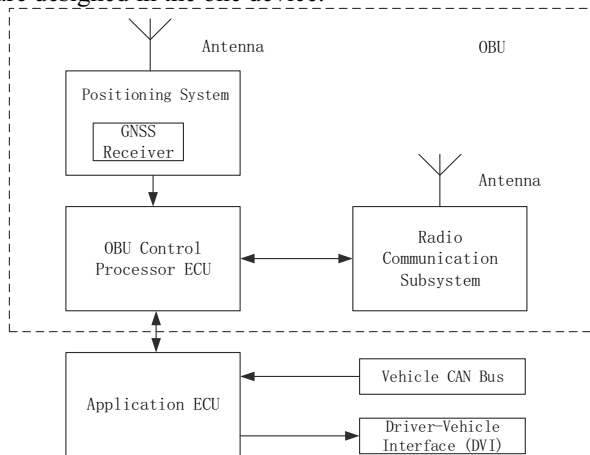
## **3. The Scope of the Standard**

This standard focuses on cooperative intelligent transportation system, composed of vehicles, roadside infrastructures, pedestrians, and etc. Traffic information can be exchanged by radio among these elements, which is the so-called V2X communication. V2X communication system is built over two kinds of equipment: On-Board Units (OBUs) and Road Side Units (RSUs). Vehicles in this system are supposed to be equipped with OBUs to communicate with others in the V2X network. RSUs, which are stationary units distributed along the road, communicate with the roaming vehicles usually by broadcasting road safety warnings and traffic efficiency information periodically. Both OBU and RSU have V2X communication ability. The typical cooperative intelligent transportation system is shown in Figure 1. Vehicles with OBUs can exchange traffic information with other vehicles and pedestrians, while they can also receive road information like the traffic signal phase and timing messages from RSUs at the same time via V2X communication.



**Figure 1.** An illustration of a V2X scenario.

For the sake of the realization of the cooperative intelligent transportation system, this standard requires that the vehicles in the system should acquire V2X communication ability, namely, vehicles should be equipped with OBUs. As is shown in Figure 2, a most typical OBU consists of a radio communication subsystem, a positioning system with a Global Navigation Satellite System (GNSS) receiver, an OBU control processor Electronic Control Unit (ECU), and several antennas for communication. Besides, the OBU is connected to the application ECU to manage the communication system applications running on the OBU. Moreover, connected to the application ECU, the Driver-Vehicle Interface (DVI) is designed to warn drivers about the danger with picture, sound, or vibration. In some cases, the OBU and the application ECU are designed in the one device.

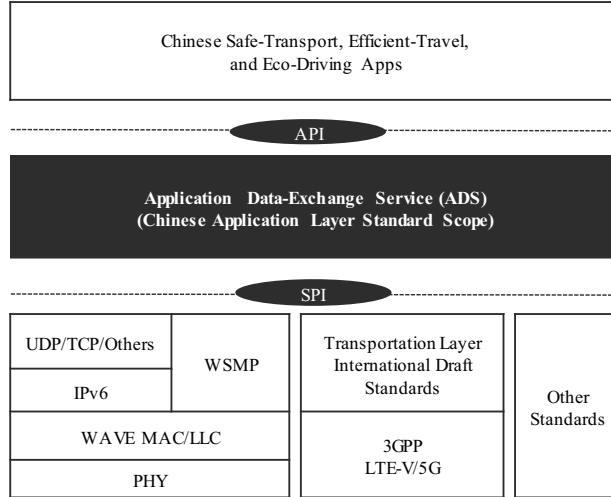


**Figure 2.** A typical OBU structure.

According to the V2X standards developed in the U.S. [3] and European Union[3][5], vehicular communication standard stack can be divided into application layer, transportation layer, network layer, datalink layer, and physical layer. Since the underlying V2X communication standards in China have not been completed yet, there are a large number of difficulties for Chinese automotive enterprises and research institutions to develop the ITS applications in large scale. Thus, C-SAE and C-ITS

decided to develop Application Layer Specification and Data Exchange Standard first to stimulate the interconnection between the devices of different manufacturers in Chinese vehicular communication network.

Referred to the V2X application layer standard in the U.S. and European Union [9][10][11], this standard focuses on the application layer and its interfaces with the neighbor layers, shown in the shaded area in Figure 3.



**Figure 3.** The scope of Chinese Application Layer Standard.

The application layer standard designs a message set for application layer data-exchange services, which is conducive for the interconnection between the devices of different manufacturers. Furthermore, it is determined that this standard is independent of the underlying communication technology (eg. DSRC, LTE-V, and 5G). Therefore, for the convenience of the developer, the standard defines specified Application Programming Interfaces (APIs) to develop applications of interconnection and interoperability without concerning about the lower layer communication protocols. Meanwhile, the standard designs Service Provider Interfaces (SPIs) for different devices to accommodate varied lower layer communication protocols and the updating communication requirements.

Table 1 summarizes the important abbreviation mentioned in the paper.

**Table 1.** The summary of the important abbreviations in this paper.

Abbreviation	Definition
API	Application Programming Interface
ASN.1	Abstract Syntax Notation One
BSM	Basic Safety Message
C-ITS	China ITS Industry Alliance
C-SAE	Society of Automotive Engineers of China
ECU	Electronic Control Unit
DME	DSRC Management Entity
DSM	DSRC Short Message
DSRC	Dedicated Short Range Communications
GB	Guo Biao (Nation Standard)
GNSS	Global Navigation Satellite System
LTE-V2X	Long Term Evolution-Vehicle to Everything
NHTSA	National Highway Traffic Safety Administration
OBU	On-Board Unit
RSM	Road Side Message

RSU	Road Side Unit
SAE	Society of Automotive Engineers
SPAT	Signal Phase and Timing Message
SPI	Service Provider Interface
UPER	Unaligned Packet Encoding Rules
V2I	Vehicle to Infrastructure
V2P	Vehicle to Pedestrians
V2V	Vehicle to Vehicle
V2X	Vehicle to Everything

#### 4. Application Scenarios and Requirements of the Standard

In order to obtain the application layer specification and data exchange standard, application scenarios and requirements should be carefully analyzed. A clear description of the application scenario helps to raise technique requirements like message frequency, maximum latency, and location accuracy for lower layers in the cooperative intelligent transportation system. Furthermore, the work in this section also helps to summarize the message set and derive APIs. Only in this way, can all the manufacturers develop the applications of interconnection without any barrier.

As mentioned above, in compilation session, all the inclusive application scenarios in this standard were solicited and voted internally. In the voting stage, the Standard Committee first collected 40 typical application scenarios, including 19 for driving safety, 12 for traffic efficiency, and 9 for information services. Then each committee member voted 15 application scenarios according to the technique maturity, application value and current technology feasibility. Finally, out of 40 collected basic application scenarios, the standard committee elected 17 typical application scenarios for further discussion.

The ultimate 17 V2X application scenarios can be broadly categorized into safety, efficiency, and information applications as described in Table 2.

**Table 2.** The Classification of the 17 V2X Application Scenarios.

Application Type	Communication Type	Application Description
Driving Safety Application	V2V	FCW: Forward Collision Warning
	V2V/V2I	ICW: Intersection Collision Warning
	V2V/V2I	TLA: Left Turn Assist
	V2V	BSW/LCW: Blind Spot Warning/Lane Change Warning
	V2V	DNPW: Do Not Pass Warning
	V2V-Event	EBW: Emergency Brake Warning
	V2V-Event	AVW: Abnormal Vehicle Warning
	V2V-Event	CLW: Control Lost Warning
	V2I	HLN: Hazardous Location Warning
	V2I	SLW: Speed Limit Warning
	V2I	SVW: Signal Violation Warning
Traffic Efficiency Application	V2P/V2I	VRUCW: Vulnerable Road User Collision Warning
	V2I	TLOSA: Traffic Light Optimal Speed Advisory
	V2I	TSC: Traffic Sign In Car
	V2I	TJW: Traffic Jam Warning
Information Service Application	V2I/V2V	EVP: Emergency Vehicle Priority
	V2I	VNFP: Vehicle Near-Field Payment

It is obvious that this standard edition mainly focuses on driving safety scenarios. All these application scenarios cover V2V, V2I and V2P communication methods,

where V2I and V2P have no appointed communication direction. Vehicles in these scenarios are supposed to possess V2X communication ability. As for exceptions, the RSUs will sense them and transmit their motion state to other vehicles on road with OBUs.

This standard provides a thorough description of each application scenario including its basic system theory, communication performance requirements and data interaction requirements. Table 3 summarizes the specifications of main application performance requirements with different classifications.

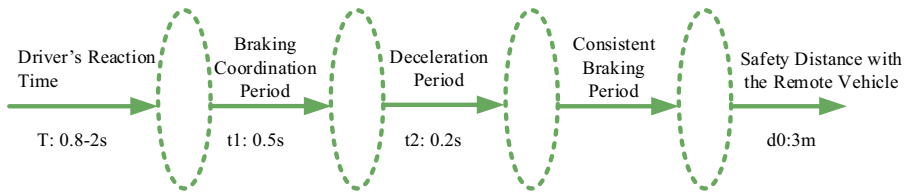
**Table 3.** Main Application Performance Requirements with Classifications.

Application Performance		Classification	Requirements
Maximum Velocity		Ordinary Road	70km/h
		Highway	130km/h
Maximum Communication Range	Communication	Bidirectional Applications	300m
		Unidirectional Applications	150m
Minimum Message Frequency		Driving Safety Applications	10Hz
		Other Applications	1Hz
Maximum System Latency		Driving Safety Applications	100ms
		Other Applications	500ms
Location Accuracy		Applications with Lane Positioning	1.5m
		Other Applications	5m

There are some notes of the application performance requirements:

- i) The specifications of the minimum SAE message frequency and maximum system latency time are referred to the SAE J2735[9], SAE J2945/1[10] and NHTSA VSC-A[12].
- ii) The information transmission time end-to-end delay refers to the delay time between the transmitter device’s application layer to the receiver device’s application layer.
- iii) The reason why the maximum communication range should be more than 300m is explained in the model of the minimum safety distance between vehicles:

Supposing a host vehicle detects a potential crash danger with a remote vehicle, the host vehicle’s braking session is shown in Figure 4.



**Figure 4.** The host vehicle’s braking session.

In this session,  $v_s$  denotes the velocity of the host vehicle,  $v_f$  denotes the velocity of the remote vehicle,  $T$  denotes the driver’s reaction time,  $t_1$  denotes braking-time delay,  $t_2$  denotes decelerate time,  $a_s$  denotes the vehicle’s braking deceleration, and  $d_0$  denotes the safety distance after braking. According to the vehicle braking dynamics theory, the minimum safety distance model can be derived as below.

$$S = (v_s - v_f)(T + t_1 + \frac{1}{2}t_2) + \frac{(v_s - v_f)^2}{2a_s} + d_0 \tag{1}$$

In this model, the driver’s reaction time  $T$  is always between 0.8 to 2s, braking-time delay  $t_1$  is set as 0.5s, decelerate time  $t_2$  is set as 0.2s, and the safety distance after braking  $d_0$  is set as 3m.

For example, suppose two vehicles’ velocity difference is 100km/h, the driver’s reaction time is 2s, and the vehicle’s braking deceleration is  $3.6 m/s^2$ . It can be inferred that the minimum safety distance is 185m according to the above model. In this example, it is under the circumstance when the velocity difference between the two vehicles is large and the driver’s reaction is slow. Hence, 300m communication distance which is much larger than 185m can meet the extreme requirements of collision avoidance.

- iv) The reasons why the location accuracy of applications with lane positioning should be less than 1.5m are explained as follows:

Chinese lane width is 3.5m. Supposing the host vehicle drives in the middle of the lane, it can be indicated that the remote vehicle and host vehicle are in the same lane if the remote vehicle drives in the transverse direction is within 1.75m of the host vehicle. Therefore, for the applications with lane positioning, the location accuracy should be less than 1.5m on the safe side.

### 5. Message Set

For the sake of the V2X application development and large scale tests among different manufacturers, a unified data exchange standard should be established for the mass implementation of the V2X communication technology. Thus, the message set is proposed as the core part of this standard.

Considering the aforementioned 17 V2X application scenarios’ requirements, the proposed message set should meet the following principles:

- i) The information in the message set should support all the applications covered in this standard.
- ii) The information in the message set should meet the requirements of the V2X applications in this standard.
- iii) The organization and packaging way in the message set should conform to the internal processing logic of the data source.
- iv) The message set should have expandability to be compatible with other mature application layer standards like SAE J2735[9]

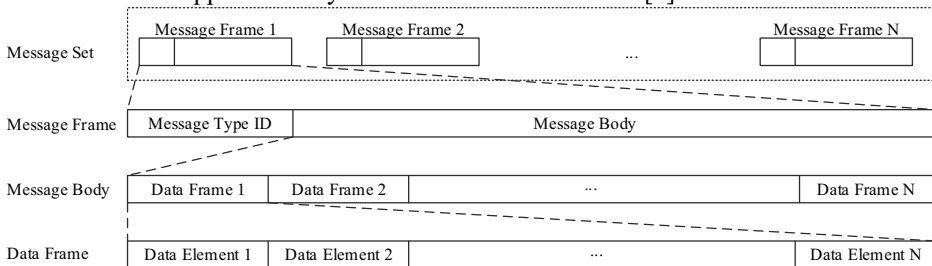


Figure 5. The logic of message design.

The logic of message design is followed as "Message Set -- Data Frame -- Data Element", as is shown in Figure 5. Message set consists of message frames of different types. Message frame is composed of a message type ID and a message body which is made up of several data frames. Moreover, data frame can be further divided into data elements which are the basic data structures and have exact physical meanings.

There are five types of message frames describe as follows.

i) BSM

BSM is the abbreviation for Basic Safety Message, which is applied most widely in the application layer. BSM is based on the V2X requirements, the vehicle's state and its sensing information. The message body includes the identification information, location and moving information, inside state information, and some extension information. BSM is used for exchanging traffic safety messages between vehicles and it supports series of the applications for traffic safety. It is usually broadcasted 10Hz periodically. Moreover, some BSMs are optimized through the standard's application field test.

ii) MAP

MAP is broadcasted by the RSUs. Passing the local map information to the nearby vehicles, MAP includes the intersection information, road information, lane information, the traffic sign information, and the connection information between roads. The MAP data structure is designed as "node - road connection - lane", while in addition there are some special features like steering information for supplementary. A single MAP message can contain more than one intersection information. The signal time information is included in SPAT and will be described below.

iii) RSI

RSI is the abbreviation for RoadSide Information. RSI is broadcasted to the nearby vehicles by RSUs. It contains traffic sign information and traffic incident messages. Traffic sign information is a notification or warning written on the roadside sign. Traffic incident messages can be announced in text, and it focuses on the dynamic and temporary traffic incidents like "Accident Ahead Warning" or "Ice Ahead Warning". When an OBU receives a RSI, it will judge if it is in its effective zone according to its own location and driving direction. The specific roadside information rules refer to *The lane, marking and traffic sign were all designed according to the Road Traffic Sign and Marking Standard of China (GB 5768-2009)* [13].

iv) RSM

RSM is the abbreviation for RoadSide Message, which is gathered by RSUs. After detecting the real time traffic participants' condition nearby, RSUs pack up the information into RSMs, then usually broadcast 1Hz periodically to the vehicles in neighbor.

The RSM is the most particular kind of message body. It is a supplementary of BSM when some traffic participants are not equipped with V2X devices. It is widely considered that RSMs will contribute to the population of V2X technology in China.

v) SPAT

SPAT is the abbreviation for Signal Phases And Time, which contains the traffic signals in one or more intersections. The SPAT data structure is designed as "traffic light - phase - color" to describe the moment's traffic light information. Coordinated with MAP, the real time and phase of the frontage traffic light can be sent to the vehicles.

The message set is all described in Abstract Syntax Notation Revision One (ASN.1) schema syntax for cross-platform operation and different-languages



compatibility. Unaligned Packet Encoding Rules (UPER) are adopted in dataset interaction encoding and decoding.

## 6. The Application Layer Data Interactive Standard and Interface Specification

With the new emerging communication technologies, this standard is designed to be independent of the applications on the top and the communication protocols at lower layers including transportation layer, network layer, datalink layer, and physical layer. Therefore, this standard specifies application layer data interactive interfaces, including Application Programming Interface (API) and Service Provider Interface (SPI) to accommodate other layers.

API is an interface for application development. As is shown in Figure 3, API is an interface to exchange data with the upper layer ITS APPs. With APIs, the developer won't have to concern about the underlying communication methods, devices, or any driver problems. The APIs can be mainly divided into 5 categories: host information, communication operation, data image service, data application service, and management service. The description of each API classification is shown in Table 4.

**Table 4.** The description of API classifications.

API Classification	Description
Host Information	Provide interfaces to get or set the host's status.
Communication Operation	Provide interfaces for the DSM communication operations.
Data Duplicate Service	Provide interfaces for the dup of BSM, MAP, SPAT, and RSM.
Data Application Service	Provide interfaces for the application service data.
Management Service	Provide interfaces for management service.

SPI is an interface for communication devices. As is shown in Figure 3, SPI is an interface to communicate with the lower layers. With SPIs, this standard can accommodate to communication devices which are based on varied different lower layer communication protocols, and thus the standard can meet the growing requirements of new communication technologies. The SPI specifications include communication configuration, message transmission and reception, management property configuration, all kinds of service requirement, and etc. The specified SPIs and the descriptions are shown in Table 5.

**Table 5.** The description of the specific SPIs.

SPI Name	Description
CommClientInit.request	Request for the initialization operation transmission
CommClientInit.confirm	Confirmation for the initialization operation transmission
CommDSMSend.request	Request for the DSM transmission
CommDSMSend.confirm	Confirmation for the DSM transmission
CommDSM.indication	Notification for the DSM reception
CommDMESetCfg.request	Request for the DME setting
CommDMESetCfg.confirm	Confirmation for the DME property setting
CommDMEGetCfg.request	Request for the DME property acquirement
CommDMEGetCfg.confirm	Confirmation for the DME property acquirement
CommDMEProviderService.request	Request for the DME provider service operation
CommDMEProviderService.confirm	Confirmation for the DME provider service operation
CommDMEUserService.request	Request for the DME user service operation
CommDMEUserService.confirm	Confirmation for the DME user service operation
CommDMENotification.indication	Indication of the DME reception notification

In this edition, the standard mainly focuses on the definition and description of the API and SPI's primitives, the specific parameter specifications and configurations will be illustrated in the further editions.

## 7. Summary and Future Directions

This paper provides a detailed description of Chinese Cooperative Intelligent Transportation System Vehicular Communication Application Layer Specification and Data Exchange Standard from several aspects, including the motivation background, compilation process, research scope, application scenario descriptions and requirements, the definition of the message set, and the interface specifications. Nevertheless, there are still several directions worth further considerations in the future standardization process. First, more application scenarios should be covered in the standard for a more complete cooperative intelligent transportation system. Then, more tests should be done to optimize the message set in practical use. Additionally, the specific parameter specifications and configurations of the application layer data interactive interfaces need to be illustrated in future standards.

## References

- [1] Zheng K, Zheng Q, Chatzimisios P, et al. Heterogeneous vehicular networking: A survey on architecture, challenges, and solutions[J]. *IEEE Communications Surveys & Tutorials*, 2015, 17(4): 2377-2396.
- [2] Abboud K, Omar H A, Zhuang W. Interworking of DSRC and Cellular Network Technologies for V2X Communications: A Survey[J]. *IEEE Transactions on Vehicular Technology*, 2016, 65(12): 9457-9470.
- [3] Kenney J B. Dedicated short-range communications (DSRC) standards in the United States[J]. *Proceedings of the IEEE*, 2011, 99(7): 1162-1182.
- [4] NHTSA.U.S. DOT advances deployment of Connected Vehicle Technology to prevent hundreds of thousands of crashes [EB/OL]. <https://www.nhtsa.gov/press-releases/us-dot-advances-deployment-connected-vehicle-technology-prevent-hundreds-thousands>, 2016-12-13/2017-4-20.
- [5] Festag A. Cooperative intelligent transport systems standards in Europe[J]. *IEEE communications magazine*, 2014, 52(12): 166-172.
- [6] Papadimitratos P, De La Fortelle A, Evenssen K, et al. Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation[J]. *IEEE Communications Magazine*, 2009, 47(11): 84-95.
- [7] Seo H, Lee K D, Yasukawa S, et al. Lte evolution for vehicle-to-everything services[J]. *IEEE Communications Magazine*, 2016, 54(6): 22-28.
- [8] Lee J, Kim Y, Kwak Y, et al. LTE-advanced in 3GPP Rel-13/14: an evolution toward 5G[J]. *IEEE Communications Magazine*, 2016, 54(3): 36-42.
- [9] Dedicated Short Range Communications (DSRC) Message Set Dictionary, SAE Std. J2735, SAE Int., DSRC Committee, Nov. 2009.
- [10] On-Board System Requirements for V2V Safety Communications, SAE Std. J2945/1, SAE Int., DSRC Committee, Mar 2016.
- [11] ETSI T. Intelligent transport systems (ITS); vehicular communications; basic set of applications [R]. Tech. Rep. ETSI TR 102 637, 2010.
- [12] Ahmed-Zaid F, Bai F, Bai S, et al. Vehicle safety communications–applications (VSC-A) final report: Appendix volume 1 system design and objective test[R]. 2011.
- [13] The lane, marking and traffic sign were all designed according to the Road Traffic Sign and Marking Standard of China (GB 5768-2009)

# Algorithm Design of GNSS/INS Integrated Navigation for Vehicle Location in Cities

Han Wang<sup>a,1</sup>, Maosong Wang<sup>a</sup>, Kun Wen<sup>a</sup>

<sup>a</sup>College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha, China

**Abstract.** With the development of modern urban intelligent transportation system, the requirement of accuracy and reliability of autonomous navigation becomes higher and higher. GNSS/INS is the most widely used method of integrated navigation at present, but in cities, satellite signals tend to be blocked by tall buildings, resulting in fewer than four visible satellite. This paper presents loosely-coupled and tightly-coupled GNSS/INS integration models using EKF and UKF algorithm respectively. The system simulation platform is built, and the dilution of precision (DOP) is adopted to help the satellite selection which ensures the effectiveness of integrated navigation. The experimental results show that when the satellite signal is more than four, both the integration models have high navigation accuracy. When the satellite signal is less than four, using the tightly-coupled model instead of the invalid loosely-coupled model with the observable satellite information can ensure the navigation precision and improve the anti-interference ability of the system.

**Keywords.** loosely-coupled, tightly-coupled, EKF, UKF, DOP

## 1. Introduction

GNSS/INS integrated navigation system is widely used in military, transportation and other fields. The integration system uses satellite information to correct the accumulated inertial navigation errors and it has the characteristics of high precision, all-weather and low cost. However, GNSS is prone to be affected by satellite clock error, ionosphere, tropospheric delay and multipath effect which leads to larger measurement errors. When the vehicle is running in the city, the satellite signals are often blocked by buildings and trees, so that the number of received satellite signal is less than four and the loosely-coupled method is invalid. The tightly-coupled navigation method with pseudo range and pseudo range ratio can make full use of GPS resources, even if one satellite signal is received, the accuracy and reliability of the system are improved [1]. The tightly-coupled navigation includes the variation of GNSS navigation parameter covariance caused by satellite geometry distribution and availability, and increases the statistical objectivity of INS errors [2].

At present, the most commonly used nonlinear filtering algorithms for integrated navigation systems are EKF and UKF. EKF uses Taylor series expansion to linearize

---

<sup>1</sup> Corresponding Author, College of Mechatronics Engineering and Automation, National University of Defense Technology, No.109, Deya Road, Kaifu District, Changsha, Hunan Province, China; E-mail: 603389841@qq.com.

the nonlinear system, and overcomes the problem that the standard Kalman filter can't be applied to nonlinear systems [3]. UKF does not require to linearize the system by calculating the Jacobian matrix. The recursive Bayesian estimation using Unscented Transformation (UT) makes the nonlinear system equation applicable to the standard Kalman filtering process under linear condition, and has higher estimation accuracy and stability than EKF.

## 2. Integrated Navigation Algorithm Design

The integrated navigation algorithm showed in Figure 1 is designed according to the requirement of the actual navigation system. Using the output of gyros and accelerometers in the INS to calculate the velocity and position of the receiver, the pseudo range and pseudo range rate are calculated by the measurements of GNSS and INS. When the number of visible satellites is more than four, select satellites by DOP and input the difference of the position and speed information calculated by the GNSS receiver and the inertial navigation system to the EKF, the pseudo range and pseudo range rate difference of the GNSS and INS are used to Unscented Kalman filtering. When the number of visible satellites is less than four, only the UKF for tightly-coupled GNSS/INS integration can be used to calculate the navigation parameter.

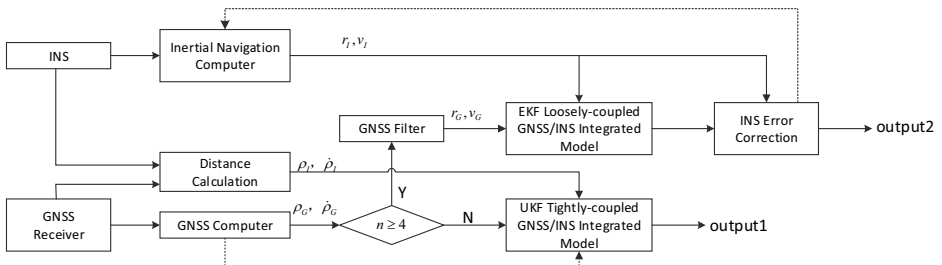


Figure 1. Algorithm Design of GNSS/INS integrated navigation

## 3. GNSS/INS Integrated Navigation Model

### 3.1. Loosely-coupled GNSS/INS Model

The inertial navigation system is the main system of the loosely-coupled GNSS/INS integration model, and the error of inertial navigation system is corrected by the position and velocity information provided by GNSS to restrain the accumulation of errors over time. In NED (north, east, down) navigation frame, the error state model of INS is showed as follow [4]:

$$\left\{ \begin{aligned} \dot{\boldsymbol{\varphi}} &= -\boldsymbol{\omega}_{in}^n \times \boldsymbol{\varphi} + \delta\boldsymbol{\omega}_{in}^n - \mathbf{C}_b^n \cdot \delta\mathbf{f}_{ib}^b \\ \delta\dot{\mathbf{v}}_e^n &= \mathbf{C}_b^n \mathbf{f}^b \times \boldsymbol{\varphi} - (2 \cdot \boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n) \times \delta\mathbf{v}_e^n - (2 \cdot \delta\boldsymbol{\omega}_{ie}^n + \delta\boldsymbol{\omega}_{en}^n) \times \mathbf{v}_e^n + \mathbf{C}_b^n \cdot \delta\mathbf{f}^b + \delta\mathbf{g}^n \\ \delta\dot{L} &= \frac{\delta v_N}{R_N + h} - \frac{v_N}{(R_N + h)^2} \delta h \\ \delta\dot{\lambda} &= \frac{\delta v_E}{(R_E + h) \cos L} - \frac{v_E \sin L}{(R_E + h) \cos^2 L} \delta L - \frac{v_E}{(R_E + h)^2 \cos L} \delta h \\ \delta\dot{h} &= -\delta v_D \end{aligned} \right. \quad (1)$$

The position and velocity difference of the GNSS and INS are used as the observation, and the error of navigation parameters is used as the state of the system, the state equation of the loosely-coupled GNSS/INS filter can be written as:

$$\dot{\mathbf{X}}(t) = \mathbf{F}(t)\mathbf{X}(t) + \mathbf{G}(t)\mathbf{W}(t) \quad (2)$$

The error state vector is written as:

$$\mathbf{X}(t) = [\boldsymbol{\varphi} \quad \delta\mathbf{v} \quad \delta L \quad \delta\lambda \quad \delta h \quad \delta\boldsymbol{\varepsilon}_{ib}^b \quad \delta\mathbf{f}_{ib}^b]^T \quad (3)$$

Where  $\delta\boldsymbol{\varepsilon}_{ib}^b$  and  $\delta\mathbf{f}_{ib}^b$  are the constant zero bias of gyros and accelerometers respectively,  $\mathbf{F}(t)$  is the state transition matrix,  $\mathbf{G}(t)$  is the noise driven matrix,  $\mathbf{W}(t)$  is the noise matrix.

