IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 3 Issue 5

May 2012

SAI

www.ijacsa.thesai.org

IJACSA

WHERE WISDOM SHARES

INTERNATIONAL JOURNAL OF
ADVANCED COMPUTER SCIENCE AND APPLICATIONS

Editorial Preface

From the Desk of Managing Editor...

IJACSA seems to have a cult following and was a humungous success during 2011. We at The Science and Information Organization are pleased to present the April 2012 Issue of IJACSA.

While it took the radio 38 years and the television a short 13 years, it took the World Wide Web only 4 years to reach 50 million users. This shows the richness of the pace at which the computer science moves. As 2012 progresses, we seem to be set for the rapid and intricate ramifications of new technology advancements.

With this issue we wish to reach out to a much larger number with an expectation that more and more researchers get interested in our mission of sharing wisdom. The Organization is committed to introduce to the research audience exactly what they are looking for and that is unique and novel. Guided by this mission, we continuously look for ways to collaborate with other educational institutions worldwide.

Well, as Steve Jobs once said, Innovation has nothing to do with how many R&D dollars you have, it's about the people you have. At IJACSA we believe in spreading the subject knowledge with effectiveness in all classes of audience. Nevertheless, the promise of increased engagement requires that we consider how this might be accomplished, delivering up-to-date and authoritative coverage of advanced computer science and applications.

Throughout our archives, new ideas and technologies have been welcomed, carefully critiqued, and discarded or accepted by qualified reviewers and associate editors. Our efforts to improve the quality of the articles published and expand their reach to the interested audience will continue, and these efforts will require critical minds and careful consideration to assess the quality, relevance, and readability of individual articles.

To summarise, the journal has offered its readership thought provoking theoretical, philosophical, and empirical ideas from some of the finest minds worldwide. We thank all our readers for their continued support and goodwill for IJACSA. We will keep you posted on updates about the new programmes launched in collaboration.

We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJACSA provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

We regularly conduct surveys and receive extensive feedback which we take very seriously. We beseech valuable suggestions of all our readers for improving our publication.

Thank you for Sharing Wisdom!

Associate Editors

Reviewer Board Members

EL JABAL AL GARBI UNIVERSITY, ZAWIA

- **Dragana Becejski-Vujaklija**
  University of Belgrade, Faculty of organizational sciences

- **Fokrul Alom Mazarbhuiya**
  King Khalid University

- **G. Sreedhar**
  Rashtriya Sanskrit University

- **Gaurav Kumar**
  Manav Bharti University, Solan Himachal Pradesh

- **Ghalem Belalem**
  University of Oran (Es Senia)

- **Gufran Ahmad Ansari**
  Qassim University

- **Hadj Hamma Tadjine**
  IAV GmbH

- **Hanumanthappa.J**
  University of Mangalore, India

- **Hesham G. Ibrahim**
  Chemical Engineering Department, Al-Mergheb University, Al-Khoms City

- **Dr. Himanshu Aggarwal**
  Punjabi University, India

- **Huda K. AL-Jobori**
  Ahlia University

- **Dr. Jamaiah Haji Yahaya**
  Northern University of Malaysia (UUM), Malaysia

- **Jasvir Singh**
  Communication Signal Processing Research Lab

- **Jatinderkumar R. Saini**
  S.P.College of Engineering, Gujarat

- **Prof. Joe-Sam Chou**
  Nanhua University, Taiwan

- **Dr. Juan Josè Martínez Castillo**
  Yacambu University, Venezuela

- **Dr. Jui-Pin Yang**
  Shih Chien University, Taiwan

- **Jyoti Chaudhary**
  high performance computing research lab

- **K V.L.N.Acharyulu**
  Bapatla Engineering college

- **K. PRASADH**
  METS SCHOOL OF ENGINEERING

- **Ka Lok Man**
  Xi'an Jiaotong-Liverpool University (XJTLU)

- **Dr. Kamal Shah**
  St. Francis Institute of Technology, India

- **Kanak Saxena**
  S.A.TECHNOLOGICAL INSTITUTE

- **Kashif Nisar**
  Universiti Utara Malaysia

- **Kayhan Zrar Ghafoor**
  University Technology Malaysia

- **Kodge B. G.**
  S. V. College, India

- **Kohei Arai**
  Saga University

- **Kunal Patel**
  Ingenuity Systems, USA

- **Labib Francis Gergis**
  Misr Academy for Engineering and Technology

- **Lai Khin Wee**
  Technischen Universität Ilmenau, Germany

- **Latha Parthiban**
  SSN College of Engineering, Kalavakkam

- **Lazar Stosic**
  College for professional studies educators, Aleksinac

- **Mr. Lijian Sun**
  Chinese Academy of Surveying and Mapping, China

- **Long Chen**
  Qualcomm Incorporated

- **M.V.Raghavendra**
  Swathi Institute of Technology & Sciences, India.

- **Madjid Khalilian**
  Islamic Azad University

- **Mahesh Chandra**
  B.I.T, India

- **Mahmoud M. A. Abd Ellatif**
  Mansoura University

- **Manpreet Singh Manna**
  SLIET University, Govt. of India

- **Manuj Darbari**
  BBD University

- **Marcellin Julius NKENLIFACK**
  University of Dschang

- **Md. Masud Rana**
  Khunla University of Engineering & Technology, Bangladesh

- **Md. Zia Ur Rahman**
  Narasaraopeta Engg. College, Narasaraopeta

- **Messaouda AZZOUZI**
  Ziane AChour University of Djelfa

- **Dr. Michael Watts**
  University of Adelaide, Australia

- **Milena Bogdanovic**
  University of Nis, Teacher Training Faculty in Vranje

(iv)

- **Miroslav Baca**
  University of Zagreb, Faculty of organization and informatics / Center for biomet

- **Mohamed Ali Mahjoub**
  Preparatory Institute of Engineer of Monastir

- **Mohammad Talib**
  University of Botswana, Gaborone

- **Mohammad Ali Badamchizadeh**
  University of Tabriz

- **Mohammed Ali Hussain**
  Sri Sai Madhavi Institute of Science & Technology

- **Mohd Helmy Abd Wahab**
  Universiti Tun Hussein Onn Malaysia

- **Mohd Nazri Ismail**
  University of Kuala Lumpur (UniKL)

- **Mona Elshinawy**
  Howard University

- **Mueen Uddin**
  Universiti Teknologi Malaysia UTM

- **Dr. Murugesan N**
  Government Arts College (Autonomous), India

- **N Ch.Sriman Narayana Iyengar**
  VIT University

- **Natarajan Subramanyam**
  PES Institute of Technology

- **Neeraj Bhargava**
  MDS University

- **Nitin S. Choubey**
  Mukesh Patel School of Technology Management & Eng

- **Pankaj Gupta**
  Microsoft Corporation

- **Paresh V Virparia**
  Sardar Patel University

- **Dr. Poonam Garg**
  Institute of Management Technology, Ghaziabad

- **Prabhat K Mahanti**
  UNIVERSITY OF NEW BRUNSWICK

- **Pradip Jawandhiya**
  Jawaharlal Darda Institute of Engineering & Techno

- **Rachid Saadane**
  EE departement EHTP

- **Raj Gaurang Tiwari**
  AZAD Institute of Engineering and Technology

- **Rajesh Kumar**
  National University of Singapore

- **Rajesh K Shukla**

- Sagar Institute of Research & Technology-Excellence, India

- **Dr. Rajiv Dharaskar**
  GH Raisoni College of Engineering, India

- **Prof. Rakesh. L**
  Vijetha Institute of Technology, India

- **Prof. Rashid Sheikh**
  Acropolis Institute of Technology and Research, India

- **Ravi Prakash**
  University of Mumbai

- **Reshmy Krishnan**
  Muscat College affiliated to stirling University.U

- **Rongrong Ji**
  Columbia University

- **Ronny Mardiyanto**
  Institut Teknologi Sepuluh Nopember

- **Ruchika Malhotra**
  Delhi Technoogical University

- **Sachin Kumar Agrawal**
  University of Limerick

- **Dr.Sagarmay Deb**
  University Lecturer, Central Queensland University, Australia

- **Said Ghoniemy**
  Taif University

- **Saleh Ali K. AlOmari**
  Universiti Sains Malaysia

- **Samarjeet Borah**
  Dept. of CSE, Sikkim Manipal University

- **Dr. Sana'a Wafa Al-Sayegh**
  University College of Applied Sciences UCAS-Palestine

- **Santosh Kumar**
  Graphic Era University, India

- **Sasan Adibi**
  Research In Motion (RIM)

- **Saurabh Pal**
  VBS Purvanchal University, Jaunpur

- **Saurabh Dutta**
  Dr. B. C. Roy Engineering College, Durgapur

- **Sergio Andre Ferreira**
  Portuguese Catholic University

- **Seyed Hamidreza Mohades Kasaei**
  University of Isfahan

- **Shahanawaj Ahamad**
  The University of Al-Kharj

- **Shaidah Jusoh**
  University of West Florida

- **Sikha Bagui**
  Zarqa University

CONTENTS

(vii)

Data normalization and integration in Robotic Systems using Web Services Technologies

Jose Vicente Berna-Martinez
Department of Computer Science
University of Alicante
San Vicente del Raspeig, Spain

Francisco Macia-Perez
Department of Computer Science
University of Alicante
San Vicente del Raspeig, Spain

Abstract— **The robotics is one of the most active areas. We also need to join a large number of disciplines to create robots. With these premises, one problem is the management of information from multiple heterogeneous sources. Each component, hardware or software, produces data with different nature: temporal frequencies, processing needs, size, type, etc. Nowadays, technologies and software engineering paradigms such as service-oriented architectures are applied to solve this problem in other areas. This paper proposes the use of these technologies to implement a robotic control system based on services. This type of system will allow integration and collaborative work of different elements that make up a robotic system**

Keywords-SOA; robots architecture; web serices; management and integration.

I. INTRODUCTION

Robotics has become one of the most active emerging areas in which converge a large number of disciplines [1]. One of the biggest changes has been the expansion of the environments where they are used, from industrial environments to service robots for professional use and or domestic environments [2]. This implies that the variety of robots has grown, the number of devices integrated has increased and diversified, the scenarios are now unpredictable, dynamic and open, and therefore the complexity and heterogeneity of the underlying information has grown. To provide a solution to this problem are being implemented proposals related to service-oriented software applications and techniques of software distributed over the Internet [3]. But because the elements that make a robot operate at different levels of technology (electro-mechanical elements, algorithms and software functions, neural networks, etc.), first standardization is required for all items, so you can see all and each of them from the same functional level. This paper presents the standardization of robotic elements as a service through a conceptual architecture based on ICT and widespread in E-Business, which enables the management of information flowing through various channels and sources of a robot. In addition to allowing homogenization of the devices involved in any robotic system also allows for standardized treatment of information, solving problems of integration of heterogeneous information, helping to define the information flows in a dynamic manner and allowing to overcome the problems caused by different frequencies and different processing requirements of work for those elements of a robotic system.

For the development of the proposal in the next section we make a brief tour of the background of major related work. Then the normalization of the components of a robotic system is presented and architecture for the transformation of elements in services is proposed. Afterwards an instantiation of the proposed architecture using Web Services to implement control systems of autonomous mobile robots are constructed, and finally the main conclusions and future lines of work are shown.

II. BACKGROUND

A robotic system consists of a set of elements that operate together to achieve a goal. The nature of these elements may be different (electro-mechanical components such as sensors or motors, software elements such as route tracing algorithms, integrated circuit or systems on chip (SoC) to implement neural networks or pattern recognizers, and so on). Moreover, these elements may vary over time to adapt to new circumstances, environments or requirements [4]. However, from a functional point of view, each of these elements can be seen as an entity that receives information, performs an action and produces results (these results can be data or may be an action on the environment). This mode of operation is similar to what we observe in the distributed software components that make up distributed applications [5], and so we can use a similar conceptual base to define each of the pieces that form a robotic system rather than seeing the robot as a rigid set of devices than should communicate between them. A centralized implementation is robust and efficient, but these applications lack the properties necessary for their maintenance, modification, modernization, adaptation or flexibility to change in the medium to long term. These deficiencies greatly influence the management of information, because changing a source of information (for example an ultrasonic sensor for a laser sensor) usually involves reconsideration or rescheduling of part or even the entire system [6].

Several proposals have emerged to provide these features. These works provide a common framework for the development of heterogeneous robotic systems using frameworks or tools like BABEL [6], CLARAty [7], LAAS [8], DAMN [9], which generally provide those features found in software distributed applications like flexibility, modularity, code reuse, management of production cycles, low-cost development, adaptation to change, and so on. However, these works make different proposals that develop technologies or

frameworks that require learning and produces that specialists in robotics are away from the world of software engineering, although the world of software engineering is which provides the desired characteristics. These characteristics are:

- Control applications must be modular to allow code reuse and rapid development.

- The control logic must be independent of hardware. The hardware provides the possibilities, but the software develops the skills.

- Support for communications should be provided by framework in which is developed the system and details should be hided for implementation of intelligence of the robot.

- Components must be able to communicate asynchronously transmitting values. If the components use references cannot be distributed independently.

- The components must be able to be linked dynamically, using modules that are necessary even in runtime.

- Reactive techniques exploit the characteristics of the real environment

- Deliberative techniques allow us to infer knowledge that is not implicit in the environment

III. NORMALIZATION OF ROBOTICS ELEMENTS

In our work we propose to rely on widely available technologies and paradigms in the development of distributed software applications, specifically service-oriented architectures. For this it is necessary that each of the elements of a robotic system is provided as a service, and each service needs a support in the form of services container. A service container provides the suitable software infrastructure to

applications will be the basis for defining our services container, the architectural model of n-levels [5]. In Fig. 1 we can see the architecture of n-Levels which reflects the elements that incorporate the service container. Fig. 1-a we can see all the software elements that make up the service container. At the user level is allowed access as services (consumers) as well as from other external systems using a view controller. At the access level, SOA and working drivers are responsible for controlling the security aspects of access. At the business level drivers and business orchestration give proper access to specific functions to be deployed on devices (calculate a path, detecting obstacles, store information from the environment, convert the movements of each system element in the current position, and so on.), and finally at the level of resources, appropriate adapters will provide access to resources such as storage, possibility of simulation on various platforms, and so on. These components are based on a common middleware services that provide common support in a generic way, as security services, service orchestration, service notification or discovery services.

Fig. 1-b shows a simplified view of architecture, where you can more easily observe that the container will comprise a series of application components that define the specific functionality provided by the container, the middleware services layer common to all components and below the layer formed by the OS and the hardware specific to each device. Through this transformation, a motor is not a physical device with which the system has to communicate in specific and concrete way, but it becomes a service that can be consulted, to which we can transmit orders and can make decisions as launch an alert to another element of control when circumstances require.

The concept of service container is easily applicable to those robotic elements of computer nature, such as pattern recognition algorithms running on a computer to identify
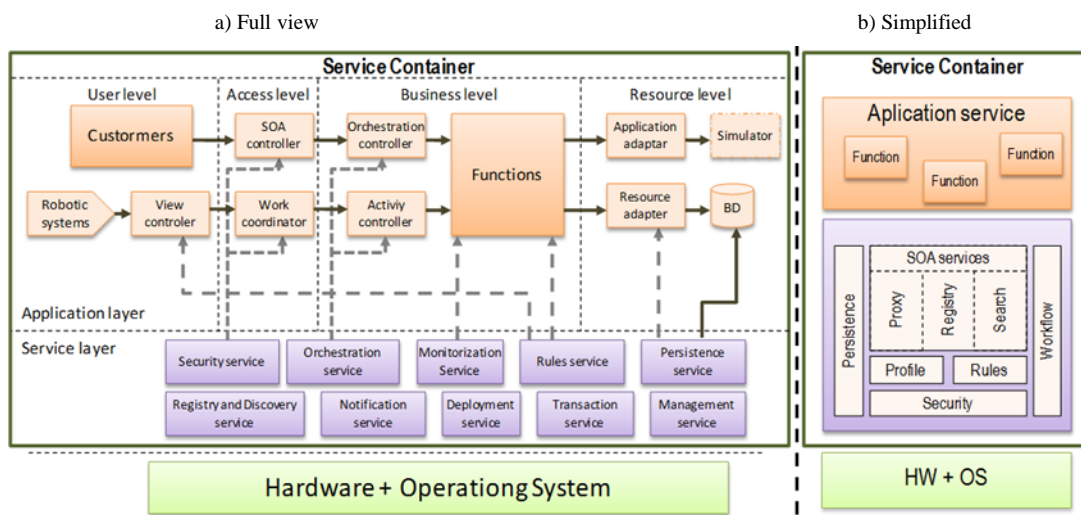
Figure 1.   Full view of the elements that make up the service container architecture. b) Simplified view of the service

deploy high-level functions on the devices. In this manner is the service that determines what function is developed and not the element or device in which is carried out. An architectural model widely used in the development of E-Business

objects in an image. However, the electro-mechanical devices of a robot (such as motors or sensors) have no basis for processing, in other words, usually have no computational or transmission elements that allow communication with other

elements. To make these devices capable of computation and communication is necessary to convert the physical devices on smart devices. To do this it is possible to incorporate the hardware necessary to bring any physical device can become a service [10].

In recent years, advances in electronics and communications have given us a range of new devices that can provide such capabilities, so-called embedded devices. These devices are characterized by their small size and low cost, allowing its integration into other devices. Through these devices we can provide advanced functionalities to electro-mechanical devices that form a robot and introduce distributed computing paradigms.
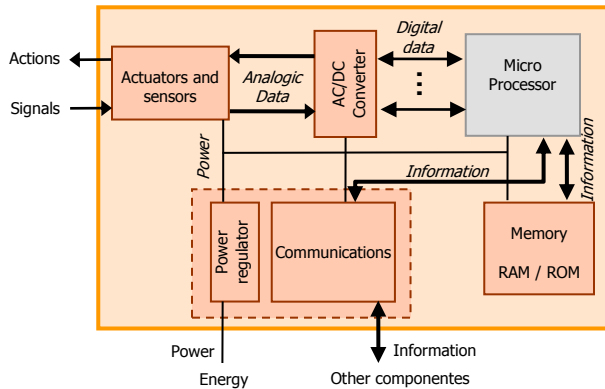
Figure 3.   Physical structure of the control

Figure 2 describes the general structure of the embedded hardware. These items can transform a passive device like a motor on a device with computing capabilities. To do this, the unit requires embedded processing unit, AC / DC converter device to communicate with actuators or sensors, internal memory and a communications module that allows to interact with a network of devices.

In fig. 3 shows the embedded device selected for our proposal, the XPort device. XPort is a compact solution which includes a 16 bit processor, RAM, Ethernet port 10/100 and serial interface that allows communication with devices such as motors or sensors. This device has already been the subject of other studies in our laboratory [11] demonstrating that the physical characteristics are sufficient for the deployment of network services.
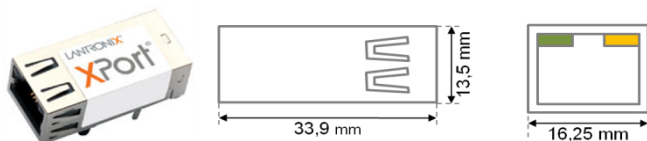
Figure 4.   XPort device description used in our experiments

IV.   TESTING AND VALIDATING

For the instantiation of our architecture we rely on autonomous mobile robots. Mobile robots are particularly interesting when used in open environments because in these environments the quantity, quality and accuracy of information is uncertain. Other reasons to tackle this type of systems is that can be highly variable: legs, wheels, chains, several sensory systems or multiple algorithms for estimation of position, which means involving a greater or lesser number of computational processes.
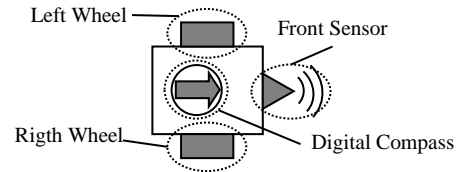
Figure 2.   Physical scheme of robot: 2 wheels, a digital compass and a front sensor

In our work we have tried two behaviors: Behavior1 (B1) - navigating through the environment from a source point to a target point, and Behavior2 (B2) – navigating through the environment from a source point to a target point with obstacle avoidance. B2 will be implemented by adding new services in B1. For our system we used a generic robot equipped with two actuators (right wheel and left wheel) from which we get the current position of the wheel (shaft encoder sensor), a digital compass that indicates the current direction and a front-sensor obstacle detection (fig. 4).

In the functional analysis of behavior we have divided each of the functions of a robot in a service, isolating each function in an independent entity [12]. Each service is executed independently (fig. 5). B1 analysis produces the following services: Sensing, services responsible for monitoring the sensing devices; Interpretation, service responsible for translating the values obtained by the sensing to consistent data

service	description
Se	sensing service: right, left, compass
In	interpretation service: right, left, compass
Si	situation service
Re	reasoner service
Pl	planner service
Mo	motion service
Em	embodiment service
Ac	actuator service: right, left
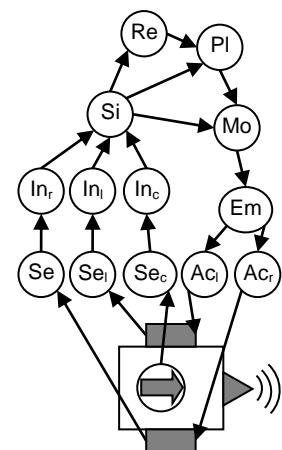
Figure 5.   Decomposition of behavior 1 in services, each sensor has sensing service and interpretation service, each

(for example floating numbers to numbers with two decimal numbers); Situation, service responsible for using the data of Interpretation to obtain an estimate of the robot's position (in this case position in the environment, but it could estimate the position of the arm, relative position, etc.); Reasoner, service responsible for determining the mission to perform, in this case lead the robot from point A to point B; Planner, service responsible for planning the robot path; Motion, service which is responsible for obtaining the next move to be performed by the robot based on planning; Embodiment, service responsible for transforming the type of motion in terms of physical structure of the robot; Actuator, services responsible for managing communication with the actuating devices.

B2 analysis incorporates new services to B1. New services are shown in fig. 6: Sensing (Sed), control service for distance sensor, Interpretation (Ind) for the sensing service, a new service, Restriction (Rc), service responsible for calculating where the obstacles based on the interpreted data, and a new service Planner (Plo) which modifies the B1 planning for obstacle avoidance.

service	description
Se	sensing service: right, left, compass, distance
In	interpretation service: right, left, compass, distance
Si	situation service
Rc	restriction
Re	reasoner service
Pl	planner service
Mo	motion service
Em	embodiment service
Ac	actuator service: right, left
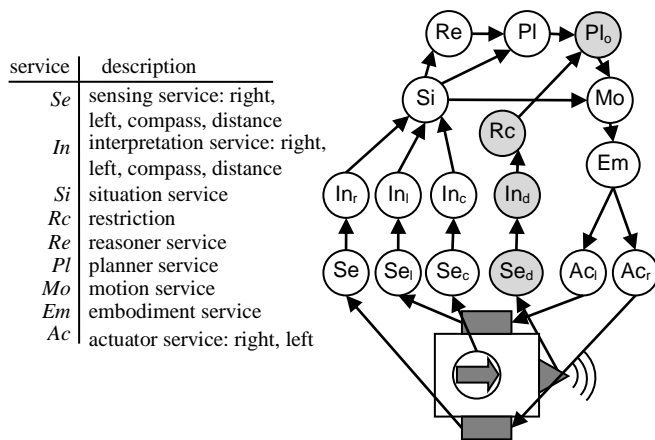
Figure 6. Decomposition of behavior 2 in services. The new services change the behavior B1.

Each of the services that integrate the control system develops a simple function, for example, Situation estimates the current position using techniques of odometry, Interpretation services translate the encoding axis of wheels into distances depending on the diameter of the wheels, and so on. Separate each system function in a service allows you to change services without changes influence the rest of the system.

For the implementation we used Microsoft Robotics Developer Studio (MRDS) because this environment provides us with an integrated development environment. NET for the design, execution and debugging robot applications scalable, concurrent and distributed, in addition to providing features such as service coordination, monitoring, configuration, deployment and reuse. RDS is built on two basic components: the Concurrency and Coordination Runtime (CCR) and the Decentralized Software Services (DSS). The CCR provides a programming model to handle multi-threaded applications and synchronization between tasks while the DSS allows to build applications based on a model of loose coupling. In addition DSS provides a lightweight model of state-oriented service that

combines the concept of Representational State Transfer (REST) with a system-level approach for building high performance scalable applications [13].

In our experiments we used the simulator MRDS, a Lego robots and a homemade root, because it demonstrates the adaptability of the control systems based on web services to any type of robot, although its components are not the most accurate. Fig. 7-a show a view of the simulated robot composed of the elements described above, and fig. 7-b show a Lego robot equipped with the same real elements and fig. 7-c show the homemade robot with the same elements.

a) Simulated robot in MRDS

b) Lego Robot   c) Homemade robot
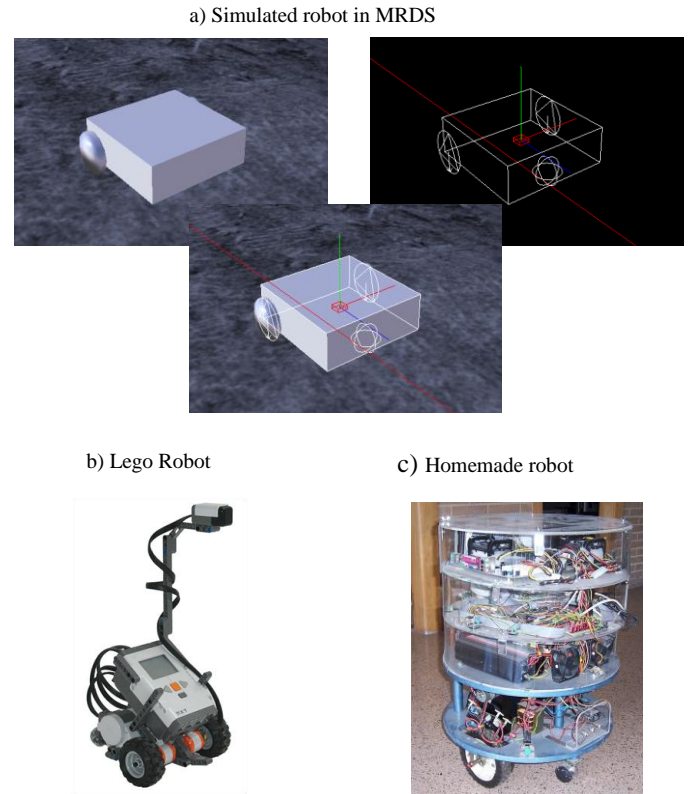
Figure 7. . a) Simulated robot. b) Lego Robot. c) Homemade robot

After deploying Web services and compose the control system according to the diagram in fig. 5 and fig. 6, we get the complete control system. For both B1 and B2, the system behaves as expected. Fig. 8-a shows the simulated robot's behavior and Fig. 8-b shows the Lego robot's behavior. Both systems use the behavior B2.

When we indicate a destination, the robotic system starts and progresses to reach the end point. Using B1, if there are obstacles in the path, the robot collides with them. Using B2, the system detects obstacles and modifies the path to avoid them. Both the simulated system as the real robot, behaviors are those specified. Most services remain common to all systems. Pass from a simulated robot to a real robot only involves modifying the services of Sensing and Actuator to connect to the appropriate resource. To use the behavior B2 only have to add the services specified in Fig. 2-b. The system thus shows its adaptability to change, flexibility to modify

capabilities and robotic devices, the ability to reuse code, and so on.

The system has the peculiarity that each Web service operates at the frequency that requires its own characteristics. For example, the services responsible for monitoring each wheel require 50ms per cycle to obtain the state of the encoder. This data is transferred to the superior services but if this information does not imply changes (for example, the robot has not moved), Interpretation services will not produce new results. Similarly, the reasoning service starts the system when the current and desired position is not equal (not reached the destination) but during the execution will not release more

Figure 8.   Simulated robot using B2 behavior. b) Lego robot using B2 behavior.

orders to planning services until it reaches the destination. Each service is independent, uses its own working frequency and its execution can influence whether or not the execution of other services and the communication is done homogeneously through message passing.

V.   ANALYSIS OF RESULTS

The experiments allow us to observe a number of features in the control system arising from the use of Web services:

- Each functional element of the robotic system has the same internal structure, all are equal.

- Each physical element of the robot is treated by the system in the same way, everyone is equal.

- The system is very flexible, can add and delete services even at runtime.

- The system is highly scalable, we can place each item in a different network node to run.

- We can reuse services or even share their implementation. For example, the obstacle detection services (Rc) may be used by other systems that require such information.

Each service can isolate units of information and its complexity, while enabling adapt each and every one of the

types of information to a common message exchange. That is, and this is one of the most important feature, different and very different information / data is shared by the system, for example, information coming from different sensing devices, with different data types and different frequency. The system allows you to isolate each unit of data, adapt it and treat it without causing other negative effects on the system. All elements of the robot, now, run a common language of communication between them.