The observation equation is written as:

$$\mathbf{Z}(t) = \mathbf{H}(t)\mathbf{X}(t) + \mathbf{N}(t) \quad (4)$$

Where  $\mathbf{Z}(t) = [\delta\mathbf{r} \quad \delta\mathbf{v}]^T$ ,  $\mathbf{H}(t)$  is the observation matrix,  $\mathbf{N}(t)$  is the noise matrix.

### 3.2. Tightly-coupled GNSS/INS Model

The tightly-coupled navigation is a method of fusing the pseudo range and pseudo range rate information measured by satellite receiver with angular velocity and acceleration information measured by gyros and accelerometers of inertial navigation system. The clock error and the clock drift of the receiver is chosen as the error state. The error state equation of GNSS is written as [5]:

$$\begin{cases} \dot{\delta t}_u = \delta t_u + \omega_u \\ \dot{\delta t}_{ru} = -\beta_{ru} \delta t_{ru} + \omega_{ru} \end{cases} \quad (5)$$

Where  $\omega_u$  and  $\omega_{ru}$  are white noises,  $\beta_{ru}$  is the correlation time.

The observation of GNSS are pseudo range and pseudo range rate provided by the receiver. The observation equation is written as:

$$\begin{cases} \boldsymbol{\rho}_{ins} - \boldsymbol{\rho}_{gps} = \|\mathbf{r}_{r,ins}^{ECEF} - \mathbf{r}_{s,gps}^{ECEF}\| - \mathbf{r}_{s,r} + \delta t_u + \nu_\rho \\ \dot{\boldsymbol{\rho}}_{ins} - \dot{\boldsymbol{\rho}}_{gps} = \|\dot{\mathbf{r}}_{r,ins}^{ECEF} - \dot{\mathbf{r}}_{s,gps}^{ECEF}\| - \dot{\mathbf{r}}_{s,r} + \delta t_{ru} + \nu_\rho \end{cases} \quad (6)$$

Where  $\mathbf{r}_{r,ins}^{ECEF}$  and  $\mathbf{r}_{s,gps}^{ECEF}$  are the position vector of the receiver and the satellite in ECEF frame,  $\mathbf{r}_{s,r}$  is the pseudo of the observable satellites measured by the GNSS receiver,  $\delta t_u$  is the distance error equivalent to the clock error,  $\nu_\rho$  is the white noise.

In this paper, the coordinates calculated in ECEF frame need to be transformed into NED frame, the transformation relation is showed as follow:

$$\begin{cases} \delta x = -(R+h)\sin L \cos \lambda \delta L - (R+h)\cos L \sin \lambda \delta \lambda + \cos L \cos \lambda \delta h \\ \delta y = -(R+h)\sin L \sin \lambda \delta L + (R+h)\cos L \cos \lambda \delta \lambda + \cos L \sin \lambda \delta h \\ \delta z = [R(1-e^2) + h]\cos L \delta L + \sin L \delta h \end{cases} \quad (7)$$

The error state equation of tightly-coupled model consists of the error state of GNSS and INS, so the error state equation of tightly-coupled model can be written as [6]:

$$\begin{aligned} \dot{\mathbf{X}}(t) &= \begin{bmatrix} \dot{\mathbf{X}}_{ins}(t) \\ \dot{\mathbf{X}}_{gps}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{ins}(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{gps}(t) \end{bmatrix} \begin{bmatrix} \mathbf{X}_{ins}(t) \\ \mathbf{X}_{gps}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{ins}(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{gps}(t) \end{bmatrix} \begin{bmatrix} \mathbf{W}_{ins}(t) \\ \mathbf{W}_{gps}(t) \end{bmatrix} \\ &= \mathbf{F}(t)\mathbf{X}(t) + \mathbf{G}(t)\mathbf{W}(t) \end{aligned} \quad (8)$$

The error state vector is represented as:

$$\mathbf{X}(t) = [\boldsymbol{\varphi} \quad \delta \mathbf{v} \quad \delta L \quad \delta \lambda \quad \delta h \quad \delta \boldsymbol{\varepsilon}_{ib}^b \quad \delta \mathbf{f}_{ib}^b \quad \delta t_u \quad \delta t_{ru}]^T \quad (9)$$

The noise vector is written as:

$$\mathbf{W}(t) = [\mathbf{w}_a \quad \mathbf{w}_f \quad w_{tu} \quad w_{ru}]^T \quad (10)$$

Where  $\mathbf{w}_a$  and  $\mathbf{w}_f$  are noises of accelerometers and gyros.

The observation equation is written as:

$$\mathbf{Z}(t) = \begin{bmatrix} \boldsymbol{\rho}_{ins} - \boldsymbol{\rho}_{gps} \\ \dot{\boldsymbol{\rho}}_{ins} - \dot{\boldsymbol{\rho}}_{gps} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_\rho(t) \\ \mathbf{H}_{\dot{\rho}}(t) \end{bmatrix} \mathbf{X}(t) + \begin{bmatrix} \mathbf{V}_\rho(t) \\ \mathbf{V}_{\dot{\rho}}(t) \end{bmatrix} \quad (11)$$

## 4. Nonlinear Filtering Algorithm Design

### 4.1. Extended Kalman Filter

EKF makes Taylor expansion of the nonlinear system at the optimal state estimation point and obtains linear model by abandoning the high-order items, the filter gain is predicted by using the prior information and the one step mean square error of the observation. The nonlinear model of the discrete system is modeled as [7]:

$$\begin{cases} \mathbf{x}_{k+1} = f(\mathbf{x}_k) + \mathbf{w}_k \\ \mathbf{z}_{k+1} = h(\mathbf{x}_{k+1}) + \mathbf{v}_{k+1} \end{cases} \quad (12)$$

Where  $\mathbf{w}_k$  and  $\mathbf{v}_{k+1}$  are the system noise and observation noise respectively,  $\mathbf{Q}_k$  and  $\mathbf{R}_{k+1}$  are the covariance.

State prediction:

$$\hat{\mathbf{x}}_{k+1,k} = f(\hat{\mathbf{x}}_k, \mathbf{u}_k) \quad (13)$$

Mean square error prediction:

$$\mathbf{P}_{k+1,k} = \mathbf{f}_{k+1,k} \mathbf{P}_k \mathbf{f}_{k+1,k}^T + \mathbf{Q}_k \quad (14)$$

Filter gain:

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1,k} \mathbf{H}_{k+1}^T (\mathbf{H}_{k+1} \mathbf{P}_{k+1,k} \mathbf{H}_{k+1}^T + \mathbf{R}_{k+1})^{-1} \quad (15)$$

State estimation:

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_{k+1,k} + \mathbf{K}_{k+1} (\mathbf{Z}_{k+1} - \mathbf{H}_{k+1} \hat{\mathbf{x}}_{k+1,k}) \quad (16)$$

Mean square error estimation:

$$\mathbf{P}_{k+1} = (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}) \mathbf{P}_{k+1,k} (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1})^T + \mathbf{K}_{k+1} \mathbf{R}_{k+1} \mathbf{K}_{k+1}^T \quad (17)$$

Where  $\mathbf{f}_{k+1,k} = \frac{\partial f(\mathbf{x}_k)}{\partial \mathbf{x}_k} \Big|_{\tilde{\mathbf{x}}_k}$ ,  $\mathbf{H}_{k+1} = \frac{\partial h(\mathbf{x}_{k+1})}{\partial \mathbf{x}_{k+1}} \Big|_{\tilde{\mathbf{x}}_{k+1}}$ ,  $\tilde{\mathbf{x}}_k$  is the optimal state estimation of

the system, the initial value of the filter is  $\hat{\mathbf{x}}_0 = E[\mathbf{x}_0]$ ,  $\mathbf{P}_0 = E\{(\hat{\mathbf{x}} - \hat{\mathbf{x}}_0)(\hat{\mathbf{x}} - \hat{\mathbf{x}}_0)^T\}$ .

#### 4.2. Unscented Kalman Filter

UKF algorithm is based on UT, the mean and variance of the posterior probabilities of the error states are approximated by a set of sample points. The filter gain is calculated using the estimation of the state and the covariance of the observations.

Sample point at time K [8]:

$$\mathbf{s}_k = [\hat{\mathbf{s}}_k \quad \hat{\mathbf{s}}_k + \gamma\sqrt{\mathbf{P}_k} \quad \hat{\mathbf{s}}_k - \gamma\sqrt{\mathbf{P}_k}] \quad (18)$$

Sample point transfer:

$$\mathbf{x}_{s_{k+1,k}} = f(\mathbf{s}_k) \quad (19)$$

Mean and variance update:

$$\hat{\mathbf{x}}_{k+1,k} = \sum_0^{2N} \mathbf{w}_i \mathbf{x}_{s_{k+1,k}} \quad (20)$$

$$\mathbf{P}_{k+1,k} = \sum_0^{2N} \mathbf{w}_i (\mathbf{x}_{s_{k+1,k}} - \hat{\mathbf{x}}_{k+1,k})(\mathbf{x}_{s_{k+1,k}} - \hat{\mathbf{x}}_{k+1,k})^T + \mathbf{Q}_{k+1} \quad (21)$$

Observation prediction:

$$\mathbf{z}_{s_{k+1,k}} = h(\mathbf{x}_{s_{k+1,k}}) \quad (22)$$

$$\hat{\mathbf{z}}_{k+1} = \sum_0^{2N} \mathbf{w}_i \mathbf{z}_{s_{k+1,k}} \quad (23)$$

Covariance of the predicted value:

$$\mathbf{P}_{zz} = \sum_0^{2N} \mathbf{w}_i (\mathbf{z}_{s_{k+1,k}} - \hat{\mathbf{z}}_{k+1,k})(\mathbf{z}_{s_{k+1,k}} - \hat{\mathbf{z}}_{k+1,k})^T + \mathbf{R}_{k+1} \quad (24)$$

$$\mathbf{P}_{xz} = \sum_0^{2N} \mathbf{w}_i (\mathbf{x}_{s_{k+1,k}} - \hat{\mathbf{x}}_{k+1,k})(\mathbf{z}_{s_{k+1,k}} - \hat{\mathbf{z}}_{k+1,k})^T \quad (25)$$

$$\mathbf{K}_{k+1} = \mathbf{P}_{xz} \mathbf{P}_{zz}^{-1} \quad (26)$$

State update:

$$\hat{\mathbf{x}}_{k+1,k+1} = \hat{\mathbf{x}}_{k+1,k} + \mathbf{K}_{k+1} (\mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1,k}) \quad (27)$$

$$\mathbf{P}_{xx(k+1,k+1)} = \mathbf{P}_{xx(k+1,k)} - \mathbf{K}_{k+1} \mathbf{P}_{zz} \mathbf{K}_{k+1}^T \quad (28)$$

### 5. Satellite Selection with DOP

In the GNSS/INS integrated navigation, the number and the geometric distribution of visible satellites directly affect the positioning error. Before the integrated navigation, four satellites are selected in all visible satellites with the best spatial position distribution by calculating the DOP to reduce the amount of computation and improve

the navigation accuracy. It is essential to calculate the Jacobian matrix G according to the observation vector  $r^n$  from the satellite to the receiver.

The geometric matrix is represented as:

$$G = \begin{bmatrix} e_x^1 & e_y^1 & e_z^1 & 1 \\ e_x^2 & e_y^2 & e_z^2 & 1 \\ e_x^3 & e_y^3 & e_z^3 & 1 \\ e_x^4 & e_y^4 & e_z^4 & 1 \end{bmatrix} \tag{29}$$

Where  $e_x^n = \frac{\partial r^n}{\partial x} \Big|_x$ ,  $e_y^n = \frac{\partial r^n}{\partial y} \Big|_y$ ,  $e_z^n = \frac{\partial r^n}{\partial z} \Big|_z$ , the satellite number is n.

The weight coefficient matrix is written as:

$$H = (G^T G)^{-1} = [h_{ij}]_{4 \times 4} \tag{30}$$

DOP is comprised of the four parameters written below:

$$\left\{ \begin{array}{l} HDOP = \sqrt{h_{11} + h_{22}} \\ VDOP = \sqrt{h_{33}} \\ PDOP = \sqrt{h_{11} + h_{22} + h_{33}} \\ TDOP = \sqrt{h_{44}} \\ GDOP = \sqrt{h_{11} + h_{22} + h_{33} + h_{44}} \end{array} \right. \tag{31}$$

HDOP is the horizontal dilution of precision, VDOP is the vertical dilution of precision, PDOP is the position dilution of precision, TDOP is the time dilution of precision, GDOP is the geometric dilution of precision.

## 6. Test Results and Analysis

In order to verify the positioning accuracy of the integrated navigation algorithm under the variable number of satellites, a vehicle test is carried out, and the trajectory of the vehicle is shown in Figure 2.

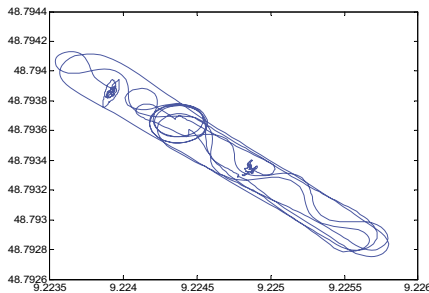


Figure 2. The route of test vehicle

Take a section of the real data for algorithm verification, the number and the DOP of the observed satellite is shown respectively in Figure 3.



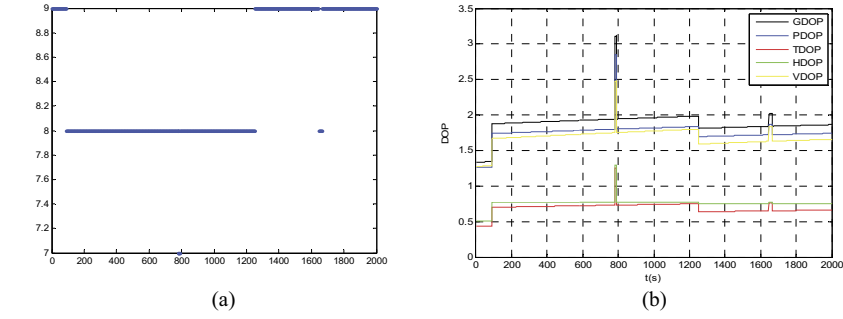


Figure 3. The number of visible satellites and the DOP value

According to the selected satellite information, the EKF algorithm is used to the loosely-coupled GNSS/INS integrated model. The initial position coordinates of the GNSS receiver is at  $48.7933^\circ$  N,  $9.2248^\circ$  E and 269.686m. The time length of the test is 1800s, the gyro random constant drift is  $0.5^\circ / h(1\sigma)$ , the gyro noise is  $0.5^\circ / h(1\sigma)$ , the accelerometer bias is  $400\mu g(1\sigma)$ , the accelerometer noise is  $400\mu g(1\sigma)$ , the filtering period is 0.5s and the update period of INS is 0.005s. The position, velocity and attitude error is shown in Figure 4.

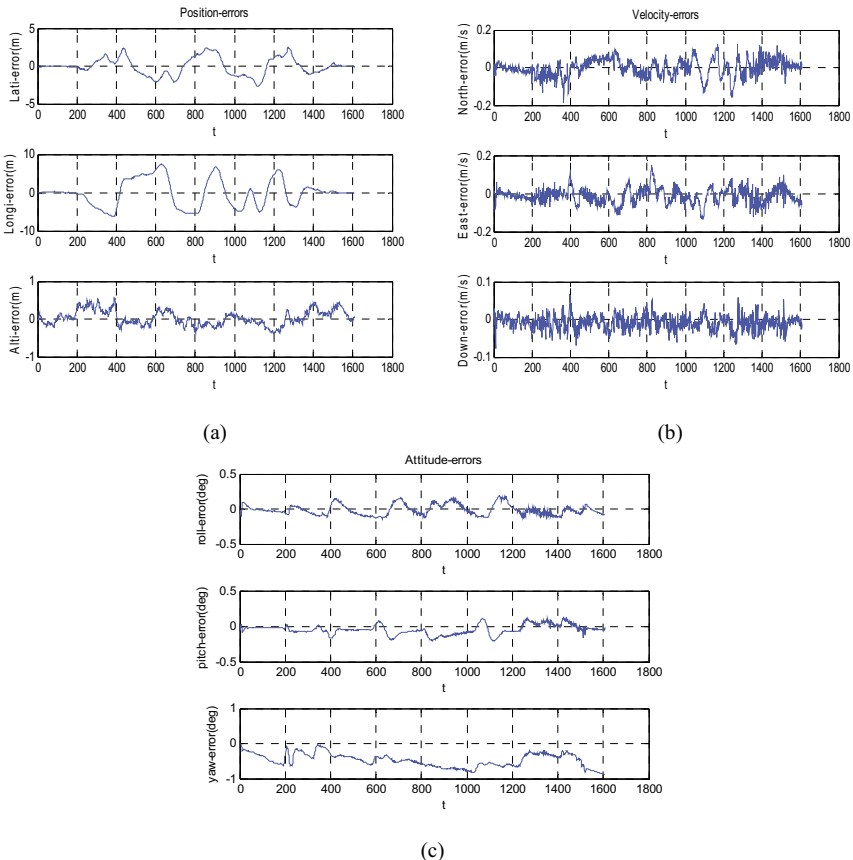


Figure 4. The position, velocity and attitude error of EKF loosely-coupled model

From the simulation results of Figure 4, the maximum position error is 7.3797m, the maximum speed error is 0.1513m/s, the maximum attitude error is 0.2041°. When the number of satellites is less than 4, the accumulative error of the loosely-coupled GNSS/INS integrated model increases gradually. This problem can be solved by using the tightly-coupled GNSS/INS integrated system.

The initial covariance matrix of the tightly-coupled GNSS/INS Integrated model is

$$P_0 = \text{diag} \left\{ \begin{matrix} (0.5^\circ)^2 & (0.5^\circ)^2 & (0.5^\circ)^2 & (0.2\text{m/s})^2 & (0.2\text{m/s})^2 & (0.2\text{m/s})^2 \\ (10\text{m})^2 & (10\text{m})^2 & (10\text{m})^2 & (0.5^\circ/h)^2 & (0.5^\circ/h)^2 & (0.5^\circ/h)^2 \\ (400\mu\text{g})^2 & (400\mu\text{g})^2 & (400\mu\text{g})^2 & (20\text{m/s})^2 & (0.05\text{m/s})^2 & \end{matrix} \right\},$$

the noise covariance matrix is  $Q_0 = \text{diag} \left\{ \begin{matrix} (0.5^\circ/h)^2 & (0.5^\circ/h)^2 & (0.5^\circ/h)^2 \\ (400\mu\text{g})^2 & (400\mu\text{g})^2 & (400\mu\text{g})^2 \\ (20\text{m/s})^2 & (0.05\text{m/s})^2 & \end{matrix} \right\}$ , the

noise covariance matrix of the observation is  $R_0 = \begin{bmatrix} (3n)^2 I_{n \times n} & 0 \\ 0 & (3n)^2 I_{n \times n} \end{bmatrix}_{2n \times 2n}$ , the

number of visible satellite is n.

When  $n \geq 4$ , the position, velocity and attitude error is shown in Figure 5.

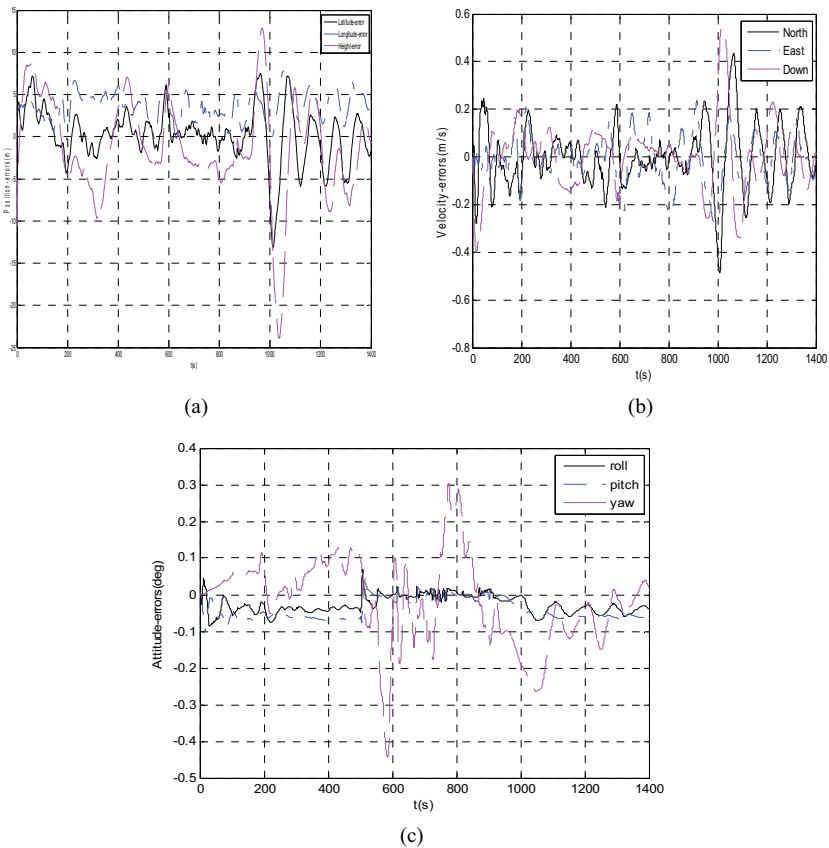
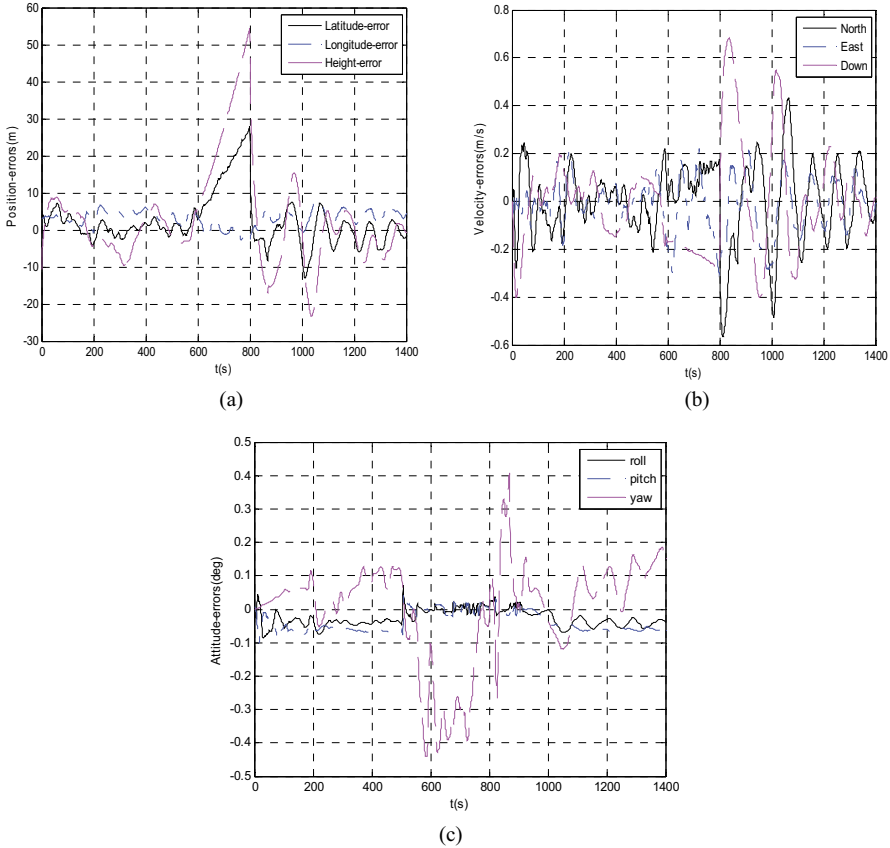


Figure 5. The position, velocity and attitude error of UKF tightly-coupled model with full satellites

It can be seen from Figure 5 that the maximum position error is 7.8213m, the maximum error of velocity is 0.1392m/s and the maximum of attitude error is  $0.3047^\circ$ . The constant bias along three axial of the carrier coordinate system of the gyro and accelerometer are  $[0.3567 \ 0.08119 \ -1.3957]$  and  $[0.00431 \ -0.006646 \ -0.01151]$ .

When  $n < 4$ , the position, velocity and attitude error is shown in Figure 6.



**Figure 6.** The position, velocity and attitude error of UKF tightly-coupled model with fewer satellites

It can be seen from Figure 6, from 600s to 800s, the number of satellites observed by the GNSS is less than four, the navigation error increases suddenly, the maximum error of position is 69.3516m, the maximum speed error is 0.6815m/s, and the maximum of attitude error is  $0.4069^\circ$ . It can provide effective navigation information for other navigation methods, and assist the positioning under the condition of lacking satellites.