VI.   CONCLUSIONS AND FUTURE WORK

This paper has proposed the development of robotic control systems based on Web Services. This proposal allows us to standardize the elements of a robotic system and enables the exchange and processing of the information produced by each of the elements. It has also shown the implementation of this system for behaviors such as autonomous navigation without/with obstacle avoidance. The resulting system performs with the requirements and desirable features such as flexibility, adaptability, short development cycles, dynamics and absorption of problems of operating frequencies and integration and management of diverse information, regardless of the source and nature of the devices. Self-adaptation of the communication provides the perfect link between the computer functions and the physical system it controls.

We are currently working on two lines. In the short term we are increasing the range of services available: services to optimize the path of roads, services for environment mapping, management services for more sensors and actuators, and so on. In the medium term we are investigating the adoption of cloud computing technologies to move services to the cloud, so that the system be independent of physical resources.

REFERENCES

[1] J.V. Berna-Martinez, F. Maciá-Perez, H. Ramos-Morillo and V. Gilart-Iglesias, *Distributed Robotic Architecture based on Smart Services*. IEEE INDIN 2006, ISBN 0-7803-9700-2, p. 480-485.

[2] IFR, Statistical Department, *World Robotics 2010 Service Robots*. 2010.

[3] S.L. Remy and M.B. Blake, *Distributed Service-Oriented Robotics*. IEEE Internet Computing. ISSN 1089-7801. Vol. 15 (2011), p. 70-74.

[4] J. Gowdy, *A Qualitative Comparison of Interprocess Communications Toolkits for Robotics.* Tech. report CMU-RI-TR-00-16, Robotics Institute, Carnegie Mellon University, June 2000.

[5] P. Harmon, M. Rosen and M. Guttman, *Developing E-business Systems and Architectures: A Manager's Guide*. Morgan Kaufmann Publishers, San Francisco, USA. ISBN 978-1558606654. 2001.

[6] J.A. Fernández-Madrigal, C. Galindo, J. Ganzález, E. Cruz-Martín and A. Cruz-Martín, *A software engineering approach for the development of heterogeneous robotic applications*. J. Robotics and Computer-Integrated Manufacturing. Vol. 24, issue 1, 2008. ISSN 0736-5845.

[7] R.   Volpe, I. Nesnas, T. Estlin, D. Mutz, R. Petras and H. Das, *The CLARAty Architecture for Robotic Autonomy*. Aerospace Conference, 2001, IEEE Proceedings. Volume: 1, pp. 121-132. ISBN 0-7803-6599-2.

[8] R. Alami, R. Chatila, S. Fleury, M, Ghallab and F. Ingrand, *An architecture for autonomy*. The Int. J. of Robotics Research, Vol. 17, No. 4, pp. 315-337 (1998).

[9] J. Rosenblatt: DAMN, *A Distributed Architecture for Mobile Navigation*. Thesis Doctoral, Tech. Report CMU-RI-TR-97-01, Robotics Institute, Carnegie Mellon University, 1997.

[10] J.V. Berná-Martinez, F. Maciá-Pérez, V. Gilart-Iglesias and D. Marcos-Jorquera, *Robotic architecture based on electronic business models. From physics components to smart services*. ICNCO 2006. ISBN 972-8865-60-0, p. 544-547.

[11] J.A. Gil Martínez-Abarca, F. Maciá Pérez, D. Marcos Jorquera, V. Gilart Iglesias. *Wake on LAN over Internet as Web Services*. Proceedings of the 11th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA'06). ISBN 1-4244-0681-1, p 1261-1268.

[12] J.V. Berná Martínez and F. Maciá Pérez, *Model of Integration and Management for Robotic Functional Components Inspired by the Human Neuroregulatory System*. IEEE ETFA 2010. ISBN 978-1-4244-6849-2.

[13] J. Kyle and T. Taylor, *Professional Microsoft Robotics Developer Studio*. Wiley Publishing, Inc. ISBN 978-0-470-14107-6.G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

AUTHORS PROFILE

J. V. Berna-Martinez was born in Spain in 1978. He received his engineering degree and the Ph.D. degree in Computer Science from the University of Alicante in 2004 and 2011 respectively. Since 2006, he is an Associate Professor at the University of Alicante. His research interests are in the area of computer networks, distributed systems, bio-inspired systems and robotics which are applied to industrial problems.

Francisco Maciá-Pérez was born in Spain in 1968. He received his engineering degree and the Ph.D. degree in Computer Science from the University of Alicante in 1994 and 2001 respectively. He worked as System's Administrator at the University of Alicante form 1996 to 2001. He was an Associate Professor from 1997 to 2001. Since 2001, he is an Assistant Professor and currently he is the Director of the Department of Computer Science and Technology at the University of Alicante. His research interests are in the area of network management, computer networks, smart sensor networks and distributed systems, which are applied to industrial problems.

A Neural Network System for Detection of Obstructive Sleep Apnea Through SpO2 Signal Features

Laiali Almazaydeh, Miad Faezipour, Khaled Elleithy
Department of Computer Science and Engineering
University of Bridgeport
Bridgeport, CT 06604, USA

Abstract—**Obstructive sleep apnea (OSA) is a common disorder in which individuals stop breathing during their sleep. These episodes last 10 seconds or more and cause oxygen levels in the blood to drop. Most of sleep apnea cases are currently undiagnosed because of expenses and practicality limitations of overnight polysomnography (PSG) at sleep labs, where an expert human observer is required. New techniques for sleep apnea classification are being developed by bioengineers for most comfortable and timely detection. In this study, we develop and validate a neural network (NN) using SpO2 measurements obtained from pulse oximetry to predict OSA. The results show that the NN is useful as a predictive tool for OSA with a high performance and improved accuracy, approximately 93.3%, which is better than reported techniques in the literature.**

Keywords- sleep apnea; PSG; SpO₂; features extraction; oximetry; neural networks.

I. INTRODUCTION

A. Background

Excessive daytime sleepiness and fatigue are the most symptoms of sleep disorders. The risk of sleepiness and fatigue lead to poor judgment and reaction time, especially for the drivers who do not take sleepiness seriously.

Sleep apnea is becoming a more common cause of sleepiness in children and adults. It is characterized by abnormal pauses of breathing or abnormally low breath during sleep. These pauses of breathing can range in frequency and duration. The duration of the pause might be ten to thirty seconds and upwards to as much as four hundred per night in those with severe sleep apnea [1].

Sleep apnea is classified into two types. The first type is Obstructive Sleep Apnea (OSA), which is generally caused by a collapse of the upper respiratory airway. The second one is Central Sleep Apnea (CSA), which is caused by inhibited respiratory drive, since the brain fails to appropriately control breathing during sleep. Out of the two sleep apnea types, OSA is more common than CSA [2].

Sleep apnea is not a problem to be taken lightly, since it is associated with, and is the possible cause of other medical conditions such as high blood pressure, heart disease, diabetes and depression.

Statistics show that over 18 million Americans suffer from sleep apnea, while an estimate of 10 million Americans remain undiagnosed [3]. Most cases go undiagnosed because of the inconvenience, expenses and unavailability of testing. Testing is inconvenient to the patient because it requires them to spend the night away from their bed, causing discomfort. It is expensive because testing is done in the hospital, causing machines and various technicians and staff to work over night. Testing is also widely unavailable due to sleep centers operating at full capacity, and those on the waiting list can be untreated for an additional 6 months.

Polysomnography (PSG) is a test commonly ordered for some sleep disorders. It records the breath airflow, respiratory movement, oxygen saturation, body position, electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG) [4].

To summarize, the mere dependency on PSG needs to be taken away from the laboratory for simpler detection and faster treatment. In this regard, we present a work based on a neural network using SpO_2 features extraction that will be used in a larger real time system for sleep apnea diagnosis. The objective of the system is to alert a patient who might be subject to an apnea attack.

B. Paper Organization

The rest of this paper is organized as follows. In Section II we glance at a variety of sleep apnea detection methods. Section III contains an overview of the system, including a description of the database of subjects, and the details of the analysis methodology including features extraction of the SpO_2 signal. The Neural Networks we used in this work is also described in the same Section. In Section IV, we detail the results of our system. Finally, Section V concludes this paper regarding the potential usefulness of our system, and highlights some directions for future research.

II. RELATED WORK

Several methods have been suggested for identification of sleep apnea over the past few years. Statistical features of different signals such as nasal air flow, the thorax and abdomen

effort signals, acoustic speech signal, oxygen saturation, electrical activity of the brain (EEG), and electrical activity of the heart (ECG) are commonly used in the detection.

Quiceno-Manrique et al. [5] proposed a simple diagnostic tool for OSA with a high accuracy (up to 92.67%) using time-frequency distributions and dynamic features in ECG signal.

The relationship between periodic changes in the oxygen saturation (SaO$_2$) profile and in the EEG pattern due to apnea events during the night was investigated in [6]. The spectral analysis of these two signals achieved 91% sensitivity, 83.3% specificity and 88.5% accuracy in OSA diagnosis.

Ng et al. [7] showed that thoracic and the abdominal signals were good parameters for the identification of the occurrence of sleep apnea. Using the mean of absolute amplitudes of the thoracic and the abdominal signals, the authors achieved good performance with a receiver operating characteristic value higher than 80%.

Wavelet transforms and an artificial neural network (ANN) algorithm were applied to the EEG signal in [8] to find a solution to the problem of identifying sleep apnea (SA) episodes. The system's identification results achieved a sensitivity of approximately 69.64% and a specificity of approximately 44.44%.

Based on spectral components of heart rate variability (HRV), frequency analysis was performed in [9] to detect sleep apnea. Using Fourier and Wavelet Transformation with appropriate application of the Hilbert Transform, the sensitivity was 90.8%. In addition, in [10] a bivariate autoregressive model was used to evaluate beat-by-beat power spectral density of HRV and R peak area, where the sleep apnea classification results showed accuracy higher than 85%.

The study in [11] assesses the analysis of various feature sets and a combination of classifiers based on the arterial oxygen saturation signal measured by pulse oximetry (SpO$_2$) and the ECG in order to evaluate sleep quality. With selected features of the SpO$_2$ and ECG signals, the Bagging with REP Tree classifier achieved sensitivity of 79.75%, specificity of 85.89% and overall accuracy of 84.40%.

III. METHODOLOGY

A. Subjects

The database of SpO$_2$ signals used in this research is available from the *PhysioNet* web site [12].

PhysioNet contains a growing collection of biomedical signals from healthy subjects and patients. The *PhysioNet* web site is a public service of the *PhysioNet* Resource funded by the National Institutes of Health's NIBIB and NIGMS. *PhysioNet* offers free access to *Apnea-ECG Database*, which we use to assess and validate our approach.

The *Apnea-ECG Database* contains 8 recordings with SpO$_2$ signals. These recordings have varying length from slightly less than 7 hours to nearly 10 hours each.

The subjects of these recordings were men and women between 27 and 63 years of age (mean: 43.8±10.8 years) with weights between 53 and 135 kg (mean: 86.3±22.2 kg). The

sleep recordings originated from 32 subjects (25 men, 7 female), who were recruited for previous studies on healthy volunteers and patients with obstructive sleep apnea [4].

B. SpO$_2$ Signal

SpO$_2$ is the amount of oxygen being carried by the red blood cell in the blood. Very simply, SpO$_2$ goes up and down
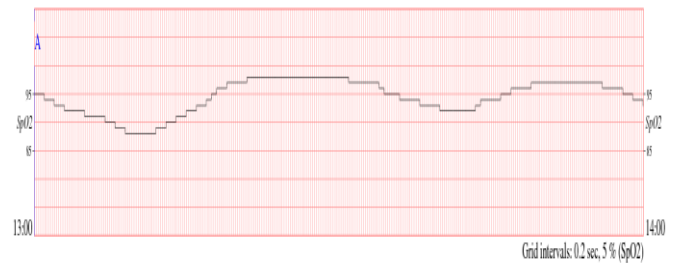
Figure 1. SpO$_2$ record with OSA negative subject [12].
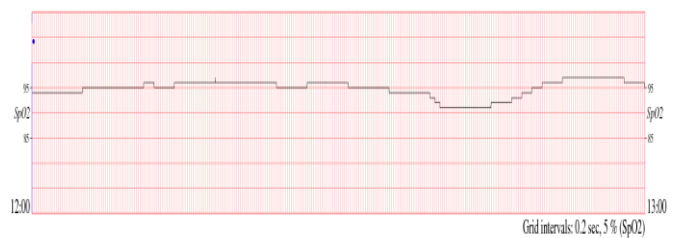
Figure 2. SpO$_2$ record with OSA positive subject [12].

according to how well a person is breathing and how well the blood is being pumped around the body [13].

SpO$_2$ measured by pulse oximetry can be useful in OSA diagnosis. Significant changes can be found in patients affected by OSA because of the recurrent episodes of apnea, which are frequently accompanied by oxygen desaturations [14].

Figure 1 depicts a common OSA negative subject, and Figure 2 shows a SpO$_2$ record with OSA positive subject. However, diagnosis of the disease is not evident by visual inspection.

C. Features Extraction

In our work, the SpO$_2$ signals are saved to separate files and processed off-line by an automated system we developed using MATLAB to compute two of the common oximetric indices and one nonlinear metric.

These three features are detailed as follows, respectively:

1) Delta index (Δ index): This is a common measure to detect the apneic events by measuring SpO$_2$ variability. Levy et al. [15] calculates Δ index as the sum of the absolute variations between two successive points, divided by the number of intervals. It is usually computed for 12-sec. intervals.

2) Oxygen desaturation indices of 3% (ODI3): This measure is obtained by calculating the number of times per hour with values of SpO$_2$ greater than or equal to 3% from the baseline. The baseline is set initially as the mean level in the first 3 minutes of recording [16].

Central tendency measure with radius 0.5 (CTM50): This measure applied in [16]. CTM50 is computed by selecting a circular region of radius 0.5 around the origin, counting the number of points that fall within the radius, and dividing by the total number of points. Given N data points from a time series, N-2 would be the total number of points in the scatter plot. Hence, CTM50 can be computed as [17]:

$$CTM = \frac{\sum_{i=1}^{N-2} \delta(d_i)}{N-2} \qquad (1)$$

where,

$$\delta(d_i) = \begin{cases} 1 & if \left[(x_{i+2} - x_{i+1})^2 + (x_{i+1} - x_i)^2\right]^{1/2} < 0.5 \\ 0 & otherwise \end{cases} \qquad (2)$$

D. Multi layer Networks Classifiers

In this research, we apply a neural network (NN) as a classifier to identify the diagnostic performance of OSA using SpO_2 features.

A neural network is used to perform a pattern classification task. NNs classifiers have been proven to be extremely helpful in assisting medical specialties in clinical diagnosis [18].

The NN described in this study is based on three layers feed-forward neural network learned with back-propagation algorithm; an input layer, an output layer, and a hidden layer. The hidden layer consists of a direct connection between the input and the output layer.

The three SpO_2 features act as inputs to a neural network, and the diagnosis of OSA is the target. This is achieved by presenting previously recorded inputs to a neural network and then tuning it to produce the desired target outputs. This process is called neural network training.

A total of 93data sets (41 with a positive diagnosis of OSA and 52 with a negative diagnosis of OSA) are used. Validation is done with the same training dataset, and test dataset has been set to 17% of the original data. With these settings, the input vectors and target vectors will be randomly divided into two sets as follows:

- 83% are used for training and validation.

- The last 17% are used as a completely independent test of network generalization.

The training set with 78 samples was used to train the network. Network parameters are adjusted through training by attempting to minimize the error between the target (t) and the actual (y) network output values. This error is expressed as the mean square error [19]:

$$E = \frac{1}{N} \sum_{n=1}^{N} \|t_n - y_n\|^2 \qquad (3)$$

where N is the number of samples in the training set.

In the training phase, the Purelin linear transfer function has been used as an activation function of the output layer of the network (for improving error derivative) [18]. Since the output space must be divided into two regions: OSA positive and OSA negative, we suggest using a single output node.
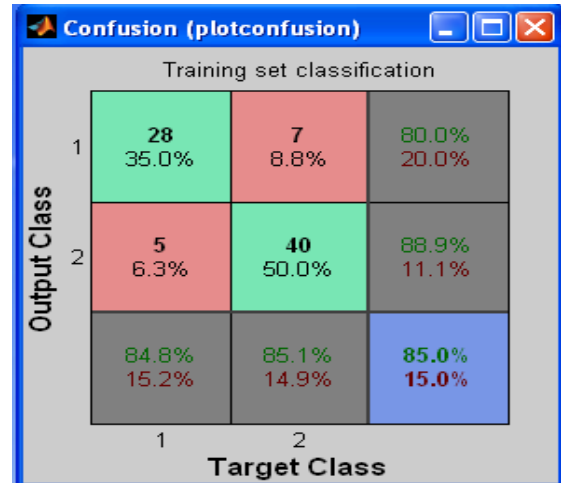
Figure 3. Confusion matrix for training set classification.
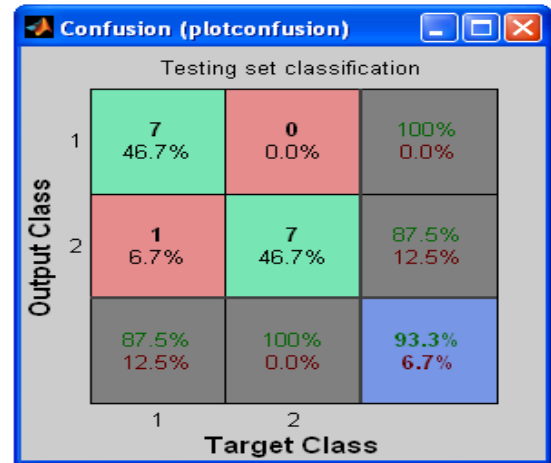
Figure 4. Confusion matrix for testing set classification.

We applied Hardlim function [18] to test the data to improve the output of the network in the validation and testing phases.

IV. RESULTS

A. Performance Evaluation

We evaluated the classification performance of the selected network configurations on the test set. Sensitivity, specificity for testing data and accuracy are computed. A confusion matrix is generated for the NN evaluation.

Figures 3 and 4 show the confusion matrix for training and testing set classification, respectively. The confusion matrix shows the total percent of correctly classified cases and the total percent of misclassified cases.

The results show very good network validation performance of 85% and high diagnostic performance with an accuracy of 93.3% correct detection rate (sensitivity 87.5%, and specificity 100%).

As a result, to reduce the dependency on complex PSG test measures, we find that NN using SpO$_2$ measurements is a practical and useful screening test to estimate whether patients have sleep apnea or not.

TABLE I.
Comparison of Sleep Apnea Detection Approaches.

Method	Ref.	Approach	*Performance* [%]		
			Se	*Sp*	*Acc.*
Quiceno-Manrique et al.	[5]	ECG signal			92.67
Alvarez et al.	[6]	SaO$_2$ and EEG signal	91	83.3	88.5
Ng et al.	[7]	Thoracic and abdominal signals	NA	NA	80
Lin et al.	[8]	EEG signal	69.64	44.44	NA
Schrader et al.	[9]	HRV Fourier and Wavelet Transformation	90.8	NA	NA
Mendez et al.	[10]	Bivariate autoregressive model of HRV	NA	NA	85
Xie et al.	[11]	SpO$_2$ and ECG	79.75	85.89	84.40
Proposed	**-**	**Features extraction of SpO$_2$ signal**	**87.5**	**100**	**93.3**

B. Comparison With other Works

We performed a comparison with other sleep apnea detection techniques. Table I represents comparative results. As can be seen, our system has achieved a comparable or better performance. This applies to the other works that rely on the SpO$_2$ signal as well as other biometric signals.

V. CONCLUSION AND FUTURE DIRECTIONS

In this work, we studied the possibility of the detection of sleep apnea from the SpO$_2$ signal variation patterns. We further developed a NN using the SpO$_2$ signal features and evaluated its effectiveness. This study has demonstrated a high performance and an improved accuracy of the NN.

A future direction to this work would be to apply our methodology to a larger population to validate the results obtained with this radius.

As another future direction, we are also planning to compute the CTM with several radii for every SpO$_2$ signal in both OSA positive and OSA negative groups. Then, we select the optimum radius that achieves the most significant differences.

Moreover, we are planning to analyze the ECG features signals, in order to use it along with the SpO$_2$ signals to build a two parameter technique and apply that as a system for automated recognition of OSA.

REFERENCES

[1] Sleep Disorders Guide. *www.sleepdisordersguide.com*.

[2] S. Isa, M. Fanany, W. Jatmiko and A. Arymurthy,"Sleep Apnea Detection from ECG Signal, Analysis on Optimal Features, Principal Components, and Nonlinearity," in *Proceedings of the 5th IEEE International Conference on Bioinformatics and Biomedical Engineering (iCBBE), pp. 1-4*, May 2011.

[3] SleepMedInc. *www.sleepmed.md*.

[4] P. Chazal, T. Penzel and C. Heneghan, "Automated Detection of Obstructive Sleep Apnoea at Different Time Scales Using the Electrocardiogram," *Institute of Physics Publishing, vol. 25, no. 4, pp. 967-983*, Aug. 2004.

[5] A.F. Quiceno-Manrique, J.B. Alonso-Hernandez, C.M. Travieso-Gonzalez, M.A. Ferrer-Ballester and G. Castellanos-Dominguez, "Detection of Obstructive Sleep Apnea in ECG Recordings using Time-Frequency Distributions and Dynamic Features," in *Proceedings of the 31st IEEE International Conference in Medicine and Biology Society (EMBS 2009), pp. 5559-5562*, Sep. 2009.

[6] D. Avarez, R. Hornero, J. Marcos, F. Campo and M. Lopez, "Spectral Analysis of Electroencephalogram and Oximetric Signals in Obstructive Sleep Apnea Diagnosis," in *Proceedings of the 31st IEEE International Conference on Engineering in Medicine and Biology Society (EMBS 2009), pp. 400-403*, Sep. 2009.

[7] A. Ng, J. Chung, M. Gohel, W. Yu, K. Fan and T. Wong, "Evaluation of the Performance of Using Mean Absolute Amplitude Analysis of Thoracic and Abdominal Signals for Immediate Indication of Sleep Apnoea Events," J*ournal of Clinical Nursing, vol. 17, no. 17, pp. 2360-2366*, Sep. 2008.

[8] R. Lin, R. Lee, C. Tseng, H. Zhou, C. Chao, J. Jiang,"A New Approach for Identifying Sleep Apnea Syndrome Using Wavelet Transform and Neural Networks," *Biomedical Engineering: Applications, Basis & Communications, vol. 18, no. 3, pp. 138-143*, 2006.

[9] M. Schrader, C. Zywietz, V. Einem, B. Widiger, G. Joseph, "Detection of Sleep Apnea in Single Channel ECGs from the PhysioNet Data Base," *Computers in Cardiology 2000, vol. 27, pp. 263-266*, Sep. 2000.

[10] M. Mendez, D. Ruini, O. Villantieri, M. Matteucci, T. Penzel and A. Bianchi, "Detection of Sleep Apnea from Surface ECG Based on Features Extracted by an Autoregressive Model, " in *Proceedings of the IEEE International Conference on Engineering in Medicine and Biology Society (EMBS 2007), pp. 6105-6108*, Aug. 2007.

[11] B. Xie, H. Minn, "Real Time Sleep Apnea Detection by Classifier Combination," in *IEEE Transactions on Information Technology in Biomedicine (in Press)*, 2012.

[12] PhysioNet, *www.physionet.org*.

[13] What is SpO$_2$. *http://www.neann.com/spo2.htm*.

[14] J. Marcos, R. Hornero, D. Alvarez, F. Campo, C. Zamrron and M. Lopez, "Single Layer Network Classifiers to Assist in the Detection of Obstructive Sleep Apnea Syndrome from Oximetry Data," in *Proceedings of the 30th IEEE International Conference on Engineering in Medicine and Biology Society (EMBS 2008), pp. 1651-1654*, Aug. 2008.

[15] P. Levy, J. Pepin, C. Blanc, B. Paramelle and C. Brambilla, "Accuracy of Oximetry for Detection of Respiratory Disturbances in Sleep Apnea Syndrome," *Chestjournal, vol. 109, no. 2, pp. 395-399*, Feb. 1996.

[16] D. Alvarez, R. Hornero, D. Abasolo, F. Campo and C. Zamarron, "Nonlinear Characteristics of Blood Oxygen Saturation from Nocturnal Oximetry for Obstructive Sleep Apnoea Detection," *Institute of Physics Publishing, vol. 27, no. 4, pp. 399-412*, Apr. 2006.

[17] J. Jeong, J. Gore and B. Peterson, "A Method for Determinism in Short Time Series, and its Application to Stationary EEG," in *IEEE Transactions on Biomedical Engineering, vol. 49, no. 11, pp. 1374-1379*, Nov. 2002.

[18] A. El-Solh, M. Mador, E. Brock, D. Shucard, M. Abdul-Khoudoud and B. Grant, "Validity of Neural Network in Sleep Apnea," *Sleep journal, vol. 22, no. 1, pp. 105-111*, 1999.

[19] J. Marcos, R. Hornero, D. Alvarez, F. Campo and Miguel Lopez, "Applying Neural Network Classifiers in the Diagnosis of the Obstructive Sleep Apnea Syndrome from Nocturnal Pulse Oximetric Recordings," in Proceedings *of the 29th IEEE International Conference on Engineering in Medicine and Biology Society (EMBS 2007), pp. 5174-5177*, Aug. 2007.

Automated Detection Method for Clustered Microcalcification in Mammogram Image Based on Statistical Textural Features

Kohei Arai, Indra Nugraha Abdullah, Hiroshi Okumura

Graduate School of Science and Engineering
Saga University
Saga City, Japan

Abstract—**Breast cancer is the most frightening cancer for women in the world. The current problem that closely related with this issue is how to deal with small calcification part inside the breast called micro calcification (MC). As a preventive way, a breast screening examination called mammogram is provided. Mammogram image with a considerable amount of MC has been a problem for the doctor and radiologist when they should determine correctly the region of interest, in this study is clustered MC. Therefore, we propose to develop an automated method to detect clustered MC utilizing two main methods, multi-branch standard deviation analysis for clustered MC detection and surrounding region dependence method for individual MC detection. Our proposed method was resulting in 70.8% of classification rate, then for the sensitivity and specificity obtained 79% and 87%, respectively. The gained results are adequately promising to be more developed in some areas.**

Keywords- Automated Detection Method; Mammogram; Micro calcification; Statistical Textural Features; Standard Deviation.

I. INTRODUCTION

Uncontrolled growth of breast cells caused by a genetic abnormality is a short meaning of breast cancer. Mostly breast cancer starts from lobules cells, glands or milk producer and duct cells, part that transporting milk from the lobules to the nipple. This cancer is exceptionally rare starts from the stromal tissues and the fatty connective tissues, but if it happens the cell changes and have the ability to divide without control and forming a tumor.

A tumor can be categorized into two types, first is benign type, which is a tumor that nearly same with the normal one in appearance, slow growth, do not spread to the other body parts and the second is malignant type, which has characteristics that vice versa from benign type.

Based on the Globocan, an international World Health Organization agency for cancer located in France, breast cancer is the most frightening cancer for women in the world, and become the most common cancer both in developing and developed regions. In 2008 estimated 1.38 million new cancer cases diagnosed, the proportion of breast cancer was 23% of all cancers.

TABLE I. SUMMARY OF BREAST CANCER INCIDENCE AND MORTALITY WORLDWIDE IN 2008

Region	Cases	Deaths
World	1384	458
Africa Region (Afro)	68	37
American Region (Paho)	320	82
East Mediterranean Region (Emro)	61	31
Europe Region (Euro)	450	139
South-East Asia Region (Searo)	203	93
Western Pacific Region (Wpro)	279	73

From the above table, we can notice to all regions, the rates of mortality are very high and obviously there is no region in the world that has not affected with this cancer. The most worrisome region is Europe region with the number of incidence cases is 450 and mortality cases is 139. That means the rate of mortality in this region is 0.308 and made this rate is nearly equal to the rate of the world region which is 0.331.

As seen below, first rank occupied by breast cancer and the portion compared to the other cancers is extremely high which represented by age-standardized mortality rates (ASR) with 38.9% for incidence and 12.4% for mortality.
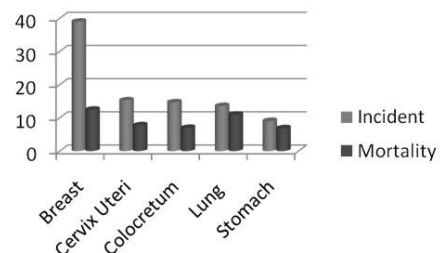
Figure 1. Age-standardized mortality rates (ASR) for women per 100000.