## 7. Conclusion

In order to solve the problem that the satellite signals are often blocked by tall buildings or trees when the vehicle is located in cities, this paper designs an integrated navigation algorithm based on EKF and UKF, and both of the algorithms are used in the loosely-coupled model and the tightly-coupled model respectively. The navigation

algorithm is verified by a vehicle test, and the results show that when the number of observable satellites is more than four, the loosely-coupled and the tightly-coupled GNSS/INS integrated system can keep high-precision positioning. When the number of satellites is less than four, the UKF tightly-coupled GNSS/INS integrated system can still keep the position, speed and attitude error not divergent while the satellite signals are blocked for 200s. It restrains the serious error accumulation of the inertial navigation system in the loosely-coupled system with the absent satellite information for a long time and improves the stability and precision of the system.

## References

- [1] Gang Xie, GPS Principle and Receiver Design, *Electronics Industry Press*, 2017, 204-206.
- [2] Groves, Paul D, Principles of GNSS, inertial, and multisensor integrated navigation systems, 2nd edition, *IEEE Aerospace & Electronic Systems Magazine* 30.2(2015), 513-516.
- [3] Sha Fu, Nonlinear Filtering Algorithm for BD/INS Integrated Navigation System, *Northwestern University Sensors*, 2015, 27-30.
- [4] Hongwei Bian, Application of Modern Information Fusion Technology in Integrated Navigation System, *National Defense Industry Press*, 2010, 101-102.
- [5] Akça, Tamer, and M. DemiRekler, An Adaptive Unscented Kalman Filter for Tightly Coupled INS/GPS Integration, *Position Location and Navigation Symposium IEEE*, 2012, 389-395.
- [6] Zhongyuan Zang, SINS/GPS Tightly-coupled Navigation System Based on Pseudo Range and Pseudo Range Rate, *Beijing Institute of Technology*, 2015, 20-24.
- [7] Badshah K, Qin Y. Tightly Coupled Integration of A Low Cost MEMS-INS/GPS System Using Adaptive Kalman Filtering, *International Journal of Control & Automation*, 2016, 9.
- [8] Julier S A, Uhlmann J. Unscented Filtering and Nonlinear Estimation. *Proceeding of the IEEE*, 2004, 92(3), 401-422.

# An Approach to Identify Travel Hotspots at Night Based on Mobile Phone Data

Guanchen DAI<sup>a,1</sup>, Linchao LI<sup>a</sup>, Jian ZHANG<sup>a</sup>, Xiaoli ZHANG<sup>b</sup>, Yi LEI<sup>c</sup>,  
Zhongyue YANG<sup>d</sup> and Qiuying PENG<sup>c</sup>

<sup>a</sup>*Jiangsu Key Laboratory of Urban ITS, Southeast University; Jiangsu Province Collaborative Innovation Center for Technology and Application of Internet of Things; School of Transportation, Southeast University, Si Pai Lou #2, Nanjing, Jiangsu 210096, China; Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies*

<sup>b</sup>*College of Civil and Transportation Engineering, Hohai University, Nanjing, P.R. China*

<sup>c</sup>*Shanghai Meihui Software Co., Ltd, Shanghai, P.R. China*

<sup>d</sup>*Jiangsu Expressway Network Operation & Management Center, Nanjing, P.R. China*

**Abstract.** To improve the urban transport planning and safety management, an approach is proposed to identify travel hotspots based on mobile phone data. The workflow including data processing, movement state recognition, computing of travel density, identification of travel hotspots and visualization is described. A case study is presented to test the proposed approach. Moreover, the spatial distribution change of travel hotspots by time interval is analyzed. The location differences of travel hotspots on weekdays and weekends are also identified accurately with thematic map. The result shows that the hotspot identification can locate spatial clustering of trip activities effectively. It indicates that this approach is a valuable supplementary way to make up for the deficiency of traditional field surveys. It can provide some significant information for urban transport management and planning.

**Keywords.** Origin-destination, travel density, spatial distribution, significance analysis

## 1. Introduction

The stable development of transportation industry not only promote regional economy growth, but also brings urban environmental problems. Public transport refers to several transportation modes which consume less fuel per capita rather than private transport. It includes bus, metro, tram and so on. Therefore, for sustainable development of cities, public transport should be encouraged for citizens as much as possible [1]. In the daytime, perfect public transport network can cover most urban areas. However, insufficient public transport resources are provided at night. For provide more comprehensive public transport service, it is necessary to get nightly travel de-

---

<sup>1</sup> Guanchen Dai, Research Center for Internet of Mobility, No. 35 Jinxianghe Road, School of Transportation, Southeast University, Nanjing, P.R. China; E-mail: daiguanchen@gmail.com.

mand information, especially travel hotspot information. It is essential for night bus planning and rescheduling of metro. In addition, the distribution of travel hotspot is also significant for urban safety management.

Traditional traffic information is always collected by artificial investigation and fixed sensors. These approaches have drawbacks of high cost and low availability. With the popularity of mobile devices, cell phone data has gradually been the concern of researchers all over the world. With advantage of low cost, big volume and high availability, cell phone data has been widely used in the field of transportation research. The distribution of population can be analyzed with long-term cell phone history data [4,5]. The cell phone users' motion trajectory can be obtained through cell phone data [2]. The identification of motion trajectory can be further applied to analyze the characteristics of urban population dynamic distribution [3].

Current research mainly focus on hotspot identification based on taxi car data [7–9]. The visual analysis is also conducted with detection hot spots [6]. The taxi car data are also widely used for transport planning with its relatively high accuracy, but taxi can't cover all areas in city. Therefore, compared with the former, mobile phone data have the advantage of more sample, wider coverage, higher efficiency and low cost. Some studies propose methods to detect hotspot for exploring the spatial structure of cities [10] or exploring spatial-temporal patterns of urban human mobility [11].

This study develops mobile phone tracking algorithms from a large mobile phone database to get spatial-temporal information of origin and destination. Then travel hotspots are identified based on it further. A case study is conducted to analysis travel demand at night. The location of travel hotspots on weekdays and weekends can be identified accurately and efficiently. Finally, the spatial distribution characteristics of travel hotspots are analyzed based on thematic map.

## **2. Mobile Phone Data**

### *2.1. Data Description*

The data in the paper contain vast information of majority mobile phone carriers, which occupy more than 60% of the total subscribers in study area. The data include mainly three aspects of information, users' identification, location and recorded time. The identification information of every user is expressed by an encrypted ID for privacy. The 14-bit timestamp, LAC, and cell number are contained in mobile phone data. With the longitude and latitude of all cells, all users' spatial-temporal trajectories are recognized.

### *2.2. Data Generating Principle*

GSM (Global System for Mobile Communication), as second generation (2G) cellular mobile communication system, has been adopted by billions of people all over the world. Cellular system comprises a number of radio base stations (BS) and a number of mobile stations (MS) [12]. The coverage of every base station is hexagonally shaped and are termed a cell. Several cells consist a location area, which are denoted with location area code (LAC). To provide continuous communication services, mobile

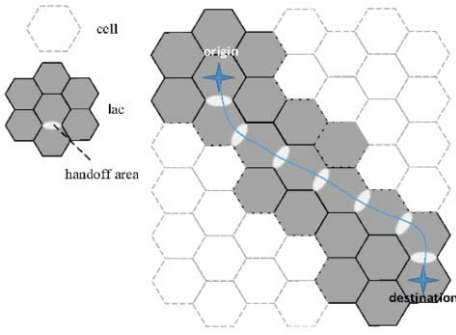


Figure 1. A moving mobile phone is calling.

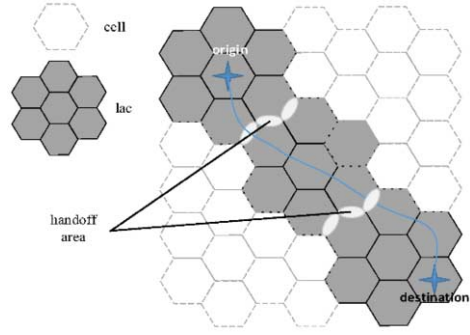


Figure 2. A moving mobile phone isn't calling.

phone users need to keep contact with one of the base stations. Therefore, when some specific events take places, a series of base station numbers containing spatial-temporal information are recorded. As a result, mobile phone data generates. These events include handoff, periodical location updating and some other events.

The communication network centrally transfers connections between different areas (handoff) when the mobile phone is moving during the service session. If a moving mobile phone is calling, handoff events will take place when the user is crossing different cells (Fig. 1). Otherwise, handoff events take place only when a mobile phone is crossing different location areas (Fig. 2). GSM also obtains the current LAC of every mobile phone actively at fixed periods (periodical location update) for keeping connection with handsets. Besides, texting, starting up and power off can also lead to generation of mobile phone data.

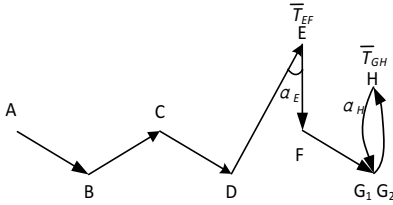
### 3. Methodology

For massiveness of mobile phone data, it is almost impossible to handle data on a personal computer. The data used in the study were stored in HDFS deployed on a server cluster and processed with Spark and Python (PySpark). All the maps were generated with ArcGIS.

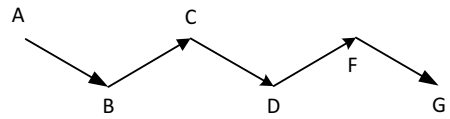
#### 3.1. Data Preprocessing

The first step is to filter the invalid and noise data. If all mobile phone data of one handset were recorded in a same cell, these data may not be used and are termed as stationary data. These data are useless for movement state recognition of travels. The stationary data should be eliminated as invalid data.

In addition, unexpected locational errors could take place. For example, users could be automatically transferred by the network from the nearest cell station to further one. The locational error can be produce by several reasons such as load restriction of cell station, signal obstruction, and so on [13]. Furthermore, users could be positioned at two adjacent cells backwards and forwards in a short period. The locational error is generated because the users are located at a common border of two cells.



**Figure 3.** Trajectory of one mobile phone user with noise data.



**Figure 4.** Trajectory of one mobile phone user after data preprocessing.

Figure 3 shows the trajectory of one specific mobile phone user with possible noise data. According to the raw mobile phone data, the sequence of trajectory points is A-B-C-D-E-F-G<sub>1</sub>-H-G<sub>2</sub>. There are two possible noisy data, which are positioned in point E and point H. For depicting more reasonable travel trajectory, appropriate judgment criteria could be used to determine whether these two data are noise data. There are two threshold values can be set. If the time difference between G<sub>1</sub> and G<sub>2</sub> is lesser than a threshold time  $T_{min}$ , which is approximately equal to the regular travel time between two different cells, the point H can be classified into noise data. If the acute angle formatted by point D, E and F (three consecutive mobile phone data) is smaller than an angle threshold  $\alpha_{min}$ , the point E can be classified into noise data. After filtering all the noise data, the trajectory (Fig. 4) is more reasonable for movement state recognition.

### 3.2. Movement State Recognition

This paper is focus on the nightly travel hotspot. It is evident that travel intensity at night is rather lower than that in the daytime. If travel demand at night is estimated simply with effective data quantity, the value will be influenced by periodical location updating data of still residents. For example, there are two traffic analysis regions (A and B). More travel activities take place in region A than region B at night, while lesser residents live in region A. It is clear that number of trips is more appropriate for estimation of travel demand, and area of these two regions also should be considered. The second step is to recognize movement state based on mobile phone data, which is the basis of O-D travel estimating. The judgment criterion is dwell time, which can be obtained by calculating the time difference between the first and the last consecutive mobile phone points appearing at the same region (cell).

For explore travel information at night, the dwell time should be long enough to avoid other disturbance such as temporal stopping caused by transfer or shopping. The other threshold value  $T_{de}$  is set to identify if the nonmoving cell phones are temporarily stopping or longtime staying users at the given region. In this paper,  $T_{de}$  is set to be a little bit more than 1 hour (one hour is the internal of periodical location update).

### 3.3. Computation of Travel Density

The third step is to compute travel density of every region based on O-D trip matrix. In this paper, travel density is the estimator of travel demand. The O-D trip matrix can be



developed through the sequence of long-time staying region. For example, if the sequence of a mobile phone user (user1) is A1-B1-C1, two pieces of O-D information can be extracted (A1-B1 and B1-C1).

Travel density  $D_i$  can be computed as:

$$D_i = \frac{t_i}{area_i} \tag{1}$$

Where  $i$  is the region,  $t_i$  is the sum of number of trips within region  $i$  per hour,  $area_i$  is the area of the region  $i$ . It is approximated by the area of voronoi diagram and the unit is square kilometer (km<sup>2</sup>). The number of trips can be acquired based on O-D trip matrix.

### 3.4. Identification of Travel Hotspots

Based on a spatial distribution of every cell’s travel density, problem of hotspot identification can be solved simply by the choice of a threshold  $\delta$  for the density  $D_i$ . However, it is clear that this method introduces some arbitrariness. The threshold is variable in different cities or in different time, and two different but close threshold value could lead to very different result. Therefore, a more universally applicable statistical analysis approach is proposed.

A human travel hotspot refers to a location with relatively higher mobility than its neighbor locations [15]. If one cell with high  $D_i$  and the local sum of  $D_i$  for the cell and its neighbors is significantly higher than the expected local sum, the location of the cell can be regarded as a statistically significant hot spot. It is clear that the local sum of  $D_i$  for the cell and its neighbors should be given spatial weight. A classical spatial statistic  $G_i^*$  is introduced for comparing local sum. The  $G_i^*$  value can be estimated by the formula:

$$G_i^* = \frac{\sum_{j=1}^n \omega_{i,j} x_j - \bar{X} \sum_{j=1}^n \omega_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n \omega_{i,j}^2 - \left( \sum_{j=1}^n \omega_{i,j} \right)^2}{n-1}}} \tag{2}$$

where

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \tag{3}$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \tag{4}$$

$x_j$  is the  $D_i$  of cell  $j$ .  $\omega_{i,j}$  is spatial weight between cell  $i$  and cell  $j$ .  $n$  is the total number of cells.

$G_i^*$  value is the z-score results of cell  $i$ . The bigger the z-score results are, the more intensely hotspot cluster.

### 3.5. Visualization

The  $G_i^*$  value enables us to recognized hotspots from a micro perspective. However, for analysis distributed information of hotspots from a macro perspective, the derived information should be visualized in an appropriate way. PDA (point density analysis) is a suitable tool for visualizing the underlying property of travel hotspot. The study area is divided into several regular grids. A parameter should be given to determine the number of times to count the point. According to the given parameter of spatial points (cells), the density of point features around each output grid is calculated. Imposed on certain color classification, the thematic map can be drawn. The map is used to support the macro exploration of hotspots and to discover spatial-temporal distribution patterns of urban nightly travel.

## 4. Case Study

The study area is a specific city, which is located in eastern China. The city is an important Railway hub connecting North China, East China and Central China. There are two big railway hub station here. It is also well served with public transport. There are 7 metro lines, 139 metro stations in the city. The daily average passenger flow is more than 2.8 million. Most of metro lines start operation at 6:00 and end at 23:00.

The mobile phone data in the nighttime of two different days (01,08,2016–02,08,2016, weekday; 06,08,2016–07,08,2016, weekend) were extracted for analyzing and visualizing travel hotspots of the travelers. After data processing, a total of 416,737 mobile phone samples can be used for excavating during the observation period.

## 5. Nightly Travel Demand Analysis

### 5.1. Origin-Destination Traffic Flow

Based on Movement state recognition, Origin-Destination matrixes were developed. The total trip from 23:00, 1st August to 6:00 2nd August (weekday) is 342,825, and the total trips from 23:00, 6th August to 6:00 7th August (weekday) is 303,275. According to O-D trip matrix, trip count are separated into two groups, trip generation count and trip attraction count. The distribution of trip generation count is shown in Fig. 5 and that of trip attraction count is shown in Fig. 6.

It is obvious that more trips occur in the weekday, whatever trip generation or attraction. Then, the trip count in weekday and weekend followed the same trend. For trip generation, the number drops gradually from 23:00 to 6:00. In contrast, the number of trip attraction meet rapidly rise. The trend of trip generation and attraction is in

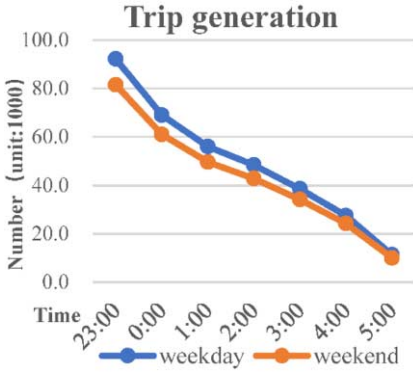


Figure 5. Trip generation count distribution.

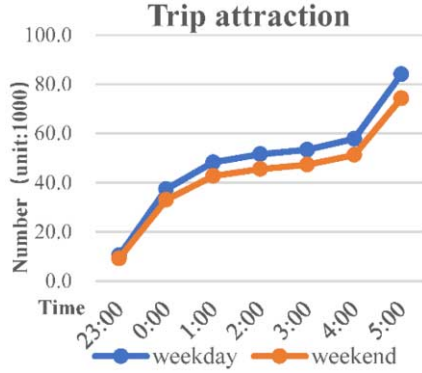


Figure 6. Trip attraction flow distribution.

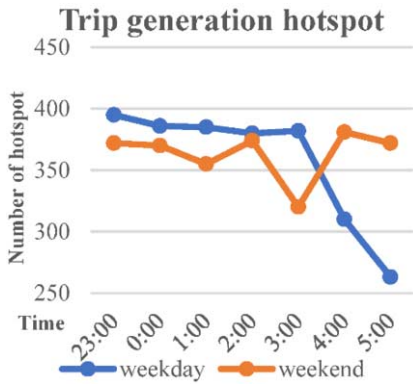


Figure 7. Trip generation hotspot distribution.

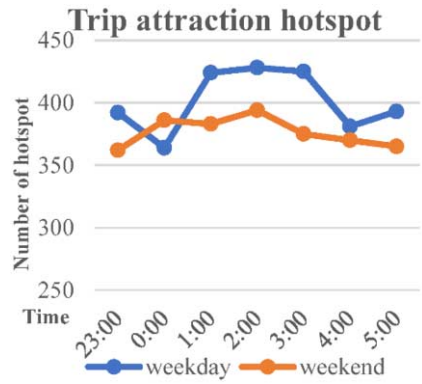


Figure 8. Trip attraction hotspot distribution.

accordance with people’s regular schedule. This confirms that the trip count estimated from the mobile phone data can offer reasonable traffic information.

### 5.2. Number of Hotspots

Based on identification of travel hotspots, the  $G_i^*$  value and p-value of every cell are acquired. Every cell with less than 0.05 of the p-value is regarded as hotspot. Based on the results, the distribution of trip generation hotspot is shown in Fig. 7 and that of trip attraction hotspot is shown in Fig. 8.

We can only observe that the number of trip generation hotspot meet fall in night. There are no notable trend for other three lines. There is also no significant difference between the trip count in weekday and in weekend. It is found that the degree of concentration did not change significantly over time in night.

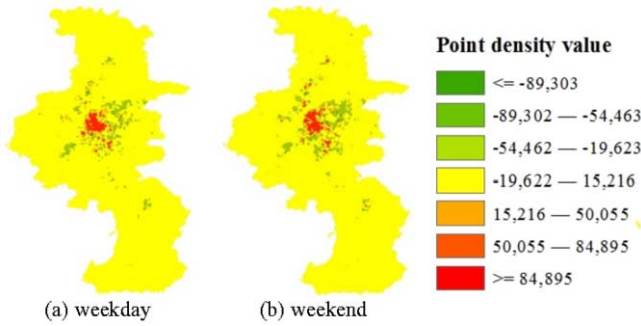


Figure 9. Thematic maps of trip attraction hotspot in different days.



Figure 10. Map of study area.

### 5.3. Spatial Structure of the Hotspots in Different Days

The spatial distribution of travel hotspot is necessary for analysis nightly travel demand. Based on visualization, the travel thematic map at night can be drawn. Figure 9 shows two thematic maps of trip attraction hotspot separately in weekday and weekend. The legend refers to the result of point density analysis, which takes the  $G_i$  value as the given parameter. Figure 10 displays locations of districts in the study area.

The urban area is a high incidence area for nightly travel whether on weekends or weekdays. In Fig. 9(a), some trip attraction hotspots distribute in suburban or rural areas (Jiangbei New Area and Lishui District). In Fig. 9(b), only a few hotspots are located in suburban areas and hardly no hotspots are located in rural areas. It is evident that night travel demand more concentrated in the city center during the working day.

### 5.4. Spatial Structure of the Hotspots in the Same Day

Figure 11 illustrates the change of spatial distribution of trip attraction hotspot from 23:00, 6th August to 6:00, 7th August. It can be found that seven maps by different time internal follow a similar pattern. Most of travel hotspots gather in the central urban area. Some of hotspots are located in Jiangning district and Qixia district. Only a few hotspots distributes in Jiangbei New Area and Lishui district. It is noteworthy that the distribution of hotspots in Fig. 11(a) is more different from that in the following six maps.

Figure 12 shows spatial distribution of trip attraction hotspots between two time internals (from 23:00, 6th August to 0:00, 7th August and from 00:00, 7th August to 01:00, 7th August). Main differences between two maps are highlighted with red boxes (Fig. 12(a)) and green circles (Fig. 12(b)). The location and land-use information of these areas are listed in Table 1. Area E, F and G are considered as industrial areas because there are some factories and warehouses of logistics company around there. It can be found that transport hub and large residential area attract more trip from 23:00 to 00:00. However, more travel activities take place in industrial areas from 00:00 to 01:00. This could possibly be due to night commute, which is more likely to occur in suburban industrial areas. The layout of bus station can be improved based on these information.

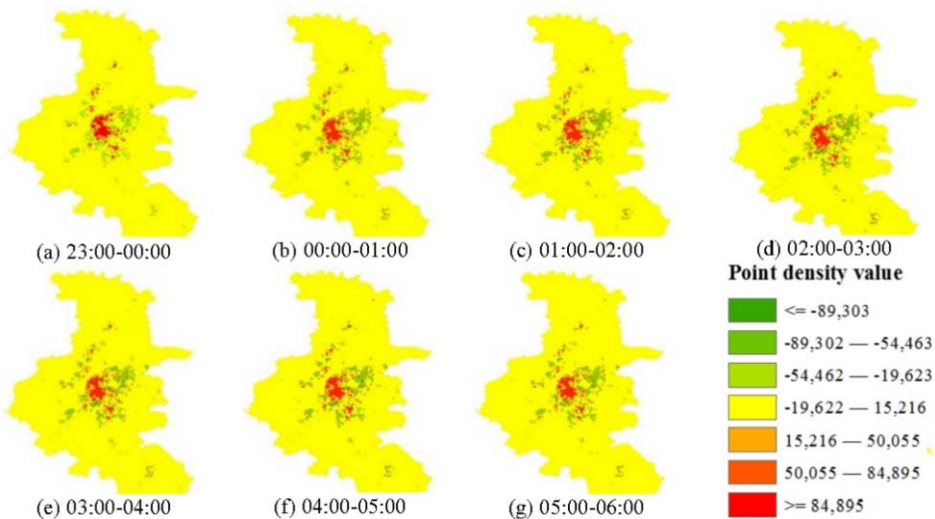


Figure 11. Thematic maps of trip attraction hotspot by time intervals.

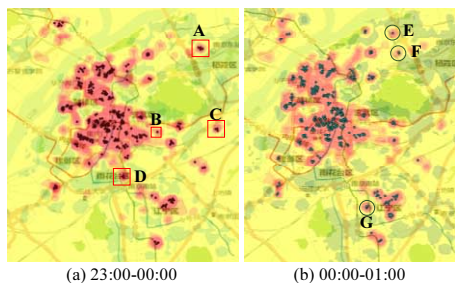


Figure 12. Local amplification maps of thematic maps by two time intervals.

Table 1. Information of distribution difference area in two time intervals.

Index	Time interval	Location information	Land-use information
A	23:00-00:00	East Railway Station	Railway transport hub
B	23:00-00:00	Muxuyuan Street	Large residential area
C	23:00-00:00	Magao Road	Large residential area
D	23:00-00:00	South Railway Station	Railway transport hub
E	00:00-01:00	Fandian village	Industrial area
F	00:00-01:00	Fendou village	Industrial area
G	00:00-01:00	Yanchun Road	Industrial area

## 6. Conclusions and Future works

In this paper, a synthetic approach for identifying travel hotspots at night is proposed. The approach describes the process to excavate spatial-temporal distribution information of human mobility based on mobile phone data. Then a case study is employed to test the proposed method. The result shows the hotspot identification can locate spatial clustering of trip activities effectively. Moreover, the spatial distribution change of travel hotspots by time interval can be analyzed. The location difference of travel hotspots on weekdays and weekends can also be accurately identified. Possible reasons of these change and differences are discussed based on location and land-use information. It indicates the approach is a valuable supplementary way to make up for the deficiency of traditional field surveys. These findings provide significant information to improve existing bus station layout. The approach can also be applied in urban security management. The places gathering a lot of travel hotspot should be of concern from enforcement authority.

However, for more intelligent urban transport planning, only travel distribution information isn't sufficient. In future works, deep learning model can be utilized to classify travel hotspots. More universal travel patterns can be explored among different spatial levels of the city. Combined with questionnaires data, the interactions between citizens and their living environments can also be studied. These findings are important theoretical basis for future traffic planning.

## Acknowledgements

Here we acknowledge the sponsors of the ITITS conference, anonymous reviewers, and authors of cited papers for their detailed comments, without which this work would not have been possible. This research is partially supported by the Science and Technology Demonstration Project of Ministry of Transport of China (No. 2015364X16030), the National Key R&D Program in China (Grant No. 2016YFB0100906) and the National Natural Science Foundation of China (No. 61620106002).

## References

- [1] Carson G, Dix M, Callaghan J, et al. Encouraging sustainable development by linking public transport accessibility, new parking standards and developer contributions [J]. *Traffic Engineering & Control*, 1999.
- [2] Yao Z F, Yang F, Cheng Y, et al. Multimode Trip Information Derivation from Personal Trajectory Data [C]//Transportation Research Board. TRB 94th Annual Meeting Compendium of Papers. Washington D.C.: Transportation Research Board, 2015:1–14.
- [3] Ran B. Daily O-D Matrix Estimation using Cellular Probe Data [C]//Transportation Research Board 89th Annual Meeting. 2010.
- [4] Hu Y K, Song L, Zhang J, et al. An Improved Method of Extracting OD Information Based on Mobile Phone Data [J]. *Journal of Transport Information and Safety*, 2015(5):84–90.
- [5] Meng H, Yu D, Hu Y, et al. Passenger Distribution Analysis for One Metro Line Based on Phone Data [C]//Cota International Conference of Transportation Professionals, 2016:932–942.

- [6] Jahnke M, Ding L, Karja K, et al. Identifying Origin/Destination Hotspots in Floating Car Data for Visual Analysis of Traveling Behavior [M]//Progress in Location-Based Services 2016. Springer International Publishing, 2017.
- [7] Krisp J M, Polous K, Peters S, et al. Getting in and out of a taxi: spatio-temporal hotspot analysis for floating taxi data in Shanghai [C]//International Symposium on Networks for Mobility. 2012.
- [8] Jahnke M, Ding L, Karja K, et al. Identifying Origin/Destination Hotspots in Floating Car Data for Visual Analysis of Traveling Behavior [M]//Progress in Location-Based Services 2016. 2017.
- [9] Lin N, Li J. Research and Implementation of Bus Route Planning Based on Floating Car Data [J]. Computer Applications and Software, 2016, 33(10):270–274.
- [10] Louail T, Lenormand M, Cantu Ros O G, et al. From mobile phone data to the spatial structure of cities. [J]. Physics, 2014, 4(2973):61–61.
- [11] Yang X, Zhao Z, Lu S. Exploring Spatial-Temporal Patterns of Urban Human Mobility Hotspots [J]. Sustainability, 2016, 8(7):674.
- [12] Balck K, Fried T, Frisell H A. Cellular mobile communication system: EP, WO 1996006512 A1 [P]. 1996.
- [13] Ran B. Use of Cellphone Data in Travel Survey and Transportation Planning [J]. Urban Transport of China, 2013, 11(1):72–81.
- [14] Wang M H, Schrock S D, Broek N V, et al. Estimating Dynamic Origin-Destination Data and Travel Demand Using Cell Phone Network Data [J]. International Journal of Intelligent Transportation Systems Research, 2013, 11(2):76–86.
- [15] Brimicombe A J. Cluster detection in point event data having tendency towards spatially repetitive events [J]. 2005.

# Online Road Boundary Detection for Sparse 3D Laser

Pengpeng Sun<sup>1</sup>, Xiangmo Zhao, Ya Sun, Haigen MIN

*Department of Traffic Information Engineering & Control, Chang'an University, Xian, Shaanxi, China, 710064*

**Abstract.** The ability to extract road boundary in the outside scene is key for autonomous vehicle navigation. This paper proposed a novel road boundary detection method for autonomous vehicle based on spatial knowledge and consistency of road boundary exploiting a 3D LiDAR. More specifically, a novel method was first used to separate the ground points so as to remove obstacles and speed up the processing of road boundary points selection. Secondly, we extract boundary candidate points from individual scan lines based on spatial knowledge, and a two-pass filter was used to eliminate fake boundary points based on consistency of road boundary. Finally, polynomial fitting and boundary update strategies were exploited to produce smoother transition and prevent line-jumping. A challenging realistic traffic scene dataset was collected to evaluate this proposed method and the experimental results demonstrate that the proposed method can detect road boundary with strong robustness at real time for both static and dynamic scenes. The proposed method has been successfully applied to our autonomous vehicle.

**Keywords.** road boundary detection; autonomous vehicle; 3D LiDAR; obstacle occlusion; spatial knowledge

## 1. Introduction

Road boundary detection is an integral and important function in autonomous vehicle navigation systems [1]. On the one hand, road boundary is a prominent feature of roads in urban environment, which can delimit the boundaries of the roads. So precisely detected road boundary can be used to reduce search area and increase object detection precision based on the fact that vehicles can only appear on the road. On the another hand, road boundary have become a commonly used feature for vehicle's location in urban environments once the GPS signal reception is limited due to the presence of trees and building etc.[2],[3],[4].

Due to the importance of road boundary detection in autonomous vehicle, there are many scholars at domestic and abroad have studied on road boundary detection technology over the past few years and different sensors have been applied in order to get better results[5], such as monocular camera[6],[7], stereo vision [8],[9],[10],[11], 2D LIDAR [12],[13],[14],[15], 3D LIDAR[16],[17],[18],[19],[20], etc.

---

<sup>1</sup> Pengpeng Sun, Institute of Traffic Information Engineering&Control, Chang'an University. E-mail: pengpeng.sun@chd.edu.cn.



Benefit from the intensively studied in image processing, camera has been applied extensively in modeling the driving environment to detect road information. The merit of monocular camera-based methods have the advantage of low cost and rich context information that camera itself can provide, such as color, texture, etc. However, they are susceptible for false positives caused by edges in the intensity images with shades and changing of illumination, which will result in unstable results in the road boundary detection. Moreover, the distance information is lacking based on monocular camera. To address the lack of distance information provided by monocular camera, stereo vision has been developed and used in autonomous driving to detect road boundary. But stereo vision has a narrower FOV. Besides, the dependency of methods to compute the image disparity to obtain depth information can be a computational bottleneck for autonomous vehicle. Compared with the stereo vision, 2D LIDAR sensors can work independently of light condition and directly return distance information, which have become one of the most popular sensors used in mobile robot navigation due to its accuracy and robustness in outdoor environment. However, 2D LIDAR can only provide points in a fixed pitch angle and a limited number of points can be detected as road boundary in each frame, which cause the performance of 2D LIDAR-based detection algorithm easily affected by observation noise and complex road structure.

3D LiDAR are employed recently years, which can suppress the lack of data and have been widely using in environment perception for autonomous vehicle. In this paper, we studied the road boundary detection using Velodyne HDL-32E 3D LiDAR scanner.

One of the most widely used methods for road boundary detection with 3D LiDAR is that of grid map-based point cloud projection, which project all the point clouds obtained from 3D LiDAR into a grid map, and detect boundary points based on local grid geometry features. Such methods have been widely used in Urban Challenge DARPA in 2007. Zhao G. et al [16] provide a grid map-based method for curb detection. In this method, the boundary candidate points were first extracted based on three local grid geometry features, then a short-term memory strategy was used to remove the obstacles causing by false positive. The work of [18] first preprocess the point cloud by interpolation and feature points were extracted based on distribution of points on the road surface. Then Iterative Gaussian Process Regression is utilized to grow road region based on the above extracted feature points. One of the advantages of boundary detection methods based on grid map is that several sensors can be easily fused, thus ensuring that any resulting mapping is straightforward. However, the precision of road boundary detection results is a little bit low due to the fact that the grid size in map is far lower than the distance precision of 3D LIDAR data. The size of grid map cannot be too big, otherwise it will affect the calculation speed and limited size of grid map can also lead to large amounts of data useless. On the other hand, many grids maintain few data due to the uneven distribution of point clouds, which will result in storage and algorithm of computing waste. Besides, due to the obstacles occluded on the road, the grid map cannot reflect the true geometry structure of road boundary. Thus, the road boundary can be easily contaminated by random noise and measurement errors.

Another major way for road boundary detection is graph-based method, which use the same data structure characteristic of 3D LiDAR. In this method, all the points are organized as undirected graph and every point can be regarded as node in undirected graph. The way of processing 3D LiDAR data can be converted to handle undirected

graph. The work of [19] detect road boundary like obstacles by analyzing the ring compression of 3D LiDAR. In [20], road boundary were detected by computing the angle between 3D LiDAR scan beams and the results show good performance. Compared with grid-based methods, graph-based methods can get a better results in detection precision and take full use of the ability in storage and computing. However, such methods have bad performance in real-time, and cannot meet the requirements of autonomous vehicle. Moreover, many road boundary detection methods do not consider the presence of dynamic obstacles or occlusions in the environment, which can impair the localization performance.

In order to detect road boundary even in occluding scenes, a novel method was proposed in this paper based on spatial knowledge using 3D LiDAR. In this method, fusion of polar grid-based and graph-based method were addressed to detect road boundary points.

The remainder of this paper is organized as follows. In Section II, we introduce our method for road boundary detection algorithm. In Section III, extensive experiments were carried out in various typical scenes to evaluate the effect of the proposed method. Finally, the conclusions and future works are presented.

## 2. Details of Road Boundary Detection

### 2.1. Ground Points Extraction

We can obtain more than 700,000 point cloud points per second from Velodyne HDL-32E LIDAR for recognition of the environment surrounding the vehicle, which results in time consuming with processing all the points for proper recognition. Thus, it's necessary to use a method that can efficiently handle such computations. Since the aim is to extract the road boundaries, so ground points can be firstly extracted with a fast but efficient ground segmentation method based on the assumption that road boundaries are part of ground. Moreover, some non-ground objects such as pedestrians, cars, vegetation also interface the detection of curbs. So it is necessary to separate the ground points first to speed up the processing and improve precision of boundary detection.

As the HDL-32E LIDAR sensor shoots 32 laser points at different vertical angles and rotates 360 degrees, so the point cloud that are close to the center of the sensor have a higher density. Our work is implemented using a projection for the construction of a 2.5 polar grid map based on polar coordinates, in such a way that we can efficiently speed up extract the ground points and minimizing the number of empty cells at the same time. We first partition a frame point cloud into a polar grid map with a given horizontal angle resolution (0.5 degree) and a radial distance resolution (0.2 m) based on the principle of 3D LiDAR sensor. Thus, each polar grid contains points belong to ground and points pertaining to objects beyond the ground such as trees, cars and pedestrians. Due to the fact that all ground targets are connected to the ground and stretching on the vertical direction and we can segment ground points according to the elevation of stretching in each polar grid. So a search strategy is carried out for each polar grid to select the highest point which stretch from the ground. We first sort out all the points in a polar grid according to the elevation of ascending order and the sorted points in a grid can expressed as below:

$$\text{PolarGrid}(m, n) = \{p_i, i = 1, \dots, N\} \tag{1}$$

Where  $N$  denotes the total number points in polar grid. Then elevation difference between two adjacent points  $p_i$  and  $p_{i+1}$  will be calculated from the lowest point. The point  $p_i$  will be labeled as the highest point if and only if the elevation difference is greater than a given threshold  $L_{th}$  and search the next polar grid until the last polar grid. If the elevation difference lies below  $L_{th}$  and the point  $p_{i+1}$  is the last point in the grid, labeling the point  $p_{i+1}$  as the highest point in the grid and search the next polar grid until the last polar grid. Otherwise, we set  $i = i + 1$ , and keep calculate the next consecutive points. End of the search until all the highest points in each grid have been labeled. Usually, we set  $L_{th} = 25cm$ . All the ground points will be extracted based on the assumption that points belong to ground are below the highest elevation in each polar grid. With this scheme, most outliers such as cars, pedestrians and trees over the ground are removed, as shown in Fig.1.

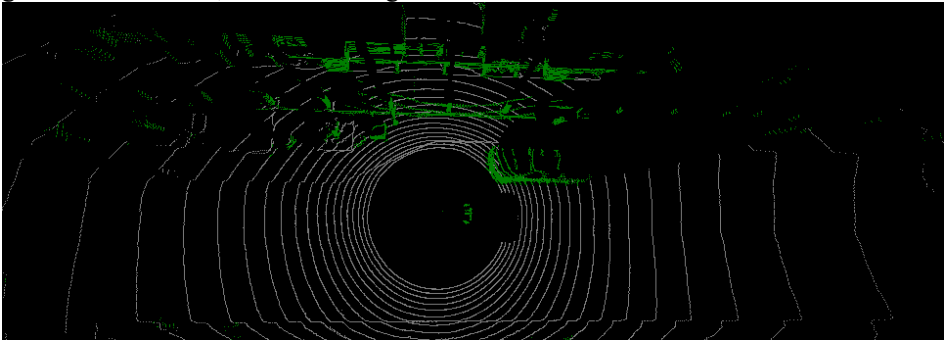


Fig.1. An example of ground extraction points result in a complex scenarios using our proposed method

### 2.2. Road Boundary Candidates Detection and Polynomial Fitting

In order to distinguish the boundary and non-boundary area, we need to select and extract some necessary features considering the geometrical structure of the environment surrounding the vehicle. The DEM can't reflect the true geometry structure of road boundary under condition of road with lower boundaries and obstacle occlusion. Besides, a lot of obstacle points have been removed through above ground point extraction method. So we take advantage of linear structures derived from LiDAR point cloud and determine the existence of road boundary points from individual scan line based on the local geometrical features.

Considering the road surface is rising and falling, and the slope of the road surface in each scan line is constrained to change slowly. So the slope cue was calculated in individual scan line as the first criterion. For a point  $p_i(x_i, y_i, z_i)$ , the slope constraint can be formulized as below:

$$S_i = \arctan\left(\frac{z_{end} - z_{start}}{\sqrt{(x_{end} - x_{start})^2 + (y_{end} - y_{start})^2}}\right) - \arctan\left(\frac{z_i - z_{start}}{\sqrt{(x_i - x_{start})^2 + (y_i - y_{start})^2}}\right) \tag{2}$$

Where  $(x_{start}, y_{start}, z_{start})$  and  $(x_{end}, y_{end}, z_{end})$  are the coordinates of start point and end points in a consecutive scan points respectively. If the slope difference between a line connecting the start and end point and a line connecting the start point

and  $p_i$  is larger than the given threshold, i.e.  $S_i > S_{thr}$ , the point  $p_i$  will be selected as boundary candidate point.

Due to the complexity of outdoor environment, it can be occur over segmentation if we just using slope value. For instance, when vehicle traveling on non-flat road, a small convex point will cause a great slope and the convex point will be identified as a boundary point based on slope cue. So elevation difference in consecutive scan points was adapted as the second criterion based on the fact that the points are the road boundaries with certain height with respective to smooth road surface. For a point  $p_i$  of in the  $j$  ring, let  $S = \{p_{i-n}, \dots, p_i, \dots, p_{i+n}\}$  represents the set of consecutive adjacent points of  $p_i$  and the extreme of elevation in the consecutive points will be calculated. Let  $H_{max}$  and  $H_{min}$  represent the maximal and minimal elevation in  $S$ , respectively. The point  $p_i$  will be selected as the curb candidate point if and only if the elevation difference within a reasonable range, i.e.

$$H_{th1} < H_{max} - H_{min} < H_{th2} \quad (3)$$

For runtime improvement of road boundary detection, the extracted ground points were first stored in a two-dimension array, in which the row direction represented the information of the 32 points obtained in one direction and stored in the order from the inside to the outside. The column direction respected the point's rotation angle in a scan line. Then we start the labeling processing from the middle column in each row, in other words, the initial labeling is entering from the road into boundary. Therefore, once a point has the characteristics of road boundary points, it means that the point reaches a possible boundary and it will be labeled as a boundary candidate point, otherwise, it will be labeled as non-boundary point. Through the experiment results and analysis, five boundary candidates closest to scanning center were selected in each side.

Through the experimental results we found that form all the boundary candidates also have some other objects such as vegetation points over the road apart from boundary. Since the largest difference between the road boundary and other obstacles is the continuity direction. So a two-pass method was implemented for further filter out the outliers considering the consistency of road boundaries. In the first pass, the RANSAC (Random Sample Consensus) algorithm was used to estimate the direction of line in left and right parts of the road boundary, respectively. The merit of RANSAC is able to perform robust estimation of the model parameters particularly when a significant percentage of data are outliers. Moreover, it considers the entire distribution of data set so as to it can give more accurate result than other algorithms. After getting the direction of road boundaries in left and right sides, a second pass method was adapted. In this step, the point cloud is divided to several slices and boundary candidate points were extracted from each slice individually based on the fit found in the first pass plus some slack to account for noisy data. More specifically, for each slice the starting  $x$  offset and ending  $x$  offset are calculated by applying the fit found in the first pass at top and bottom  $y$  coordinates of the current slice. In order to get consecutive and smooth road boundaries, a second order polynomial was used to fit the boundary candidate points in the left and right road boundary that detected in previous step based on its simplicity and reliability in the experiments. The Equation is given as below:

$$x = a_0 + a_1 * y + a_2 * y^2 \quad (4)$$

### 2.3. Boundaries Update

Compared with residual error, the wrong road boundaries detection have more serious consequences for autonomous vehicle decision. For instance, if the detected road boundaries in previous frame have a great difference compared with road boundaries detected in current frame, it will cause line-jumping. So in real environment application, we have to evaluate the road detection results.

In the experiment, the number of efficient point accounts for the proportion of the total detected boundary candidate points was used as the first criterion to evaluate the road detection result. The efficient point denotes the boundary candidates after two-step pass filter. That is to say, more points lying in the boundary with stronger boundary features and less model error can lead to a higher proportion for the detection result. If the proportion is greater than 50% in left or right road boundaries, the current fitting is likely to be effective. Another evaluation criterion that we using in the experiment is the difference between the ending points for the x value. For this purpose, the x position at the ending points is calculated for both the previous and the current fits and then compared.

The current fit for left or right road boundary will be accepted if and only if the proportion of efficient point is greater than 50% and the difference between the end points for the x value in the previous and the current fits is below a threshold. If the current fit is accepted, the average with the previous fit is calculated to produce smoother transitions and prevent line-jumping. Otherwise, we will discard the current fit.

### 3. Experimental Evaluation and Analysis

To evaluate the proposed boundary detection method, extensive experiments were carried out in various urban scenes with our autonomous vehicle platform. The platform is a modified BYD Speed Sharp as shown in Fig 2. The sensor set and perception modules can be found in the platform, which include a Velodyne HDL-32E



Figure 2. The autonomous vehicle platform

LIDAR, 3 SICK LIDAR, 2cameras, a binocular vision system, a RT3000 Inertial and GPS Navigation System (GPS/INS) and some other sensors. The Velodyne is mounted on the top of our platform to obtain 3D point clouds.

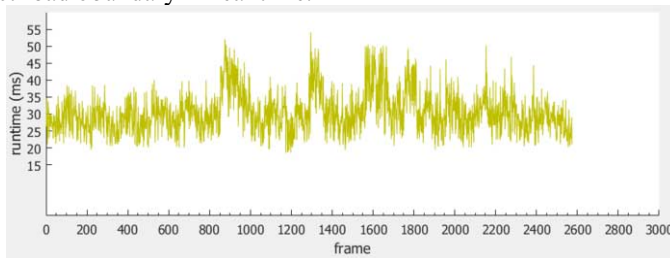
In order to evaluate the reliability of our algorithm quantitatively, the experiment was carried out on our campus. The travelled distance of the vehicle is about 2.1 km and received 2573 frames of laser point clouds. The test environment contains various road condition such as straight road, curved road and crossroad etc.

To assess computational efficiency, the proposed algorithm was tested on our on-board industrial PC. The hardware and software in our system is shown in Table 1.

**Table 1.**The hardware and software configuration in our system.

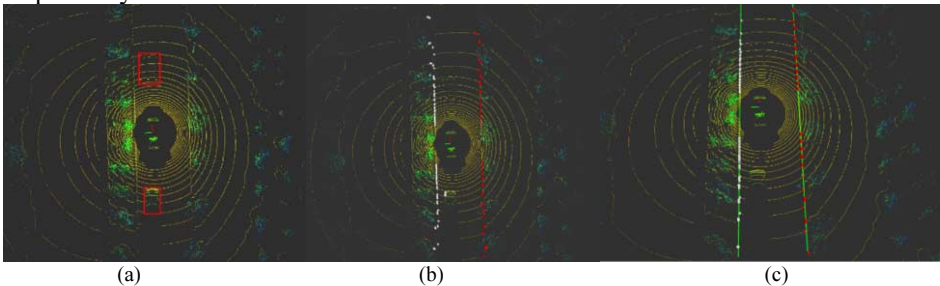
Configuration	Parameter
CPU	Intel(R) Core 4
Dominant Frequency	3.1 GHz
Memory	8GB
Operating System	Linux+ROS
Programming language	C++

The execution time of our algorithm in the experiment is shown in Fig 3. Through operations in the road boundary detection processing are mostly performed on 3D point cloud, but the average period is less than 35ms. The process in computation is efficient due to the most of experimental roads are straight or curved and no data indexing structures required for data querying and searching. Thus, our algorithm is a feasible means to detect road boundary in real time.



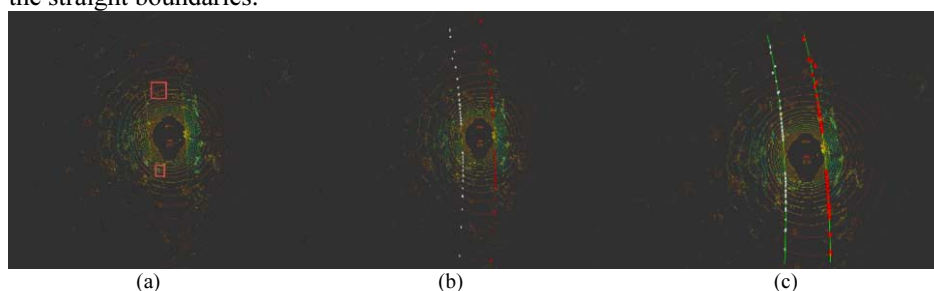
**Figure 3.**The runtime of our algorithm

In order to verify the effectiveness of the proposed method quantitatively. Three typical scenes from the whole data set were selected to evaluate the performance of our algorithm. The experimental results of three scenes will be introduced in the following, respectively.



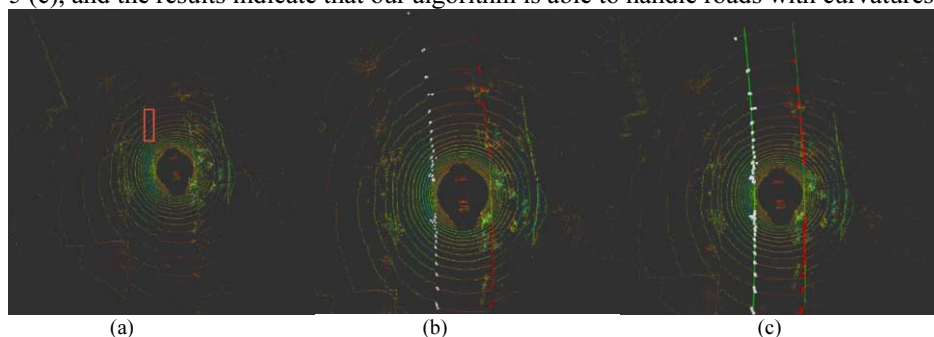
**Figure 4.** The result of the straight road boundary detection. (a)The first scene. (b) The results of boundary candidate points detected. (c) The fitting results of road boundaries.

The first scene is a typical campus environment with straight road boundary, but it is more complex as shown in Fig 4 (a). There are six cone-shaped barrels in front our vehicle and a car behind our vehicle as can be seen in the red rectangular boxes, and those presence of obstacles occluded the left road boundary. Besides, those obstacles can be easily identified as curbs by mistaken due to the similar local geometry features. The detected boundary candidate points after two-pass filter as shown in Fig 4 (b). The white points denote the boundary points in the left boundary and the red points denote the boundary points in the right boundary. We can find that our method correctly detect the boundary point and no obstacle points were recognized as boundary points by mistaken. The main reason is that our method uses several spatial property of boundary and only detect boundary in the identified ground region. The obstacle region can be filtered out through ground extraction method, which makes our method robust to the obstacle occlusion. The fitting result as shown in Fig 4 (c), and our method can well fit the straight boundaries.



**Figure 5.** The result of the curved boundary detection. (a) The second scenes. (b) The results of boundary candidate points detected. (c) The fitting results of curved road boundaries.

The curved boundary are included in the second scene, as shown in Fig 5 (a). In the scene, there are many weeping willows on both sides of road and even some willow branch have been touched the road surface as can be seen in red rectangular boxes. In this situation, most of boundary points can still be accurately extracted as shown in Fig 5(b). The boundary were fitted into two smooth boundaries of the road as shown in Fig 5 (c), and the results indicate that our algorithm is able to handle roads with curvatures.



**Figure 6.** The result of the curved boundary detection in more complex scene. (a) The third scene. (b) The results of candidate points detected. (c) The fitting results of irregular road boundaries with curvatures.

The third scene is still a typical crooked road scene, but the road boundaries are more irregular as shown in Fig 6(a). The left edge of road is made up of grass and the right edge of road is composed of bar store, so the boundary geometry features are not obvious on the left edge of road as seen in the red rectangle box. However, our method

detected most of boundary points as shown in Fig 6(b). The main reason is that detect the boundary points from individual scan line, which can extract the share of road boundary with non-obvious road conditions, such as low road boundary, irregular road boundary etc. The detected boundary points were fitted as shown in Fig 6(c), the results verify the efficient of our algorithm to deal with irregular road scenes.

#### 4. . Conclusion and Future Work

Against road boundary irregularly and susceptible to interference by obstacles within road, this paper presents a novel boundary detection algorithm for self-driving platform using Velodyne HDL-32E 3D LIDAR. This method has been verified by experiments for various road environments. Both the quantitative and qualitative experiments demonstrate the accuracy and robustness of the proposed method and the method has been successful used in our self-driving platform for navigation.

The algorithm proved to be fast and robust in most continuous road boundary conditions. But in crossroad the continuous information can't be utilized to detect all the lines that represent the road boundary. So the crossroad curb detection is a key issue to be addressed in the future work. Besides, we are going to apply this boundary detection method for other self-driving applications, such as vehicle localization and map building etc.

#### Acknowledgement

The authors would like to thank the members of XINDA self-driving team in Department of Traffic Information Engineering and Control. This work was also supported in part by the 111 Project on Information of Vehicle-Infrastructure Sensing and ITS Under Grant No B14043.

#### References

- [1] Wijesoma W S, Kodagoda K R S, Balasuriya A P. Road-boundary detection and tracking using lidar sensing[J]. *IEEE Transactions on Robotics & Automation*, 2004, 20(3):456-464.
- [2] Kim S H, Roh C W, Kang S C, et al. Outdoor navigation of a mobile robot using differential GPS and curb detection[C]//*Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007: 3414-3419.
- [3] Qin B, Chong Z J, Bandyopadhyay T, et al. Curb-intersection feature based monte carlo localization on urban roads[C]//*Robotics and automation (ICRA), 2012 IEEE international conference on*. IEEE, 2012: 2640-2646.
- [4] Hata A Y, Wolf D F. Feature Detection for Vehicle Localization in Urban Environments Using a Multilayer LIDAR[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(2): 420-429.
- [5] Hillel A B, Lerner R, Levi D, et al. Recent progress in road and lane detection: a survey [J]. *Machine vision and applications*, 2014, 25(3): 727-745.
- [6] Lu W, Wang H, Wang Q. A synchronous detection of the road boundary and lane marking for intelligent vehicles[C]//*Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, 2007. *SNPD 2007. Eighth ACIS International Conference on*. IEEE, 2007, 1: 741-745.
- [7] Seibert A, Hähnel M, Tewes A, et al. Camera based detection and classification of soft shoulders, curbs and guardrails[C]//*Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013: 853-858.