In order to overcome this problem, every woman needs to concern about their health through several continuous tests; Breast cancer tests covering screening tests, diagnostic tests, and monitoring tests.

In this study, we will focus on the test in screening tests called Mammograms, this test has been using for almost 40 years and the most valuable tool not only to screen the cancer, and also to diagnose and evaluate. The cooperation between mammography technician and radiologist can be involved to help the doctor increasing the accuracy of the final decision. Mammogram can read any signs of abnormality such as asymmetry of shape, irregular areas, clusters of small micro calcification (MC) and area of skin thickening. Commonly, the radiologist also operates a Computer Aided Diagnosis (CAD) system. This system will analyze the digital format of mammogram, and the result is a mammogram with any markers in the suspicious areas. The difficulty for the system is to detect clustered extra small calcifications in the form of clusters called with clustered MC.

Many researchers conducted to find the best method detecting the clustered MC. Yu and Guan [5] made a CAD that consist of two steps, first was the detection of MC pixel through classification of wavelet features and gray-level statistical features, and the second was the detection of individual MC objects, surely that the system needs a large amount of time and memory. Then Abdallah et.al [3] reported the efficient technique to detect the ROI using multi-branch standard deviation analysis and resulting the promising result which more than 98% of true positive (TP) cases. The most current one is Tieudeu et.al [1] detect the clustered MC based on the analysis of the their texture. Selection process has done via labeling method of the image that obtained from subtraction the smoothing image from the contrast enhance image, and classification of features successfully completed by neural network.

This method was resulting superfine sensitivity equal with 100% and 87.7% of specificity with proper classification rate 89%.

Therefore in this study we propose to make a system that can automatically detect the clustered MC based on the strengths from the Tieudeu et.al [1] with different enhancement image algorithm combine with detection of individual MC as done by Kim and Park [4] which employed the statistical features to detect the MC.

II. PROPOSED METHOD

A. Segmentation

The data set comes from the Japanese Society of Medical Imaging Technology, and each image has size 2510x2000 pixels and each pixel consists of 10 bits. Three categories can be found in this data set, namely calcification, normal and tumor categories. Before enter to the main process, the data should be preprocessed. The objective is to gain efficiency of time and/or memory processing, in consideration of the large size of image and size of each pixel.

Many studies have been implementing the Otsu threshold method when they want to form a binary image from the gray scale image. The main reasons are both the time processing is remarkably short and provides a satisfaction result. In this study, the segmentation operation is not only the Otsu method itself but also morphological operation being involved.

Otsu threshold method is a binarization method that calculates a measure of spread of the pixel value and iterates all possible values as a threshold. The objective is to find the threshold value based on a minimum value of within class variance and the equation described as below:

$$\sigma_W^2 = W_b \sigma_b^2 + W_f \sigma_f^2. \tag{1}$$

Where σ_W^2 is within class variance, W indicating weights, σ is a variance, b and f are background and foreground, respectively.

As a deficient result from Otsu threshold method from this data set, we need to improve the segmentation method to gain the better result of segmented image. In this study, we are applying one of morphology operations that called erosion operation. This is not ordinary erosion operation but erosion operation with small modification. There still remaining noise in the previous segmented image that must be removed which is the patient tag number, through this method that noise easily be removed. In spite of need much time to process, yet, will produce a satisfied result. The algorithm of our special erosion operation can be seen as below:

```
Input ROW, COL, MAX_ITER
Input N[ROW][COL]
For x=1to MAX_ITER do
        For i = ROW/2 to ROW do
                For j=1toCOLdo
                If N[i][j] = 0 then
                        N[i − 1][ j] = 0
                end if
                if j < COL/2 and N[i][j] = 0 then
                        N[i][ j − 1] = 0
                end if
                if j > COL/2 and N[i][j] = 0 then
                        N[i][ j + 1] = 0
                end if
                end for
        end for
end for
```

Algorithm 1. Our erosion algorithm

B. Detection of Clustered MC

1) Breast Tissue Detection Based on Texture-based Analysis

In this study, we are applying the method that has developed by Tieudeu et.al [1] with modification in one specified area. They are developed the main method by utilizing three methods. First is enhancing the contrast of the original image then produce an image called with contrast enhance image (CI) and the way to get this image become a point of modification. The second is smoothing the original image then produce an image called with smoothed image (SI). The last is subtraction the smoothed image from enhanced image then called with difference image (DI).

This adoption motivated by clustered MC that allied with breast mass can be concluded as a benign or even premalignant cancer. Frequently, MC only associated with extra cell growth inside the breast. Different with the previous study when forming the CI, we are using the histogram equalization method with an aim to spread the most frequent intensity values that make the lower contrast reach a higher contrast. The details represented by the equation below:

$$prob_n = \frac{\#pixels\ intensity\ n}{\#pixels};\ \ n = 0, 1, 2\ ...\ L. \qquad (2)$$

$$M_i = floor\left((L)\sum_{n=0}^{i} prob_n\right). \qquad (3)$$

Where $prob_n$ denotes the normalized histogram for each gray level value, n is gray level values, L is maximum gray level value and M is image matrix.

2) Multi-branch Standard Deviation Analysis

MC related with local maxima values in the image. This idea became a point to find up a correlation between the local maxima and its neighboring pixels. In this study, we conduct an analysis with make use of standard deviation method to find that correlation as reported by Abdallah et.al [3]. Based on visual observation for calcification category, there is not only one or two clustered MC in one image but even more than five clusters of MC can be found. In relation of that problem, developing a multi-branch point of view become something primary needs. It because highly possible if we find a local maxima in one direction and after take a look in a different direction that point is not a local maxima. That critical point provides promising solution to find the clustered MC in one small area. The illustration provided as below:
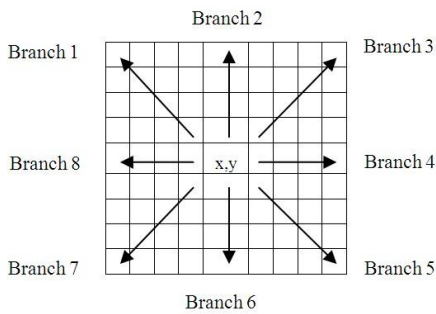
Figure 2.   Multi-branch standard deviation analysis to find MC.

Where x, y point is an ideal local maxima if from all branches seen as a local maxima, branch direction move clockwise start from branch 1, branch 2, branch 3, branch 4, branch 5, branch 6, branch 7 and branch 8. At the time that we want to know one point is local maxima from one branch, the threshold value and the counter needed. While calculating the threshold between the central pixel and its neighbor pixels if the standard deviation greater than the threshold value the counter will be increasing by one, whereupon an ideal local maxima is the point that has a counter value equal with eight. Described with the following equation:

$$STD_i = \sqrt{\frac{\sum_{i=1}^{n}(Center-x_i)^2}{n}};\ \ i = 1, 2, ...,8. \qquad (4)$$

Where:

STD_i	=	Standard Deviation at branch i
Center	=	Cluster center
x_i	=	Gray level value at the specified position i
n	=	Number of pixels

As said before the counter will have a maximum value 8, that value is equal with a total of branches in this method. Size

of the detection window in this method is 9x9, and that size obtained from the reference that MC in mammogram image can be captured through that size of the mask. ROI as a final result of this section has size 128 x 128 which matched with the most clustered MC size. In this study, one mammogram image represented by one ROI although there is more than one clustered of MC can be found. It because this system's purpose is giving assistance to the doctor and the radiologist when they are facing the final decision, at the moment only one representation of clustered MC is found still means the patient categorized as calcification and need further treatment. Moreover, selection criterion of ROI is the area with the highest number of suspicious local maxima pixels.

C.   Detection of Individual MC

1) Surrounding Region Dependance Method

In this part, we will talk about detection of individual MC through the method that previously used by Kim and Park [4]. The method is Surrounding Region Dependence Method (SRDM) which utilizing rectangular and threshold in order to obtain the distribution matrix. This matrix represents a characteristic of the ROI image that related to calcification case or not. Consider two rectangular windows are centered in *x, y* pixel, with largest window has size 5, and intermediate is 3. As shown with the image below:
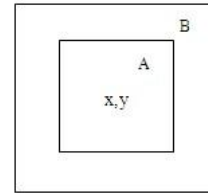
Figure 3.   Surrounding Region Dependence Method.

A is inner surrounding region, and B is outer surrounding region.

SRDM involves a *M(q)* or matrix of a surrounding region dependence obtained from transformation of an ROI image and *q* is a given threshold value. The details presented as below:

$$M(q) = [\alpha(i,j)], 0 \le i \le n. \qquad (5)$$

$$\alpha(i,j) = \#\{(x,y)|c_A(x,y) = i\ and\ c_B(x,y) = y, (x,y) \in L_x \times L_y\}. \qquad (6)$$

Where $L_x \times L_y$ is two dimensional image space and c_A, c_B are inner count, outer count, respectively.

Feature extraction is an essential part when dealing with the classification term. Hereafter horizontal, vertical, diagonal, and grid-weighted sums are extracted from the characteristics of the element distribution in the SRDM matrix as textural features.

The distribution for a positive ROI will tend to the right and/or lower right of the matrix and indicate us if subtraction neighbor values in inner and outer rectangles from the center value more than the threshold, those values will be located at

the right part of the matrix. For negative ROI has a contrary description, the distribution will tend to other location of the positive ROI.

III. EXPERIMENTS

A. Segmentation

Segmentation process in this study has an aim to remove the noise which called mammogram's tag number and the backlight. Otsu threshold method successfully removed the backlight from the image and the remaining noise is tag number, this noise removing process is handling by erosion method. We have 65 images in the data set and only three images that categorized as dissatisfied results. The reason of the negative appearance is because breast size of those patients classified as extraordinarily large size and has a round shape that made on both corners of the mammogram image have a less visible area. The satisfy segmented image and dissatisfy segmented image presented respectively as below:

Figure 4. Satisfy segmented images.

Figure 5. Dissatisfy segmented images.

B. Detection of Clustered MC and Individual MC

Through the described method, we obtained all images called the CI, SI and DI. From below DI image we can obviously see the breast tissue area and hereafter this area will be the main concern when finding the clustered MC. As an example, shown with the images below:
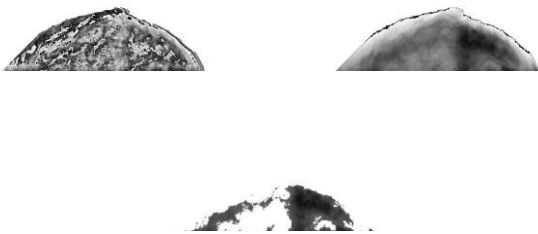
Figure 6. Sample of the CI, SI and DI images.

The naming format of below images is category plus image number in data set, for example, C5 means ROI image that categorized as calcification with number image is 5, in the sequel example are C6U and C6L shown us that U has originality from the upper part of C6, hence, the C6L from the lower part. The others categories denote with T for tumor category, and N for normal. Mostly the MC detected on this

category and obviously showed that this method was suitable to detect the clustered MC. From the experiment, threshold value for clustered MC detection equal with 8 was the maximum threshold value. Hereupon, best threshold value for individual MC detection was 3. According to the proposed method, resulting ROI images as presented below:
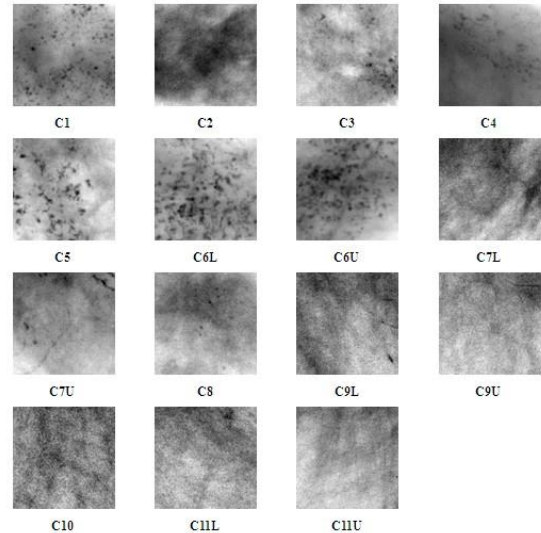
Figure 7. Region of Interests (ROI) from calcification category.

In this part, the data set separated into two parts that are training and testing parts with the data proportion were 50% and 50%, respectively. For training data, we were adding ideal output in the form of ROI from all categories manually to train the classifier and then extracted their features. Manual observation of all data passed, and we acquired the information that in category tumor also found clustered MC. At least, four tumor images possessing clustered MC and those were T1, T2, T4 and T8 images. That finding guidance based on sketch images that provided inside the data set, as seen in the following image:
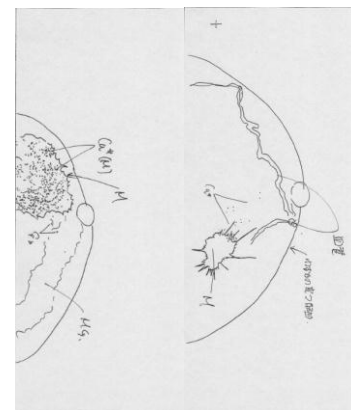
Figure 8. Sketch images, calcification category (left), tumor category contains MC (right).

Classification result for this system was good enough pointed by the classification rate that was 70.8%.

Mostly, true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are the options for

diagnosis decision. TP means similarity cancerous of judgment from an expert and system, TN means similarity a non-cancerous judgment from an expert and system, FP means a non-cancerous classified as cancerous, and last is FN which means a cancerous classified as a non-cancerous. After the experiment, the results shown with the following table:

TABLE II.   CONFUSION MATRIX

TP	FP
15	6
FN	**TN**
4	40

Hereafter let we talk about other parameters that could indicate the system whether is acceptable or not which are sensitivity and specificity. Both parameters shown as below:

Sensitivity = TP/(TP+FN)
Specificity = TN/(TN+FP)

Then obtained those values equal with 79% and 87%, respectively. Regarding the sensitivity value was deficient, there is a primary reason, because we were trying to find the MC which had a round shape. In fact on few images the shape of MC included round as well as long shape. The system could not find that shape of calcification precisely. The reason is the window for detecting local maxima pixel that has identified as MC was a small size rectangular. On account of that reason, the value of four for a false negative was appeared. Described with the following image:
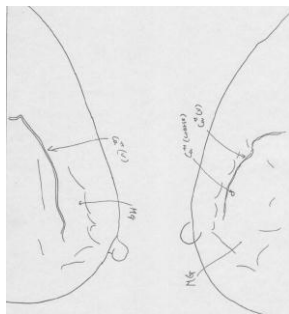
Figure 9.   Sketch image of different calcification shape.

We can obviously see the long shape denotes with $Ca^+(v)$. That shape also became a barrier for detector of clustered MC to detect the correct shape.

IV.   CONCLUSION

This study is developed exclusively to detect clustered MC. We have reasons why this system could not gain the perfect classification rate, first is the textural features that became an input of the network had a lack of proper characteristics to discriminate a clustered of MC and nonclustered MC, and second because we worked on small data only consists of 65 images. On the contrary, we realized to publish this kind of data should have a permit for their own information. However, from the gained results are adequately promising to be more developed in some areas, parallel with important thing for a human being is to help each other.

V.   FUTURE WORK

The future work that can be developed from this current progress is the detection of clustered MC to determine a mammogram image is included as benign or malignant. Conduct another localized and efficient method when forming contrast enhance image.

REFERENCES

[1]   A. Tieudeu, C. Daul, A. Kentshop, P. Graebling, D. Wolf, "Texture-based analysis of clustered microcalcifications detected on mammograms,"Digital Signal Processing,Vol 22, pp. 124-132, 2011.

[2]   Lucio F. A. Campos, A. C. Silva, A. K. Barros, "Diagnosis of breast cancer in digital mammograms using independent analysis and neural network," CIARP, pp. 460-469, 2005.

[3]   M. H. Abdallah, A. A. Abubaker, R. S. Qahwaji, M. H. Saleh, "Efficient technique to detect the region of interests in mammogram image," Journal of Computer Science, Vol 4(8), pp. 652-662, 2008.

[4]   Jong Kok Kim and Hyun Wook Park,"Statistical textural features for detection of microcalcifications in digitized mammograms," IEEE Transactions on Medical Imaging, Vol 18(30), pp. 231-238, 1999.

[5]   Songyang Yu and Ling Guan,"A cad system for the automatic detection of clustered microcalcifications in digitized mammograms films," IEEE Transactions on Medical Imaging, Vol 19(2), pp. 115-125, 2000.

AUTHORS PROFILE

KOHEI ARAI received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a counselor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology from 1998 to 2000. He was a counselor of Saga University for 2002 and 2003. He also was an executive counselor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008.

INDRA NUGRAHA ABDULLAH was born in Bogor, Indonesia on June 18th 1987. Finished his bachelor degree in Bogor Agricultural University and graduated from Saga University for master degree in the field of Information Science on March 2011. He is currently pursuing to get Ph.D. Degree from the same university with specialization in image processing area. Leaf identification becomes his interest in his latest degree.

HIROSHI OKUMURA was born at Kyoto, Japan in 1964. He received B.E.S.E. and M.E.S.E. degree from Hosei University in 1988 and 1990, respectively, and Ph.D degree on environmental engineering from Chiba University in 1993. He became a research associate at   Remote Sensing and Image Research Center, Chiba University first in 1993. Next, he became a research associateand a lecturer at the Department of Electrical Engineering, Nagaoka University of Technology in 1995 and 2000, respectively. He is now an associate professor at the Department of Information Science, Saga University. His research interests are in image and speech processing and remotesensing.

Eye Detection Based-on Color and Shape Features

Aryuanto Soetedjo

Department of Electrical Engineering
National Institute of Technology (ITN)
Malang, Indonesia

Abstract—**This paper presents an eye detection technique based-on color and shape features. The approach consists of three steps: a rough eye localization using projection technique, a white color thresholding to extract white sclera, and an ellipse fitting to fit the ellipse shape of eye. The proposed white color thresholding utilizes the normalized RGB chromaticty diagram, where white color objects are bounded by a small circle on the diagram. Experimental result shows that the proposed technique achieves a high eye detection rate of 91%.**

Keywords-eye detection; color thresholding; ellipse fitting.

I. INTRODUCTION

Nowadays researches on facial features detection and recognition attract attention for researchers, due to their significant contribution in the Human Computer Interface (HCI) applications. One of the interesting topics is an eye detection technique which is employed for activating mobile phone [1]; phoning, reading and browsing [2]; monitoring driver fatigue [3]; and improving human-computer dialogue [4].

Basically, eye detection technique is divided into three methods [5]: a) Shape-based, b) Appearance-based; c) Hybrid shape and appearance-based. In shape-based method, the eye shape such as eye edges, pupil, and eye corners are used as the features. In appearance-based method, the photometric appearance of eye is used for detection. Hybrid method combines both shape and appearance methods to exploit their advantages.

In [6], the eye is detected by extracting the pixels darker than surrounding and grouping them according to their geometric centers. The best eyeball is selected by applying a set of geometric constraints. In [7], an ellipse fitting method is employed to detect the potential eye region. The antropological characteristics of human eye, such as the width and height ratio of the eye and the orientation of the major axis are used to verify the detected eye. The horizontal projection is employed in [8] to find the eyebrows. Further the region below is extended to locate the pupil. The generalized projection function is proposed in [9]. It combines the integral projection, variance projection, and hybrid projection functions. In the method, pupil is located by applying the vertical and horizontal projections.

In [10], a cascaded classsifier trained by AdaBoost algorithm is employed to detect multi-view face and eye. To detect eye, they proposed four-step detection, i.e. eye's candicate area restriction, eye's candidates detection using an eye classifier, eye pair candidates using an eye pair classifier, and an eye pair decision based on features extracted from previous steps. Instead of rectangle feature which is used in [10], a pixel-pattern-based texture feature is used in [11]. They employed Adaboost and Support Vector Machine classifier to classify the eye and non-eye image patches. In [12], wavelet and Neural Network is used to classify an eye or non-eye region.

In [13] and [14], shape and appearance are combined to generate an eye model which is invariant to scale and rotation changes. In [13], a color model is employed for coarse-scale eye tracking and grayscale appearance is used for precise localization.

Previous works as decribed above mostly utilize intensity image for detecting eye. Color is usually employed for skin color modeling [13]. In eye detection, the Infrared light (IR) is the dominant use of color [15], [16]. This method relies on the fact that when IR light falls on the eye, it produces the bright pupil effect. However, it has a drawback that a special geometric arrangement of IR light emitting diodes (LEDs) should be prepared for producing the desired bright pupil effect.

In this paper, a color thresholding technique is proposed to detect eye. It takes an advantage of the white color of eye's sclera [17]. Further the shape feature, i.e. ellipse shape, is utilized to find the precise location of eye. The method assumes that a face region has been obtained by a face detection technique. The proposed eye detection consists of three steps: a) Once a face is detected, the rough location of eye is localized by horizontal projection of the gradient image; b) White color thresholding is employed to extract eye's sclera; and c) An ellipse detection technique is applied to detect the precise eye location.

The rest of paper is organized as follows. Proposed technique is described in Section 2. Section 3 discusses the experimental results. Finally conclusion is covered in Section 4.

II. PROPOSED TECHNIQUE

A. Overview

Fig. 1 illustrates the overview of proposed system. Prior to eye detection, the face detection is applied to locate a face. It employs the popular Viola-Jones face detector [18]. Since the Viola-Jones detector requires a grayscale image as the input, it needs to convert the RGB color image to the grayscale image

as shown in the figure. After face is detected, the eye detection algorithm starts by localizing the rough eye location using horizontal projection technique of the gradient image. By analyzing peaks of the horizontal projection, the rough location of eye is obtained. Then the original color image is cropped according to the rough eye location. In the cropped image, white color of eye's sclera appears as the significant clue for detecting eye. Thus white color thresholding is applied to detect the sclera. It creates a blob image (binary image) contains the white and non-white objects. Since the shape of eye (sclera) is ellipse then an ellipse detector is applied to detect the boundary of eye's sclera.

images with their horizontal projections of the upper-half part. Fig. 3(b) shows that maximum peak of the projection denotes the location of the eyelid. However this situation does not always hold. In Fig. 3(d), maximum peak of the projection denotes the location of eyebrow. Fortunately, from the experiments, the location of eyelid could be found as follows. It first finds the two maximum peaks of the projection. From these two maximum peaks, find the one which the vertical position is lower than the other one, and assign it as the vertical position of eyelid.

Figure 2. (a). Grayscale image; (b) Vertical gradient image.
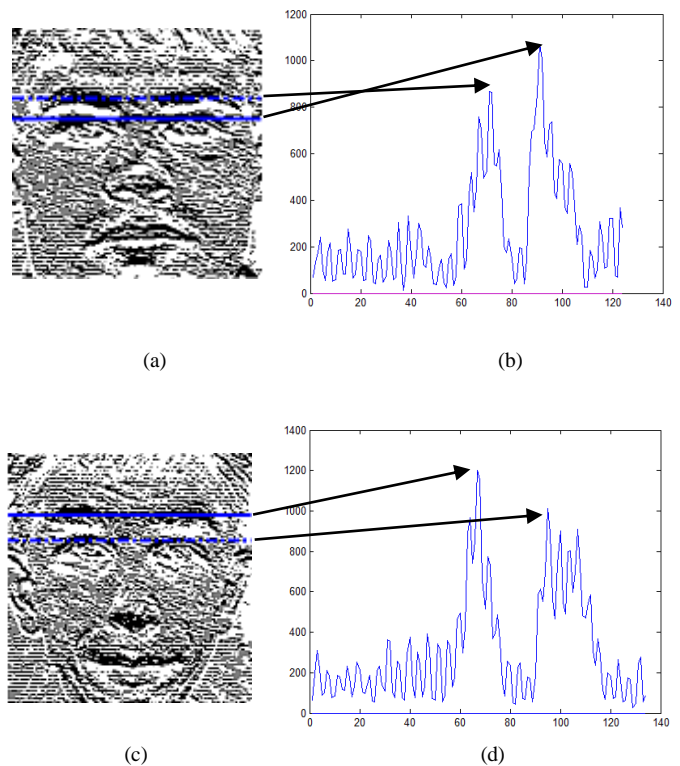
Figure 3. (a), (c) Gradient image; (b), (d) Horizontal projections of the image (a) and (b) respectively.

Once the vertical position (*vp*) is found, the upper boundary (*vu*) and lower boundary (*vl*) of eye are defined as

$$vu = vp - (0.1 \times hface) \qquad (1)$$

$$vl = vp + (0.1 \times hface) \qquad (2)$$

RGB Color image

Grayscale conversion

Face detection

Rough eye localization

Image cropping

White color thresholding

Ellipse detection

Detected eye

Figure 1. Overview of proposed eye detection.

B. Rough Eye Localization

It is observed from the face image that vertical gradient of the image is dominated with horizontal edges, due to the appearance of the facial's components such as eyebrow, eyelid, and lip as shown in Fig. 2(b). Therefore, if the gradient image shown in the figure is projected onto vertical axis (horizontal projection), then there will exist peaks on the projection represent the location of eyebrow, eyelid, and lip. Since the aim here is to localize eye, thus the projection is applied to the upper-half face only. Fig. 3 shows the gradient

where *hface* is the height of face image. The boundary of left and right eyes are defined as

$$Left\_x = 0.2 \times wface \tag{3}$$

$$Left\_y = vu \tag{4}$$

$$Left\_w = 0.2 \times wface \tag{5}$$

$$Left\_h = vl - vu \tag{6}$$

$$Right\_x = 0.6 \times wface \tag{7}$$

$$Right\_y = vu \tag{8}$$

$$Right\_w = 0.2 \times wface \tag{9}$$

$$Right\_h = vl - vu \tag{10}$$

where *wface* is the width of face image, *Left(Right)_x*, *Left(Right)_y*, *Left(Right)_w*, *Left(Right)_h* are the *x*-coordinate of left-top corner, *y*-coordinate of left-top corner, width and height of bounding box of the left(right) eye respectively.

C. White Color Thresholding

Color thresholding is usually used for skin color segmentation [13], [19]. Previous work [19] shows that the normalized RGB color space is effectively used for skin color segmentation. Skin region is defined on the normalized RGB chromaticity diagram as shown in Fig. 4. In the figure, skin region is bounded by five lines i.e.: a) *line g=2*, b) *line g=r-0.4*, c) *line g=0.4*, d) *line g=-r+0.6*, e) *line-c*. A small circle denoted as *line-c* is used to exclude white pixels. Contrary, in the proposed eye detection, the aim is to extract the white color of eye's sclera. Therefore the white color thresholding could be considered as extracting the pixels which the *r* and *g* components are inside on the small circle. The proposed method is described in the following.
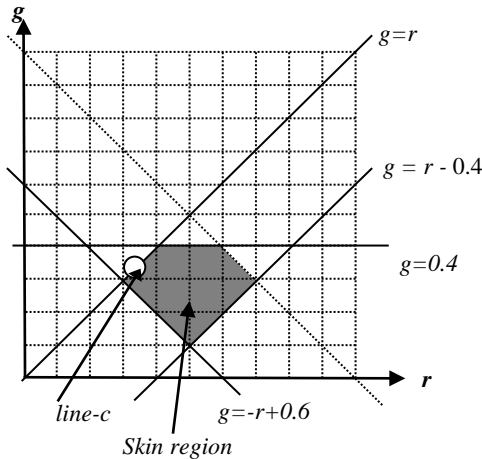
Figure 4.   Skin region on normalized RGB chromaticity diagram [19].

Let the red (*r*), green (*g*) and blue (*b*) components of the normalized RGB color are expressed as:

$$r = \frac{R}{R+G+B} \tag{11}$$

$$g = \frac{G}{R+G+B} \tag{12}$$

$$b = \frac{B}{R+G+B} \tag{13}$$

The binary image of extracted white pixels is denoted as *Ithr* and obtained by the following equations:

$$Ir(x,y) = (r(x,y) - 0.33)^2 \tag{14}$$

$$Ig(x,y) = (g(x,y) - 0.33)^2 \tag{15}$$

$$Ic(x,y) = Ir(x,y) + Ig(x,y) - 0.0009 \tag{16}$$

$$Ithr(x,y) = \max(sign(Ic(x,y)), 0) \tag{17}$$

where

R is red component of *RGB* color

G is green component of *RGB* color

B is blue component of *RGB* color

I(x,y) is the intensity value at pixel (*x,y*)

sign(a) is equal 1 if *a>0*, 0 if *a=0*, -1 if *a<0*

max(a,b) is the maximum value of *a* and *b*

Fig. 5 shows the thresholded image obtained by above equations. Using Eqs. (14)-(17), the binary image (*Ithr*) is obtained as the shown in Fig. 5(b), where the black color represents the extracted objects, in this case white objects, while white color represents non-extracting object. From the figure, it could be seen that the method also extracts non-white objects outside the face. It is confirmed by [19] and could be observed from Fig. 4 that the above formulas will not extract skin color. Thus the skin-face region shown in Fig. 5(b) is in white color (not extracted). Since the white color thresholding is applied on rough eye region where only eye's sclera is in white color, the thresholding method will work appropriately.