- [8] Oniga F, Nedeveschi S. Polynomial curb detection based on dense stereovision for driving assistance[C]//Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on. IEEE, 2010: 1110-1115.
- [9] Siegemund J, Franke U, Förstner W. A temporal filter approach for detection and reconstruction of curbs and road surfaces based on conditional random fields[C]//Intelligent Vehicles Symposium (IV), 2011 IEEE. IEEE, 2011: 637-642.
- [10] Kellner M, Bouzouraa M E, Hofmann U. Road curb detection based on different elevation mapping techniques[C]//2014 IEEE Intelligent Vehicles Symposium Proceedings. IEEE, 2014: 1217-1224.
- [11] Sodhi D, Upadhyay S, Bhatt D, et al. CRF based method for curb detection using semantic cues and stereo depth[C]//Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing. ACM, 2016: 41.
- [12] Wijesoma W S, Kodagoda K R S, Balasuriya A P. Road-boundary detection and tracking using lidar sensing[J]. IEEE Transactions on robotics and automation, 2004, 20(3): 456-464.
- [13] Liu Z, Wang J, Liu D. A new curb detection method for unmanned ground vehicles using 2D sequential laser data[J]. Sensors, 2013, 13(1): 1102-1120.
- [14] Kellner M, Hofmann U, Bouzouraa M E, et al. Laserscanner based road curb feature detection and efficient mapping using local curb descriptions[C]//17th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, 2014: 2602-2609.
- [15] Xia Y, Pengbo S, Jie L, et al. Traversable detection in semi-structured environment by using 2D sequential laser data for a small mobile robot[C]//Control and Decision Conference (CCDC), 2016 Chinese. IEEE, 2016: 5308-5313.
- [16] Zhao G, Yuan J. Curb detection and tracking using 3D-LIDAR scanner[C]//2012 19th IEEE International Conference on Image Processing. IEEE, 2012: 437-440.
- [17] Fernández C, Llorca D F, Stiller C, et al. Curb Detection Method Based on Curvatures for Urban Autonomous Navigation[J].
- [18] Chen T, Dai B, Liu D, et al. Velodyne-based curb detection up to 50 meters away[C]//2015 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2015: 241-248.
- [19] Hata A Y, Osorio F S, Wolf D F. Robust curb detection and vehicle localization in urban environments[C]//2014 IEEE Intelligent Vehicles Symposium Proceedings. IEEE, 2014: 1257-1262.
- [20] Levinson J, Thrun S. Robust vehicle localization in urban environments using probabilistic maps[C]//Robotics and Automation (ICRA), 2010 IEEE International Conference on. IEEE, 2010: 4372-4378.

# Application and Structure Design of Polyvinylidene Fluoride Membranes in Road Energy Collection

Xiao Jian<sup>1</sup>, Zhang Yufei, Zou Xiang, Li Pengfei, Gen Hongyang, Huang Linlin  
*School of electronics and control engineering, Chang'an University, Xi'an 710064*

**Abstract.** With the promotion of green energy, road piezoelectric system is causing more and more people's attention. In this paper, the finite element simulation of polyvinylidene fluoride (PVDF) composites is studied. Four different types of protective measures are designed and tested on asphalt pavements. The piezoelectric output of the piezoelectric sensor is evaluated. The road piezoelectric performance of each protection measure is studied and analyzed. The experimental results show that the PVDF polymer film with length of 20 mm, width of 20 mm and thickness of 190  $\mu\text{m}$  could meet the requirements of asphalt pavement power generation system. The epoxy resin encapsulation structure has better piezoelectric performance.

**Keywords.** polyvinylidene fluoride, piezoelectric, asphalt pavement, energy collection

## 1. Introduction

In recent years, with the rapid development of society, the use of piezoelectric materials on the collection of energy more and more people's attention, energy recovery has also been seriously attention. China's highway mileage and the number of motor vehicles is also growing, over the years the vehicle has accumulated a lot of frequent mechanical vibration energy. More and more people are beginning to focus on converting the energy of road mechanical strain into electricity. Piezoelectric materials have the ability to convert mechanical energy into electrical energy, and there are various vibration modes, the more commonly used electromechanical coupling mode is 31 modal and 33 modal.

More and more people from the piezoelectric principle, structure, simulation and so do in-depth study. Zhi-Min Dang <sup>[1]</sup> and so on mainly from carbon nanotubes and titanium oxide barium synthesis of composite materials to study its dielectric properties. Sang Goo Lee <sup>[2]</sup> studied the effect of calcium fluoride (CaF<sub>2</sub>) on polyvinylidene fluoride (PVDF). Sun <sup>[3]</sup> et al. Investigated the effect of the multilayer structure of the pure PVDF layer on the insertion of carbon nanofiber (CNF) / polyvinylidene

---

<sup>1</sup> Corresponding Author, Associate Professor, School of electronics and control engineering, Chang'an University, Middle Section of Nan Erhuan Road, Xi'an 710064, Shaanxi Province, China. E-mail: [xiaojian@chd.edu.cn](mailto:xiaojian@chd.edu.cn)

This material is based upon the work supported by International Science & Technology Cooperation and Exchanges Plan in Shaanxi Province of China under Grant Number 2016KW-044

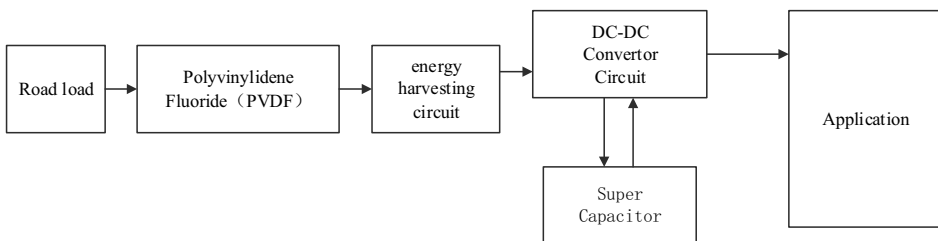
difluoride (PVDF) composites on the dielectric properties. At present, there are a variety of piezoelectric transducer structure in line with the road piezoelectric demand, Such as multilayer<sup>[4]</sup>, Moonie<sup>[5]</sup>, Cymbal<sup>[6]</sup>, Bimorph<sup>[7]</sup>, PZT piles<sup>[8]</sup> and so on.

A wide range of piezoelectric materials, in addition to commonly used piezoelectric ceramic materials, some polymer films have excellent stretch extensibility, Has a good piezoelectricity after the electric field polarization treatment. At present found the polymer piezoelectric film has PVF<sub>2</sub>, PVF, PVC, PMG, Polycarbonate and so on. Which PVF<sub>2</sub> higher piezoelectric constant it is a polymer organic semi-crystalline polymer, Crystallinity of 50%, according to the need to make thin film, tubular and so on. Lou<sup>[9]</sup> et al. Used the finite element method to analyze the induction and linear ability of PZT piezoelectric microcantilever. Han<sup>[10]</sup> et al. Analyzed the force-electric coupling of PVDF by ANSYS finite element method. PVDF piezoelectric sensitivity than the binary system piezoelectric ceramic PZT ten times higher, its high mechanical strength, soft and easy to process, with good winding and ductility.

PVDF piezoelectric film is a kind of polymer piezoelectric material, it has a unique dielectric effect, piezoelectric effect, thermoelectric effect. Compared with the traditional piezoelectric materials (such as piezoelectric ceramics) with high power conversion sensitivity, high mechanical properties, light weight, impact resistance and other advantages. In recent years in the mechanical, acoustic, optics, electronics, measurement, infrared, safety alarm, health care, military, transportation, information engineering, office automation, marine development, geological exploration and other technical fields are widely used.

## 2. Road energy collection system program

Compared to ordinary piezoelectric materials, the structure and dielectric properties of PVDF have been greatly improved. Thence, the PVDF material is applied to the road energy collection system. The road energy collection system is mainly composed of piezoelectric transducers(PVDF), energy harvesting circuit, charge storage (Super Capacitor). The flow chart is shown in Figure 1.



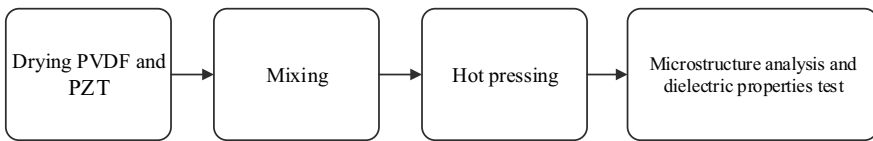
**Figure 1.** PVDF road energy collection system overall design block diagram.

When the car through the road will have some pressure on the road and vibration, although the pressure and vibration for our study of road energy harvest is very small, it is of great significance. The piezoelectric material is laid on the road surface or directly with the piezoelectric material as a pavement component, Through the road when the vehicle vibration to generate electricity, the resulting electrical energy is circuit-adjusted to provide operating power for the road test sensor, or to the energy

storage device for storage and utilization. The system comprises a power generating device, a piezoelectric conversion and an energy storage device including a piezoelectric material placed in a road surface. The piezoelectric energy collecting device is connected with an energy collecting circuit, the energy collecting circuit is connected with the energy storage device, the energy storage device is connected to and supplied with a road sensor or other external load. This study can effectively solve the mountain and long distance road along the road detection sensor network power shortage problem. In addition, we also on the road surface vibration frequency and stress level to do a study, Found a tire pressure of about  $0.7 \text{ MPa}$ , vibration frequency of about  $10 \sim 20 \text{ Hz}$  between.

### 3. Synthesis of PVDF Piezoelectric Materials

PVDF powder first dried. Pbbased Lanthanumdoped Zirconate Titanates (PZT) powder was dried at  $120 \text{ }^\circ\text{C}$ . PVDF and PZT grinding  $6 \text{ h}$ , so that its particle size small enough, The size of the particles was observed by electron microscopy (SEM). The dried PZT powder was mixed with the appropriate amount of alcohol for ultrasonic treatment for  $6 \text{ h}$ . After the proportion of PZT added to the PVDF powder, a dimethylformamide (DMF) in a mass ratio of 1: 9 was weighed into a round bottom flask, Then stirred, followed by sonication at room temperature for  $30 \text{ min}$ , After standing for quantitative coating, the oven temperature was set to  $200 \text{ }^\circ\text{C}$ , the time was set to  $3 \text{ h}$ , Finally, drop to room temperature after the film removed, packed back. The specific process shown in Figure 2.



**Figure 2.** PVDF piezoelectric film production process.

### 4. Simulation based on COMSOL

Finite element simulation is a very important part of the entire engineering design field. A lot of important engineering case analysis must first be designed with finite element simulation.

#### 4.1. Simulation of the design ideas

We used COMSOL Multiphysics software to simulate the PVDF material. First establish the force-electric coupling model, select the three-dimensional modeling, the model for the disc-shaped film. A thickness of  $190 \mu\text{m}$  and a radius of  $20 \text{ mm}$ . Ignore silver electrode and internal loss.



### 5.2. Structure Design of PVDF

Considering that the piezoelectric element is in direct contact with the stone, Particularly, under the load rolling, the piezoelectric element structure is liable to be damaged due to the stress concentration. As shown in Figure 4, the life of the unprotected PVDF piezoelectric material is too low to meet the requirements of road power generation.

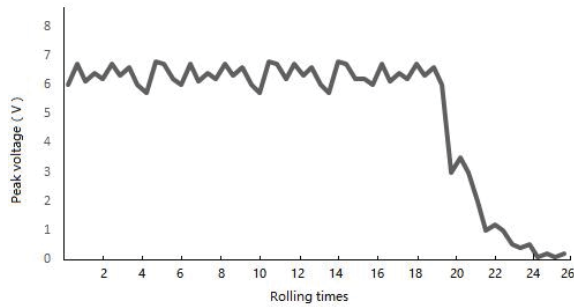


Figure 4. Unprotected PVDF output voltage.

So, we designed the protection of piezoelectric components. Taking into account the piezoelectric components in the power conversion requires a certain deformation of space. So we chose the high temperature corrosion-resistant elastic rubber as a flexible material, the use of excellent mechanical properties of phosphor bronze as a protective material, At the same time also tested the curing of epoxy resin. It is determined that rubber, phosphor bronze and epoxy resin are used as encapsulating materials. Epoxy resin packaging materials using epoxy resin and polyamide resin 2: 1 ratio, placed at 80 °C for 10 hours, to make a full chemical reaction, it should be placed at room temperature for 48 hours.

A PVDF piezoelectric film with a length of 20 mm, a width of 20 mm and a thickness of 190 μm was selected as the research object. Aiming at the problem of poor compression resistance of PVDF piezoelectric film, several improved processing schemes are proposed.

- PB-ER type protection measures: Phosphor bronze and elastic rubber structure, the rubber is located on the upper surface of PVDF piezoelectric film, copper film is located on the lower surface of PVDF piezoelectric film.
- D-ER type of protection measures: Epoxy resin double protection, PVDF piezoelectric film placed in the middle of the epoxy resin, up and down the thickness of 1-2 mm.
- D-R type of protection measures: Double rubber protection, PVDF piezoelectric film placed between the two elastic rubber, and sealed.
- D-PB type of protection measures: Double-layer phosphor bronze protection, with hot melt adhesive PVDF material fixed between the two pieces of copper.

### 5.3. Experiments under different protective measures

PB-ER, D-ER, D-R, D-PB four different packaging PVDF materials were placed on the rutting instrument on the experiment, the experimental results in Table 1.

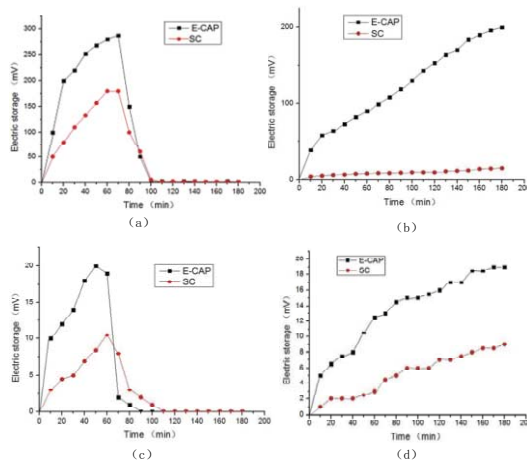
**Table 1.** Experimental test results.

Types of protection measures	The number of times the load is applied	The structural state after the experiment is completed
PB-ER	2500	Destruction
D-ER	5000	Complete
D-R	1800	Destruction
D-PB	5000	Complete

It can be seen from Table 1, after the completion of the experiment PB-ER, D-R type protection measures basically disappeared piezoelectric effect, D-ER, D-PB type protection measures in the rolling after the structure is complete, have voltage output. Experiments show that PB-ER, D-R type protection measures against damage performance, can not meet the asphalt pavement power generation requirements. And D-ER, D-PB type protection measures, with good resistance to destruction of piezoelectric materials can play a protective role.

*5.4. Analysis of experimental results*

In the energy collection circuit part, the PB-ER, D-ER, D-R, D-PB four kinds of protection mode of piezoelectric materials were used aluminum electrolytic capacitor (E-CAP) and super capacitor (SC) for energy collection.



**Figure 5.** with a super capacitor and aluminum electrolytic capacitors on different packaging structure of the piezoelectric film for electricity collection.

Under the same conditions, used automatic rutting test machine, made 200 min energy collection for PB-ER, D-ER, D-R, D-PB, the experimental results in turn correspond to Figure 5 *a, b, c, d*. The experimental results showed that the collection performance of aluminum electrolytic capacitor is superior to that of super capacitor in the collection of road piezoelectric, and the super capacitor attenuation is higher than that of aluminum electrolytic capacitor. At the same time, it can be seen that the power collected by the D-ER type protection measures is much higher than that of the D-PB type protection structure. Considering, D-ER structure protection is the best structural protection mode in four kinds of protection structure design.

## 6. Conclusion

The electric potential distribution of PVDF surface under ideal conditions is studied by COMSOL simulation. Four PVDF external protection structures are designed, the piezoelectric performance and the anti-damaging properties are tested by an automatic rutting test machine. The results show that the D-ER structure can ensure better power output and stronger anti-damage ability. This paper verifies the feasibility of PVDF materials in the field of road energy collection, and puts forward new possibilities in the field of road energy collection.

## References

- [1] Zhi-Min Dang, Sheng-Hong Yao, Jin-Kai Yuan, Jinbo Bai. Tailored Dielectric Properties based on Microstructure Change in BaTiO<sub>3</sub>-Carbon Nanotube/Polyvinylidene Fluoride Three-Phase Nanocomposites [J]. American Chemical Society, 2010, 114, 13204.
- [2] Sang G L, Ha J W, Sohn E H, et al. Enhancement of polar crystalline phase formation in transparent PVDF-CaF<sub>2</sub> composite films[J]. Applied Surface Science, 2016, 390:339-345.
- [3] Sun L L, Li B, Zhao Y, et al. Structure-induced high dielectric constant and low loss of CNF/PVDF composites with heterogeneous CNF distribution[J]. Nanotechnology, 2010, 21(30):305702.
- [4] A. Heinzmann, E. Hennig, B. Kolle, D. Kopsch, S. Richter, H. Schwotzer and E. Wehrsdorfer, ACTUATOR 2002, 8th Int. Conf. on New Actuators, Bremen, Germany (2002).
- [5] A. Dogan, "Flexensional 'Moonie and cymbal' actuators", a thesis in materials program submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, (1994): 63-66 .
- [6] J. F. Tressler, A. Dogan, J. F. Fernandez, J. T. Fielding, Jr, K. Uchino and R.E. Newnham, Proceedings of the 1995 IEEE Ultrasonics Symposium, Nov. 7-10, Seattle, U.S.A.(1995):897-900.
- [7] S. Roundy, "Energy scavenging for wireless sensor nodes with a focus on vibration to electricity conversion", a dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Engineering-Mechanical Engineering in the graduate division of the University of California, Berkeley, California (2003).
- [8] Hongduo ZHAO, Luyao Qin, Yujie Tao, Jianming Ling. Study on Structure of PZT piles Based Transducer for Harvesting Energy from Asphalt Pavement, 2013 IJPC-International Journal of Pavements Conference, São Paulo, Brazil, (2013):1-9.
- [9] Lou L F, Yang Y T, Zhang J Q, et al. Application of Ansys to the Piezoelectric Analysis of PZT Piezoelectric Thin Film Microsensor[J]. Mechanical Science & Technology, 2005.
- [10] Han J, Wan Z, Cheng L, et al. The Coupling Structure Modal Analysis of PVDF Piezoelectric Wafer Oscillator[C]// Third International Conference on Instrumentation, Measurement, Computer, Communication and Control. IEEE, 2014:1253-1256.



# Research on Setting and Benefit Evaluation of Integrated Waiting Area Based on VISSIM Platform

XinChao CHEN<sup>a,1</sup>, Si QIN<sup>b</sup>, Jian ZHANG<sup>a</sup> and Hong PAN<sup>c</sup>

<sup>a</sup>*Jiangsu Key Laboratory of Urban ITS, School of Transportation, Southeast Univ.; Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies; Jiangsu Province Collaborative Innovation Center for Technology and Application of Internet of Things; Research Center for Internet of Mobility, Southeast Univ., Si Pai Lou #2, Nanjing 210096, China*

<sup>b</sup>*School of Highway, Chang'an Univ., South section of S.2nd Ring Rd., Xi'an, Shaan'xi Province 710064, China*

<sup>c</sup>*Shanghai Mei Hui Software Co., Ltd*

**Abstract.** This paper research on the setting condition and benefit of the integrated waiting area (IWA) using VISSIM platform. The traffic organization and detailed setting conditions of the IWA is analysed. At the end, the capacity, delay is selected as the index to evaluate the performance. The results show that the IWA can effectively improve the traffic capacity and reduce the vehicle delay.

**Keywords.** Traffic organization, integrated waiting area, capacity, delay, VISSIM

## 1. Introduction

Recent years, integrated waiting area (IWA) has been applied in Shanghai, Xiamen and other cities, which make fully use of the advantages of waiting area, tidal lane and other basic principle of traffic control method, and then, the traffic flow through the intersection more quickly and efficiency (Wang J Y, Wei W et al, Zhong Z, Liu H, Ma W, Jiang J S, Dong L Y). This control model effectively improve the intersection space resource utilization, traffic capacity and reduce traffic delays. As one of the effective intersection channelization methods, more and more IWA is conducted in the urban intersection for reduced the confluence, diversion, weaving characteristic of traffic flow.

This paper proposed a method for setting IWA where left and straight traffic flow can smoothly switch rather than division by the lane line and special phase. Compared with the general waiting area, the IWA can fully use of the resources of the intersection, improve the utilization rate of road. This paper is organized as follows: the following

---

<sup>1</sup> Corresponding author. Tel.: +86 15720615913; E-mail address: cxc964809270@gmail.com.

section presents the current situation and development trend of the comprehensive waiting area. Then, in Section 3, the setting method and benefit evaluation index was conducted. In Section 4, an example is conducted to verify the correctness of the algorithm described previously. The final section then concludes the paper and recommends further research issues.

## 2. Literature Review

Sun and Chen (2015) study the setting conditions of IWA and the influence on capacity and delay by VISSIM platform based on the investigation of the specific characteristics of the IWA. Pillai et al. (1994) proposed -MILP methods for obtaining the maximum wire bandwidth (Mixed Integer Linear Programming). Shebeeb et al. (1995) study the left-turn lane of the intersection and the analysis results show that the intersection can effectively organize traffic, improve the traffic efficiency of vehicles of the left turn traffic flow with the special left turn lanes. K. Larry et al. (1997) conducted the dynamic programming based on the concept of a general algorithm, real-time control on the algorithm can be applied in the intersection, and a series of indicators of delay and queue length are optimized. Ma (2001) research the single signal control intersection using the optimization of traffic design snag signal optimization by the simulization platform, the traffic organization method of the signal intersection is proposed and the performance is pretty good. Gao et al. (2003) optimize the Webster theory and put forward the formula of signal for urban traffic intersection in china. Wang et al. (2009) analyzed the intersection geometry, sets up a queue model for left turn vehicles using the wave theory and cumulative curve. Current research on the existing signal intersection is mainly focused on the left turn waiting area and straight waiting area. In terms of signs and markings, the relevant reference standard is not given. In spect of signal phase, the adjustment of signal period is ignored. Although the IWA has been widely used in our country, there are still many details need to be studied.

## 3. Methodology

### 3.1. Notations and Assumptions

The notations used throughout the paper are listed as follows unless otherwise specified.

$g$	The length of the green signal in a cycle (s)
$G$	Main signal straight green time in a cycle (s)
$Q$	number of arriving straight traffic vehicles (pcu/h)
$s$	saturation flow of a through lanes (pcu/h)
$n_T$	number of straight lanes after the the first stop line (pcu)
$N_T$	number of straight lanes after the second stop line (pcu)
$t$	the difference between pre straight green signal time and the main straight green signal time (s)

- L the length of integrated waiting length (m)
- $L_T$  the length of integrated waiting area for straight traffic flow (m)
- $L_L$  the length of integrated waiting area for left turn traffic flow (m)
- d Queue length of straight vehicle arriving second stop lines in a signal cycle (m)
- v average travel speed in integrated waiting area (km/h)
- $w_1$  speed of aggregation wave of straight traffic flow after the second stop line (km/h)
- $w_2$  speed of dissipation wave of straight traffic flow after the second stop line (km/h)
- $\Delta t_1$  The time of the last straight vehicle arrives the end of the queue in pre signal (s)
- $\Delta t_2$  The time of dissipation wave of straight traffic flow to the end of the queue (s)
- $k_j$  Blocking density of traffic flow (pcu/km)
- K Initial Vehicle density in integrated waiting area (pcu/km)

### 3.2. Length of the Integrated Waiting Area

Traffic flow wave theory will be applied to discuss the length of the IWA, that is, the queue of straight vehicles in the second parking line. first, the length of the waiting area need to determined according to experience, and then whether the length of the area meet the waiting length need to be examed according to the following method.

When pre signal start, the straight traffic flow travels to the second stop line from the first stop line with at speed v. Then, the aggregation wave formed. The speed of the aggregation wave is described as:

$$w_1 = \begin{cases} \frac{-sn_T}{(N_T - n_T)K_j}, & \text{when the traffic flow arriving the first stop line is saturated} \\ \frac{-Q}{(N_T - n_T)K_j}, & \text{when the traffic flow arriving the first stop line is unsaturated} \end{cases} \tag{1}$$

When the main signal start, the straight traffic flow begins to dissipate, and then the dissipating wave formed. The speed of the dissipate wave is described as:

$$w_2 = \frac{-sN_T}{N_T(K_j - K)} \tag{2}$$

The required time  $\Delta t_1$  for the straight vehicle to reach the end of the queue in the green time of pre signal is described as:

$$\Delta t_1 = \begin{cases} \frac{L_T - \left(g - \frac{L_T}{v}\right) |w_1|}{v + |w_1|}, & \frac{L_T}{v} \leq g \\ \frac{L_T}{v} + \frac{g |w_1|}{v + |w_1|}, & \frac{L_T}{v} > g \end{cases} \tag{3}$$

The required time  $\Delta t_2$  for dissipate wave of the straight vehicle to reach the tail of queue is described as:

$$\Delta t_2 = \begin{cases} \frac{L_T - \Delta t_1 v}{w_2}, & g + \Delta t_1 < t + \frac{\left(t - \frac{L_T}{v}\right) |w_1|}{|w_2| - |w_1|} \\ \frac{\left(t - \frac{L_T}{v}\right) |w_1|}{|w_2| - |w_1|}, & g + \Delta t_1 \geq t + \frac{\left(t - \frac{L_T}{v}\right) |w_1|}{|w_2| - |w_1|} \end{cases} \tag{4}$$

The queue length  $d$  of the straight vehicle after the second stop line is described as:

$$d = \Delta t_2 \cdot w_2 \tag{5}$$

Therefore, the length of the IWA for straight vehicles is described as:

$$L_T \geq d \tag{6}$$

### 3.3. Phase




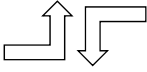
In order to predict the speed, there are some presumptions before it kicks off.

- Because the right turning vehicle can be released in the same phase with the straight traffic flow, and the whole control strategy is not affected, so to simplify the expression, only straight and left turning vehicles are considered in the model.
- acceleration time and deceleration time of the vehicle and the lane changing time is Ignored in the model.
- All the drivers are follow with the control plan, and automatically queue in the waiting area, so that each lane can be fully utilized in the IWA.