(a)

(b)   (c)

Figure 5.   (a) Color image, (b)Thresholded image using Eqs. (14)-(17), (c) Thresholded image using Eqs. (14)-(18)

The normalized RGB color has drawback that *r* and *g* component of the dark and bright color may have the same value such as addressed in [19]. Suppose a dark pixel with the value of *R*=0.1, *G*=0.1, and *B*=0.1. The chromaticity values of this pixel are *r*=0.33 and *g*=0.33. In other hand, a bright pixel with the value of *R*=0.9, *G*=0.9, and *B*=0.9 will have the value of *r*=0.33 and *g* =0.33.

Therefore those two pixels will be considered as the same color in the chromaticity diagram. From Fig. 5(b), it is observed that the black color of the hair is also extracted. To overcome the problem, the following rule is introduced. The pixels extracted by Eqs. (14)-(17) are assigned as the white pixels if the following formula is satified.

$$\frac{R(x,y)+G(x,y)+B(x,y)}{3} > 0.4 \qquad (18)$$

Using Eq. (18) the black hair is not extracted as shown in Fig. 5(c).

D. *Ellipse Detection*

The proposed white color thresholding extracts the white color of eye, thus the extracted blob image does not make solid ellipses, but they contain holes (i.e. pupils) as shown in Fig. 6. To find the precise boundary of the eye, an ellipse detection technique is employed. The ellipse detection works with edge images, thus it needs to detect edge pixels of the blob. A Sobel edge detector is employed to find the edge pixels. After edge pixels are extracted, ellipse fitting method [20] is applied to detect the ellipses.
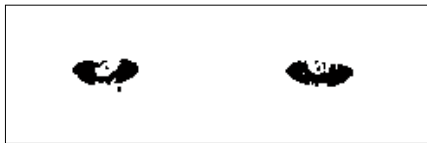
Figure 6.   Blob image of extracted eye's sclera

The ellipse fitting method [20] uses the least square criterion to fit an ellipse from edge pixels. Let an ellipse is expressed as

$$a_1 x^2 + a_2 xy + a_3 y^2 + a_4 x + a_5 y + f = 0 \qquad (19)$$

The five coefficients a_1, a_2, a_3, a_4, a_5 are estimated by the least square estimator from the given edge pixels (points) $(x_1,y_1),...,(x_N,y_N)$ by minimizing the squares sum of an error between the edge pixels and the ellipse using the following formula

$$err = \sum_{i=1}^{N}(a_1 x_i^2 + a_2 x_i + a_3 y_i^2 + a_4 x_i + a_5 y_i + f)^2 \qquad (20)$$

The estimated five coeffects are then converted to the parameters of ellipse, i.e.: orientation, center coordinate, semi-major axis, and semi minor-axis.

Fig. 7 shows the detected ellipses on the left and right eyes superimposed on the blob image. From the figure, it is shown that the method is able to detect the boundary of eye. It is worth noting that blob image contains hole inside it, thus the edge pixels contain the outer and inner part of the blob. However the ellipse fitting method fits the outer ellipse appropriately.

Figure 7.   Detected ellipse on the left and right eyes.

III.   EXPERIMENTAL RESULTS

To verify the proposed eye detection method, the algorithm is implemented using MATLAB and tested on two hundred face images taken from [21]. The dataset contains frontal face images with varying lighting.

In the experiments, three approaches are evaluated: a) Eye detection based-on Haarcascade Classifier [18] (**HC**); b) Proposed rough eye localization (**PL**); c) Proposed eye detection (**PD**). It is noted here that eye detection based-on Haarcascade Classifier detects the bounding box of eye, but the proposed eye detection detect the precise location of eye (eye's sclera) acoording to the shape. Therefore these two methods could not be compared directly. In addition to the proposed eye detection, the proposed rough eye locatization is evaluated here for fair comparison to the eye detection based-on Haarcascade Classifier.

The true detection rate and false detection rate are calculated for comparison. The true detection rate is defined as the total number of detected eye's pair divided by the total number of tested images, while the false detection rate is defined as the total number of detected non-eye objects divided by the total number of tested images. By this definition, if only one eye (not a pair of eye) is detected, then it is considered as not detected. Table I shows the comparison results.

The true detection rate of **HC** is the highest, but the false detection is also high. It is a typical result obtained by the Haarcascade classifier, i.e. a high true detection rate, but a false detection occurs. The true detection rate of **PL** and **PD** are lower compared to the **HC**, but they have no false detection. From the observation, the lower true detection of **PL** is caused by the misdetection of eye due to the position of left and right eyes which are not in a horizontal line. This result relates to the approach employed in the proposed rough eye localization which assumes that the position of two eyes should be in a horizontal line as described in Section B.

TABLE I.   COMPARISON RESULTS

No	Method	True Detection	False Detection
1.	Eye detection based-on Haarcascade classifier (**HC**)	96.5%	43%
2.	Proposed rough eye localization (**PL**)	93.5%	0%
3.	Proposed eye detection (**PD**)	91%	0%

From the table, there is a deviation between **PL** and **PD**. It implies that the some eye images obtained by the rough localization technique could not be detected by white color thresholding and the ellipse detection techniques properly. From the observation, it is caused by two cases : a) The white color thresholding fails to extract eye pixels, thus there are no

ellipse detected; b) The white color thresholding extract the spurious pixels which yield the wrong ellipses. The high true detection rate of **PL** is comparable to the one of **HC**. It shows that the proposed white color thresholding works effectively. The advantages of the proposed method compared to **HC** are: a) There is no false detection; and b) The eye is detected precisely according to the eye's shape.

Fig. 8 illustrates some of the detection results, where the figures in the left, center, and right columns represent the results of **HC**, **PL**, and **PD** respectively. In the first and third rows, all three methods detect or localize the eye properly. But the **PD** localizes eye precisely, in the sense that it detect the ellipse boundary of eye, while the detected boundaries of **HC** and **PL** are in rectangle boxes. From the third row, it is shown that the detected eyes obtained by the **HC** are not in a horizontal line due the skew position of the left and right eyes. However the detected boundaries of left and right eyes are always in the horizontal line. In the second and fourth rows, the non-eye objects are detected by **HC** (false detection), but not detected by **PL** and **PD**. In the fourth row, there is an eye glasses in the image. The **HC** and **PL** locate the eyes properly, but **PD** could not detect eye properly, i.e. the detected ellipses do not fit the eye properly. As shown in the figure, the glasses create the white shadows which result the wrong ellipse detection.
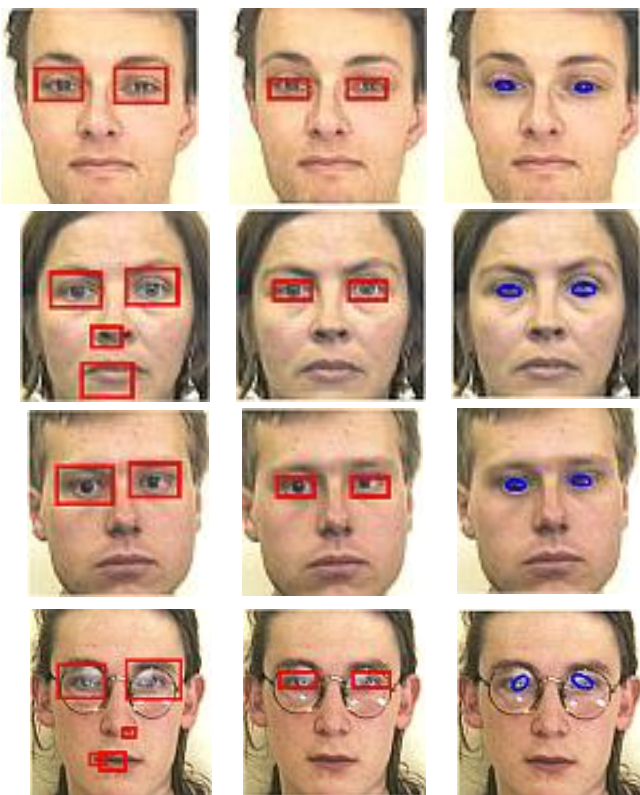
Figure 8.   Eye localization and detection results.

IV.   CONCLUSION

In this paper, an eye detection technique is presented. It first localizes the rough position of eye using projection technique. Then a white color thresholding based-on the normalized RGB

chromaticity diagram is employed. It extracts the white color of eye's sclera. Finally the boundary of eye is found by an ellipse fitting method. The method shows a high performance in rough eye localization of 93.5 % and the true eye detection rate of 91% is achieved.

In future, the performance of proposed method will be improved and will be extended for detecting the driver's fatigue by measuring the degree of openess of the eye. Further the real time implementation using video camera will be conducted.

REFERENCES

[1]   E. Miluzzo, T. Wang, and A.T. Campbell, "EyePhone: activating mobile phones with your eyes," Worshop on Networking, Systems, Applications on Mobile Handhelds (MobiHeld). ACM, 2010.

[2]   K. Arai and R. Mardiyanto, "Eye-based human computer interaction allowing phoning, reading e-Book/e-Comic/e-Learning, Internet browsing, and TV information extraction," International Journal of Advanced Computer Science and Application, Vol. 2, No. 12, pp. 26-32, 2011.

[3]   Q. Wang, "Eye location based on structure feature for driver fatigue monitoring," World Academy of Science, Engineering and Technology, Vol. 72, pp. 368-372, 2010.

[4]   J. Machrouh, F. Panaget, P. Bretier, and C. Garcia, "Face and eyes detection to improve natural human-computer dialogue," Proceedings of the Second IEEE-EURASIP International Symposium on Control, Communication, and Signal Processing, Morocco, 2006.

[5]   D.W. Hansen and Q. Ji, "In the eye of the beholder:a survey of models for eyes and gaze," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 3, pp. 478-500, 2010.

[6]   J. Ren and X. Jiang, "Fast eye localization based on pixel differences," Proceedings of the 16th International Conference on Image Processing, Cairo, pp. 2733-2736, 2009.

[7]   D. Sidibe, P. Montesinos, and S. Janaqi, " A simple and efficient eye detection method in color images," Proceedings of International Conference Image and Vision Computing New Zealand, Great Barrier Island, 2006.

[8]   M.A. Kashani, M.M. arani, M.R. Ramezanpour, "Eye detection and tracking in image with complex background," Journal of Emerging Trends in Computing and Information Sciences, Vol. 2, No. 4, pp. 182-186, 2011.

[9]   Z.H. Zhou and X. Geng, "Projection functions for eye detection," Pattern Recognition, Vol. 37, No. 5, pp. 1049-1056, 2004.

[10]   J.B. Kim, S.C. Kee, and J.Y. Kim, "Fast detection of multi-view face and eye based on cascaded classifier," Proceedings of IAPR Conference on Machine Vision Applications, Tsukuba, 2005.

[11]   H. Lu, W. Zhang, and D. Yang, "Eye detection based on rectangle features and pixel-pattern-based texture features," Proceedings of International Symposium on Intelligent Signal Processing and Communication System, Xiamen, pp. 265-268, 2007.

[12]   M. Motwani, R. Motwani, and F.C. Harris, "Eye detection using wavelets and ANN," Proceedings of Global Signal Processing Conferences and Expos for Industry, California, pp. 27-30, 2004.

[13]   D.W. hansen, J.P. Hansen, M. Nielsen, A.S. Johansen, and M.B. Stegman, "Eye typing using markov and active appearance models," Proceedings of IEEE Workshop Applications on Computer Vision, pp. 132-36, 2003.

[14]   T. Ishikawa, S. Baker, I. Matthews, and T. Kanade, "Passive driver gaze tracking with active appearance models," Proceedings of 11th World Congress Intelligent Transportatiom Systems, Nagoya, 2004.

[15]   Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," Real-Time Imaging, Vol. 8, pp. 357-377, 2002.

[16]   D.W. Hansen, "Using colors for eye tracking," Color Image Processing: Method and Applications, pp. 309-327, CRC Press, 2006.

[17] Aryuanto and F.Y. Limpraptono, "A vision-based system for monitoring driver fatigue," Proceedings of TEKNOIN, Yogyakarta, 2009.

[18] P. Viola, M. Jones, "Robust real-time face detection," International Journal of Computer Vision, Vol. 57, No. 2, 2004.

[19] A. Soetedjo and K. Yamada, "Skin color segmentation using coarse-to-fine region on normalized RGB chromaticity diagram for face detection," IEICE Transactions on Information and System, Vol. E91-D, No. 10, pp. 2493 -2502, 2008.

[20] http://www.mathworks.com/matlabcentral/fileexchange/3215-fitellipse

[21] http://pics.psych.stir.ac.uk/

Bins Formation using CG based Partitioning of Histogram Modified Using Proposed Polynomial Transform 'Y=2X-X^2' for CBIR

H. B. kekre

Professor Department of Computer Engineering
MPSTME,
NMIMS University,
Vileparle (W), Mumbai, India

Kavita Sonawane

Ph. D Research Scholar Department of Computer
Engineering MPSTME,
NMIMS University,
Vileparle (W), Mumbai, India

Abstract— **This paper proposes a novel polynomial transform to modify the original histogram of the image to adjust the pixel density equally towards the high intensity levels so that uniform distribution of the pixels can be obtained and the image can be enhanced. We have shown the efficient use of this modified histogram for Content Based Image Retrieval. According to the CBIR system described in this paper each image is separated into R, G and B plane and for each plane a modified histogram is calculated. This modified histogram is partitioned into two parts by calculating the Center of gravity and using it 8 bins are formed on the basis of R, G and B values. These 8 bins are holding the count of pixels falling into particular range of intensity levels separated into two parts of the histogram. This count of pixels in 8 bins is used as feature vector of dimension 8 for comparison to facilitate the image retrieval process. Further these bins data is used to form the new variations of feature vectors ; Total (sum) and Mean of pixel intensities of all the pixels counted in each of the 8 bins. These feature vector variation has also produced good image retrieval. This paper compares the proposed system designed using the CG based partitioning of the original and histogram modified using the polynomial transform for formation of the 8 bins which are holding the Count of pixels and Total and Mean of intensities of these pixels. This CBIR system is tested using 200 query images from 20 different classes over database of 2000 BMP images. Query and database image feature vectors are compared using three similarity measures namely Euclidean distance, Cosine Correlation distance and Absolute distance. Performance of the system is evaluated using three parameters PRCP (Precision Recall Cross-over Point), LSRR (Length of String to Retrieve all Relevant images) and Longest String.**

Keywords- Polynomial Transform; Euclidean Distance; Cosine Correlation Distance; Absolute Distance; Modified Histogram; CG based partitioning; Bins formation; PRCP; LSRR; Longest String.

I. INTRODUCTION

Content Based Image Retrieval is the promising approach to search the desired images from the large image databases using the image contents like color, shape, texture and their representation as feature vectors in various other formats [1][2][3]. Digital images are the best and convenient media for describing and storing spatial, temporal, spectral and physical components of information from various domains (e.g.

satellite images, biomedical images). Today's advanced technology made it easy to capture and store the large no of images. These images can be made available to the users for many fields like scientific, educational, medical, industrial, and other applications. To manage and make efficient use of these large image databases is an important issue [4][5]. CBIR is one of the important area to be studied today to overcome the drawback of text based image retrieval techniques [6][7]. It is vast area for researcher's to find out new approaches to retrieve the similar images from database with very good accuracy and less computational complexity[8][9][10][11]. We have taken a step towards the same problem with a new approach to extract the image features from the spatial domain. There are various techniques based on histograms are used for CBIR [12-16].

Color histogram depicts color distribution using a set of bins. Using the global color histogram an image will be encoded with its color histogram, and the distance between two images will be determined by the distance between their color histograms. This increases the size of the feature vector and also the time required to calculate the distance between these histograms. Size of the feature vector decides the time required for comparing feature vectors for retrieval. It is an important factor to be considered while designing an efficient CBIR system [17][18][19]. In our system we have solve this issue by exploring the new technique to form the bins so that color details of the image will be separated properly, feature vector size can be reduced and the comparison will take less time. We have first modified the image by modifying the image histogram using the new polynomial transform introduced in this paper. Then to get the uniform pixel distribution we have used CG i.e Center of Gravity to divide the histogram in two equal parts. This process is applied separately to each R, G and B plane of the image for extracting its features. This partitioning will lead to the formation of 8 bins. After preparing the feature vector databases for 2000 BMP image database we have tested the system performance using 200 query images .i.e 10 images from each of the 20 classes of database [20]. The comparison between query and database is images performed using Euclidean, Cosine Correlation and Absolute distance measures. Performance of the system is evaluated using three parameters PRCP, LSRR

and Longest String [21][22]. Organization of the paper is as follows: Section II describes the polynomial transform introduced to modify the histogram. Section III elaborates on the partitioning process using CG and formation of 8 bins. Results are discussed in detail in Section IV followed by conclusions in Section V.

II. POLYNOMIAL TRANSFORM TO MODIFY THE HISTOGRAM

The polynomial transform equation (1) designed to modify the histogram such that if the image has large pixels having low intensity levels, they will be shifted towards high level intensities. This adds good enhancement in the image. The polynomial equation is given as follows.

$y = (2x - x^2)$	(1)
$y = 0$; IF $x = 0$ $y = 1$; IF $x = 1$ $y > x$ for $0 < x < 1$	

As shown in the following Figure.1. The *X* are transformed to *Y* such that Y are always greater than *X* for all values in range 0 to 1; i.e. *for $0 < x < 1$*
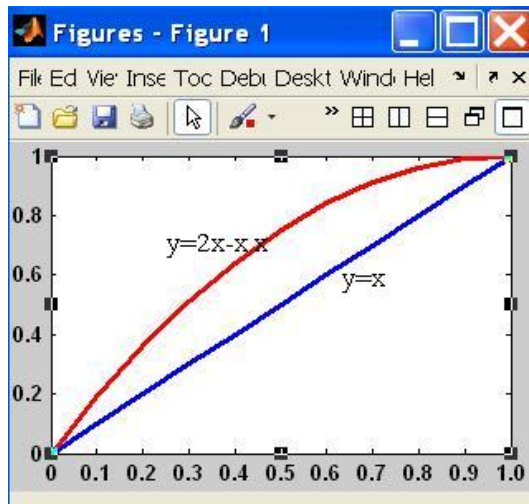
Figure 1.   Polynomial Transform y = 2x-x^2

We have used this transform to modify the original image histogram of the R, G and B planes of the images. In above fig we can see that y=x in the blue curve, shifting of pixels from lower side to higher side can be observed in the red curve which is obtained for the given polynomial function.

III. CG BASED PARTITIONING OF MODIFIED HISTOGRAM

A. Modified Histograms: R, G and B Planes

The image in feature extraction process is separated into R, G and B planes. Histogram of each plane is obtained and modified using the given polynomial transform. The effect of polynomial transform is reflected in the Figure.2. It shows effect of modified histogram

Original Image
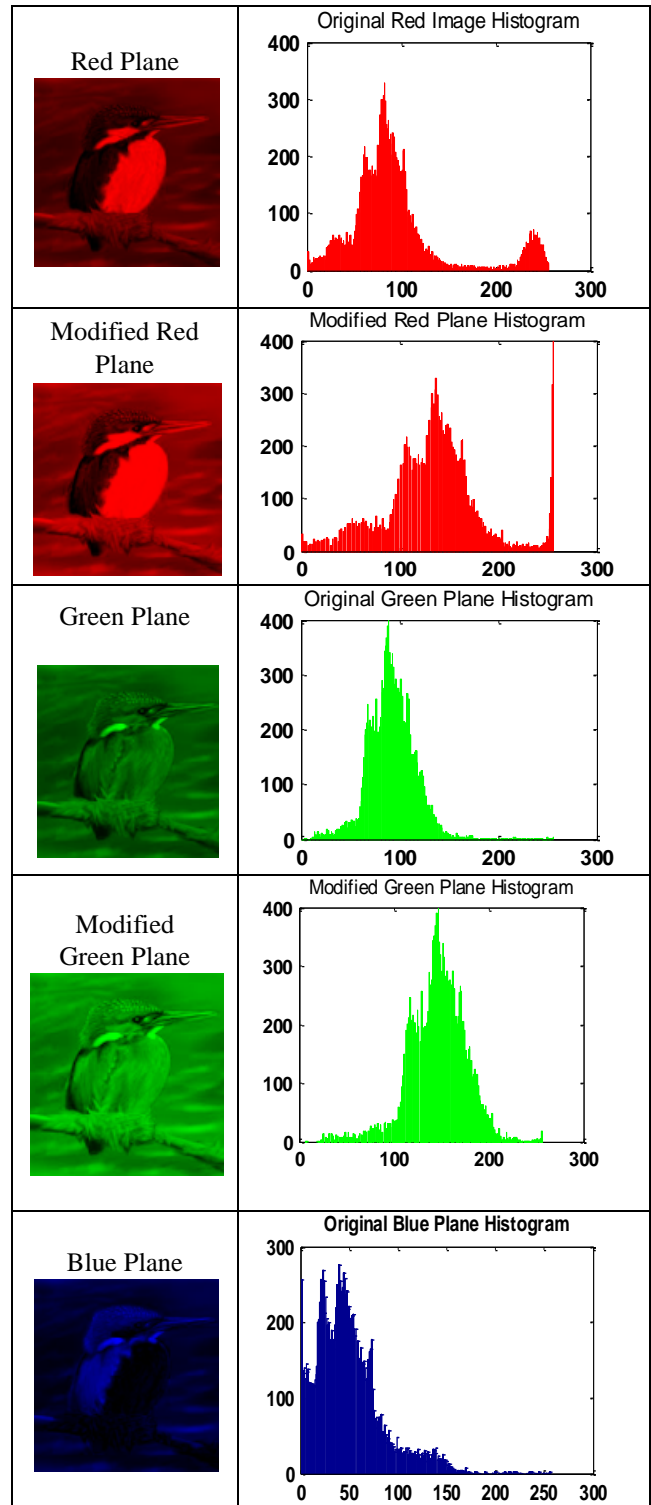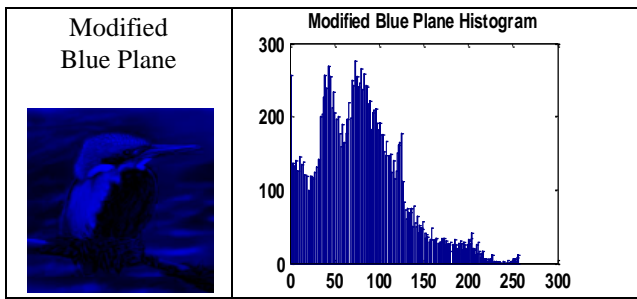
Figure 2.   Kingfisher Image : R, G, B Original and Modifed Histograms

Here the image is sperated into R, G and B Planes and orginal histogram of each plane is obtained. These planes are modified by modifying their histograms after applying the polynomial transform. We can observe in original histogram, that the pixels from low level intensities are shifted to high intensities in the modified histogram. The information can be seen clearly in the modified image planes.

B. Partitioning

Modified Histogram of three planes are partitioned into two parts by calculating the Center of Gravity (CG) using equation 2. Center of gravity finds out the point in the scale of intensity levels such that the moments of image intensities above and below CG are equal. It is the balancing point which exactly divides the histogram into two parts.

$$CG = \left[\frac{(L_i W_i + L_{i+1} W_{i+1} + \cdots + L_n W_n)}{\sum_{i=1}^{n} W_i} \right]$$	(2)
Where L_i is intensity Level and W_i is no of pixels at L_i	
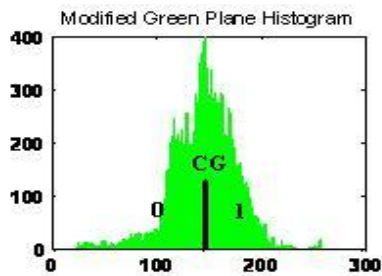
Figure 3.   CG Partitioning

Once the modified histogram is divided into two parts we could form the bins as explained in the following section.

C. Bins Formation: Feature vector Generation

The partitions of histogram are named with id's '0' for intensities below CG and '1' for above CG. In the process of formation of bins we have followed the following sequence of steps:

1. Separate the image into planes : R, G and B
2. Calculate the histogram of each plane , and modify it using the polynomial transform given in equation 1.

3. Partitioned the modified histogram using the CG given in equation 2.
4. Assign id 0 and 1 to two parts of each plane as explained above.
5. Now consider the pixel from the image for which the feature vector is being extracted, check its R, G and B values whether they fall in part 0 or 1 of their respective histogram. For e.g. if the R value falls in part '0', G in part '1' and B falls in part '0' then flag of that pixel is set with these three values '0 1 0' which determines the address of that pixel to be pushed into that bin, i.e. here the pixel will be counted in 'Bin 2'. As we have three planes and each is divided in two parts we could form 8 bins out of it , i.e Bin 000 to 111 (Total 8 Bins).
6. Same process is applied to each pixel and its bin address where that pixel has to reside will be calculated. Total 16384 pixels of an image under feature extraction process (image size is 128 x 128) are segregated by taking their count into 8 bins from bin addresses 000 to 111.
7. Feature vector: 'Count of pixels', 'Total Intensities', and 'Average intensities'. Here these set of 8 bins holding the count of pixels for one image is considered as feature vector of dimension 8 representing that image. Further these bins are directed to calculate the sum or Total and Average of Red, Green and Blue intensities of the pixel counts in each of the 8 bins. The 8 bins of red, green and blue color are maintained separately for each image and these are considered as our feature vectors of the image.

Applying the same process explained above to all database images; we have extracted three features of each image as Count of pixels, Total of intensities and Average of intensities and we have prepared the feature vector database separately for each type of feature vector. We have total 7 feature vector databases prepared each having feature vector of dimension 8.

Count of Pixels : One feature vector database for 2000 database images.

Total of Intensities: R, G and B – Three feature vector databases

Average of Intensities: R, G and B – Three feature vector databases.

IV.   RESULTS AND DISCUSSION

Once the feature vector databases are prepared a query can be fired to the system. Working scenario of the system ready to face the query and generate the retrieval result is explained through the following block diagram shown in Figure.4

Once the query image is fired it will proceed through all different stages shown in the figure, its feature will be extracted and will be compared with the database image feature vectors by means of similarity measure.
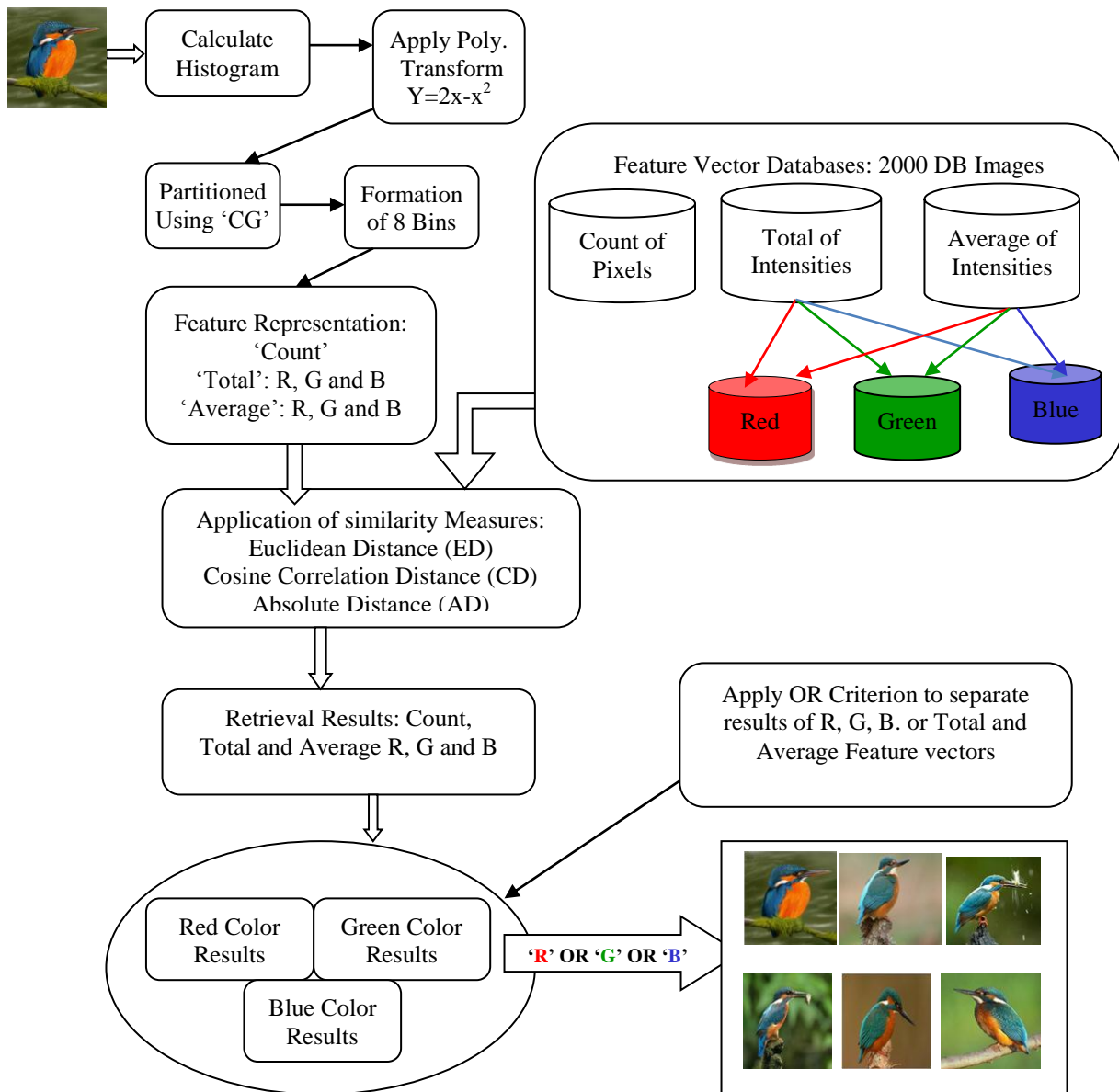
Figure 4.   Block Diagram of the proposed system ready to accept Query and Produce the Retrieval Result

A. Database and Query Image

We have experimented this system for 2000 BMP images from 20 different classes each having 100 images: : Flower, Sunset, Mountain, Building, Bus, Dinosaur, Elephant, Barbie, Mickey, Horse, Kingfisher, Dove, Crow, Rainbow rose, Pyramid, Plate, Car, Trees, Ship and waterfall. We have randomly selected 10 sample query images from each class. Means the system is tested and verified using 200 query images.

B. Application of Similarity Measure

Once the query image enters in the system, feature vector for the same will be extracted and the distance between the query and database images will be calculated to find out the match between similar images. To accomplish this task the

system is designed to use three different similarity measures namely Euclidean, Absolute and Cosine correlation distance given in equation 3, 4 and 5 [23][24]. Three result sets are obtained for each query for each type of feature vector separately.

C. Retrieval Results and Performance Evaluation[25][26]

On application of three similarity measures three distance sets are obtained for the given query, these distances are the sorted in ascending order from min to max distance. Now, we have considered first100 images with respect to first 100 minimum sorted distances. Out of these 100 we have taken images relevant to query as a retrieval result; as we have 100 images of each class in database, this results are termed as PRCP value i.e. Precision recall Cross over Point. The values obtained for PRCP are shown for feature vector type 'Count of

pixels' in Table 1. We can observe that modified histogram based bins are producing very good results as compared to original histograms for all three similarity measures.

Euclidean Distance : $$D_{QI} = \sqrt{\sum_{i=1}^{n}\left|(FQ_i - FI_i)\right|^2}$$	(3)
Absolute Distance: $$D_{QI} = \sum_{1}^{n}\left|(FQ_I - FI_I)\right|$$	(4)
Cosine Correlation Distance $$\frac{(D(n))\bullet(Q(n))}{\sqrt{\left[\left|D(n)\right|^2\left|Q(n)\right|^2\right]}}$$ Where D(n) and Q(n) are Database and Query feature Vectors resp.	(5)

TABLE I. RESULT FOR COUNT OF PIXELS INTO 8 BINS

Count	CD ORG	CD MOD	ED ORG	ED MOD	AD ORG	AD MOD
Flower	204	**238**	217	**258**	232	**265**
Sunset	355	**406**	352	**404**	383	**438**
Mountain	114	**146**	105	**125**	117	**143**
Building	140	**150**	139	**149**	149	**161**
Bus	360	**361**	362	353	380	**406**
Diansour	478	444	462	406	493	462
Elephant	118	**171**	115	**158**	119	**153**
Barbie	301	**335**	265	**347**	287	**355**
Mickey	253	216	257	242	274	264
Horses	290	278	333	**338**	370	363
Kingfisher	103	**105**	111	**116**	109	**122**
Dove	542	539	589	545	617	576
Crow	244	**253**	251	234	260	**261**
Rainbowrose	176	**179**	206	**214**	212	**217**
Pyramids	231	222	232	213	256	241
Plates	219	**255**	229	**254**	228	**267**
Car	109	**116**	126	113	141	132
Trees	291	**302**	284	**312**	299	**332**
Ship	177	**199**	170	**190**	196	**217**
Waterfall	172	161	186	168	197	181
TOTAL	4877	**5076**	4991	**5139**	5319	**5556**

Remark: Modified Histogram gives better performance in all cases with similarity measures CD (14), ED (12) and AD (13) Out of 20.

Chart 1 and 2 are showing the results for the other types of feature vectors that are Total and Average of R, G and B intensities in 8 bins separately. These charts are showing the result as total PRCP obtained for 200 query images. Here we found for all three colors and all distance measures modified histogram is performing better as compared to original histogram. If we take average PRCP result of 200 queries we found it in between 0.3 to 0.4. To improve these results further, we have applied 'OR' Criterion to the results obtained for R, G and B colors separately [22][23].

Application of OR criterion combines the results of R, G and B and produces single result which has achieved good height of retrieval for parameter PRCP. It has reached to 0.5 as average for 200 query images.

Results after OR criterion are shown in Charts 3 for Total and Average of intensities Observing the chart 3 obtained for criterion OR over R, G B results of Total and Average of intensities, we can say that the PRCP is reached to good height that is from 0.35 to 0.5 for average results of 200 queries. One more observation is that among three distances CD, ED and AD CD and AD are performing better as compared to ED at many places.

Between two feature vector types i.e Total and Average of intensities feature vector of 8 bins containing 'Average of intensities performing far better as compared to Total of intensities.

Chart I   Comparing the results of original histogram based bins with modified histogram bins, we found that before applying OR criterion, results obtained for both types of feature vectors (Total and Average) with respect to all three distance measures (CD, ED and AD) and three colors R, G and B bins formed using modified histogram are always betters in all the cases.

Chart 1.Total PRCP for 200 queries fired on database of Total of Intensities feature vector for R, G and B colors with CD, ED and AD

	RED CD	GREEN CD	BLUE CD	RED ED	GREEN ED	BLUE ED	RED AD	GREEN AD	BLUE AD
ORG	4716	4379	3933	5080	4739	4546	5448	5188	4951
MOD	4891	4507	4261	5216	4820	4431	5694	5263	4854
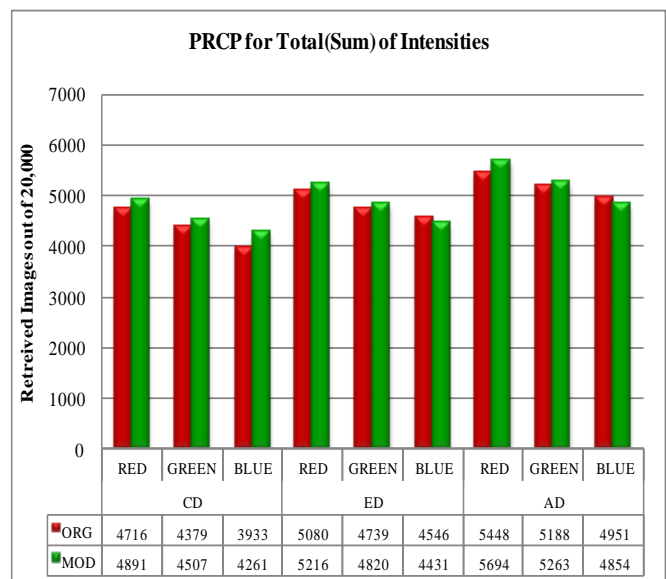
Remark: Modified Histogram gives better performance in 7 out of 9.

Chart 2. Total PRCP for 200 queries fired on database of Average of Intensities feature vector for R, G and B colors with CD, ED and AD.

PRCP for Average of Intensities

	RED	GREEN	BLUE	RED	GREEN	BLUE	RED	GREEN	BLUE
		CD			ED			AD	
ORG	5722	5802	5236	5573	5448	5309	5749	5480	5421
MOD	5990	6435	6253	5693	5712	5753	5773	5724	5794
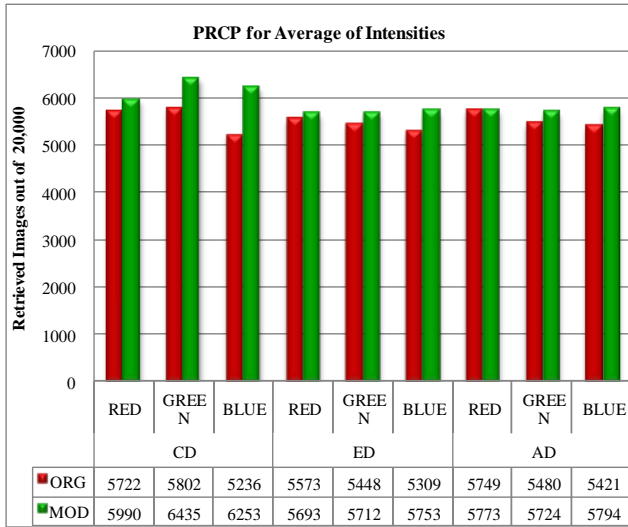
Remark: Modified Histogram gives better performance in all cases.

Chart 3. Total PRCP after application of 'OR' Criterion over PRCP results obtained for R, G, B Colors with CD, ED and AD

Application of Criterion 'OR'

	TOTAL	AVG	TOTAL	AVG	TOTAL	AVG
	CD		ED		AD	
ORG	6565	8978	7316	8911	7582	9141
MOD	6752	9604	7043	8888	7480	9098
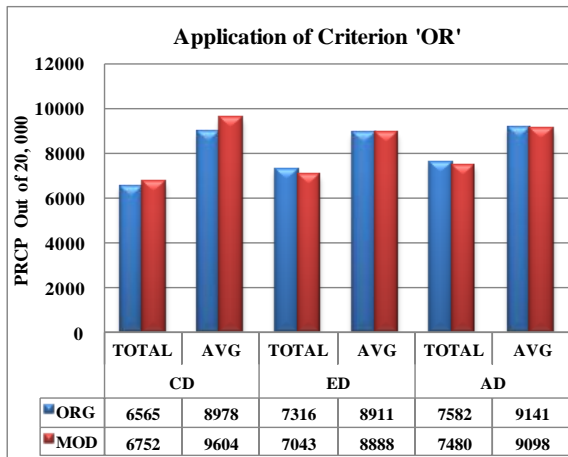
Remark: Modified Histogram gives better performance in 2 cases.
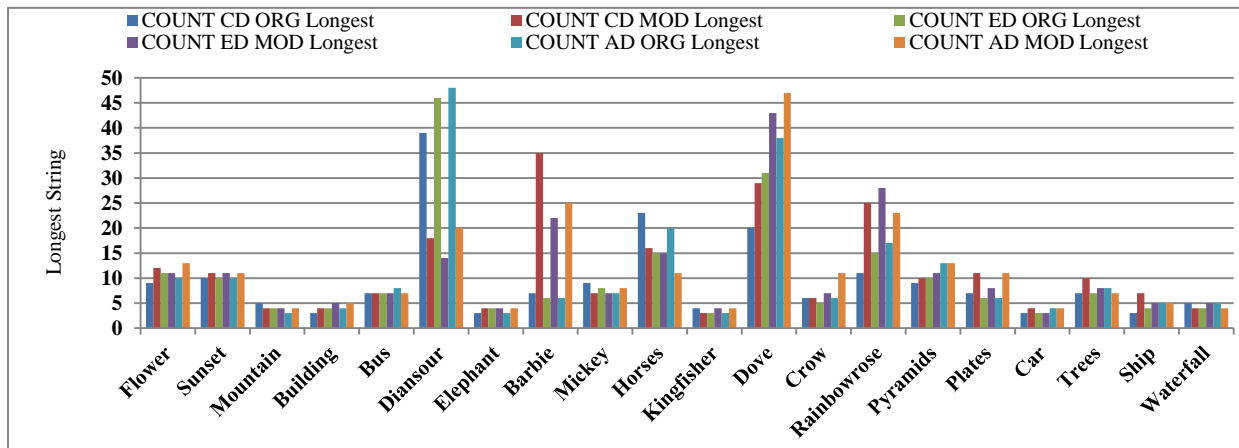
D. Performance Evaluation: Longest String and LSRR[21]

Longest string parameter identifies the longest continuous string of images relevant to query from all database images sorted to retrieve according to their distances with query which are sorted in ascending order (minimum to maximum distance). Longest string results are obtained for each of the 200 queries separately for all types of feature vectors, for all three distance measures and for three colors R, G and B.

Here we have taken the 'maximum longest string' out of this whole result set obtained for each type of feature vector and distance measure irrespective of the three colors R, G and B and are shown in the following charts 4, 5 and 6 for 'Count' of pixels, 'Total' of intensities and 'Average' of intensities respectively.

In chart 4, 5 and 6 we observed that CD and AD are performing better as compared ED measure. Between CD and AD, AD is producing good results at maximum places. When we look at the count of longest string we found 48 as max longest string for Dinosaur and Dove class for Count Pixel as a feature vector. Next we found 35 as best value for dinosaur class for total of intensities and for the feature vector Average of intensities we found best result for longest string as 89 for Barbie class. Overall observation says that feature vector formed using 8 bins holding the average of pixel intensities is giving best performance for all classes and for all distance measures as compared to other two feature vectors.

The next parameter we used for evaluating the performance of the system is LSRR i.e 'Length of the String to Retrieve all Relevant' images. Retrieval of all images of query class from the database is indication of the parameter recall reached to 1 which is the ideal value of recall for CBIR. According to this we are evaluating the performance of our algorithms on the basis of the length of the sorted distances to be traversed by the system to retrieve all relevant images from database. We expect the LSRR should be as low as possible.

The LSRR parameter is applied to check performance of the system for all feature vector databases using the same set of 200 queries applied to system.

Chart 4. Longest String for Count of Pixels With CD, ED, AD for MOD and ORG Histograms

Remark: Modified Histogram gives better performance in 16 out of 20.

Chart 5. Longest String Total Of Intensities With CD, ED, AD for MOD and ORG Histograms
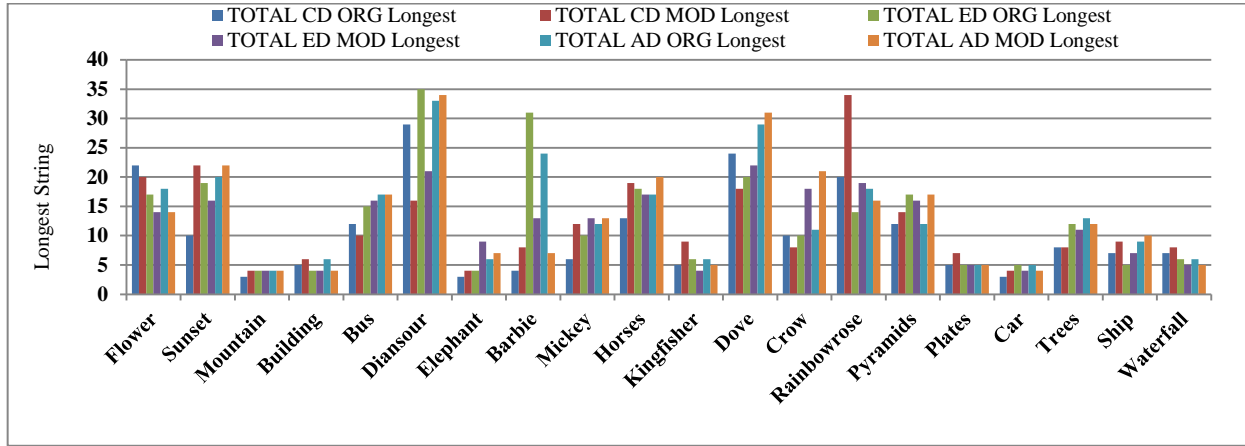
Remark: Modified Histogram gives better performance in 12 out of 20.

Chart 6. Longest String Average of intensities With CD, ED, AD for MOD and ORG Histograms
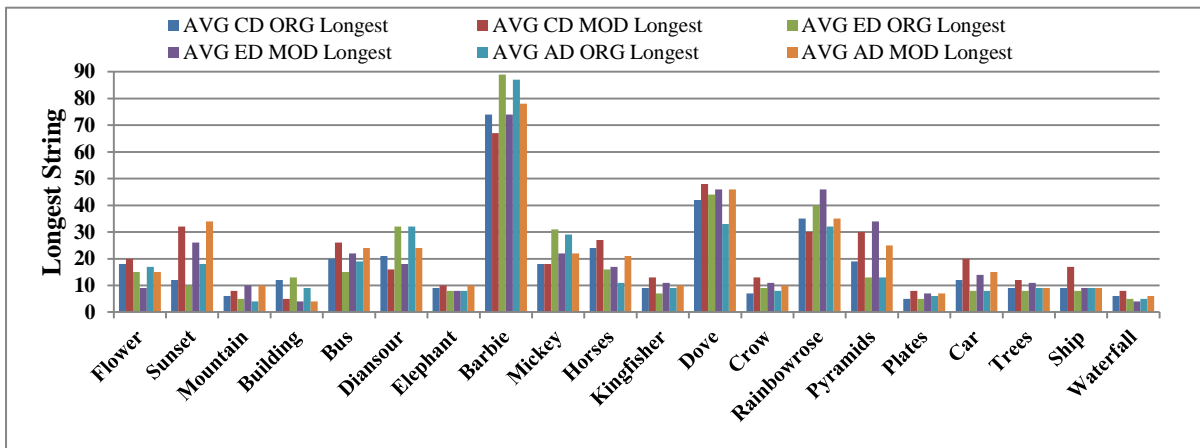
Remark: Modified Histogram gives better performance in 15 out of 20.

Chart 7. % LSRR for Count of Pixels With CD, ED, AD for MOD and ORG Histograms
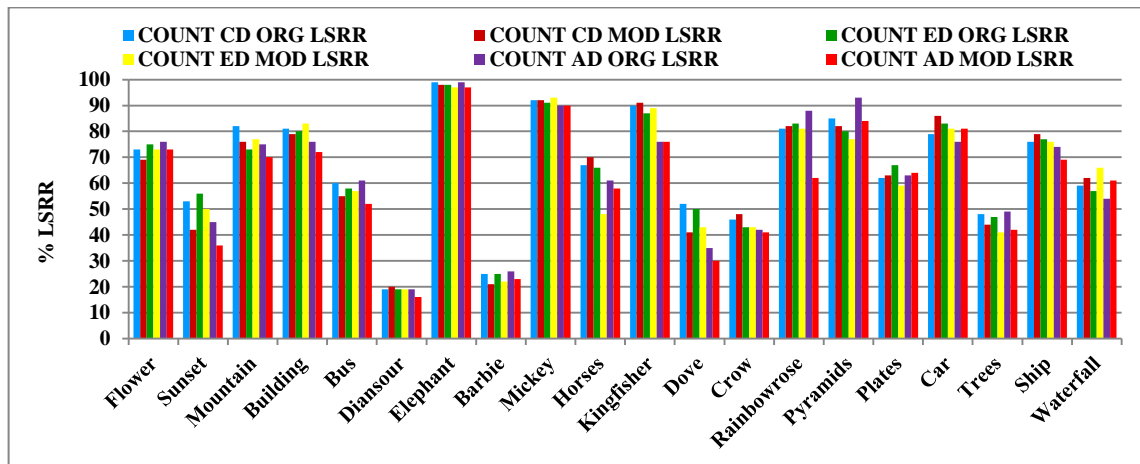
Remark: Modified Histogram gives better performance in 13 out of 20.

Chart 8. % LSRR for Total of intensities With CD, ED, AD for MOD and ORG Histograms
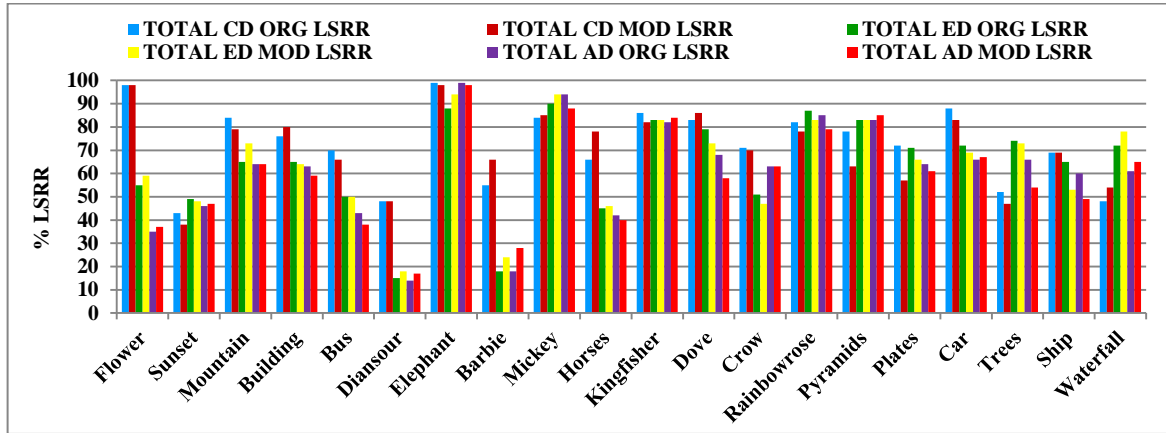
Remark: Modified Histogram gives better performance in 10 out of 20.

Chart 9. % LSRR for Average of intensities With CD, ED, AD for MOD and ORG Histograms
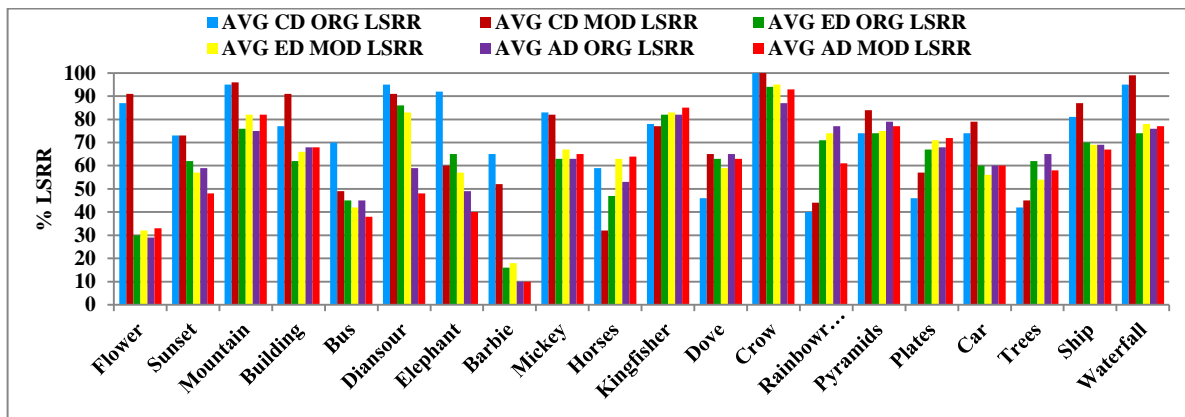
Remark: Modified Histogram gives better performance in 9 out of 20.

Results obtained are analyzed for Count of pixels, Total of intensities and Average of intensities as shown in charts 7, 8 and 9 respectively.

Observing these graphs 7, 8 and 9 we can say that all the results shown in graph are traveling less than 100% length of string of images to retrieve all relevant. For count of pixels feature vector out of 20 query classes, for 15 classes we obtained the LSSR between 10% to 70%, for dove Crow and Tree it is in range 10% to 40% LSRR to make recall 1. Among these results best LSRR obtained is for classes Barbie and Dinosaur in range 10% to 20% only. For feature vector type 'Total of intensities', 15 classes we obtained the LSRR in range 10% to 70%, classes Flower, Sunset, Dinosaur, Barbie and Horses obtained LSRR in range 10 to 40%. Best results for this feature vector are again for class Dinosaur and Barbie in range 10% to 15%.

For next type feature i.e Average of intensities we got LSRR in range 10% to 70% for 15 classes. Flower, Bus and Barbie class results are in range 10% to 40%. The best results here we could obtained is for class Barbie, only 10% traversal of sorted distances gives us the 100% recall of images similar

to query which is very good and desirable performance of any CBIR system.

V. CONCLUSION

The proposed CBIR system gives new feature extraction method into 8 Bins formed using CG based partitioning of the histogram which is modified using the new polynomial transform introduced as 'Y= $2X-X^2$'. The system's performance tested using 200 query images from 20 classes fired over all different feature vector databases formed using 'Original' and 'Modified' histograms and comparisons are made successfully using three similarity measures CD, ED and AD where each of them has yield different sets of results. After the detail analysis of the results presented and discussed in section IV i.e Results and Discussion. Few conclusions can be made which are given as follows. The new polynomial transform shifts the pixels form lower intensities to higher side and gives the new image of modified histogram which generates enhanced image. After this the image details can be seen clearly and used effectively for the feature extraction. Modifed histogram based feature vectors have produced better results as compared to the original histogram based results.

The CG can be used effectively to partition the histograms in two equal parts which produces the 8 bins to form the feature vectors.

Three types of features are extracted into 8 bins by representing the R, G, B intensities in three different forms Count of pixels, Total of intensities(R, G and B Separately), and Average of intensities(R, G and B Separately), for both Original and Modified histogram. Among this Feature vector type Count of pixels and Average of intensities produced good retrieval as compared to Total of intensities.

Among the three similarity measures we found CD and AD are producing very good results as compared to ED. We observed in the analysis of results that the images of query class are shown at larger distance using ED measure; which are shown at smaller distance using CD measure.

The results are evaluated using parameters PRCP, Longest String and LSRR. PRCP gives delineates the performance point where the system generates the results such that the precision and recall both are at same level. According to conventional parameters precision and recall if both are closer to 1 indicating that system is the ideal CBIR system. The Average of 200 queries, after applying OR criterion we could achieve good height for PRCP to **0.5** for CD and for Modified histogram which is high as compared to original histogram result.

LSRR best value among all results is for 10% for Barbie class for modified and original histogram for feature vector Average of intensities. For 'Longest String' we got best result as 89 for class Barbie. Observing the PRCP, LSRR and Longest String results we can say that for each feature vector with different similarity measure we got different set of results which are covering different image classes each time as the best result. It indicates that, the variation in representing the 'image contents' and variation in the comparison process by changing the distance measures the proposed system is generating positive variations in the results by giving good results for different categories for each change either of feature vector, similarity measure or contents of Modified histogram.

Overall performance of the system is compared with respect to all evaluation parameters for original and modified histogram; we can conclude that histogram modified using polynomial transform function gives better results in all cases as compared to original histogram.

REFERENCES

[1] Kato, T., "Database architecture for content based image retrieval in Image Storage and Retrieval Systems" (Jambardino A and Niblack W eds),Proc SPIE 2185, pp 112-123, 1992.

[2] John R. Smith and Shih-Fu Chang, "VisualSeek: A Fully Automated Content-Based Image Query System", Proceedings of the ACM International Conference on Multimedia, 1996.

[3] J. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries", IEEE transaction on PAMI, Volume 23, No 9,Sept 2001, pp. 947-963.

[4] Rui, Y., Huang, T., Chang, S.F.: Image Retrieval:Current Techniques, Promising Directions and Open Issues. J. Visual Communication and Image Representation, Vol. 10. (1999),39–62.

[5] Sameer Antania, Rangachar Kasturia, Ramesh Jainb, "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video, Pattern Recognition 35 (2002) 945–965.

[6] Datta R, et al (2008), Image Retrieval: Ideas, Influences, and Trends of the New Age, Vol 40, No 2, ACM Computing Surveys, The Pennsylvania Stqte University.

[7] W. Niblack, R. Barber, "The QBIC Project: Querying images by content using color, texture, shape," in Proc. SPIE Storage & Retrieval for images & video database, Feb 1994.

[8] V.N.Gudivada and J.V. Raghvan, "Special issues on content based image retrieval system," IEEE Com. Magzine, 1985.

[9] J. P. Eakins and M. E. Graham, "Content-based image retrieval: A report to the JISC Technology Applications Programme, Institute for Image Data Research, University of Northumbria at Newcastle, available online at http://www.unn.ac.uk/iidr/research/cbir/ report.html," January 1999.

[10] Q. Iqbal and J. K. Aggarwal, "Retrieval by classification of images containing large manmade objects using perceptual grouping," Pattern Recognition, vol. 35, pp. 1463–1479, July 2002.

[11] A. Smeulders, M.Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 1349 – 1380, December 2000.

[12] Wei-Min Zheng, Zhe-Ming Lu, and Hans Burkhardt, "Color Image Retrieval Schemes Using Index Histograms Based On Various Spatial-Domain Vector Quantizers", International Journal of Innovative Computing, Information and Control ICIC International Conf. 2006 ISSN 1349-4198, Volume 2, Number 6, December 2006.