### 3.4. Phase Setting Rule

In order to fully reflect the function of the IWA, special left-turn phase should be utilized. The main signal has four phases and the initial phase sequence is shown in Ta-

**Table 1.** phases diagram of the IWA.

Phase				
Intersection signal	Straight north-south traffic	Straight east-west traffic	Turn left north-south traffic	Turn left east-west traffic
Waiting area pre-signal	Straight traffic	Left turn traffic	Left turn traffic	Straight traffic

ble 1. The signal period of the pre signal is consistent with the straight phase and left-turn phase.

### 3.5. Phase Length Constraint

When the main signal and the pre signal are performing, each phase start time of the main signal and the pre signal is not necessarily same with each other.

#### 1. Cycle constraint

The cycle time of main and pre signals should be equal.

#### 2. Empty constraint in the IWA

In order to ensure the coordination between the pre signal and the main signal as well as all the vehicles in the area to be emptied, the pre signal ahead time  $t$  must be satisfied.

$$g + \frac{L_T}{v} < t + G \tag{7}$$

#### 3. Traffic capacity constraint

In order to ensure that the vehicle is not stranded in the integrated area, the main signal capacity should be greater than the pre signal capacity, the relationship is described as:

$$sn_T \frac{g}{C} < sN_T \frac{G}{C} \tag{8}$$

### 3.6. Selection of Evaluation Index

The mainly benefit of the IWA is reflected as high capacity and low delay, this paper selects the capacity, delay as the evaluation index, and then, the comparative analysis is adopted to comparative the advantages and disadvantages of the method.

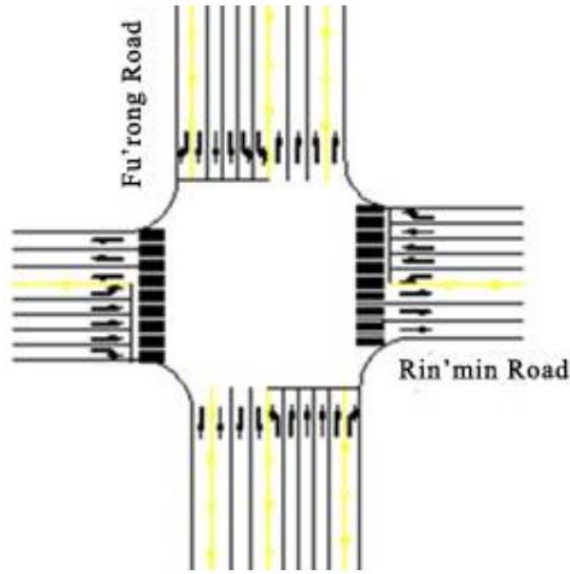


Figure 1. The lane layout of the intersection.

## 4. Case Study

### 4.1. Data Source

The cross intersection is located in Tian'xin district, Changsha City, which is one of the main intersection between the main expressway named Ren'min Road and the main distributed Road named Fu'rong road. The details of the lane layout is described as Fig. 1.

### 4.2. Phase Parameter

Compared with other entrance, the capacity of the north is extremely giant than others, which lead to a large number of twice parking. This paper choose the IWA to release the delay and enhance the capacity. According to the above, The length of signal design method of the IWA has already been analyzed, from the formulas (1)–(3), the length of the IWA is described as:

$$L = 50 \text{ m}$$

Then, The Webster method is adapted to calculate main and pre signal cycle of the control model. In order to coordinate with the main-phase, the straight and the turn left phase of the pre phase must satisfy the formulas (5)–(8). The calculation results are described as follows:

$$C = 177s; t_T = 29s; t_L = 27s$$

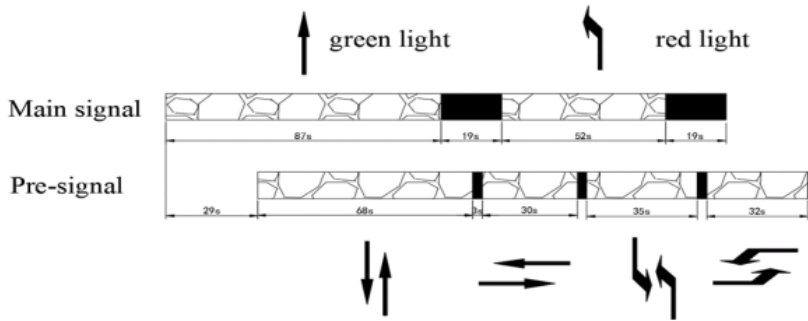


Figure 2. The signal timing diagram of the intersection.

Table 2. Simulation result of the evaluation before and after setting the IWA.

Direction of traffic flow	Evaluating index	Present situation	Optimization with IWA	Optimization rate
Left turn of north entrance	Delay (s)	51.4	42.6	-17.1%
	Capacity (pcu/h)	554.0	846.0	52.3%
Straight of the north entrance	Delay (s)	24.7	18.6	-24.6%
	Capacity (pcu/h)	1470.0	2495.0	69.7%
Left turn of south entrance	Delay (s)	50.6	65.9	30.2%
	Capacity (pcu/h)	169.0	172.0	1.8%
Straight of the south entrance	Delay (s)	27.2	21.6	-20.1%
	Capacity (pcu/h)	1960.0	1996.0	1.8%
Left turn of east entrance	Delay (s)	47.6	41.6	-12.6%
	Capacity (pcu/h)	284.0	252.0	-11.3%
Straight of the east entrance	Delay (s)	43.5	29.5	-32.1%
	Capacity (pcu/h)	873.0	774.0	-11.3%
Left turn of west entrance	Delay (s)	32.2	33.6	4.3%
	Capacity (pcu/h)	310.0	275.0	-11.3%
Straight of the west entrance	Delay (s)	37.2	24.8	-33.3%
	Capacity (pcu/h)	873.0	774.0	-11.3%
All of the entrance	Delay (s)	33.2	26.4	-20.4%
	Capacity (pcu/h)	6493.0	7584.0	16.8%

Meanwhile, the main phase time and the pre phase time scheme are described as shown Fig. 2.

The control model mentioned before is conducted by VISSIM platform, the simulation results are shown in Table 2.

From the Table 2, the performance of the IWA is pretty good than the previous. The delay of the intersection is reduced by 20.4% and the capacity is enhanced by 16.8%. At last, the assessment indices of the IWA are verified through case studies and the results are acceptable.

## 5. Conclusion

In order to make full use of road space-time resources, reduce vehicle delays, improve intersection capacity, this paper proposed integrated waiting area, which is important to improve the utilization rate of road, reduce traffic jam, and maximize the safety of driving. However, due to the limitations of the test conditions, the study has not considered the interference factors of un-motor vehicles and pedestrians, which will be of great value for future research.

## Acknowledgements

This study is financially supported by the Information Technology Research Project of Ministry of Transport of China (No. 2015364X16030), Fundamental Sciences of Southeast University (2242015K42132), and the National Natural Science Foundation of China (No. 51308115).

## References

- Wang J Y, Wei W. Approach function split method at signalized intersection [J]. *Jilin Daxue Xuebao*, 2007, 37(6):1278–1283.
- Zhong Z, Liu H, Ma W, et al. An Optimization Method of Dynamic Lane Assignment at Signalized Intersection [C]//*Intelligent Computation Technology and Automation (ICICTA)*, 2008 International Conference on. IEEE, 2008:1277–1280.
- Jiang J S, Dong L Y. Modeling and Simulation of Versatile Waiting-Area at Isolated Signalized Intersection [J]. *Journal of Shanghai University*, 2012.
- Pillai R S, Rathi A K, Cohen S L. A restricted branch-and-bound approach for generating maximum bandwidth signal timing plans for traffic networks [J]. *Transportation Research Part B Methodological*, 1998, 32(8):517–529.
- Shebeeb O. Safety and efficiency for exclusive left-turn lanes at signalized intersections [J]. *Ite Journal*, 1995, 65(7).
- Sen S, Head K L. Controlled Optimization of Phases at an Intersection [J]. *Transportation Science*, 1997, 31(1):5–17.
- Inoue T, Futamura M, Yokota T, et al. Traffic signal offset design using event scanning method simulation of traffic congestion [J]. *Electronics & Communications in Japan*, 1998, 82(8):1–13.
- Kim J, Kim B M, Huh N C. Genetic algorithm approach to generate rules an membership functions of fuzzy traffic controller [C]//*The IEEE International Conference on Fuzzy Systems*. IEEE, 2001:525–528.
- Fhwa B U S. Federal Highway Administration. Manual on Uniform Traffic Control Devices for Streets and [J]. 2014.
- Al-Salman H S T, Salter R J. Control of right-turning vehicles at signal-controlled intersections [J]. 1974.
- Abdulhai B, Pringle R, Karakoulas G J. Reinforcement Learning for True Adaptive Traffic Signal Control [J]. *Journal of Transportation Engineering*, 2003, 129(3):278–285.
- Chen H. Simulation and Evaluation of the Left-turn Waiting Zone at Signalized Intersection [J]. *Traffic & Transportation*, 2012.
- Wang D, Lili L I, Chen Y. Critical Condition for Setting Left-turn Waiting Area [J]. *Journal of Highway & Transportation Research & Development*, 2009, 26(11):132–135.



# Implementation and Adoption of E-HRM in Small and Medium Enterprises of Pakistan

ABDUL WAHEED<sup>a1</sup>, SALMA WAHEED<sup>a2</sup>, JAWAD KARAMAT<sup>a3</sup>, NAVEED  
AHMAD<sup>a</sup>, ABDUL MAJEED<sup>b</sup>

<sup>a</sup> School of Management, Northwestern Polytechnical University, Xi'an, China

<sup>b</sup> School of Economics and Management, Lahore University, Pakistan

**Abstract.** The purpose of this study is to investigate the predictors of e-HRM implementation in Small medium enterprise (SMEs) in manufacturing sector. Three main predictors including conventional HRM practices, availability of resources and employees attitude towards e-HRM were analyzed. Data was collected through a questionnaire survey and 500 employees participated in this survey. Results reveal that e-HRM heavily depends on employee attitude towards e-HRM and availability of resources. Conventional HRM practices including training and development, compensation and benefits, and performance appraisal also have an important role. The readiness of implementing e-HRM practices in small and medium enterprises heavily dependent on expertise, financial and technical resources. This study will help managers to develop strategy for effective implementation of e-HRM in Small medium enterprises (SMEs) in manufacturing sector.

**Keywords:** E-HRM, Conventional HRM, Small Medium Enterprises, Pakistan SME's, Implementation.

## 1. Introduction

Electronic Human Resource Management (e-HRM) is a buzz in recent years due to its continuous expansion, implementation and application. Many surveys conducted on HR development and they advocated that plenty of organizations accepted e-HRM and its applications in their business constantly [1]. Hence, an escalating quantity of practitioner reports give subjective proof that e-HRM is going to be a common and may guide to be extraordinary changes [2]. As the world is growing to be complicated and organizations are overcoming these complications by using Information Technology (IT) applications in recent years [3, 4]. The addition of IT gradually become very famous in the Human Resource Management (henceforth called HRM) field, which has concluded in a wide development of the findings into Electronic Human Resource Management (henceforward called e-HRM). Although rapid growth

---

<sup>1</sup> PhD students in School of Management, Northwestern Polytechnical University, Xi'an, China, Email: waheed\_2506@mail.nwpu.edu.cn

<sup>2</sup> PhD students in School of Management, Northwestern Polytechnical University, Xi'an, China, Email: salma@mail.nwpu.edu.cn

<sup>3</sup> PhD students in School of Management, Northwestern Polytechnical University, Xi'an, China, Email: naveedahmad@mail.nwpu.edu.cn

of the researches have given an indication of the worth construction of e-HRM.[3, 5] . An investigated that many organizations use specific amount in their finances on the investment of e-HRM solution every year [6, 7]. In this research paper needs to be known that whether applications and implications of e-HRM are booming for Pakistan manufacturing business. As researches investigated and proofs that some organizations are booming in managing the implementation of the e-HRM solution and permit it contribution in HRM and business performance, whereas rest of other organizations are a smaller amount or not booming in running the implementation to the advantage of the organization. The motive for these deviations remains often unclear. The use of IT for HRM department has progressively more grown to get administrative- and strategic benefits over the past few years [8-10]. Need to know that as transaction processing systems IT implementations started in the 1950s and 1960s and are now extended to cloud-based enterprise wide systems[11]. IT intended to assist with human resources administrative functions. With e-HRM has been recognized as a catalyst towards attaining business strategies, there is little significance among local business vendors of Pakistan to admit it. Some small medium enterprises (SMEs) assert that this fact will reduce the number of employees in the HR department. While some SMEs analysis that e-HRM would reduce costs and motivate a more strategic approach in human resource management (HRM), others views that the implementation and maintenance of e-HRM systems engage massive investment. Thus, to what level has e-HRM already been implemented in SMEs manufacturing business in Pakistan? This study anticipates exploring the keenness and feasibility of implementing e-HRM in the small and medium sized enterprises in Pakistan. Firstly, it attempts to find out if a large number of small and medium sized enterprises (SMEs) in Pakistan are involving conventional HRM than e-HRM. To what level have SMEs in Pakistan employ e-HRM tools in the administration of its HR functions? The paper will focus on main areas of conventional HRM, which are considered to have significant impact on the competitiveness of the business namely, recruitment, compensation and benefits, performance appraisal, communication, and training and development. It is believed that the viability of implementing e-HRM is very much dependent on the availability of resources. With the availability of expertise, technological and financial resources, companies would maintain the appearance of e-HRM. E-HRM will modify the role of HR. When e-HRM launched, employees will have to adjust to be added responsibilities they have for HR issues. Thus, the approach and readiness of the employees to adjust to this modification would affect the implementation of e-HRM.

## **2. Literature Review**

### *2.1 E-HRM*

E-HRM is defined as “the integration of IT and the HRM field of scholarly inquiry. This focuses on all the HRM content that is shared through IT that aims to make HRM processes distinctive and consistent, more efficient, high in quality and which create long-term opportunities within and across organizations for targeted users” [12]. It provides information to employees through the Internet. So E- HRM is a way of implementing the HR strategies, policies and practices with the help of technology. The e-HRM technology sustains the HR functions through web technology based channels [13, 14].

## 2.2 Small and Medium Sized Enterprises (SMEs) in Pakistan

Abbreviation of SMEs is small and medium sized enterprises. SMEs universally are not defined. In Pakistani point of views, SMEs lie on the number of employees up to 250 people, paid-up capital up to Rs.25 million and annual sales up to Rs.250 million [15]. This definition was on behalf of a review process of spanning over two years go after by scrutiny and refining at different levels of government before its conclusions and authorization by the Federal Cabinet in 2007 [16].

In Pakistan different institute have different definition of SMEs, such as State bank of Pakistan (SBP) refers SME can be categorized into three classes, micro, small, and medium enterprises (State Bank of Pakistan, 2010). And it also characterizes SME as, “any private economic establishment engaged in manufacturing, trading or service providing business with net annual turnover or sales up to Rs.300 million in the current fiscal year; or any manufacturing entity having total assets up to Rs.100 million excluding land and buildings with maximum 250 employees or any trading or service concerning total assets up to Rs.50 million excluding land, buildings and with maximum 50 employees”. In the perspective of Pakistan, no consistent and average definition of SMEs has been accessible that resulted the progress and success of the SMEs [17]. Practitioners observed that HRM studies primarily focused on large firms or organizations in the past. SMEs seldom obtained a point out in HRM literature regardless of the fact that SMEs in the majority countries contribute towards employment [18]. According to the Economic Survey of Pakistan (2008-09) (Government of Pakistan - Ministry of Finance, 2009) SMEs are rising faster, contributed almost 99.06% in 2008-2010, dependable for the 30% of GDP. SME considers in Pakistani companies when it has employees up to 250, paid up capital of Rs. 25 million and annual sales of Rs. 250 million [19]. Despite of its entire growth still SMEs needs to be more improvement [20]. Moreover, the details about the composition of SMEs in other countries are given in table 1.

## 2.3 Conventional HRM in Pakistan's SMEs

Many studies show that the local or wide area networks and internet or IT system unlocked the doors for the HRM development. These assimilated system made HRM functions more reliable and efficient for the business [21]. HR efforts in SME sector participates a fundamental role in economic development without any gap for their personal growth [22]. Research investigated that large business adopt more refined and socially receptive HR practices [23] because they are more observable and are under more strain to gain authenticity [23] Production, finance, marketing and accounting as personnel practices in SMEs were judged by managers to be insignificant compared with other areas [24]. In addition, SMEs in Pakistan have great approach for greater economic and social development [25] but the lack of employ of best practices, lack of sophistication, and lack of attention to the documented relationships that have been revealed between HRM practices and organizational results in larger firms [26] Above literature helps to formulate the following hypothesis:

H1: There is a large number of SMEs in Pakistan are practiced conventional HRM rather than e-HRM.

**Table 1.** Definition of SMEs in various Asian countries

Country	Definition of SMEs	Measurement
People's Republic of China	Varies with industries, less than 100 employees	Employment
Indonesia	Less than 100 employees	Employment
Korea	Manufacture: less than 300 employees Service: less than 300 employees	Employment
Malaysia	Varies, turnover: less than RM25 million and 150 employees.	Shareholders, Funds and Employment
Pakistan	Less than 250 employees and less than 100 million PKR assets for manufacturing. Less than 50 employees and less than 50 million PKR for trade/services. Net sales less than 300 million PKR	Employment and assets
Thailand	Less than 200 employees, less than 200 million Baht assets	Assets and Employment

#### 2.4 Availability of Resources

A well-supported and improved small business sector is likely to maintain the economic development process in the same manner as a large business and SMEs give a variety of benefits [27, 28]. Many findings suggested that smaller entrepreneurial firms usually have key tackles with HRM for the motive that the small size of the firm often does not assure appoint professionals completely dedicated to HRM proceedings [29]. According to Brand and Bax [30] many small firms meet severe HR problems, at the same time as; HRs is taking a critical part in rising and maintaining their competitive benefits. They have also concluded that the existing knowledge on HRM in small firms is particularly explanatory and fragmented Institutionalisms. Hall and Soskice [31] Dispute that national factor such as economics, governance, financial and legal systems and trade unions, which together form the national business system. SMEs have traditionally taken as an important part in contributing to economic progress of many countries around the world [32]. SMEs practically rapid growth and resource shortages are also more dedicated to e-HRM schemes. In such a position, manual processes would reason a backlog in HR functions and would impede further improvement and growth. The rapidity and effortlessness at which e-HRM could handle HR functions validate its introduction. In base of the quick technological changes, the implementation of e-HRM plans offers enormous opportunities to recover on HRM practices. A survey done in 2005 showed that the major problems faced by the SMEs in Pakistan are lack of technical acquaintance, financial restraints, specific market, linkage between fasten companies and incapability to enter the export market [33]. Accordingly, it is reasonable to create the hypothesis that:

H2: There is a strong relationship between the availability of resources and the feasibility of implementing e-HRM.

Thus, the keenness of HR practitioners to squeeze technology and apply it to HR functions is one of the important factors in the beginning of e-HRM initiatives. Given the augmented highlighting in the exploit of information and communication

technology (ICT) in the working situation, it is rational to formulate the hypothesis that:

H3: Employees are ready and receptive to the implementation of e-HRM.

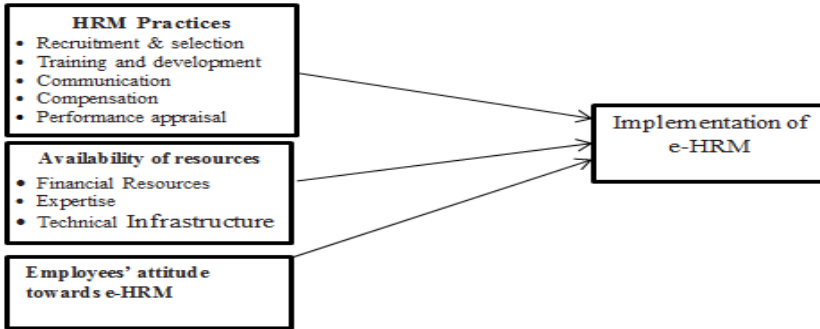


Figure 1. Conceptual Framework

### 3. Research Methodology

The collection of data for this research engaged both the gathering of primary and secondary data, but the basic method was primary data survey. The study began with a literature review of books, as well as articles in journals related to the research topic. Company records, newspaper articles, handbooks and magazines were other sources of secondary data. After searching the World Wide Web for information on the Internet, also served as a source for secondary data. Secondary data gathered through literature review and views of various writers on issues relating to HRM policies provided an in depth consideration of the study. A theoretical framework that obviously identified the variables was then formulated for the research.

The implementation of e-HRM is dependent on various factors, among them, the accessibility of resources and the approaches of the employees. Constructing companies of Pakistan are more likely to implement e-HRM if resources are available and employees view e-HRM positively. Therefore, the implementation of e-HRM is the dependent variable, and the availability of resources and employees' attitude the independent variables. Although this relationship can be said to embrace true generally, it is never the less conditional on the communication of e-HRM. Management's effectiveness in communicating e-HRM to the rest of the organization has a contingent effect on the independent variables and dependent variable relationship. The moderating variable in this relationship is conventional HRM. Primary data was collected from observations, administered questionnaires, and from individuals who provide information when interviewed. The Internet also served as a primary data source as questionnaires were administered over it. 500 Questionnaires written in English were sent through the mail, and electronically administered. A 5-point Likert scale allowed respondents to indicate how strongly they agree or disagree with the statements relating to the research question. Before analyzing the survey questionnaires, a pilot survey was conducted with a few selected constructing organizations and individuals. The collected raw data from the respondents was then transformed into

readable information and its relation to the research hypotheses were analyzed using SEM methodology [34]. The results of the survey only described the characteristics of Pakistan employees in the constructing companies at a septic point of time.

### 3.1 Results of Hypothesis

The first hypothesis examines whether a large number of SMEs in Pakistan are practicing conventional HRM rather than e-HRM. Five variables, namely recruitment and selection, training and development, communication, compensation and performance appraisal were used for the analysis. The mean for the five variables used to analyze if a large number of SMEs in Pakistan are practicing conventional HRM rather than e-HRM are tabulated as in Table below.

**Table 2.** Mean and Reliability Analysis

Serial	Variables	Mean	Cronbach's Alpha
	<b>HRM Practices</b>		0.78
1	Recruitment & selection	2.7	
2	Training and development	2.7	
3	Communication	3.8	
4	Compensation	2.9	
5	Performance appraisal	2.2	

Table 2 shows that the mean for four of the five variables is less than 3, that is towards the left side of the 5-point scale, denoting low usage of e-HRM among the respondent companies. This is especially so in training and development as well as in performance appraisal. Among the five variables, the use of electronic methods in companies' HR practices is most popular in communication, followed by recruitment and selection, and compensation. The general decision to reject the hypothesis that a large number of SMEs in Pakistan are practicing conventional HRM rather than e-HRM is if the mean for all the five variables is more than 3. As the average mean (2.91) for all the five variables is less than 3, the hypothesis that a large number of SMEs in Pakistan are practicing conventional HRM rather than e-HRM is substantiated. However, as the average mean for all the five variables is close to 3, it can be concluded that SMEs in Pakistan are moving towards digitizing their HRM practices. In general, priority would be given to recruitment and selection as well as compensation.

**Table 3.** Mean and Reliability Analysis

Serial	Variables	Mean	Cronbach's Alpha
	<b>Lack of resources</b>		0.84
1	Financial resources	3.2	
2	Expertise	3.1	
3	Technical infrastructure	3.8	

The second hypothesis examines if there is a strong relationship between the availability of resources and the feasibility of implementing e-HRM. There are three variables for the second hypothesis, namely, financial resources, expertise and technical infrastructure. Four items were formulated to test the feasibility to implement e-HRM.

Two items examine if the implementation of e-HRM is held back by financial constraints. One item was included to find out if the respondent companies have the expertise to implement e-HRM. Another two items were used to see if the respondent companies lack suitable technical infrastructure to implement e-HRM. The three variables used to analyze whether a relationship exists between the availability of resources and the feasibility of implementing e-HRM are summarized in Table 6 and further discussed below. From the above table, it can be concluded that most respondent companies agreed that it is feasible to implement e-HRM if the necessary resources such as financial resources, expertise and technical infrastructure are available. The average mean of the three variables (3.47) to the right side of the 5-point scale supports this. As far as the availability of resources is concerned, most of the respondent companies lack financial resources and expertise to implement e-HRM. The results also show that technical infrastructure is available in most of the respondent companies and is not a hindrance in the implementation of e-HRM. Moreover, the average mean of the three variables is slightly above 3 at 3.01, indicating that the lack of available resources for the implementation of e-HRM is not at the extreme end. With measures to overcome this, it may be feasible for more SMEs to implement e-HRM in the near future. To prove the third hypothesis that employees are ready and receptive to the implementation of e-HRM, seven items were included in the survey. The aim is to gauge the attitude of the employees towards the implementation of e-HRM and their readiness to adapt to the new system. On the first item as to whether e-HRM is crucial to their organizations, 38.3 per cent of the respondent companies is of the opinion that e-HRM is not crucial as opposed to 33.3 per cent that disagreed. Those that disagreed would be more receptive if e-HRM is implemented as they feel that e-HRM is crucial. The other companies would be less receptive as they seem contented with the current HRM practices. Hypothesis 1 proposed a relationship between HRM practices and e-HRM implementation in SMEs. The results showed significant positive relationship between HRM practices and e-HRM implementation (H1:  $\beta = 0.1479$ ,  $t=4.678$ ,  $p<0.001$ ). Hypothesis 2 predicted a relationship between availability of resources and e-HRM implementation. Results revealed a significant positive relationship between

**Table 4.** Mean and Reliability Analysis

Serial	Variables	Mean	Cronbach's Alpha
<b>Employee Attitude towards e-HRM</b>			0.93
1	E-HRM is not crucial to my organization	3.2	
2	The advantages of e-HRM are outweighed by the cost implication	2.9	
3	Our staff does not have relevant skills for e-HRM	2.9	
4	My organization lacks the technology needed for e-HRM	2.8	
5	I have concerns related to the lack of security of e-HRM	3.6	
6	E-HRM is too complex to comprehend/understand	2.4	
7	It is important that my organization introduce e-HRM in the next 12 months	3.5	

Availability of resources and e-HRM implementation (H1:  $\beta = 0.3487$ ,  $t=8.674$ ,  $p<0.001$ ). Hypothesis 3 proposed a relationship between employee's attitude towards e-HRM and e-HRM implementation. The results showed significant positive relationship

between employees' attitude towards e-HRM and e-HRM implementation (H1:  $\beta = 0.1479$ ,  $t=4.678$ ,  $p<0.001$ ).

**Table 5.** Mean, Standard Deviation and Correlation

Variables	HRM Practices	Availability of resources	Employees attitude towards e-HRM
HRM Practices	1		
Availability of resources	0.34*	1	
Employees attitude towards e-HRM	0.19*	0.27*	1

**Table 6.** Results of Hypothesis

Hypothesis	Variables	$\beta$	S E	t	p
H1	HRM Practices	0.1470	0.0358	4.678	0.000** *
H2	Availability of resources	0.3478	0.0346	8.674	0.000** *
H3	Employee Attitude towards e-HRM	0.2709	0.0679	6.467	0.000** *

#### 4. Discussion

This study found availability of resources and employees attitude towards e-HRM as the main predictors of e-HRM implementation. SMEs in Pakistan are laggards in the implementation of e-HRM. Traditional HRM practices are still backbone of HR departments in many industries. However, implementation of e-HRM in SMEs can give competitive edge to get some good staff. The results are contradictory with the previous result in other countries like Malaysia. Due to acceptance of hypothesis 1, it can be concluded that SMEs in Pakistan are moving towards digitizing their HRM practices. It will furnish e-HRM in SMEs of manufacturing industries in Pakistan. Companies lack suitable technical infrastructure to implement e-HRM. Moreover, due to results of hypothesis 2, it can be concluded that most respondent companies agreed that it is feasible to implement e-HRM if the necessary resources such as financial resources, expertise and technical infrastructure are available. The results of hypothesis 3 are consistent with the previous research. Employee's attitude towards implementation of e-HRM is positive. However, manufacturing companies would be less receptive as they seem contented with the current HRM practices. It strengthens the roots of e-HRM in SMEs in Pakistan.

#### 5. Conclusion

The purpose of this study is to check the predictors of e-HRM implementation in SMEs of manufacturing industries. This study is a pioneer attempt in implementing e-HRM practices in SMEs of Pakistan. Empirical results found that employees attitude towards e-HRM strongly influences the implementation of e-HRM. These results indicate that by integrating new technologies and knowledge, firms can increase of capability of improving existing HRM practices. In addition, utilizing e-HRM capability and



knowledge consistently, contribute for the enhancement of HR performance. In short, future is all about e-HRM.

## 6. Practical Implications

Implementation of e-HRM in the SMEs of manufacturing industry is the necessity. Companies need to transform from traditional HRM to e-HRM practices. Government has an important role for transforming and can act as a moderator by including the implementation of e-HRM in the future plans. Different linkage programs of SMEs with technical institutes can also help to easy implementation of e-HRM. Moreover, government can give incentives and subsidies to SMEs for implementation of e-HRM. Training programs can be organized to promote the e-HRM effectiveness in organizations. Government should organize conference and invite international HR professionals to transfer the HR skills and provision of training for effective implementation of e-HRM. E-HRM should be included as a subject in the university curriculum for preparing the fresh graduates in an effective manner.

## 7. Limitation and Future Research Direction

This study is based on a case study of SMEs in manufacturing sector in Pakistan and sample was selected from major SMEs of Pakistan. Future research can be expanded to other industries and types of firms. Therefore, application of this model needs to be tested in other types of firms. Data obtained in this study is cross sectional. Time series study can be potential research in the future. This study checked the effect of conventional HRM practices, availability of resources and employees attitude towards e-HRM. However, other emerging variables like work place democracy, organizational politics, Government support, adaptation of digital HRM techniques and employee relation climate can be potential predictors of e-HRM implementation.

## References

- [1] C. Crestone, "The Cedar Crestone 2005 Workforce Technologies And Service Delivery Approaches Survey," *8th Annual Edition*, 2005.
- [2] S. Strohmeier, "Research in e-HRM: Review and implications," *Human resource management review*, vol. 17, pp. 19-37, 2007.
- [3] D. Démeijer, "Making digital HRM work: A study in changes in perceived consequences of e-HRM in the past decade," University of Twente, 2017.
- [4] A. Waheed, M. Xiaoming, N. Ahmad, S. Waheed, and A. Majeed, "New HRM Practices and Innovation Performance; the moderating role of Information Technology Ambidexterity," 2017.
- [5] H. Ruël, T. Bondarouk, and J. K. Looise, "E-HRM: Innovation or irritation. An explorative empirical study in five large companies on web-based HRM," *Management revue*, pp. 364-380, 2004.
- [6] H. Ganzeboom, P. d. Graaf, and M. Kalmijn, "De culturele en de economische dimensie van beroepsstatus," *Mens en Maatschappij*, vol. 62, pp. 153-175, 2014.
- [7] N. Ahmad, N. Iqbal, R. Kanwal, H. Javed, and K. Javed, "The mediating role of employee engagement in relationship of internal branding and brand experience: Case of service organizations of Dera Ghazi Khan," *International Journal of Information, Business and Management*, vol. 6, pp. 26-41, 2014.
- [8] T. Bondarouk and H. Ruël, "Electronic Human Resource Management: challenges in the digital era," *The International Journal of Human Resource Management*, vol. 20, pp. 505-514, 2009.

- [9] S. Strohmeier, "Concepts of e-HRM consequences: a categorisation, review and suggestion," *The International Journal of Human Resource Management*, vol. 20, pp. 528-543, 2009.
- [10] T. Bondarouk, E. Parry, and E. Furtmueller, "Electronic HRM: four decades of research on adoption and consequences," *The International Journal of Human Resource Management*, vol. 28, pp. 98-131, 2017.
- [11] R. D. Johnson, K. M. Lukaszewski, and D. L. Stone, "The evolution of the field of human resource information systems: Co-evolution of technology and HR processes," *Communications of the Association for Information Systems*, vol. 38, pp. 533-553, 2016.
- [12] T. Bondarouk, R. Harms, and D. Lepak, "Does e-HRM lead to better HRM service?," *The International Journal of Human Resource Management*, vol. 28, pp. 1332-1362, 2017.
- [13] F. Sabir, M. Abrar, M. Bashir, S. A. Baig, and R. Kamran, "E-HRM IMPACT TOWARDS COMPANY'S VALUE CREATION: EVIDENCE FROM BANKING SECTOR OF PAKISTAN," *International Journal of Information, Business and Management*, vol. 7, p. 123, 2015.
- [14] N. Haider, N. Ahmad, O. Farooq, I. Rasheed, and S. Parveen, "The role of organizational resources and environment in Organizational performance and customer loyalty; service climate as mediator: A Study of Telecommunication Sector of Pakistan," 2014.
- [15] N. Kureshi, R. Mann, M. Khan, and M. Qureshi, "Quality management practices of SME in developing countries: a survey of manufacturing SME in Pakistan," *Journal of Quality and Technology Management*, vol. 5, pp. 63-89, 2009.
- [16] S. Saleem, "SMEDA SME Policy paper 2007-A Critical Review (An analytical commentary upon SME policy proposed by SMEDA Pakistan)," 2008.
- [17] M. K. Hashim and I. Osman, "An evaluation of the business practices in Malaysian SMEs," *Malaysian Management Review*, vol. 38, pp. 1-18, 2003.
- [18] S. P. Deshpande and D. Y. Golhar, "HRM practices in large and small manufacturing firms: A comparative study," *Journal of Small Business Management*, vol. 32, p. 49, 1994.
- [19] R. Bennett, "SME policy support in Britain since the 1990s: what have we learnt?," *Environment and Planning C: Government and Policy*, vol. 26, pp. 375-397, 2008.
- [20] S. W. H. Naqvi, "Critical success and failure factors of entrepreneurial organizations: Study of SMEs in Bahawalpur," *Journal of Public Administration and Governance*, vol. 1, pp. 17-23, 2011.
- [21] A. Moomal and M. Masrom, "ICT Development and Its Impact on e-Business and HRM Strategies in the Organizations of Pakistan," *Journal of Advanced Management Science Vol.*, vol. 3, 2015.
- [22] N. Khan, "HRM significance and SME sector," *Business Recorder*, vol. 11, 2011.
- [23] M. A. Bawa and J. Ali, "The challenges of globalization and the role of human resources," in *International Conference on Challenges of Globalization, Bangkok Thailand [On-line]*. Available from: <http://econ.tu.ac.th/iccg2.htm>, 1999.
- [24] G. M. McEvoy, "Small business personnel practices," *Journal of Small Business Management (pre-1986)*, vol. 22, p. 1, 1984.
- [25] N. R. Khan, S. M. Taha, A. M. Ghouri, M. R. Khan, and Y. C. Ken, "The impact of HRM practices on supply chain management success in SME," 2013.
- [26] M. A. Huselid, "The impact of human resource management practices on turnover, productivity, and corporate financial performance," *Academy of management journal*, vol. 38, pp. 635-672, 1995.
- [27] J. Abor and P. Quartey, "Issues in SME development in Ghana and South Africa," *International Research Journal of Finance and Economics*, vol. 39, pp. 215-228, 2010.
- [28] N. Ahmed, A. K. Khattak, N. Iqbal, O. Farooq, and J. Iqbal, "Role of celebrity endorsement upon consumer vanity with mediating role of materialism: Evidence from business students of Pakistan," *Journal of Business and Management Research*, vol. 4, pp. 98-107, 2014.
- [29] D. F. Kuratko, R. D. Ireland, and J. S. Hornsby, "Improving firm performance through entrepreneurial actions: Acordia's corporate entrepreneurship strategy," *The Academy of Management Executive*, vol. 15, pp. 60-71, 2001.
- [30] M. J. Brand and E. H. Bax, "Strategic HRM for SMEs: implications for firms and policy," *Education+ Training*, vol. 44, pp. 451-463, 2002.
- [31] P. A. Hall and D. W. Soskice, *Varieties of capitalism: The institutional foundations of comparative advantage* vol. 8: Wiley Online Library, 2001.
- [32] M. Kongolo, "Job creation versus job shedding and the role of SMEs in economic development," *African Journal of Business Management*, vol. 4, p. 2288, 2010.
- [33] N. A. Ismail, "Factors influencing AIS effectiveness among manufacturing SMEs: Evidence from Malaysia," *The Electronic Journal of Information Systems in Developing Countries*, vol. 38, 2009.
- [34] F. Shaheen, N. Ahmad, M. Waqas, A. Waheed, and O. Farooq, "Structural Equation Modeling (SEM) in Social Sciences & Medical Research: A Guide for Improved Analysis," *International Journal of Academic Research in Business and Social Sciences*, vol. 7, 2017.

# The Relationship Between Product Modularization and Open Innovation Based on Sports Industry

Min Zhang<sup>1</sup> and Xun Wen

*Department of Physical Economics and Management, Xi'an Physical Education University, Xi'an 710064, P.R. China*

**Abstract.** This article defines incremental innovation in modular organization as the independent innovation within the module and the synergetic innovation among modules. Intensity of knowledge interaction increases along with the transformation of independent innovation into synergetic innovation, while empirical data validate inverted U-shaped relationship between intensity of knowledge interaction and modular knowledge autonomy. The process of open innovation in modularization complies with the path of spiral and amplified evolution. Basing on case studies of Chinese sport industry, we analyzed the actual operation of open innovation model in sport modularization. Standards of Communication and operating system were both influenced by independent and synergetic open innovation.

**Keywords.** Product Modularization, Open Innovation, Information Transfer Process, Sports Industry

## 1. Introduction

Many scholars stressed the same problem that the capability creating and utilizing knowledge is a sustainable source of competitive advantage. For business, the companies need to integrate resources which engender a brand-new integration mechanism to ensure continuous and rapid innovation in order to survive in competition <sup>[1]</sup>. On the other hand, continual and intensive investment in R & D can spontaneously excitate firm's defense against competitor [2]. It raises questions about distinction of open innovation in modularization, and furthermore, how these open innovation advances evolution of whole systematic standard. Such a shift in general orientation will involve a re-conceptualization of the open innovation in modularization. We need integrated innovation when the product segment must be in-depth subdivided in order to meet system performance requirement [3]. The modular innovation will be appropriate when the product segment boundaries have been extended and covered more technical parameters [4]. This study focuses on open innovation in product modularization. We compare the integrated innovation with modular innovation under the background of complex knowledge, and then summarize the knowledge type and knowledge creation within the product architecture system by a literature review.

---

<sup>1</sup>Corresponding Author: Min Zhang, Department of Physical Economics and Management, Xi'an Physical Education University, NO.65 Hanguang Road, Xi'an China; E-mail: 37742769@qq.com

## **2. Categories of Open Innovation**

### *2.1 Independent open innovation*

The former refers to systematic architecture and interface standard while the latter refers to the autonomy of a subsystem. The module is a black box. The key of innovative ability within functional module is to improve performance through independent open innovation. Under the obedience of design rules, the module is free to innovate without the restriction from other modules. Accordingly, each module within the entire system becomes a separate unit and exploits the innovative potential adequately. Modularization is the critical step to affect integration of a variety of innovative resources, especially for knowledge integration. Due to the realization of a specialized division of knowledge, each module focus only on their own field, thus greatly improves the professional level. On the other hand, each module develops independently without paying attention to other modules, which greatly reduces both the complexity of innovation and the risk of failure. Modularity also enables the possibility that different knowledge modules benefit from parallel development. Knowledge innovated by module within the modular organization also creates a stream of fresh information, which might then bring changes in wider knowledge systems.

### *2.2 Synergetic open innovation*

Recently, scholars have put forward the concept of modularity across the boundaries of the modules, and shifted their focus towards the development of modular value networks <sup>[5]</sup>. Many small-medium enterprises are engaged in modularity production, especially in developing countries. In view of avoiding the weakness in knowledge accumulation, the open innovation realized by networked organization, such as cooperative R&D, joint knowledge alliance, the copartner in the same or similar industries are the first choice. Synergetic open innovation occurs when innovative activities are implemented across module borders. Knowledge is transferred easily in the dynamic environment. Independent open innovation within single module is insufficient to sustain an advantage permanently, thus transferring and exchanging knowledge to other modules in whole modular system might be a strategy to prolong knowledge life cycle <sup>[6]</sup>. In conclusion, the synergetic open innovation provides a platform which solves the problem of knowledge complexity, thereby, enhances the knowledge compatibility.

## **3. Open Innovation and Standard Evolution**

The standards in modularity is for testing a module's conformity to the design rules and for measuring one module's compatible performance with others. It aims to counterbalance the performance advantage of synergistic effect. Furthermore, synergistic specificity can keep the modular system away from disaggregation, which easily caused by the fact that some combinations of components function better than other configurations. Under the guidance of modular design rules, open innovation develops independently internal modules in accordance with their commitment. Sometimes, the design rules will be regarded as a successful fixed paradigm when the

whole system achieves the desired return. Independent or synergetic open innovation in modularity will form new technology, and sometimes it will violate the original design rules [7]. When the former system is not suited to the development of modular innovation, a new amendment is necessary, which means a new round of major innovations. Consequently, open innovation within single module is spilled over the interface boundary. Fig. 1 depicts the process of open innovation in modularization which includes module innovation and standard innovation which accelerate the pace of industry upgrading altogether.

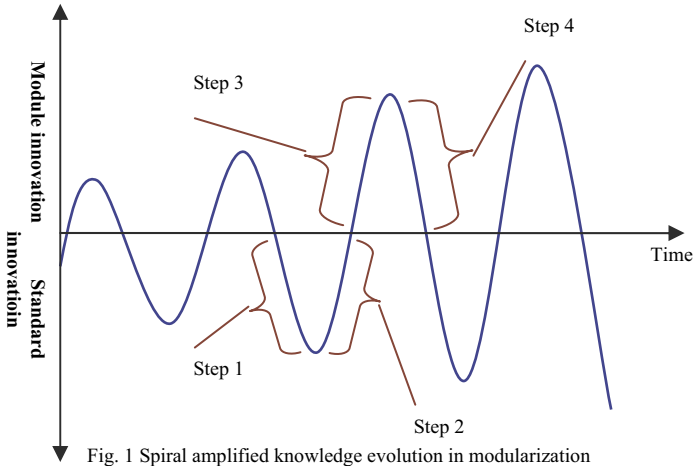


Fig. 1 Spiral amplified knowledge evolution in modularization

Fig. 1 longitudinal axis up direction represents the degree of open innovation within functional module, including technologic up-grade and performance improvement, mainly caused by independent open innovation. The downward direction represents the degree of standard innovation, including the adjustment of design rules or even technique protocol, mainly caused by synergetic innovation. The horizontal axis represents the time. Open innovation curve shows a trend of gradual spiral amplification evolution with time. It is an orbicular innovation cycle between each of the two peaks, or denoted by a whole wave.

#### 4. Case Studies in Chinese Sport Industry

##### 4.1 Purpose and method

We would also like to validate the correlation between intensity of knowledge interaction and the modular knowledge autonomy in order to meets the theoretical analysis in fourth section. We surveyed the representative company in sport industry with questionnaire designed by liket-5 scale method. We survey and score the intensity of knowledge interaction from higher to lower and grade from 5 to 1 based on three dimensions as: (1) frequency of knowledge interaction with partner; (2) continuity of knowledge interaction with partner; (3) scale of partner during knowledge interaction. The modular knowledge autonomy is identified by: (1) rate of standard accessory production; (2) degree of interdependency in R&D. This study selected 150 enterprises in sport industry for further research including the 50 production-oriented enterprises in

downstream supply chain, 50 intermediate product manufacturing enterprises, and 50 R & D enterprises upstream, a total of 94 valid questionnaires were received.

4.2 *Modular structure of the sport industry*

Based on the advanced technology, sport provides users with vast screen and bulky bandwidth, provides a big stage for software and content service providers, such as Internet browsing, downloading, e-mail, video, online games, bank payment, remote control, identification, and entertainment. Sport modular production increases significantly in evidence by its powerful complex applications, supported by systematic architecture which is the precondition for modularity.

With the industrialization of sport, enlargement of special knowledge accelerated the pace of modularization. The different working procedures deviated from the industrial chain and gradually formed multitudinous modules<sup>9</sup>. According to their functions, these modules are divided into three camps, the first camp is specialized in hardware, chipset, handset design and the operating system; the second camp including mainly modular integrator which are responsible for the final product integration based on different modular suppliers; the third camp is the operating platform suppliers which ensure sports actualize communication, Internet and other functions, including telecom operators, network operators and Internet content providers. Indicate in Fig. 2.

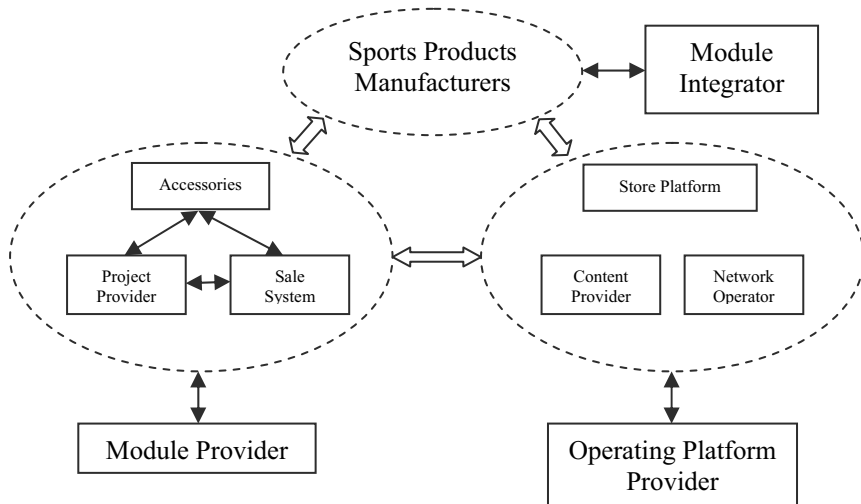


Fig. 2 Modular network of sport Industry

Open innovation in modularization of sport industry comprises independent innovation, depicted by innovation within the module, and synergetic innovation depicted by collaborated innovation among different functional modules.

The designer acts as the leader of modular organization, the makers of design rules, responsible for modular integration. There are two types of companies playing the designer role in the sport industry. The first is sports product operators with responsibility for product mode standard; the second is the system platform developers. In fact, the product mode is established by the negotiation of technologic protocol between modular supplier and the industry association. The operating system platform suppliers who often allied with manufacturer, assures system is applied efficiently. For example, the largest system platform supplier is also the major shareholder of Nike and Adidas, jointly shares the highest market value.

#### *4.3 Intensity of knowledge interaction in sport modularization*

Characteristics of independent open innovation. Independent open innovation in sport industrial chain is under the guidance of design rules including communication mode and operating system. The modules focus on the accumulation of their core knowledge resulting in improvement of local function in whole system. A large number of independent modules are eventually combined into diversified products. In addition, the independent open innovation results in the appearance of sub-modular outsourcing, which is recapitulated in two aspects as follows: Characteristics of synergetic open innovation. Synergetic open innovation in modularization refers to complete collaborative innovation, especially when breakthrough innovation emerges. In sport industry, operating system platform improvements lead to more external software support, speed up meanwhile the transmission of high-capacity information. This intensifies the innovation pressure of software developers and content service providers. Synergetic open innovation in sport modularization is mainly reflected in two aspects. Amendments of design rules within the module Technological innovation of functional module will affect the performance of final product. In order to match with the interface standards, design rules in subsystem need amendments. In the sport industry, the capacitive touch buttons are widely employed, whose sensitivity and stability will largely affect consumers' experience. Nike sports watch is the pioneer introducing this technology, but errors happen frequently (for example, accidental contact often cause hanging up by the contact of ears, cheeks or other parts of the body), which not only arose consumer's complaints, but also hold back the technology spread. After redesigning the interface parameters, the touch button matches the sport smoothly and the technology of capacitive touch buttons is recognized again in market. After quantity analysis about the features of open innovation in sport industry, we test the correlation between intensity knowledge interaction and modular knowledge autonomy based on the obtained validity data. The statistic output neither significant positive nor negative relationship. Further regression analysis was tested, the statistic output inverted U-shaped relationship between intensity of knowledge interaction and modular knowledge autonomy as shown in Fig. 3. The result implicates that excessive independent or synergetic relationship between modules during open innovation both led to lower degree of modular knowledge autonomy. Modules in medium states represents the most modular units in sport industry whose most enterprises comply with design rules at the same time but have a lot of knowledge autonomy about the technical parameters.

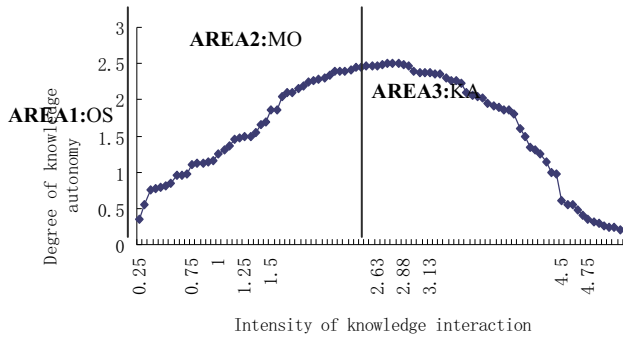


Fig. 3 Inverted U curve in modular open innovation

We divide the curve into three areas in which AREA1 represents the simple production, often referred as outsourcing (OS); AREA3 represents the excessive synergetic open innovation, often referred as knowledge alliance (KA) while AREA2 represents the apt modular open innovation, often referred as modular innovation (MI). Modular components in AREA1 engaged in simple outsourcing with lower degree of knowledge autonomy which exists simultaneously in AREA3 caused by the constraints of knowledge conformity. AREA2 represents the middle part of the business community in sport industry which takes on a mixed open innovation involve independent open innovation and synergetic innovation. Enterprises in this area possess a higher degree of knowledge autonomy.

### 5. Conclusions

We distinguish the type of open innovation in modularization from the modular internal and external level, referred as independent open innovation and synergetic open innovation, which playing the key role in modular upgrade and system optimization. We study the properties of independent open innovation, synergetic open innovation and the standard evolution driven by them which can be abstracted as the curve of knowledge spiral amplified evolution in modularization, proof-tested by Chinese sport industry. As attached questions, we also discuss the low-level efficiency and resistance of open innovation in modularization. The main conclusions are:

- (1) The independent open innovation is induced by the hidden design parameters, which offer a platform for module innovative practice under the background the architected conformity. Independent open innovation is accelerated by back to back competition among modules, which is also the stimulation of modular autonomy. Outcome of independent open innovation enlarges and amplifies the content of knowledge in whole system.
- (2) The synergetic open innovation is defined as the collaborative innovation between different modules which belong to the same layer or different layers. Synergetic open innovation is critical in reducing design problems and improving final product quality. In addition, the independent innovation is accelerated by doing this during the introduction of heterogenous information.



(3) There is an inverted U-shaped relationship between intensity of knowledge interaction and modular knowledge autonomy. Modular organization is a flexible form compared with higher standardization industry organization or corporative alliance organization. Apt modularity protects the knowledge autonomy.

## **Acknowledgments**

This research is financially supported by the Shaanxi Social Science Foundation (No. 12P048), title is research on policy transformation based on integration of sports industry chain in Shaanxi undertaking the transfer of eastern sports industry.

## **References and Notes**

- [1] Argote, L., Ingram, P. Knowledge Transfer: A basis for competitive advantage in firms. *Organizational Behavior and Human Decision Process.*,82, 150–169, (2000).
- [2] Raubitschek, R.S. Product Sequencing: Co-Evolution of Knowledge, Capabilities, and Products. *Strategic Management Journal.* 21, 961-979, (2002).
- [3] Schilling, MA., Paparone, CR. Modularity: an application of general systems theory to military force development. *Defense Acquisition Review of Journal.* 34(2005), 21-7.
- [4] Mark, Lehrer, Michael Behnam. Modularity vs. programmability in design of international products: Beyond the standardization–adaptation tradeoff? *European Management Journal.* 7, 281-292, (2009).
- [5] Krikke H, Le Blanc I, Van de Velde S. Product modularity and the design of closed-loop supply chains. *California Management Review.* 46, 23–39, (2004).
- [6] Mikko, Paananen. Exploring the relationships between knowledge sources in the innovation process: evidence from finish innovators. *Technology Analysis & Strategic Management.* 21, 711-725, (2009).
- [7] Hibata, T., Yano, M., Kodama, Empirical analysis of evolution of product architecture. *Research Policy.* 34, 13-31, (2005).