[13] V. Vijaya Kumar, N. Gnaneswara Rao, A.L.Narsimha Rao, and V.Venkata Krishna.(2009) IHBM: Integrated Histogram Bin Matching For Similarity Measures of Color Image Retrieval, Pattern recognition vol 2 issue 3 pages 109 -120.

[14] Ole Andreas Flaaten Jonsgård.(2005): Improvements on colour histogram-based CBIR. Master's Thesis ,Gjøvik University College.

[15] H. B. Kekre , Kavita Sonawane, "Feature Extraction in Bins Using Global and Local thresholding of Images for CBIR" International Journal Of Computer Applications In Applications In Engineering, Technology And Sciences, ISSN: 0974-3596, October '09 –March '10, Volume 2 : Issue 2.

[16] H. B. Kekre, Kavita Sonawane. (2012) : Bins Approach to Image retrieval using Statistical Parameters Based on Histogram Partitioning of R, G, B Planes, IJAET, Jan 2012, ISSN 2231-196.

[17] Gi-Hyoung Yoo, Beob Kyun Kim, and Kang Soo You "Content-Based Image Retrieval Using Shifted Histogram", Y. Shi et al. (Eds.): ICCS 2007, Part III, LNCS 4489, pp. 894–897, 2007.© Springer-Verlag Berlin Heidelberg 2007.

[18] P.S.Suhasini , Dr. K.Sri Rama Krishna, Dr. I. V. Murali Krishna, "Cbir Using Color Histogram Processing", Journal of Theoretical and Applied Information Technology, 2005 - 2009 JATIT. All rights reserved. www.jatit.org.

[19] Esin Guldogan, "Improving Content Based Image Indexing and Retreival Performance" Book, Tampere University Of Technology, January 2009.

[20] D. N. F. Awang Iskandar James, A. Thom S. M. M. Tahaghoghi "Content-based Image Retrieval Using Image Regions as Query Examples", Copyright c 2008, Australian Computer Society, Inc. This paper appeared at the Nineteenth Australasian Database Conference (ADC2008), Wollongong, Australia, January 2008

[21] H. B. Kekre , Kavita Sonawane, "Feature Extraction in the form of Statistical Moments Extracted to Bins formed using Partitioned Equalized Histogram for CBIR." International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-3, February 2012.

[22] H. B. Kekre , Kavita Sonawane, "Query based Image Retrieval using Kekre's, DCT and Hybrid wavelet Transform over 1st and 2nd moment", International Journal of Computer Applications (0975 – 8887)Volume 32– No.4, October 2011.

[23] H. B. Kekre , Kavita Sonawane, "Retrieval of Images Using DCT and DCT Wavelet Over Image Blocks", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 10, 2011.

[24] H. B. Kekre , Kavita Sonawane, "Standard Deviation of Mean and Variance of Rows and Columns of Images for CBIR". International Journal of Computer, Information, and Systems Science, and Engineering 3:1 2009.

[25] Henning Muller Wolfgang Muller_ David McG. Squire "Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals" Technical Report, Universite De Geneve, Centre Universitaire D'informatique, Groupe Vision, December 1, 1999.

[26] Julia Vogela, Bernt Schieleb "Performance evaluation and optimization for content-based image retrieval", 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958,M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S. Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, Mumbai.

He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. He has guided 17 Ph.D.s, 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networks. He has more than 450 papers in National / International Conferences / Journals to his credit. Recently twelve students working under his guidance have received best paper awards. Five of his students have been awarded Ph. D. of NMIMS University. Currently he is guiding eight Ph.D. students. He is member of ISTE and IETE.

Ms. Kavita V. Sonawane has received M.E (Computer Engineering) degree from Mumbai University in 2008, currently Pursuing Ph.D. from Mukesh Patel School of Technology, Management and Engg, SVKM's NMIMS University, Vile-Parle (w), Mumbai, INDIA. She has more than 8 years of experience in teaching. Currently working as a Assistant professor in Department of Computer Engineering at St. Francis Institute of Technology Mumbai. Her area of interest is Image Processing, Data structures and Computer Architecture. She has 7 papers in National/ International conferences / Journals to her credit.She is member of ISTE.

Defect Diagnosis in Rotors Systems by Vibrations Data Collectors Using Trending Software

Hisham A. H. Al-Khazali[1]

Institute School of Mechanical & Automotive Engineering,
Kingston University,
London, SW15 3DW, UK.

Mohamad R. Askari[2]

Institute School of Aerospace & Aircraft Engineering,
Kingston University,
London, SW15 3DW, UK.

Abstract-**Vibration measurements have been used to reliably diagnose performance problems in machinery and related mechanical products. A vibration data collector can be used effectively to measure and analyze the machinery vibration content in gearboxes, engines, turbines, fans, compressors, pumps and bearings. Ideally, a machine will have little or no vibration, indicating that the rotating components are appropriately balanced, aligned, and well maintained. Quick analysis and assessment of the vibration content can lead to fault diagnosis and prognosis of a machine's ability to continue running. The aim of this research used vibration measurements to pinpoint mechanical defects such as (unbalance, misalignment, resonance, and part loosening), consequently diagnosis all necessary process for engineers and technicians who desire to understand the vibration that exists in structures and machines.**

Keywords- vibration data collectors; analysis software; rotating components.

I. INTRODUCTION

A. Vibration Analysis Software

In the past, diagnosis of equipment problems using vibration analysis was mostly dependent on the ability of the maintenance technician or the plant engineer [1, 2]. However, today's vibration analysis equipment utilizes software that has greatly enhanced the analysis of vibration measurements and the prediction of equipment performance [3, 4]. Before computer-based data collectors, most vibration programs consisted of recording overall velocity, displacement, and acceleration measurements on a clipboard [5, 6]. Those with IRD instruments had a mysterious measurement available called Spike Energy [7, 8]. The measurements were transferred to charts for trending. Obviously, this was very labour intensive, but it was successful [9, 10]. Those fortunate enough to have spectrum analyzers with plotters, would take spectra and paste them to a notebook with the overall trends [11,12& 13]. This whole process of manual storage and trending of overall and spectral data was not very cost effective, but many of those programs were successful [3].

The advent of computer-based data collectors and trending software made this whole trending process much more cost-effective. The first system only recorded overall measurements [14,15& 16]. Spectral data still had to be taken with a spectrum analyzer. Those programs based only on overall measurements were usually successful in identifying a machine developing a problem [3, 17].

Vibration analysis software is an essential tool for professionals who are analyzing the vibratory behaviour of structures and machines. There are two types of vibration analysis investigations that are commonly done today [4, 18]. The first involves analyzing the mechanical vibration of new products that are being designed or tested. The second involves analyzing the vibration that exists in rotating machinery such as compressors, turbines, and motors [19, 20]. Advances in computing power and software have greatly enhanced the vibration analysis process for both new and existing products. The most common software package for analyzing new products is finite element analysis software. Vibration in existing machinery is analyzed with integrated software that enhances the vibration analysis equipment [14, 21].

B. Vibration Testing

Vibration testing is used to determine how well a product will withstand its expected service and transportation environments. Equipment that must withstand vibration testing includes automotive, aerospace, machinery, electrical, medical and power [22].

Performing a vibration test reproduces one of the most severe real-world environmental conditions that equipment will encounter. Since vibration testing is crucial during product development, selecting and using vibration testing equipment is an important step for engineers and product managers [23, 24].

II. METHODS AND EXPERIMENTAL PROCEDURES

A. Vibration Testing Equipment

Vibration testing equipment includes accelerometers, controllers, analyzers, amplifiers, shakers, and vibration test fixtures. While each industry utilizes vibration testing in a unique manner, the most important components of a vibration test system are basically the same [25,26& 27]. Vibration is measured and controlled using displacement transducers or accelerometers. Then, do the experimental testing using the electromagnetic shaker test, installed two accelerometers (model 339A32, SN 4851, sensitivity 96.5 & 99.6 mV/g) in Y& Z direction, it was attached to the test structure with creating a computer when taking readings in file that was dimensions and introducing it with the data within the program (smart office). Vibration monitoring equipment includes PC-based controllers and analyzers. Vibration is transmitted to the test product using stiff, lightweight test fixtures. Vibration shakers are available in electrodynamics or hydraulic versions. Electrodynamics' shakers are used for smaller products that require smaller displacements and a larger frequency range. It was powered

by amplifiers that may rival a radio station for electrical   power output [26, 28].
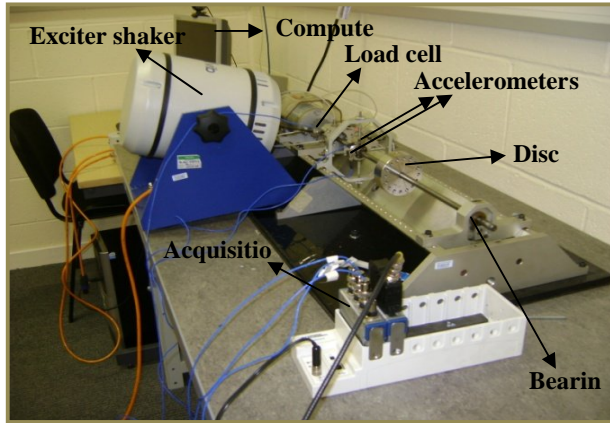
Figure 1. Experimental setup for the modal testing. (Different types of accelerometers, 2012).
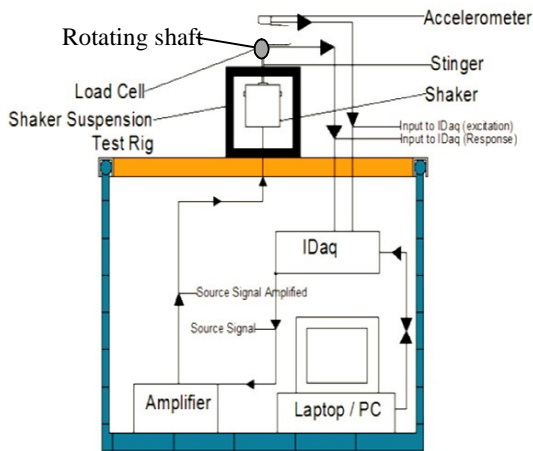
A . Conceptual design show method of the outcome results using electromagnetic Shaker test.
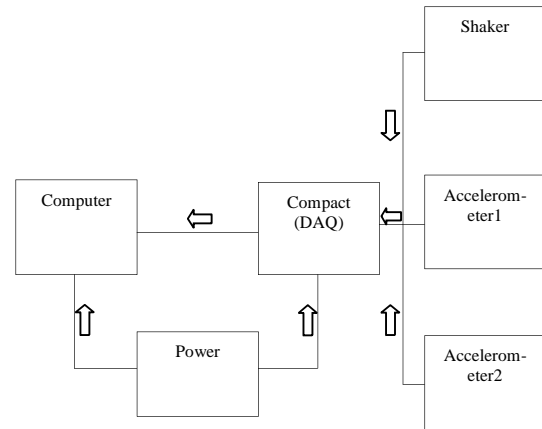
B. Block diagram of test processing controller.

Scheme 1. Test setup;

B. Mathematical Models

(i) Equation of motion in rotor system for passive structures [23, 29], (no self-excitation).

$$[M]\left\{\ddot{q}(t)\right\} + [C]\left\{\dot{q}(t)\right\} + [K]\left\{q(t)\right\} = \left\{F(t)\right\}$$

(1)

All the above matrices are normally symmetric, after solved it we get:

$$\alpha_{jk}(\omega) = \sum_{r=1}^{n} \frac{\varphi_{jr}\,\varphi_{kr}}{M_r\left[\omega_r^2 - \omega^2 + 2\,i\,\omega\,\xi_r\,\omega_r\right]}$$

(ii) For (active) rotating structures [17,23& 30], includes the gyroscopic effect [G], however in machinery where there are rotating parts as well as the supporting structure, the matrices are generally non-symmetric, and therefore need to be treated differently.

$$[M]\left\{\ddot{q}(t)\right\} + [(C+G)(\Omega)]\left\{\dot{q}(t)\right\} + [K(\Omega)]\left\{q(t)\right\} = \left\{F(t)\right\}$$

The non-homogeneous part of the above equation 3, may be solved for free vibration analysis, we obtain:

$$\alpha_{jk}(\omega) = \sum_{r=1}^{n}\left(\frac{2\omega_r\left(\zeta_r\,\mathrm{Re}(_rG_{jk}) - \sqrt{1-\zeta_r^2}\left(\mathrm{Im}(_rG_{jk})\right)\right) + i\left(2\omega\,\mathrm{Re}(_rG_{jk})\right)}{\left(\omega_r^2 - \omega^2 + 2i\omega\omega_r\zeta_r\right)}\right)$$

(4)

(iii) Equation 4, represents the receptance between two coordinates j and k for a system with n degree of freedom. The denominator is identical to the denominator of the receptance expression for an n degree of freedom system with symmetric system

matrices (equation 2), but the numerator is very different.

III. RESULTS

A. *The Orbit Analysis Postprocessing Wizard*

This orbit analysis tool has been designed for bearing and shaft analysis. It post-processes time history file that must include a once-per-rev key-phasor signal to provide a speed and reference for top dead centre (TDC). Various filters are available to smooth the data over a number of revolutions [4, 10].
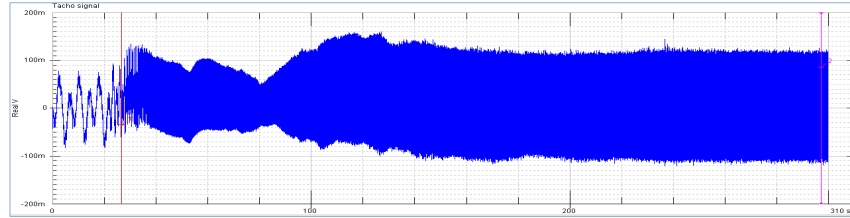
A. Real versus time (sec.)
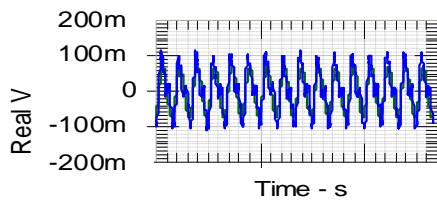
B- y, axis measurement.
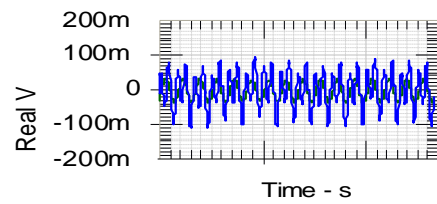
C- z, axis measurement.

Figure 2. Signal detection in (y & z) axis measurement direction;

A-

B-

C-

D-

E-
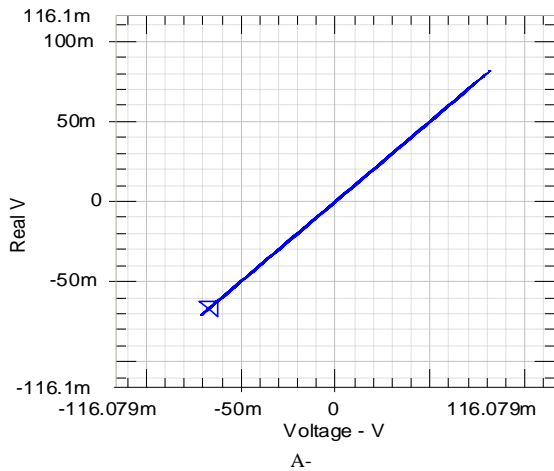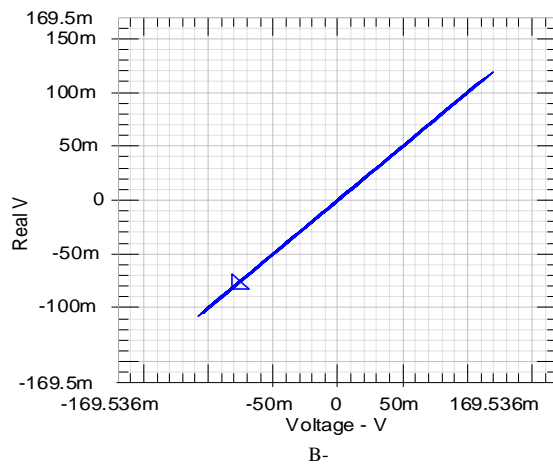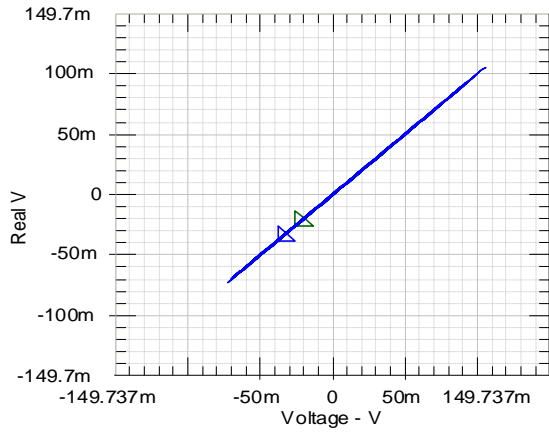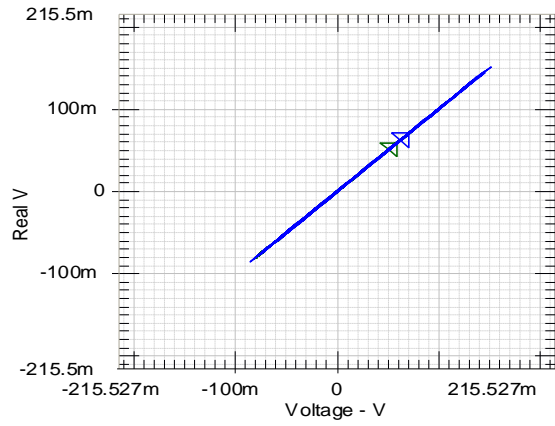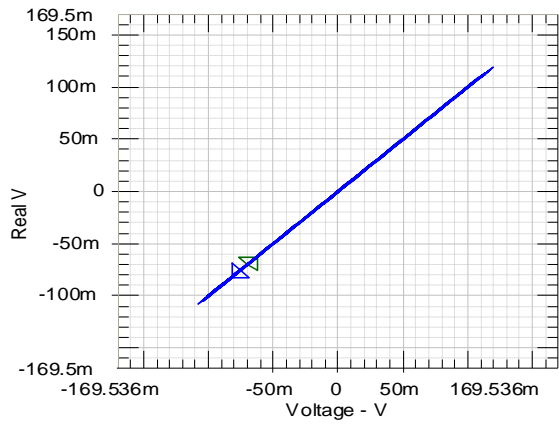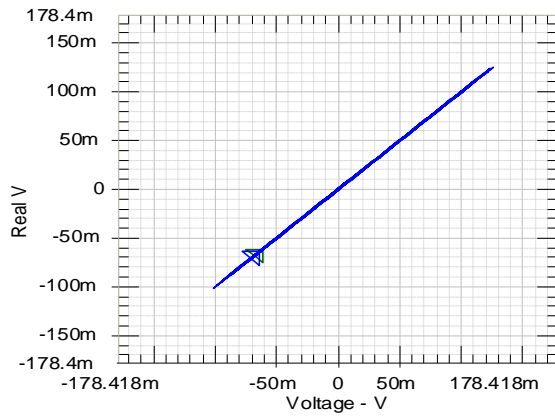
Experimental orbit analysis direction in rotor dynamics;

Figure 4.

F-
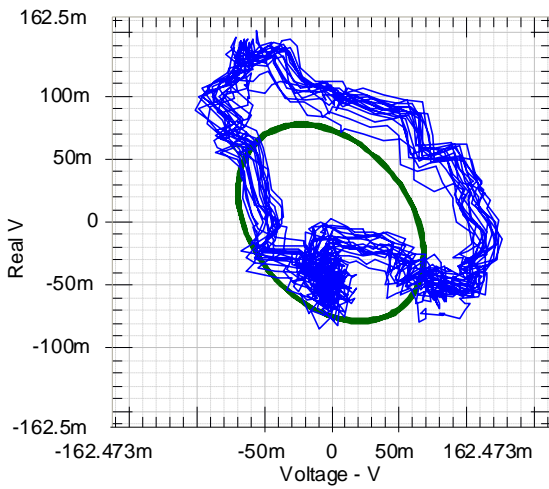
A-
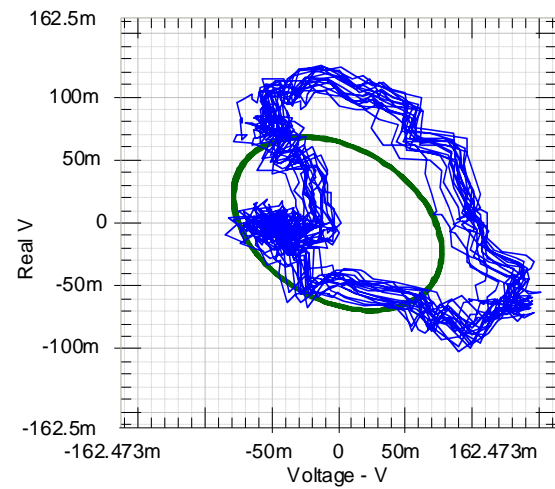
B-
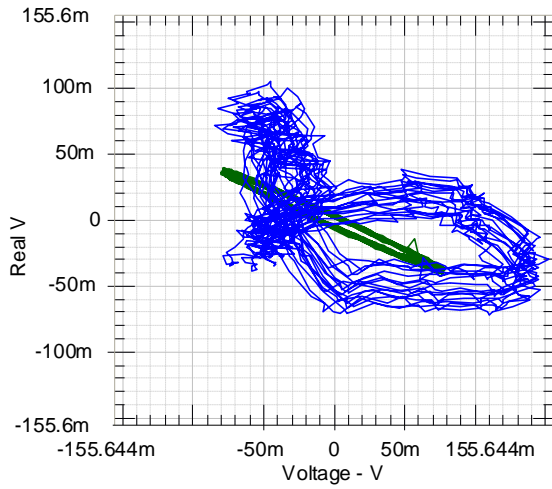
C-
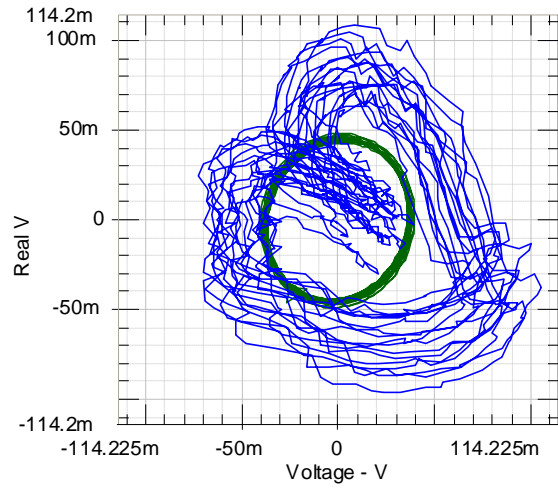
D-
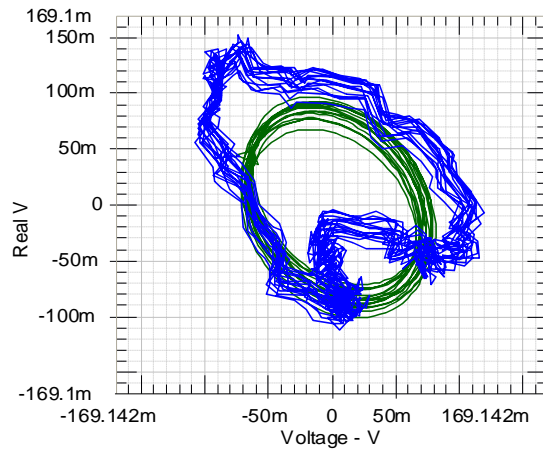
E- Without filter (HP filter).
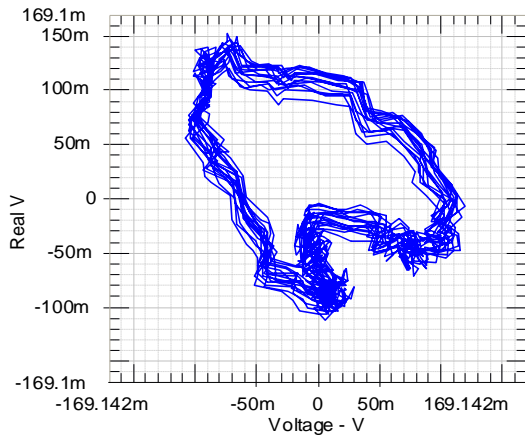
F- Without filter.
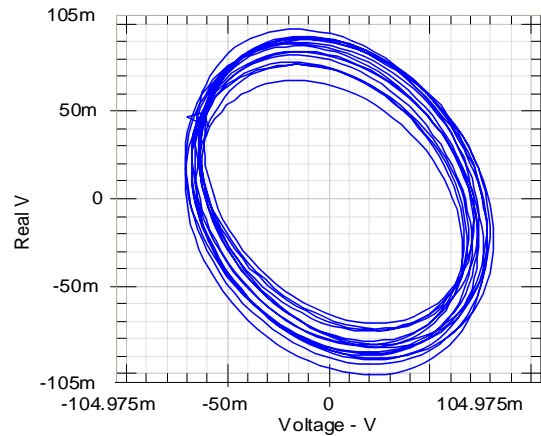
G- Without show row data filter.
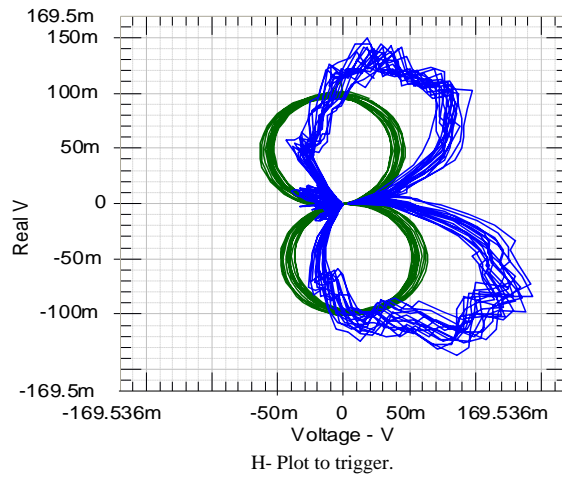
H- Plot to trigger.

Figure 4. Experimental orbit analysis fundamentals behaviour for key-phasor;
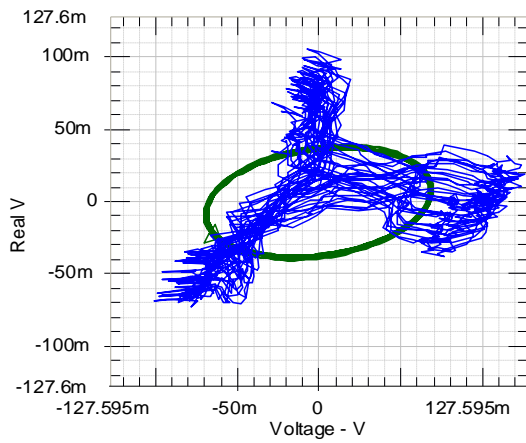
A-
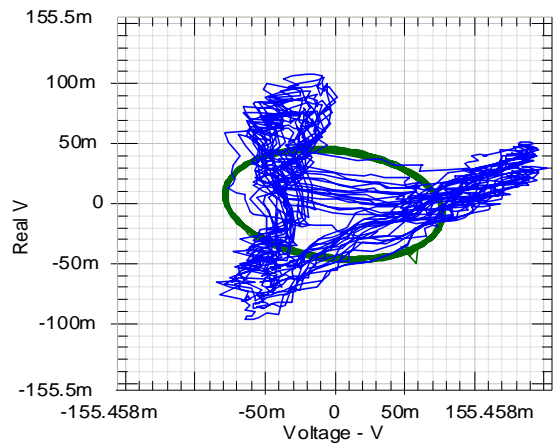
B-
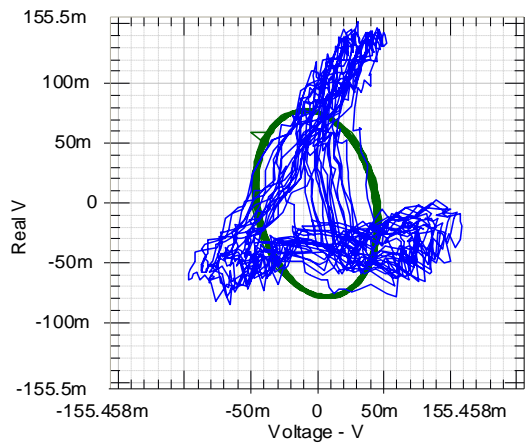
C-
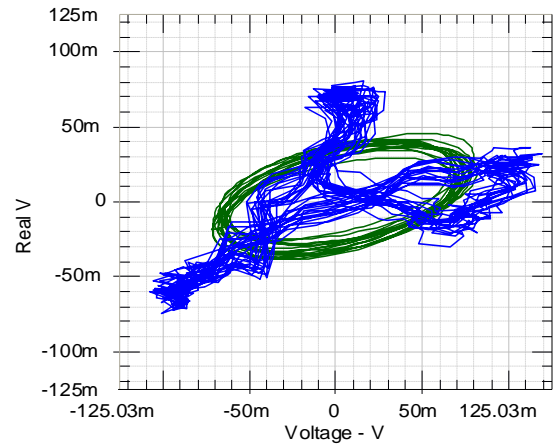
D- Without filter (HP filter).
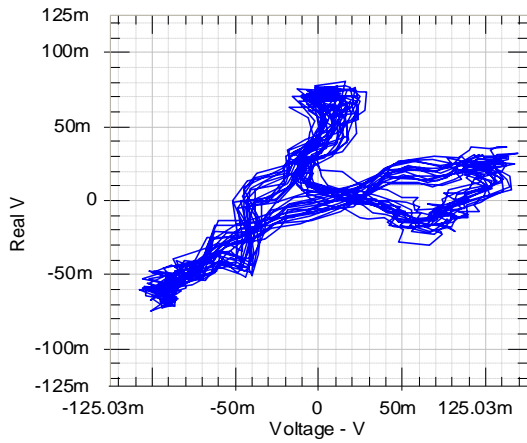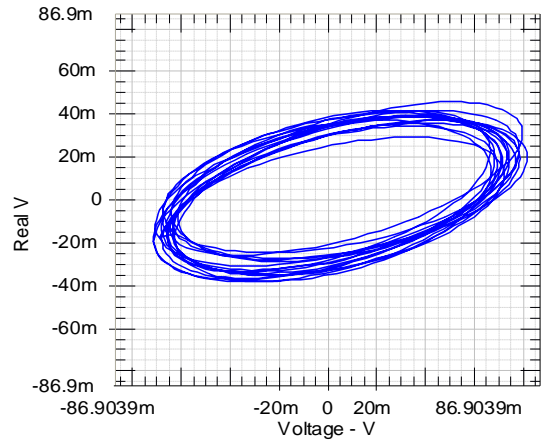
E- Without filter.