# Study on Prediction of the Shanghai Composite Index Based on EMD and NARX Neural Network

Yan XIU<sup>a,1</sup> and Xinye CHEN<sup>b</sup>

<sup>a</sup> *Tianjin Chengjian University, Tianjin China, 300384*

<sup>b</sup> *Party School of CPC Tianjin Municipal Committee, Tianjin China 300191*

**Abstract.** The prediction of Shanghai composite index is one of the hot topics in financial research. In this paper, based on empirical mode decomposition (EMD) algorithm combined with nonlinear autoregressive models with exogenous inputs (NARX) neural network theory, an EMD-NARX combination forecasting model is proposed. Firstly, the EMD algorithm is used to decompose the time series data into the intrinsic mode function components of different scales, and then the NARX neural network is used to predict the future index with multi step rolling forecast. The experiments show that the presented method not only can make accurate short-term prediction of Shanghai stock index, but also can predict the Shanghai index directly on the long-run trend. The method breaks through the limitations of traditional forecasting methods in long-term prediction.

**Keywords.** empirical mode decomposition, NARX neural network, Shanghai composite index, prediction.

## 1. Introduction

With the continuous development of China's economy, standardize the operation of the stock market, the stock price index has gradually become one of the important indicators to measure the level of national economic development, also affect the development of national economy to a certain extent. Since the Shanghai composite index covers a wide range, its trend reflects the economic performance. Since its establishment for more than 20 years, it has been the focus of attention of government, economic management departments and institutions and investors. The prediction of Shanghai stock index has become one of the focuses in the field of financial securities. In recent years, the researchers used GARCH model and neural network model, support vector machine model based on wavelet analysis and chaos theory and empirical mode decomposition method to predict the Shanghai index of multi angle [1-5]. However, due to the Shanghai index price influenced by various economic factors, the trend of the Shanghai index shows a very strong non-stationary and nonlinear characteristic, is very difficult to predict accurately the short term, and to predict its future trend is also very difficult. To solve this problem, this paper first uses the method of empirical mode decomposition (EMD) of Shanghai stock index data smooth

---

<sup>1</sup> Corresponding Author. Yan XIU (1973 -), female, Ph. D., associate professor, main research interests: chaos theory and its applications. E-mail: xiuyan1973@163.com

processing, then applied to nonlinear system of NARX neural network, innovative NARX neural network prediction model is established based on the combination of empirical mode decomposition (EMD-NARX model). The results show that the prediction model established in this paper improves the Shanghai index of short-term prediction accuracy, reliability and prediction of the long-term trend, can make an effective reference for economic management departments and investors.

## 2. Shanghai stock index EMD-NARX prediction model

### 2.1. EMD decomposition algorithm of Shanghai Stock Index

EMD method is a data processing method used to analyze nonstationary nonlinear real change processes [6]. This method is put forward by NASA in the United States the work of Dr. Huang et al in 1998. This method is based on the data of the local characteristic time to decompose, the complex nonlinear time series data is decomposed into a finite and usually several small number of intrinsic mode functions (IMF). Due to the nonlinear and non-stationary characteristics of the Shanghai Stock Index timing data, the EMD method can be used to decompose the time series data of Shanghai stock index, and then some IMF components with different fluctuation scales can be obtained. The interference or coupling between these components is minimal, while the remainder represents the trend or mean of the time-series data of the Shanghai Composite index.

Specific algorithms are as follows: set the Shanghai Stock Index timing data for  $S(t)$ .

The first step: first calculated each local maxima, and then use the three spline function connecting these points envelope composition written for  $u(t)$ , also found minima to connect these points to form the lower envelope is called  $d(t)$ , the mean curve envelope on the record  $m(t)$ , i.e.

$$m(t) = [u(t) + d(t)] / 2 \quad (1)$$

The second step: using the time series data of the Shanghai Stock Index  $S(t)$  minus the mean of the envelope  $m(t)$ , the first difference is obtained

$$h(t) = S(t) - m(t) \quad (2)$$

The third step: to determine whether the  $h(t)$  meets the two conditions of IMF. First, its zero number is equal to the pole number, or at most one difference. Second, the mean of the upper and lower envelope determined by its maximum and minimum is zero. If the above two conditions are not satisfied, then  $h(t)$  is used instead of  $S(t)$  to repeat the above steps until the resulting  $h(t)$  satisfies the IMF condition. If the above two conditions are satisfied, the first IMF component  $imf_1(t)$  and the remainder  $r_1(t)$  of the Shanghai stock index time series data are obtained,

$$imf_1(t) = h(t) \quad (3)$$

$$r_1(t) = S(t) - imf_1(t) \quad (4)$$

The fourth step: the Shanghai stock index time series data for the remainder of  $r_1(t)$  continue to be decomposed by EMD from the first step to the third step, until the rest are single signal or its value is less than the set value, complete decomposition. Finally, all the IMF components and the residual quantities of the Shanghai stock index time series data re obtained. It can be expressed as:

$$S(t) = \sum_{i=1}^n imf_i + r_n \tag{5}$$

2.2. Introduction of NARX neural network

The NARX (Nonlinear Auto-Regressive models with Exogenous Inputs) neural networks are called nonlinear autoregressive models with external inputs [7-8]. This network is a dynamic neural network with feedback, and it is a neural network with memory function. The output of this network depends not only on the current input, but also on the output of the past, the current input and the past output, which determine the current output. A typical NARX recurrent neural network is mainly composed of input layer, hidden layer, output layer and input and output delay. Before application, the order of delay of input and output and the number of neurons in hidden layer should be determined in advance. The NARX model not only inherits the advantages of the traditional time series model, but also makes the model more adaptive to nonlinear data through training. Therefore, the NARX model is very suitable for the prediction of complex, non-stationary and nonlinear time series of Shanghai Stock index. The mathematical model of the NARX neural network can be expressed as:

$$y(t) = f[y(t-1), y(t-2), \dots, y(t-n_y), x(t-1), x(t-2), \dots, x(t-n_x)] \tag{6}$$

$f(\cdot)$  means nonlinear functions implemented by neural networks. The output data  $y(t)$  is used as input in the feedback function, and the computational accuracy of the neural network is improved by open loop training or closed cycle training. The network structure is shown in figure 1.

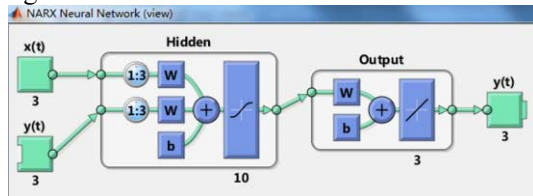


Figure 1. Structure diagram of NARX neural network

2.3. EMD-NARX multi step rolling forecasting model of Shanghai Stock Index

According to the nonlinear and non-stationary characteristics of the time-series data of Shanghai Composite Index, the data of Shanghai stock index series are firstly decomposed by EMD algorithm, that is, the system of timing data of Shanghai Stock Index

$$S(t) = f(imf_1(t), imf_2(t), \dots, imf_i(t)) \tag{7}$$

$imf_1(t), imf_2(t), \dots, imf_i(t)$  is the IMF component data obtained by the decomposition, and  $t$  is the data length of the time series data of the Shanghai Composite index, and  $f(\cdot)$  is the nonlinear function mapping. The IMF components decomposed by the EMD algorithm have better stationary, and the interference or coupling between each IMF component is very small. These components have different scales and represent the stock price volatility caused by traders with different investment time levels [11]. Then, each  $imf_i(t)$  component obtained from the EMD decomposition is pretreated separately. According to the ratio of 70%, 15% and 15%, the sample data is divided into training data, verification data and test data. NARX neural network is used to carry out rolling multi-step prediction for each  $imf_i(t)$  component. The results of each prediction are added to the input data, and a most remote input data is removed. In this way, we can use the existing data to predict the Shanghai stock index in the next few days at the same time, and make a more accurate prediction of the long-term trend of the Shanghai Stock index. Finally, the predicted values of each  $imf_i(t)$  component are simply aggregated. And the final multi step prediction value of Shanghai stock index time series data  $S(t)$  is obtained.

### 3. Empirical research

#### 3.1. Data sources and EMD decomposition

The experimental data of this paper selected Shanghai Composite Index (000001) from October 10, 2011 to March 10, 2015. The daily closing price of the 828 trading days was used as the object of study. The closing price of the first 728 trading days is used as the original data for EMD decomposition and prediction modeling, and the latter 100 data are used as prediction contrast data. By the EMD decomposition algorithm in the second part of this paper, the data of the Shanghai stock index's daily closing price in the first 728 trading days are decomposed, and the results are shown in figure 2. As you can see from Figure 2, after the EMD decomposition, the raw data is decomposed into 8 IMF components from high frequency to low frequency and a residual quantity representing the general trend of the raw data  $R$ . Although the high frequency components still have some non-stationary, their non-stationary has been greatly weakened relative to the original data. Low frequency components are more robust and each component is independent of each other and does not affect each other.

#### 3.2. Multi step rolling forecast of daily closing price of Shanghai stock index

Eight IMF components and one residue  $r$  decomposed by EMD are modeled and rolling multi-step prediction respectively using EMD-NARX multi step rolling forecasting model of Shanghai Stock Index. Firstly, eight IMF components and one residue  $r$  are respectively tested by the expert experience and the method of trial and error. The number of hidden nodes and the order of the input and output delay of the NARX neural network are tested several times. Then, through the comparison of the error and the network performance, the parameters of the hidden layer nodes and the

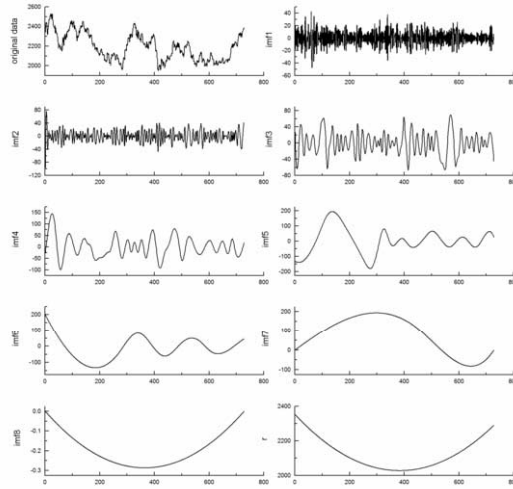


Fig 2. EMD decomposition result of Shanghai Stock Index daily closing price

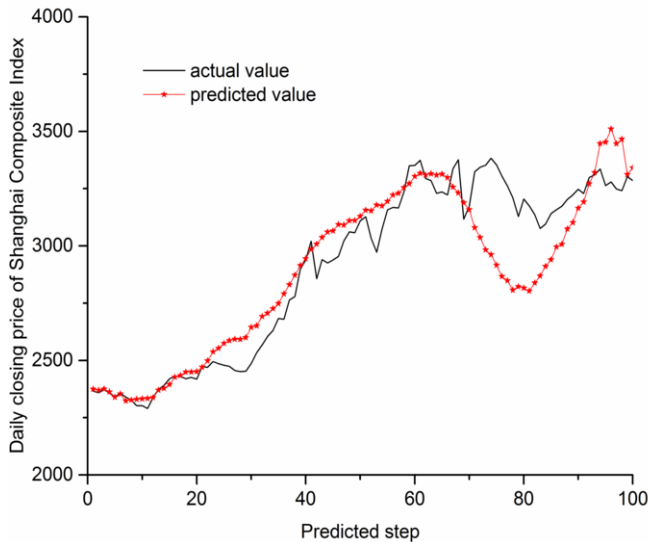
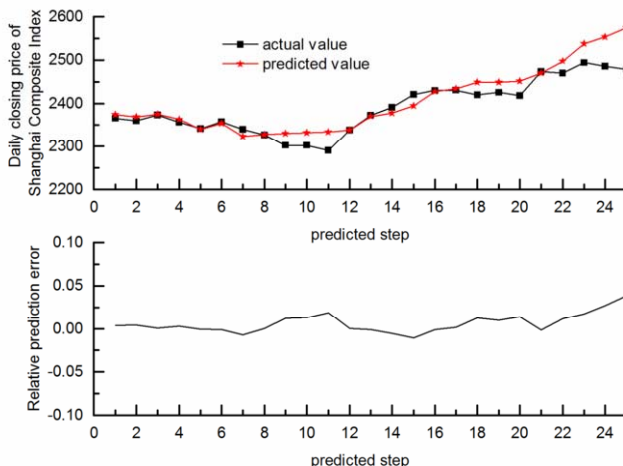


Figure 3. the Shanghai Stock Index closing price of the next 100 trading days chart

input and output delay order of the final NARX neural network are selected to determine the network structure, and then the multi-step prediction of rolling extrapolation is carried out. Finally, the predicted values of eight IMF components and the predictive value of one residue  $r$  are reconstructed to obtain the final prediction value. Figure 3 compares the forecast and actual values of the next 100 trading days of the Shanghai stock index's daily closing price. As can be seen from Figure 3, the EMD-NARX prediction model established in this paper can predict the fluctuation trend of the Shanghai stock index in the next 100 trading days more accurately.

As can be seen from Figure 4, the EMD-NARX prediction model established in this paper can make a relatively short and accurate prediction of the Shanghai Composite Index, and the relative forecast error in the next 23 trading days is less than 5%.



**Figure 4.** forecast for the next 25 trading days of the Shanghai stock index's daily closing price

In order to verify the performance of the EMD-NARX prediction model proposed in this paper, NARX neural network model is used to predict the same experimental data. A comparison of the forecast error indicators for the next 23 trading days is shown in table 1. It can be found that the prediction effect of the EMD-NARX model established in this paper is obviously better than the NARX model of daily closing price data of Shanghai stock index without EMD decomposition.

**Table 1.** Comparison of model error indexes

Error index	RMSE	MAPE	MAE
<b>EMD-NARX model</b>	21.027	0.403	15.611
<b>NARX model</b>	42.553	1.044	24.702

#### 4. Conclusions

The empirical results show that, although the Shanghai composite index time series data has strong nonlinear and non-stationary characteristics, but by the decomposition of the EMD algorithm, can reduce the impact of noise on the data, the volatility of Shanghai stock index time series data is greatly reduced. The IMF component obtained by the decomposition can reflect the internal fluctuation of the Shanghai Stock Index time-series data, and thus help to improve the prediction accuracy of noisy data. In addition, due to the NARX neural network is more suitable for modeling of nonlinear dynamic systems as the predictive model, the Shanghai index EMD-NARX multi-step prediction model in this paper can not only make accurate forecast of the Shanghai index, but also can predict the Shanghai index directly on the long-run trend, breaking the traditional prediction method the limitations in the long-term forecast. Because in determining the parameters of NARX neural network hidden layer node number and the order of the input and output delay, no mature theoretical basis, only by the experience of selection, which affects the efficiency of using NARX neural network modeling, this problem in the future research should be strengthened.

## Acknowledgment

This research has been supported by the General Project of Humanities and Social Science Research of Education Ministry (11YJCZH202), the Tianjin municipal philosophy and Social Sciences Program (TJYY13-031), and the School level research project of grade 2016 of Party School of Tianjin Municipal Committee of CPC (DXKT1620)

## References

- [1] Yang Jianhui, Zhang Ranxin. statistics and decision HP prediction filter and GARCH model of stock price trend, 2013 (5): 84-87.
- [2] Zhang Ning, Zhou Jiali, Sun Wujun. Predict and optimize the echo state network shares of management research project, 2014, 28 (1): 94-101.
- [3] Yu Zhijun, Yang Shanlin, Wang Xiaojia, et al. Error correction method of GM (1,1) based on data transformation, Systems Engineering Theory and Practice, 2015 (9): 2339-2347.
- [4] Hou Liqiang, Yang Shanlin, Wang Xiaojia, et al. Stock index fluctuation of Shanghai Composite Index: a study based on fuzzy FEGARCH model and different distribution hypotheses, China Management Science, 2015, 23 (6): 32-40.
- [5] Tan Zhengxun, Zhang Qian. Long term memory and trend prediction of Chinese stock market, Statistical Research, 2016, 33 (10): 57-66.
- [6] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. The Royal Society, 1998, 454(1971): 903-905.
- [7] Wong C X, Worden K. Generalised NARX shunting neural net-work modeling of friction. Mechanical Systems and Signal Processing, 2007, 21(1):553-572.
- [8] Math Works. Matlab R2013b Neural Network Toolbox User Guide. 2013



## Subject Index

3D LiDAR	548	clustering	1
60 GHz	441	collision avoidance	255
absorbance	238	combination effect	117
acceleration	275	commercial vehicle	91
accounting information quality	331	complex network	150
AdaBoost	499	computer aided diagnosis	509
Adjacent city	100	computer vision	212
ADS-B	8	conflict	8
ADVISOR	47	constancy	238
air transport network	180	container terminal	168
android	66	control strategy	47
ant colony algorithm	396	conventional HRM	573
APP management	125	convolutional neural network	350
application layer specification	516	cooperative control	255
ARCF	108	cooperative intelligent transport system	516
array antenna	441	correlation	8
artificial neural network	490	cropped function	429
asphalt pavement	558	cryptography	450
attacks	450	cuckoo search	490
autonomous vehicle	548	data exchange standard	516
aware-map	350	de-noising	469
background difference method	222	defense range	82
bacterial cell	238	dehaze	371
battery	193	delay	565
Bayesian filter	21	delay propagation probability	180
behavior recognition	21	deplaning	29
bidirectional DC/DC converter	193	discrete time simulation	29
BING	212	distributed drive electric vehicle	299
bluetooth	66	DOP	527
boarding	29	dSPACE	283
bounded variation space	480	dynamic hybrid	404
C/FD-GMRES	244	dynamic model	299
cabin	29	E-HRM	573
capacity	565	E-REV	47
car-following model	117	early warning broadcast	404
case base	142	earnings response coefficient	331
catastrophe theory	134	EKF	527
CCD camera	222	emission control area	168
channel coding	160	empirical mode decomposition	590
channel estimation	160	energy collection	558
characteristic parameter	480	energy management	413
Chinese research papers	343	energy saving	91
classification	469		

EPI	480	intelligent transportation	371, 388
EPnP	499	intelligent transportation systems	290
EPS	299	IT ambidexterity	228
expressway	134	IT flexibility	228
facial images	429	IT standardization	228
fault diagnosis	321, 460	K-means clustering	212
fine-grained classification	350	Kalman filter	311
flight delay propagation	180	knapsack problem	203
floating car	266	knowledge management	361
forest type	490	knowledge management system	361
formation control	255	lane change	21
forward error correction	160	large data-base	429
fractional dimension	480	LCL filter	193
FRID	388	leaf recognition	469
fuel economy	413	least squares	160
gain	441	leveling	58
gender recognition	429	license plate location	371
gray level co-occurrence matrix	509	license plate recognition	371
grey correlation matching model	379	liver cirrhosis	509
GUI	429	location	150
$H\epsilon$ gate	203	loosely-coupled	527
$H_\infty$	299	low-cost sensors	311
handling robot	396	main fan	460
hardware-in-the-loop simulation	283	map matching	266
health care	361	McCulloch's Method	490
health care management	361	microdroplet	238
heterogeneous traffic	117	microscopic image	509
Hidden-Markov Model	266	missile control system	283
high-speed train lateral damper	321	mixed sensitivity robust control	58
hill	91	MRCF	108
hybrid vehicular	413	multi agents	168
image analysis	469	multi-scale wavelet transform	275
image processing	469	multi-target	8
image watermarking	450	multiple candidates softmax	
implementation	573	regression	350
improved particle swarm		multiscale entropy	321
optimization	460	NARX neural network	590
inference engine	142	network security	82
information	343	neural network	469
information sharing service	37	new HRM practices	228
information space	142	object detection	350
information technology	361	object recognition	499
information transfer process	583	object tracking	499
information tunnel	388	objects detection	212
innovation performance	228	obstacle occlusion	548
integrated waiting area	565	online reservation	125
intelligent port	379	online training	76
intelligent riding	404	online trajectory planning	244
intelligent traffic	66	open innovation	583

OPENCV	222	simulation	168
operation system	168	single-damaged	108
optimal control	244	Situational Awareness	142
origin-destination	537	Small Medium Enterprises	573
orthogonal frequency division multiplexing	160	spatial distribution	537
Pakistan SME's	573	spatial knowledge	548
parameter matching	47	Sports Industry	583
particle swarm optimization algorithm	396	Steganography	450
passenger flow forecast model	100	support vector machine	21, 509
path planning	396	susceptible-infected-recovered model	180
PCA	429	system development	424
people identification	404	technology mechanism	37
piezoelectric	558	tightly-coupled	527
polyvinylidene fluoride	558	TLD	76
port informatization	379	tracker	76
position estimation	255	traffic incident	134
power monitor	82	traffic light control	290
power performance	47	traffic monitoring	244
prediction	590	traffic organization	565
product modularization	583	traffic signals	388
propagating decovolution	371	traffic weather	424
PSNR	480	travel density	537
public transportation	100	UAV	244
pulverized coal optimization	480	UKF	527
Q-gate	203	unmanned aerial vehicles	255
QR code scanning	125	urban rail transit	150
quantitative analysis	343	urban traffic	266
quantum evolutionary algorithm	203	V2X	516
qubit	203	variational mode decomposition	321
queuing theory	66	vehicle handling stability	299
RCF	108	vehicle navigation	311
real-time systems	290	vehicle velocity constraint	311
recognition	429	vibration displacement	275
region proposal	350	vibration signal	460
road boundary detection	548	visible watermark	450
ROC curves	21	VISSIM	565
RRCF	108	VS2010	222
rule base	142	VSR	193
safety	91	vulnerability	150
sea wave disturbance	58	wavelet neural network	460
semantics	1	wavelet transform	413
service discovery	1	web service	1
severity	134	WebGIS	424
Shanghai composite index	590	wireless communication	441
significance analysis	537	wireless mesh network	37
similarity matrix	1	WordNet	1
simply supported beam	108	XBRL	331
		ZigBee	66, 388

This page intentionally left blank

## Author Index

Ahmad, N.	228, 573	Han, S.	424
Al Quthami, R.A.	469	Hao, L.	203
Al-Nufaei, A.A.	469	Hao, Z.	429
Al-Otaibi, M.B.	469	He, A.	21
An, Y.	266	He, D.	516
Ashour, A.S.	450, 469, 490, 509	He, L.	396, 404
Bai, Y.	404	Hou, J.	168
Bao, J.	125	Hou, Y.	396, 404
Bei, S.	413	Hou, Z.	142
Bei, S.-y.	47	Hu, X.	311, 499
Borra, S.	450	Hu, Z.	108
Cai, Q.	238	Huan, J.	1
Cao, L.	509	Huang, C.	238
Chatterjee, S.	490	Huang, L.	558
Chen, M.	212	Huang, M.	193, 222
Chen, W.	160	Huang, Y.	160
Chen, X.	565, 590	Jatoi, M.A.	429
Chen, Y.	283	Jiao, W.-d.	8
Chen, Y.-l.	331	Jin, W.	321
Cheng, X.	266	Joya, C.	350
Cui, C.	460	Karamat, J.	361, 573
Dai, G.	537	Lei, M.	203
Dey, N.	450, 469, 490, 509	Lei, Y.	537
Dharejo, F.A.	429	Li, B.	413
Dong, C.	255	Li, C.	321, 404
Du, J.	311	Li, D.	396
Fan, W.	160	Li, H.	134
Fan, X.	299	Li, J.	66
Fang, B.	244, 275	Li, L.	537
Fei, C.	66	Li, M.	212
Feng, X.-f.	275	Li, P.	558
Feng, Y.	255, 516	Li, Q.	222
Fong, S.J.	490	Li, S.	134, 350
Gao, J.	21	Li, X.	212, 321
Gao, Q.	499	Li, Z.	117
Gen, H.	558	Liang, Z.	388
Gou, X.	321	Liao, Z.	66
Guan, Y.	516	Liu, H.	91
Guo, C.-w.	37	Liu, H.-m.	47
Guo, J.	160	Liu, J.-h.	275
H.R., Lakshmi	450	Liu, Q.	413
Hai, Y.	91	Luo, D.	1
Han, H.	142	Ma, Q.	283

Mahmood, K.	361	Wang, R.	371
Majeed, A.	228, 573	Wang, S.	37
Meng, F.	266	Wang, X.	499
Miao, H.	283	Wang, Y.	244, 509
Miao, L.	290	Wang, Y.-b.	8
Miao, X.	228	Wang, Z.	238, 480
Min, F.	424	Wei, M.	82
Min, H.	548	Wen, C.	125
Na, Q.	203	Wen, K.	527
Ni, Z.	299, 413	Wen, X.	583
Niu, L.	516	Witlox, F.	180
Pan, H.	565	Wu, W.	180
Pang, Q.	100	Wu, Y.	29
Peng, L.	142	Xia, B.	424
Peng, Q.	537	Xiao, J.	558
Qi, Y.	76	Xiu, Y.	590
Qian, Y.	117	Xu, F.	343
Qiao, X.	238	Xu, J.	117
Qiao, X.-w.	58	Xu, L.	290
Qin, S.	565	Xu, M.	142
Qin, Y.-y.	58	Xu, N.	66
Sang, A.	212	Xu, P.	388
Sen, S.	490	Xu, S.	117, 244
Shan, Y.	21	Xu, X.-t.	37
Shen, X.-y.	8	Yang, J.Y.	244
Shi, F.	450, 469, 490, 509	Yang, L.	266, 343
Song, C.	396	Yang, M.	516
Song, D.	21	Yang, Y.	460
Su, Y.	212	Yang, Y.-c.	331
Sun, P.	548	Yang, Z.	537
Sun, Y.	548	Yao, B.	441
Tang, C.	331	Ye, C.	117
Tang, J.	441	Yuan, X.	460
Tang, K.	311	Zeng, W.	66
Tang, X.	1	Zhang, H.	180
Tao, J.	125	Zhang, J.	125, 537, 565
Tong, S.	361	Zhang, L.	499
Tunio, M.A.	429	Zhang, M.	379, 583
Waheed, A.	228, 361, 573	Zhang, N.	150
Waheed, S.	228, 573	Zhang, P.	76
Wang, C.	21	Zhang, R.	21
Wang, G.	142, 283, 371	Zhang, W.	266
Wang, H.	29, 142, 527	Zhang, X.	134, 537
Wang, H.-j.	8	Zhang, Y.	396, 558
Wang, H.-m.	58	Zhang, Z.-f.	58
Wang, J.	66	Zhao, J.	193, 413
Wang, L.	331	Zhao, J.-b.	47, 299
Wang, M.	29, 480, 527	Zhao, X.	134, 266, 548
Wang, Q.	255	Zheng, Q.	441

Zhou, B.	8	Zhu, H.	134
Zhou, Q.	142	Zhu, J.	180
Zhou, T.	413	Zhu, L.	150
Zhou, Y.	100	Zou, X.	160, 558

This page intentionally left blank