F- Without show row data filter.
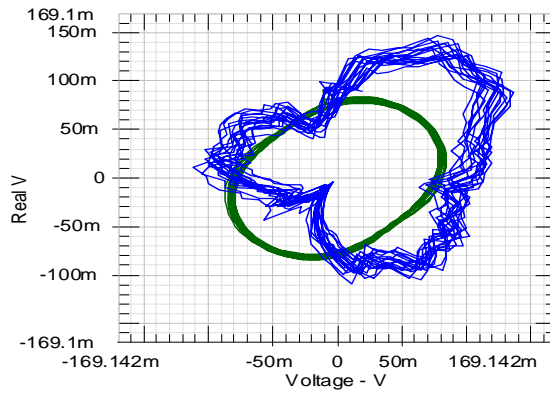
G- Plot to trigger.

Figure 5.   Experimental orbits analysis fundamentals behaviour of the rotor at the different measured planes in first critical speed for channel one;

A-

B-

C-

D- Without filter (HP filter).

E- Without filter.

F - Without show row data filter.
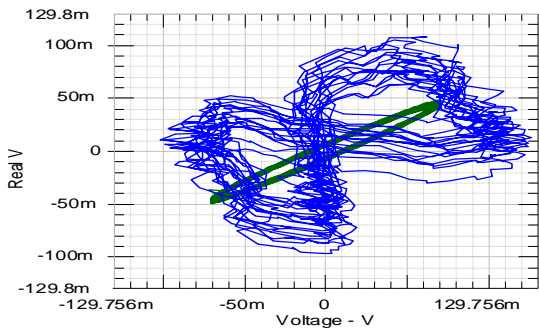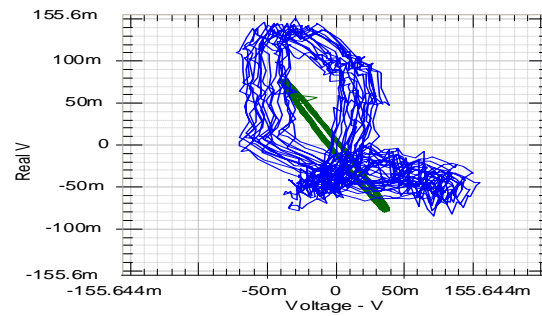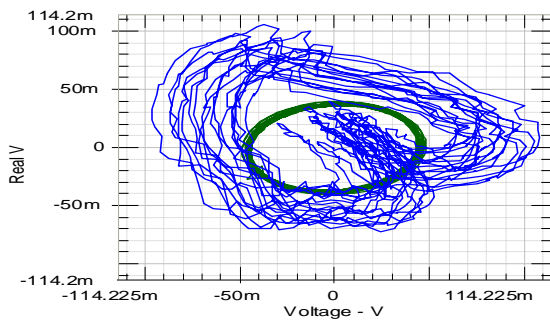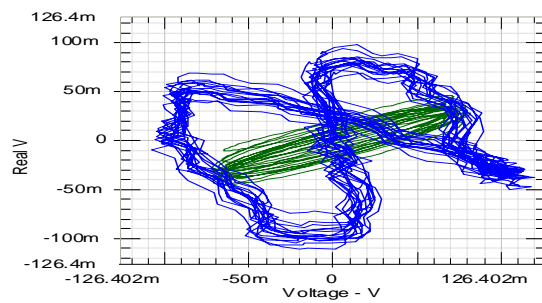
G- Plot to trigger.
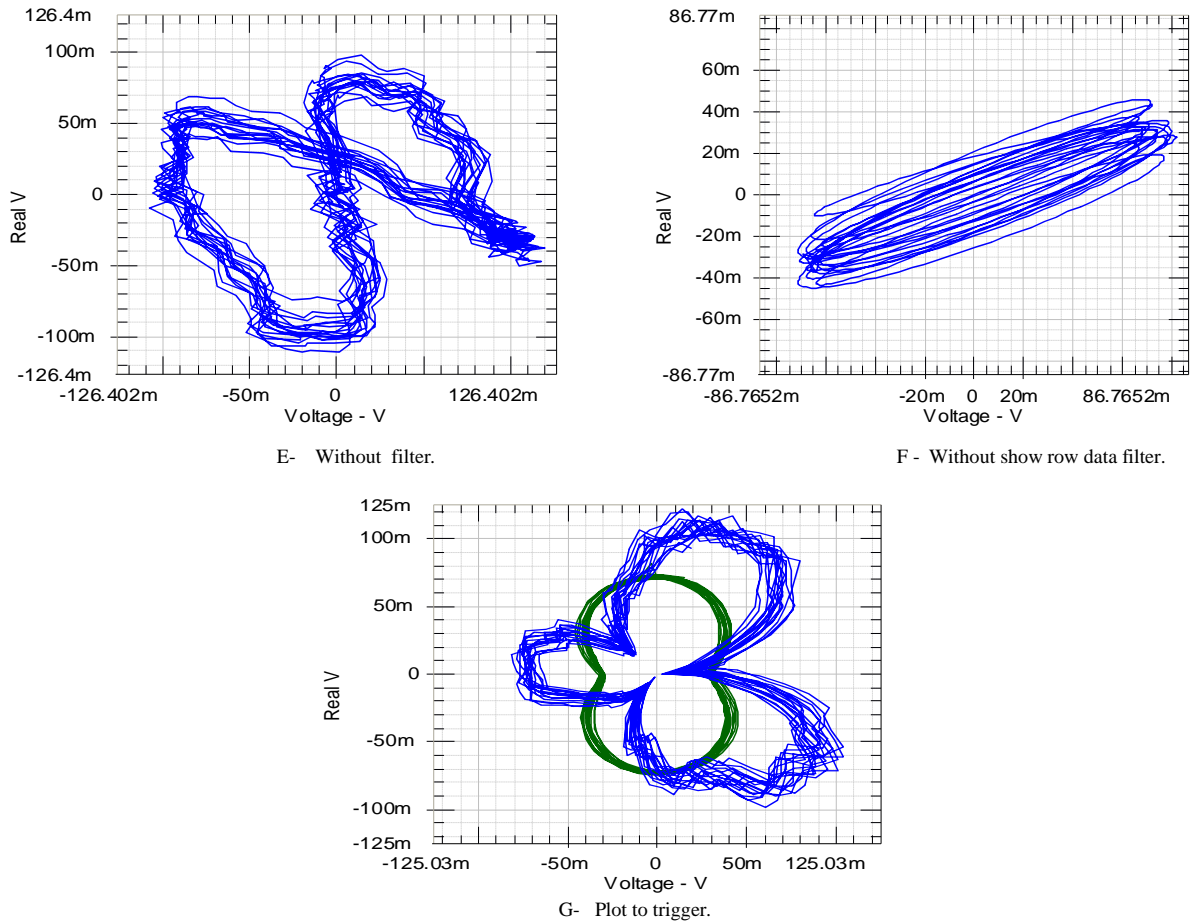
Figure 6.   Experimental orbits analysis fundamentals behaviour of the rotor at the different measured planes in first critical speed for channel two;

B. The Shock Capture Module

The shock capture module provides shock capture, data validation and reporting. It allows you to use (m + p) library of standard test limit overlays, capture data from any number of channels, including triax's, filter the data, automatically adjust overlays for best fit [31, 32].

You will see in the display that you get an immediate readout of key pulse parameters and the whole test is controlled from one simple window for fast and efficient use even by user less familiar with the SO Analyzer and details of the application, we can see that clear in Fig.,(6).

A-Tacho signal.
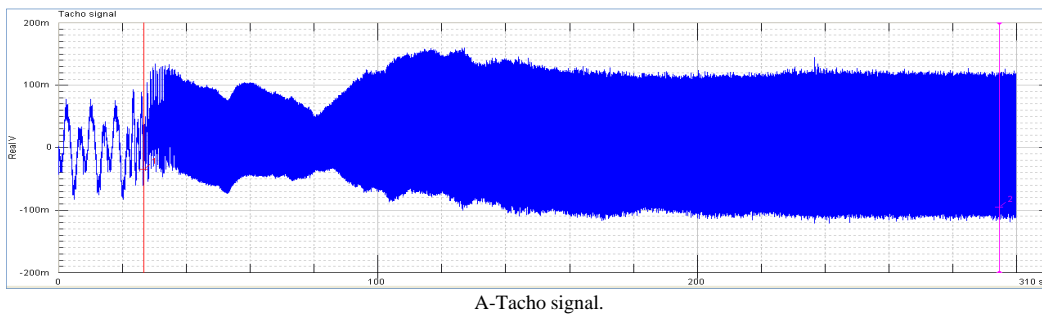
B- Before use filter.

C- After used window filter with increasing the speed.
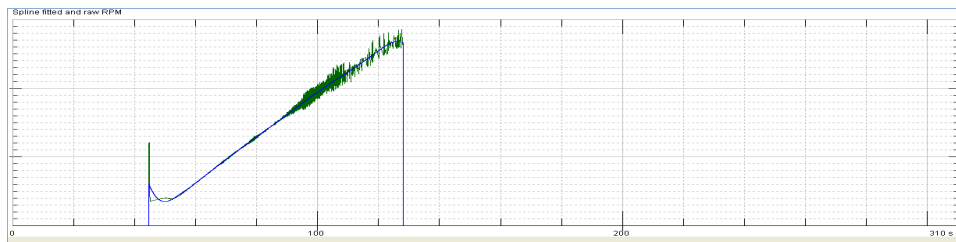
D- Improved or decay shock capture after used window filter.

Figure 7.   Shock capture module;

Hybrid Denoising Method for Removal of Mixed Noise in Medical Images

J UMAMAHESWARI

Research Scholar, Department of Computer Science
Dr.G.R.D. College of Arts and Science
Coimbatore, Tamil Nadu, India.

Dr.G.RADHAMANI

Director, Department of Computer Science
Dr.G.R.D. College of Arts and Science
Coimbatore, Tamil Nadu, India.

Abstract— **Nowadays, Digital image acquisition and processing techniques plays a very important role in current day medical diagnosis. During the acquisition process, there could be distortions in the images, which will negatively affect the diagnosis images. In this paper a new technique based on the hybridization of wavelet filter and center weighted median filters is proposed for denoising multiple noise (Gaussian and Impulse) images. The model is experimented on standard Digital Imaging and Communications in Medicine (DICOM) images and the performances are evaluated in terms of peak signal to noise ratio (PSNR), Mean Absolute Error (MAE), Universal Image Quality Index (UQI) and Evaluation Time (ET). Results prove that utilization of center weighted median filters in combination with wavelet thresholding filters on DICOM images deteriorates the performance. The proposed filter gives suitable results on the basis of PSNR, MSE, UQI and ET. In addition, the proposed filter gives nearly uniform and consistent results on all the test images.**

Keywords- Gaussian noise; impulse noise; UQI; Wavelet filter; CWM; hybrid approach.

I. INTRODUCTION

Many scientific datasets are contaminated with noise, either because of the data acquisition process, or because of naturally occurring phenomena. Pre-processing is the first step in analyzing such datasets. There are several different approaches to denoise images. The main problem faced during diagnosis is the noise introduced due to the consequence of the coherent nature of the image capture. In image processing applications, linear filters tend to blur the edges and do not remove Gaussian and mixed Gaussian impulse noise effectively [7], [8]. Inherently noise removal from image introduces blurring in many cases. These noises corrupt the image and often lead to incorrect diagnosis. Gaussian noise is an additive noise, which degrades image quality that originates from many microscopic diffused reflections leads to discriminate fine detail of the images in diagnostic examinations [1], [2], [3]. Thus, denoising these noises from a noisy image has become the most important step in medical image processing.

The most common type of noise is generated by the detector [6],[8]. Thermal fluctuations is one type of disturbance occurred due to many interconnected electronics components which has a Gaussian Probability Density

Function (PDF), from the contributions of many independent signals, PDF results in a signal with a Gaussian PDF [11],[12]. Many methods have been available for noise reduction [13], [14], [9]. The existing filters used for mixed noise reduction techniques includes median filter, center weighted median filter and wavelet filters. Nowadays, the uses of wavelet based denoising techniques have gained more attention by researchers [10].

In this work a fusion technique is proposed to find the best possible solution, so that after denoising PSNR, MSE, UQI and ET values of the image are optimal. The proposed method is based on wavelet transform and center weighted median filtering, which exploits the potential features of the combination of both wavelet and center weighted median.

This paper is organized as follows. Section 2 discusses the wavelet based thresholding method for denoising Gaussian noise. Section 3 describes the center weighted median filter for denoising impulse noise. The proposed methodology is explained in Section 4. Experimental results are given in Section 5. Finally Conclusion and reference are discussed in Section 6.

II. WAVELET BASED THRESHOLDING

The following figure 1 shows the wavelet denoising steps:

- Apply wavelet transform to the noisy image to produce the noisy wavelet coefficients.

- Select best appropriate threshold limit at each level by using threshold method (hard or soft thresholding) to remove the noises. Here soft thresholding is used for removal of noise.

- Inverse wavelet transform [5] is applied to thresholded wavelet coefficients to obtain a denoised image.
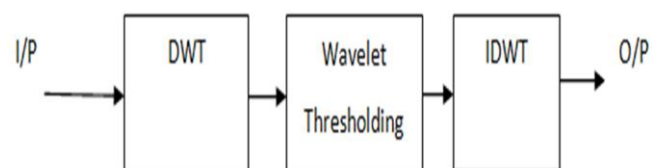
Figure 1.   Denoising using wavelets transform filtering

A. Wavelet Representation of Image

During transmission the image f is corrupted by white Gaussian noise with independent and identically distributed by mean, and standard deviation. The noisy image received is $g_{ij} = f_{ij} + sn_{ij}$. When estimate the image signal f from noisy observations g_{ij}, PSNR and UQI is maximum as well as MSE and ET should be minimum. These methods use a threshold and determine the wavelet coefficients. There are two types of thresholding for removal of noise, namely the hard thresholding method and the soft thresholding method [2],[10].

1) Soft Thresholding Method

Generally Hard thresholding is discontinuous, to overcome this Donoho [8] introduced the soft thresholding method. If the absolute value of a coefficient is less than a threshold, then is assumed to be 0, otherwise its value is shrunk by threshold.

This removes the discontinuity, but degrades all the other coefficients which tend to blur the image. In the soft thresholding method, there are deviations between image coefficients and threshold coefficients which directly influence the accuracy of the reconstructed image.

III. Center Weighted Median Filter

When one give more weight to the central value of the window a special case of weighted median filters called the Center Weighted Median filter will be produced, and thus it is easier to design and implement the general weighted median filters. For the discrete-time continues-valued of K input samples in $M \times N$ window W at point (n_1, n_2), $n_1 = 1, ..., N_1$, $n_2 = 1, ..., N_2$, $u(n_1, n_2) = [u(1), ..., u(K)]$, the output y of center weighted Median filter of K samples is given by [5],

$$y(n_1, n_2) = MED\left[u(1), ..., u(K), 2l....u(1) K \in W\right]$$

Where, l is a non-negative integer. When $l=0$, the CWM filter becomes the median filter, and when $2l+1$ is greater than or equal to the window size, it becomes the identity filter. The statistical properties of center weighted median filters have been studied to evaluate the noise suppression, edge and detail, e.g. fine lines, preservation characteristics, while the study of the deterministic properties includes root sets and convergence behavior of the filters in time domain. For identically and independently distributed inputs $F(n)$, the output distribution function $P_{cwm}(n)$ of the center weighted median with K number of samples, $K = 2k+1$, and center weight $L = 2l+1$ is given[9],

$$P_{CWM} = \sum_{k+1-l}^{2k} \binom{2k}{i} F(n)^{i+1} \left(l - F(n)\right)^{2k-i} + \sum_{k+1+l}^{2k} \binom{2k}{i} F(n)^{i} \left(l - F(n)\right)^{2k+1-l}$$

Obviously, a center weighted median filter with a larger central weight performs better in detail preservation than with a smaller central weight. The central weight should be carefully selected depending on the characteristics of the input image and its noise. The advantage of center weighted median filter is to reduce noise and to preserve fine details.

IV. PROPOSED METHODODLGY

In the hybrid work, two techniques namely, wavelet thresholding and center weighted median filters are combined to form a hybrid denoising model. These techniques are used
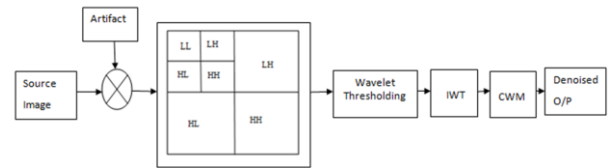
to suppress the mixed noise (Gaussian and impulse noise). The figure 2 shows the proposed method for denoising mixed noise in medical domain.

1. Gaussian and impulse noises are estimated.

2. Perform multi-scale decomposition of the image corrupted by guassian noise using wavelet transform.

3. For each level, compute the scale parameter.

4. For each sub-band

    a) Compute the standard deviation. .

    b) Compute threshold.

    c) Apply soft thresholding to the noisy coefficients.

5. Invert the multi-scale decomposition to reconstruct the denoised image.

6. The denoised image obtained from wavelet method is again processed for reducing impulse noise.

7. The denoised output produced by wavelet method still contains some artifacts.

8. To reduce this artifact and to preserve fine details the CWM filter is applied.

9. Finally the denoised image is obtained.

Figure 2. Proposed methods for denoising mixed noise in medical images

Wavelets have made quite a splash in the field of image processing. Proposed model is the newly designed hybridized one as shown in figure 2. In this model, the image is denoised first with wavelet decomposition into four sub-bands using haar wavelet filters. In the next level the wavelet based soft thresholding is applied on all the sub-bands. The shrink wavelet co-efficient is a soft thresholding is applied. It is used for suppressing the Gaussian noise. Resultant coefficients are used for image reconstruction with IWT. The results obtained after thresholding are then used to reconstruct the image. In the last level, center weighted median filter is used to remove impulse noise present in the image during transformation. The final denoised image is obtained.

Wavelets work for decomposing signals (such as images) into hierarchy of increasing resolutions. The advantage of wavelet denoising is possible to remove the noise with little loss of details. The wavelet mode denoises only the Gaussian type of noise. So when multiple noise present in the image it will remove only Gaussian the remaining noise are unremoved. So for removing the remaining noise and to preserve the fine details CWM filter is applied. The advantage of center weighted median filter can denoise the large window size. The proposed method consist of the following process

V. EXPERIMENTAL RESULTS

DICOM medical images are taken as test images for evaluating results. Here the average of ten images is taken for

evaluation. The algorithm is tested in MATLAB (7.8 Version).The reconstruction of an image has the dimensions of 256 pixel intensity. The DICOM images contain a wide variety of subject matters and textures. Most of the images used are brain images with defect and without defect images. The PSNR and UQI value must be high for a medical image, MSE and ET must be less value for a better filtering algorithm.

A. Peak Signal to Noise Ratio

The PSNR is defined in logarithmic scale, in db (decibels). The table1 shows the parametric evaluation for mixed noise removal **.**The image metric PSNR is defined as:

$$PSNR = 20.\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$
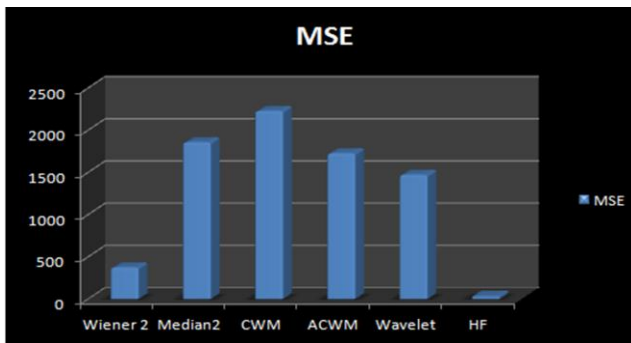
Figure 4. MSE value for Denoising Mixed noise

TABLE I.   PARAMETRIC EVALUATION OF MIXED NOISE REMOVAL

Method	PSNR	MSE	UQI	ET
Wiener2	20.42063	375.333	0.7898	3.042129
Median2	15.47557	1857.474	0.27858	2.929375
CWM	14.6842	2228.741	0.21503	3.62574
ACWM	15.79021	1727.661	0.35028	3.01658
Wavelet	71.2354	1472.231	0.231	3.2456
HF	21.49234	34.84196	0.98692	2.434407
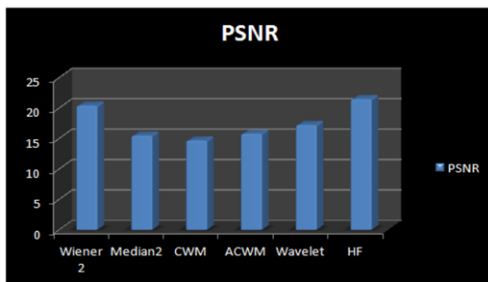
Figure 5. PSNR value for Denoising Mixed noise

The figure 4 shows the PSNR value for denosing mixed noise using didfferemt filtering technique. The PSNR value must be high for an image. The proposed method gives desire results when compare to other filtering technique.

B. Mean Square Error (MSE):

*T*he metric MSE is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|I(i,j) - K(i,j)\|^2$$

For two $m \times n$ monochrome images *I* and *K*, one of the images is considered a noisy approximation of the other. The figure 5 shows the MSE value for denosing mixed noise using didfferemt filtering technique. The MSE value must be low for an better quality image. The propsed method gives suitable results when compare to other filtering technique.
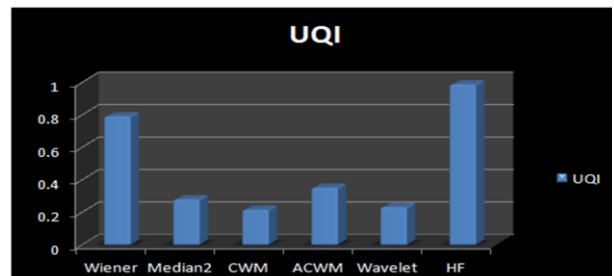
Figure 6. UQI value for Denoising Mixed noise

C. UQI

UQI measures image similarity across distortion types. Distortions in UQI are measured as a combination of three factors; Loss of correlation, Luminance distortion and Contrast distortion. Let $\{x_i\}$ and $\{y_i\}$ =1,2,...,N be the original and the test image signals, respectively. The universal quality index is defined as

$$UQI = \frac{4\sigma_{xy}\overline{x}\overline{y}}{\left[\sigma_x^2 + \sigma_x^2\right]\left[\left(\overline{x}\right)^2 + \left(\overline{y}\right)^2\right]}$$

The above figure 6 shows the UQI value for denosing mixed noise using didfferemt filtering technique

The UQI value must be high for an good image.

The propsed method gives most suitable results when compare to other filtering technique.The dynamic range of UQI is [-1, 1].

D. Evaluation Time

Evaluation Time (ET) of a filter is defined as the time taken by a digital computing platform to execute the filtering algorithms. The execution time taken by a filter should be low for online and real-time image processing applications.

Hence, a filter with lower ET is better than a filter having higher ET value when all other performance-measures are identical.
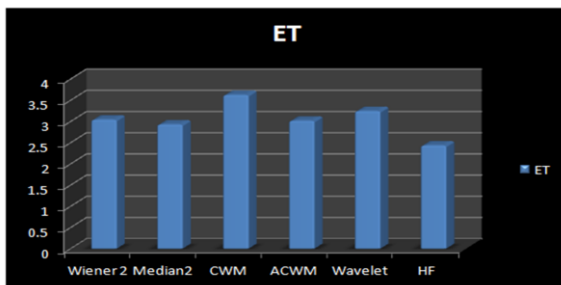
Figure 7.   ET value for Denoising Mixed noise

The above figure 7 shows the ET value for denosing mixed noise using differemt filtering technique. The proposed method gives desirable results when compare to other filtering algorithm.

V.   CONCLUSION

In this paper, an important research challenge is to improve the visual quality of CT brain images through image processing in order to detect abnormal brain at an early stage. This paper describes new methods for brain image pre-processing for noise suppression based on the wavelet transform. The image pre-processing was designed to suppress the noise present in local dense regions adaptively.

A new technique based on the hybridization of wavelet and center weighted median filters for denoising of variety of noisy images is presented in this paper. The model is experimented on standard Digital Imaging and Communications in Medicine (DICOM) images and the performances are evaluated in terms of peak signal to noise ratio (PSNR), Mean Absolute Error (MAE), Universal Image Quality Index (UQI) and Evaluation Time (ET). Results prove that utilization of center weighted median filters in combination with wavelet thresholding filters on DICOM images deteriorates the performance. The proposed filter gives suitable results on the basis of PSNR, MSE, UQI and ET. In addition, the proposed filter gives nearly uniform and consistent results on all the test images.

REFERENCES

[1]   Prof. Syed Amjad Ali, Dr. Srinivasan Vathsal and Dr. K. Lal kishore," CT Image Denoising Technique using GA aided Window-based Multiwavelet Transformation and Thresholding with the Incorporation of an Effective Quality Enhancement Method", International Journal of Digital Content Technology and its Applications, Vol. 4, N0. 4, P.75-87, 2010.

[2]   Susmitha Vekkot, and Pancham Shukla, "A Novel Architecture for Wavelet based Image Fusion," World Academy of Science, Engineering and Technology, No. 57, P.372-377, 2009..

[3]   Jenny Rajan, M.R Kaimal †Dept. Of Computer Science, University of Kerala, Trivandrum, India "Image Denoising Using Wavelet Embedded anisotropic Diffusion", Appeared in the Proceedings of IEE International Conference on Visual Information Engineering, P.589 – 593, 2006

[4]   Mukesh C. Motwani, Mukesh C. Gadiya, Rakhi C. Motwani "survey of image denoising techniques", Proceedings of GSPx 2004, Santa Clara Convention Center, Santa Clara, CA. P.27-30, 2004 .

[5]   Sudipta Roy, Nidul Sinha & Asoke K. Sen," A New Hybrid image denoising method," International Journal of Information Technology and Knowledge Management, Vol. 2, No. 2, P. 491-497,2010.

[6]   Mamta Juneja, Parvinder Singh Sandhu, "Design and Development of an Improved Adaptive Median Filtering Method for Impulse Noise Detection," International Journal of Computer and Electrical Engineering, Vol. 1, No. 5, P. 1793-8163,2009.

[7]   S.Satheesh, Dr.KVSVR Prasad, "Medical Image Denoising using Adaptive Threshold based on Contourlet Transform," Advanced Computing: An International Journal (ACIJ), Vol.2, No.2,P.52-58, 2011.

[8]   Geoffrine Judith.M.C1 and N.Kumarasabapathy, "Study and Analysis of Impulse Noise Reduction Filters," Signal & Image Processing : An International Journal(SIPIJ) Vol.2, No.1,P. 82-92, 2011.

[9]   Rastislav Lukac_, Bogdan Smołka, "Application of the Adaptive Center–Weighted Vector Median Framework for the Enhancement of CDNA Microarray Images," Int. J. Appl. Math. Comput. Sci., Vol. 13, No. 3, 369–383, 2003.

[10]   Amir Beck, Marc Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm with Application to Wavelet-based Image Deblurring," IEEE conference, ICASSP 2009,P.693-696,2009.

[11]   Naga Sravanthi Kota, G.Umamaheswara Reddy ," Fusion Based Gaussian noise Removal in the Images Using Curvelets and Wavelets With Gaussian Filter," International Journal of Image Processing (IJIP), Vol.5, No.4 ,P.456-468, 2011.

[12]   Gnanambal Ilango and R. Marudhachalam, "New Hybrid Filtering Techniques for Removal of Gaussian Noise From Medical Images," ARPN Journal of Engineering and Applied Sciences, VOL. 6, NO. 2, P.8-12, 2011.

[13]   Florian Luisier, Thierry Blu, Senior and Michael Unser, "Image Denoising in Mixed Poisson–Gaussian Noise," IEEE Transactions on Image Processing, Vol. 20, No. 3,P. 696-707, 2011.

[14]   Pierre Gravel, Gilles Beaudoin, and Jacques A. De Guise, "A Method for Modeling Noise in Medical Images," IEEE Transactions on Medical Imaging, Vol. 23, No. 10,P.1221-1232, 2004.

AUTHORS PROFILE

Ms. J.Umamaheswari Research Scholar in Computer Science, Dr. G.R.D college, Coimbatore. She has 5 years of teaching experience and two years in Research. Her areas of interest include Image Processing, Multimedia and communication. She has more than 3 publications at International level. She is a life member of professional organization IAENG.

Dr. G. Radhamani Professor & Director in School of Computer Scienec, Dr.G.R..D College, Coimbatore. She was awarded "Bharat Joythi" in lieu of the yeomen services rendered with excellence in their respective fields. She has worked for the Malaysia Venture Capital Management Ph.d, and has been a reviewer for IEEE Wireless

A Discriminant Model of Network Anomaly Behavior Based on Fuzzy Temporal Inference

Ping He

Department of Information
Liaoning police Academy
Dalian 116036, China

Abstract—**The aim of this paper is to provide an active inference algorithm for anomalous behavior. As a main concept we introduce fuzzy temporal consistency covering set, and put forward a fuzzy temporal selection model based on temporal inference and covering technology. Fuzzy set is used to describe network anomaly behavior omen and character, as well as the relations between behavior omen and character. We set up a basic monitoring framework of anomalous behaviors by using causality inference of network behaviors, and then provide a recognition method of network anomaly behavior character based on hypothesis graph search. As shown in the example, the monitoring algorithm has certain reliability and operability.**

Keywords- network anomaly behavior; anomalous omen; fuzzy temporal; set covering; hypothesis graph.

I. INTRODUCTION

Network anomaly behavior monitoring is a hotspot in reaches of network security. Up till now the basic idea of network anomaly discriminant lies in anomaly detection method, provided by Denning in 1987[1]. That is to say, according to abnormality situations of audit statement in monitoring system, the bad behaviors (i.e. events of violating safety norms) in network can be detected. The most study on network anomaly detection is based on the theory of data analysis. For example, probability and statistical method [2], data mining method [3], artificial immune algorithm [4], and corresponding artificial intelligence method [5] and so on. But the above methods have common constraint condition-data completeness.

To a certain extent, this condition is implementable. Meanwhile, a superior false alarm rate exists relatively. So it can not satisfy the requirement about reliability of network monitor. Literature [5] introduced an analysis method based on knowledge diagnosis. Literature [6] put forward a method for anomaly detection based on direct inference. Literature [7] brought forward an extrapolation inference diagnosis model based on set covering (GSC). It is one of knowledge diagnosis models with many advantages, such as intuition, parallel, leading into heuristic algorithm easily. Probability causality was led into the model by Peng in literature [7，8].

In order to solve problems about anomaly detection for fuzzy network behavior, literature [9] combined intrusion detection model with fuzzy theory. Among the above mentioned methods, it had no consideration of causality between anomalous behaviors and data in network, as well as temporal constraint relationship with each other. Because data features come into being with network anomaly behaviors, the interaction discriminant method based on anomalous behavior-data is an active network monitor and defensive strategy.

In this paper, we propose a fuzzy temporal inference method, and describe inaccuracy for network temporal knowledge by using of fuzzy set, and then constitute a model of network temporal generalized behaviors covering (NTGSBC). Finally, discriminant results are satisfactory.

II. NETWORK TEMPORAL GENERALIZED BEHAVIORS COVERING (NTGSBC)

A. Fuzzy Temporal of Network Behaviors

Fuzzy temporal analysis method has reached satisfied result in the research of fault diagnosis [3, 4]. In this section, we will set up an inference method for network anomaly behavior by using the analysis method in literature [4]. As we know, the main character of complexity in network behaviors is time fuzziness of its behavior state. In fact, occurring temporal of network behavior is an uncertain time based on interval transition.

It is a fuzzy period of time, so definition is as follows:

Definition 2.1 [4]. Network behavior happened in a network fuzzy time interval (N.F.T.I), suppose I be a trapezoid fuzzy number (T.F.N) defined on the network behavior time axis T，$I = (\theta_t, t_s, t_f, \theta_h)$, the membership degree $\mu_I(t)$ is as follows：

$$\mu_I(t) = \begin{cases} 0, t < t_s - \theta_l \\ (t - t_s + \theta_l) / \theta_l, t_s - \theta_l \leq t \leq t_s \\ 1, t_s \leq t \leq t_f \\ (t_f + \theta_h - t) / \theta_h, t_f \leq t \leq t_f + \theta_h \\ 0, t > t_f + \theta_h \end{cases} \quad (1)$$

The start time in I is expressed by start (I), and it is defined as:

$$\mu_{start(I)} = \begin{cases} 0, t < t_s - \theta_l \\ (t - t_s + \theta_l)/\theta_l, t_s - \theta_l \le t - t_s \\ 1, t = t_s \\ 0, t > t_s \end{cases} \quad (2)$$

The end time in I is expressed by end (I), and it is defined as:

$$\mu_{end(I)} = \begin{cases} 0, t < t_f \\ 1, t = t_f \\ (t_f + \theta_h - t)/\theta_h, t_f < t \le t_f + \theta_h \\ 0, t > t_f + \theta_h \end{cases} \quad (3)$$

When $t_s = t_f = t$, I can be reduced to a network fuzzy time point(F.T.P), $I = (\theta_l, t, t, \theta_h)$, so it turns into a triangular fuzzy numbers(T.F.N). When $\theta_l = \theta_h = 0$, I can be reduced to an accurate time. F.T.I and F.T.P are shown as Figure 1.
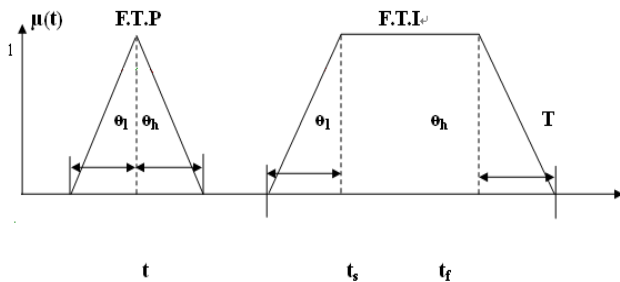
Figure 1. F.T.I and F.T.P

Fuzzy difference between two (T.F.N) reflect a kind of fuzzy temporal in the real network, and it exits

$$I = I_1 \simeq$$
$$I_2 = (\theta_l^{(1)}, t_s^{(2)}, t_f^{(1)} - t_f^{(2)}, t_f^{(1)} - t_s^{(2)}, \theta_h^{(1)} + \theta_l^{(2)}) \quad (4)$$

B. Detection model for Anomalous Behavior

Definition 2.2 A discriminant model of network anomaly behavior is a relation mapping inversion space of data – behavior with fuzzy temporal, the formal representation can be shown as $P = <A, D, R, DEL, D^+, DOCT>$, where

(1) $A = \{a_1, a_2, ..., a_m\}$, it is a non-empty finite set with anomalous behavior;

(2) $D = \{D_1, D_2, ..., D_n\}$, it is a non-empty finite set with anomalous data;

(3) $R \subseteq A \times D$ is a relationship of network anomaly behavior defined on $A \times D$. $<a_i, D_j> \in R$, if and only if behavior a_i reflects data D_j in network, and $\forall D_j \in D$,

there exists an element $a_i \in A$ at least, such that $<a_i, D_j> \in R$;

(4)DEL is a delay matrix on $|A| \times |D|$. For $<a_i, D_j> \in R, DEL(i, j) = (\theta_l^d, a^{(i,j)}, A^{(i,j)}, \theta_h^d)$ is a F.T.I. It expresses the delay time approximately "from $a^{(i,j)}$ to $A^{(i,j)}$" between "the beginning of anomalous behavior" and "the beginning of data D_j"; for $<a_i, D_j> \notin R$, $DEL(i, j)$ has no definition.

(5) $D^+ \subseteq D$ expresses a known anomalous data set of the discriminent target P. DOCT expresses |D| dimensional vector. For $D_j \in D^+$, DOCT $(j) = (\theta_l^m, t_s^{(j)}, t_f^{(j)}, \theta_h^m)$ is a F.T.I. It expresses the appearance time of the known anomalous data D_j approximately "from $t_D^{(j)}$ to $t_D^{(j)}$". For $D_j \notin D^+, DOCT(j)$ has no definition. $cause(D_j) = \{a_i | a_i \in A, <a_i, D_j> \in R\}$ is an all possible anomalous behavior set caused D_j. $cause(D^+) = \bigcup_{D_j \in D^+} cause(D_j)$ is an all possible anomalous data set caused D^+.

$effect(a_i) = \{D_j | D_j \in D, <a_i, D_j> \in R\}$ is a data set caused by anomalous behavior a_i. $effect(A_I) = \bigcup_{a_i \in A_I} effect(a_i)$ is an all possible data set caused by $A_I \subseteq A$. For $A_I \subseteq A, D^+ \subseteq D$, D_I is a covering of D^+, if and only if $D^+ \subseteq effect(A_I)$. Network anomaly behavior $a_i \in A$ is relative to the known anomalous data $D_j \in D^+$, the start time of anomalous behavior $a_i \in cause(D_j)$ is as follows:

$$begin - OCT(a_i | D_j) = begin\ DOCT\ j(\simeq DEL(i, j) \quad (5)$$

The end time of anomalous behavior a_i, relative to D_j, is as follows:

$$end - OCT(a_i | D_j) = end(DOCT(j)) \quad (6)$$

Definition 2.3 For $a_i \in A, D^{+(i)} \subseteq D^+$, $\{a_i\}$ is a temporal consistency covering on $D^{+(i)}$, if

(1) $\{a_i\}$ is a covering of $D^{+(i)}$.

(2) $\min(\max_{i \in T} \mu_{begin}(a_i \mid D^{+(i)})(t), \max_{i \in T} \mu_{end}(a_i \mid D^{+(i)})(t)) \geq \delta$

$$(7)$$

$$begin(a_i \mid D^{+(i)}) = \tilde{I}_{s_j \in S^{+(i)}} begin-OCT(a_i \mid D_j) \qquad (8)$$

$$end(a_i \mid S^{+(i)}) = \tilde{I}_{D_j \in D^{+(i)}} end-OCT(a_i \mid D_j) \qquad (9)$$

$0 \leq \delta \leq 1$ is a threshold constant with temporal consistency,

$\tilde{I}$ expresses intersection of fuzzy sets.

Different from literature [1,4], the solution of discriminant target in NTGSBC is not only pointed out anomalous behavior set A_I covered consistency data set D^+, but also ensured $D_j \in D^+$ caused by $a_i \in A_I$, because of temporal consistency requirements.

Definition 2.4 Suppose a complete explanation of discriminant target P in network behaviors *be*

$Co-\exp(P) = \{(a_i, D^+(a_i)) \mid a_i \in A, D^+(d_i) \subseteq D^+\}$,

and satisfying the following conditions:

(1) For $(a_i, D^+(a_i)) \in Co-\exp(P)$, $\{a_i\}$ is a temporal consistency covering on $D^+(a_i)$.

(2) $\bigcup_{(a_i, D^+(a_i)) \in Pa-\exp(P)} D^+(a_i) = D_1^+$ expresses a known

anomalous omen set.

$A_I = \{a_i \mid ((a_i, D^+(a_i)) \in Co-\exp(P)\}$ is called an explanation omen set of complete explaining $Co-\exp(P)$. Obviously A_I covers D^+.

Definition 2.5 Suppose a partial explanation of discriminant target *P* in network behaviors be

$Pa-\exp(P) = \{(a_i, D^+(a_i)) \mid a_i \in A, D^+(a_i) \subseteq D^+\}$,

and satisfying the following conditions:

(1) For $(a_i, D^+(a_i)) \in Pa-\exp(P)$, $\{a_i\}$ is a temporal consistency covering on $D^+(a_i)$.

(2) $D_1^+ = \bigcup_{(a_i, D^+(a_i)) \in Pa-\exp(P)} M^+(a_i) \subset D^+$,

and for $D^+ - D_1^+$,

$cause(D_2^+) \subseteq A_I = \{a_i \mid (a_i, S^+(a_i)) \in Pa-\exp(P)\}$.

D_2^+ is called a non-covering anomaly data set of $Pa-\exp(P)$, A_I is called an explanation anomaly behavior set of $Pa-\exp(P)$.

In the definition 2.5, due to $cause(D_2^+) \subseteq A_I$, it has $D_2^+ \subseteq effect(A_I)$, namely A_I is a covering of D_2^+, so A_I is a covering of the whole known anomalous data set $D^+ = D_1^+ \cup D_2^+$. But there exists a constraint condition of temporal consistency, all the a_i in A_I only can explain (or cover) D_1^+, a part of D^+.

A complete solution $Co-\exp(P)$ of discriminant target *P* is a complete explanation of *P*. and the cardinality of anomalous data set D_I, explained $Co-\exp(P)$, is minimum. A partial solution of *P* is a partial explanation $Pa-\exp(P)$, and the cardinality of non-covering anomaly omen set D_2^+ is minimum.

III. SOLVING DISCRIMINANT TARGET IN NETWORK BEHAVIORS

A. Hypothesis Graph

Solving problems of a discriminant target in network behaviors is based on a search method of hypothesis graph. Hypothesis graph is defined by network nodes and successor.

Definition 3.1 Suppose a node n_i be

$n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)})$,

in the hypothesis graph G(P) of discriminant target *P*, where

$A_I^{(i)} \subseteq A, D_1^{+(i)} \subseteq D^+ I\ effect(A_I^{(i)}),$

$D_2^{+(i)} = D^+ - D_1^{+(i)}$

$n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)})$

in G(P) is divided into three types:

1) Complete end node: for $n_i, D_2^{+(i)} = \phi$。

2) Partial end node: $n_i, D_2^{+(i)} \neq \phi$ *and* $cause(D_2^{+(i)}) \subseteq A_I^{(i)}$.

3) Non-end node: the other nodes except the above two types nodes.

Definition 3.2 *For the node* $n_k = (A_I^{(k)}, D_1^{+(k)}, D_2^{+(k)})$ in the hypothesis graph G(P), if it has $D_j \in D_2^{+(k)}, d_i \in cause(D_j) - A_I^{(k)}$, then a successor node of n_k is as follows:

$succ(n_k, a_i) = (A_I^{(k)} U\{a_i\}, D_1^{+(k)} U\ some(D_2^{+(k)} I\ effect(a_i)), D_2^{+(k)} - some(D_2^{+(k)} I\ effect(a_i)))$

Where $\{a_i\}$ is a temporal consistency covering on $some(D_2^{+(k)} I\ effect(a_i)) \subseteq D_2^{+(k)} I\ effect(a_i)$.

When constructed G (P), it began from an initial node $n_0 = (\phi, \phi, D^+)$, then continually expanded nodes and

generated its successor node, until to the expansion node translated into non-expansion complete nodes or partial nodes. The following two theorems point out corresponding relationship in G (P) between a path and discriminant targets.

Theorem 3.1 *Suppose a path in hypothesis graph* G (P), from an initial node $n_0 = (\phi, \phi, D^+)$ to some complete end node $n_l = (A_I^{(l)}, D_1^{+(l)}, D_2^{+(l)})$, be (without loss of generality) $n_0, n_1, ..., n_i, n_{i+1}, ..., n_l$, where n_{i+1} is a successor node of n_i, $0 \le i \le l-1, n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)})$, then

$$D^+(a_k)) \mid a_k \in A_I^{(i+1)} - A_I^{(i)}, D^+(a_k) = D_1^{+(i+1)} - D_1^{+(i)},$$
$$0 \le i \le l-1\}$$

is a complete explanation $Co - \exp(P)$ of the target *P*, and the set of anomalous behavior explanation $Co - \exp(P)$ is $A_I = A_I^{(l)}$.

Theorem 3.2 Suppose a path in hypothesis graph *G(P)*, from an initial node $n_0 = (\phi, \phi, D^+)$ to some partial end node $n_l = (A_I^{(l)}, D_1^{+(l)}, D_2^{+(l)})$, be (without loss of generality) $n_0, n_1, ..., n_i, n_{i+1}, ..., n_l$, where n_{i+1} is a successor node of n_i, $0 \le i \le l-1, n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)})$, then

$$\{S^+(a_k)) \mid a_k \in A_I^{(i+1)} - A_I^{(i)}, D^+(a_k)$$
$$= D_1^{+(i+1)} - D_1^{+(i)}, 0 \le i \le l-1\}$$

is a partial explanation $Pa - \exp(P)$ of the target *P*, and the set of non-covering anomaly omen $Pa - \exp(P)$ is $D_2^+ = D_2^{+(l)}$.

The proof of above two theorems is shown in appendix.

B. Operating process of the discriminant system

According to theorem 3.1and 3.2, the operating of monitor system is based on the search of hypothesis graph. If it exists complete end nodes in the final hypothesis graph G(P), then the complete explanation $Co - \exp(P)$ of the target *P* can be obtained. And then a minimum in $|A_I|$ is taken for a complete solution $Co - \exp(P)$ of the target *P*, where A_I is a set of explanation anomaly behavior $Co - \exp(P)$. If it only exits partial end nodes in G(P), then the partial explanation $Pa - \exp(P)$ of the target *P* can be obtained. And then a minimum in $|S_2^+|$ is taken for a partial solution $Pa - \exp(P)$ of the target *P*, where S_2^+ is a set of non-covering anomaly omen $Pa - \exp(P)$.

Solution procedure is as follows:

(1) Algorithm Solve-TGSC(A, D, R, DEL, D^+, MOCT)

(2) Variable n_i: node, D_j: data, a_k: anomalous behavior, table OPEN, table CLOSE

(3)Begin OPEN: $= \{n_0 = (\phi, \phi, D^+)\}$

(4) CLOSE: $= \phi$: {Initializing table OPEN, CLOSE}

(5) While there are non-terminal nodes in OPEN do

(6) Begin n_i : = POP(OPEN)

{Removing non-end node $n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)})$ from table OPEN, and add to table CLOSE}

(7) $D_j := select(D_2^{+(i)})$; {Selecting D_j from $D_2^{+(i)}$ }

(8) For each $a_k \in cause(s_j) - A_I^{(i)}$ do

(9) Begin SUB: = choose $(D_2^{+(i)} I\ effect(a_k))$; {Constructing set SUB}

(10) For each $sub_i \in SUB$ do

(11) Begin such（n_i, a_k）： =
$$(A_I^{(i)} U\{a_k\}, D_1^{+(i)} U\ sub_i, D_2^{+(i)} - sub_i);$$

{Constituting successor node of n_i }

(12) INSERT （such（n_i, d_k）, OPEN）;{ renewing table *OPEN*}

(13)   End for
         End for
      End for

(14) If there are complete terminal nodes in OPEN

(15) Then solving the complete solution of problem *P*, return $Co - sol(P)S$; {theorem 3.1}

(16) Else solving the partial solution of problem *P*, return $Pa - sol(P)S$; {theorem 3.2}

(17) End.

In the step (9), it constructs a set by using function choose, satisfying the following conditions

$$SUB = \{sub_i \mid sub_i \subseteq D_2^{+(i)} I\ \ effect(a_k), sub_i \ne \phi\}$$

(a) For $sub_i \in SUB, \{a_k\}$ is a temporal consistency covering.

(b) For $sub_i \in SUB$, It does not exist

$$sub_i \subseteq D_2^{+(i)} I\ \ effect(a_k),$$

such that $sub_j \supset sub_i$, and $\{d_k\}$ is a temporal consistency covering on sub_i.

In the step (12), for successor node of n_i,

$$n_j = succ(n_i, a_k) = (A_I^{(j)}, D_1^{+(j)}, D_2^{+(j)}),$$

INSERT revises table OPEN as follows:

(i) If n_j is a complete or partial end node, n_j will be put into the table OPEN directly. Otherwise n_j is a non-end node, and taken by the following steps:

(ii) If it has $n_t = (A_I^{(t)}, D_1^{+(t)}, D_2^{+(t)})$ in the table OPEN and *CLOSE*, such that $D_2^{+(j)} = D_2^{+(t)}, A_I^{(t)} \subseteq A_I^{(j)}$, then the node n_j is abandoned, the table *OPEN* is not change.

(iii) If it has $n_t = (A_I^{(t)}, D_1^{+(t)}, D_2^{+(t)})$ in the table OPEN, such that $D_2^{+(j)} = D_2^{+(t)}, A_I^{(t)} \supset A_I^{(j)}$, then the node n_t is deleted from the table OPEN, and put into the table CLOSE, meanwhile n_j is put into the table OPEN.

(iv) Otherwise, n_j is put into the table OPEN.

In afore-mentioned algorithm, the table OPEN is using to deposit expanding nodes, and the table CLOSE is using to deposit expanded nodes. In the process of constructing G(P), successor nodes is generated in basis of causality and temporal constraint(temporal consistency) between anomalous behavior and anomalous omen. Suppose the graph is acyclic graph, and the number of nodes is limited. The complete or partial nodes can be obtained through successor expanded $succ(n_i, a_k)$ after undergoing limited steps (no more than |D|). Therefore termination of algorithm is quite obvious.

IV. INSTANCE ANALYSIS

A discriminant target of network behaviors

$P = < A, D, R, DEL, D^+, DOCT >$ has the following definition :

$D^+ = \{D_1, D_2, D_3, D_4, D_5\}$, $A = \{a_1, a_2, a_3\}$,

then the relation matrix of behavior-data in network is as follows :

$$R = \begin{array}{c} \\ D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{array} \begin{array}{ccc} a_1 & a_2 & a_3 \\ \left[\begin{array}{ccc} 1 & & 1 \\ 1 & & \\ 1 & 1 & \\ & 1 & 1 \\ & 1 & 1 \end{array}\right] \end{array}$$

Based on the theorem 3.1 and 3.2, it can be obtained the following values:

DOCT(1)=(1,8,10,1), DOCT(2)=(1,9.8,9.8,1),
DOCT(3)=(1,12,13,1),DOCT(5)=(1,10,11,1)

$DEL(1,1) = (1,3,4,1), DEL(1,3) = (1,3,3,1),$

$DEL(2,4) = (1,7,7,1), DEL(3,1) = (1,8,8,1),$

$DEL(1,2) = (1,4.6,5.7,1), DEL(2,3) = (1,4,5,1),$

$DEL(2,5) = (1,5,6,1), DEL(3,4) = (1,9,10,1)$

$DEL(3,5) = (1,6,$

It takes a temporal consistency threshold $\delta = 0.6$, it is taken the first in first out strategy the table OPEN, and generated a hypothesis graph, it is shown as Figure 2.
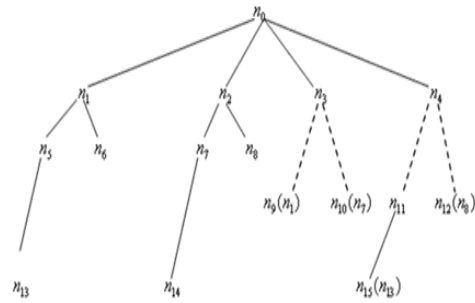
Figure 2.   Hypothesis graph of discriminant target *P*

Where

$n_0 = (\phi, \phi, D^+), n_1 = (\{a_1\}, \{D_1, D_2\}),$

$n_2 = (\{a_1\}, \{D_3\}), n_3 = (\{a_3\}, \{D_1\}),$

$n_4 = (\{a_3\}, \{D_5\}), n_5 = (\{a_1, a_2\}, \{D_1, D_2, D_3\}),$

$n_6 = (\{a_1, a_2\}, \{D_1, D_2, D_5\}), n_7 = (\{a_1, a_3\}, \{D_3, D_1\}),$

$n_8 = (\{a_1, a_3\}, \{D_3, D_5\}), n_9 = (\{a_3, a_1\}, \{D_1, D_2\}),$

$n_{10} = (\{a_3, a_1\}, \{D_1, D_3\}), n_{11} = (\{a_3, a_1\}, \{D_5, D_1, D_2\}),$

$n_{12} = (\{a_3, a_1\}, \{D_5, D_3\}), n_{13} = (\{a_1, a_2\}, \{D_1, D_2, D_3, D_5\}),$

$n_{14} = (\{a_1, a_3, a_2\}, \{D_3, D_1, D_5\}),$

$n_{15} = (\{a_1, a_3, a_2\}, \{D_5, D_1, D_2, D_3\})$

Because of $D_1^{+(i)} \cup D_2^{+(i)} = D^+$, n_i is written $n_i = (A_I^{(i)}, D_1^{+(i)})$ for short. In the beginning of algorithm, it is taken $D_1 \in D_2^{+(0)} = D^+$ from a initial node $n_0 = (\phi, \phi, D^+)$, then $a_1, a_3 \in cause(D_1) - A_I^{(0)}$. Based on a behavior a_1, it can construct sub-set $SUB = \{\{D_1, D_2\}, \{D_3\}\}$, and generate successor nodes n_1, n_2 in the basis of n_0, then put it into table *OPEN*. In the same way, based on a_3, it can construct sub-set $SUB = \{\{D_1\}, \{D_3\}\}$, and then obtain the successor nodes n_3, n_4 in the basis of n_0.

After generated $n_9 = (\{a_3, a_1\}, \{D_1, D_2\}, \{D_3, D_5\})$ in the step (4), due to $A_I^{(1)} = \{a_1\} \subset A_I^{(9)}, D_2^{+(1)} = D_2^{+(9)}$ in n_1, according to principle of INSERT in the step (ii), the node n_9 could be abandoned, namely it may be pruning correspondingly. Based on INSERT principle, n_{10} and n_{12} are treated in the same way. The corresponding complete solution of paths $n_0 \to n_1 \to n_5 \to n_{13}$ and $n_0 \to n_4 \to n_{11} \to n_{15}$ is

$$Co - sol(P) = \{(a_1, \{D_1, D_2\}), (a_2, \{D_3\}), (a_3, \{D_5\})\}$$

. Partial end nodes n_6 and n_{14} are corresponding to partial solutions respectively. For example n_6 corresponds to $Pa - sol(P) = \{(a_1, \{D_1, D_2\}), (a_2, \{D_5\})\}$, it is known that the omen m_3 does not explain. The occurrence time of each known omen and corresponding anomalous behavior is shown in Figure1.

V. CONCLUSION

The solution for network anomaly detection in this paper is a further development based on literature [5，6]. Actually, it is breadth- first search method. So the search cost is still greater, especially for a large amount of data, though the method presented in this paper makes pruning, in order to decrease the number of network nodes, by using of temporal consistency in the step of SUB, INSERT etc. Trying to resolve this conflict, a possible method is to convert the original method into depth-first search method by introducing node evaluation function, such as literature [7，8，9]. But in the model of NTGSBC, node evaluation function must reflect causality and temporal constraint between anomalous behavior and data at the same time. It is more complex than pure probability causality in literature [2, 3]. It is yet to be further studied about how to seek appropriate node evaluation function in the model of NTGSBC. The other possible solution is problem decomposition. For example, in total behavior detection system modeling of a large website, we will divide the total detection process into many subsystems according to structure and function of website system. And define the causality among subsystems. Moreover the subsystem itself is defined by the model of NTGSBC. Anomaly behavior detection process is separated into inner inference for NTGSBC monitor system and anomaly causality diffusion among subsystems. The advantage of this method is as follows: (1) the detection scale of subsystems, obtained after decomposition, is smaller. It is fit for NTGSBC modeling and problem solving. (2) Through problem decomposition and defining the diffusion causality among subsystems, multi-layered causality model about total monitor targets will be built up, based on two layer causality from anomalous behavior to data. Multi-layered causality model is more suitable for detection target describing and problem solving.

Temporal consistency set covering is defined in this paper. And based on this definition, we described the basic framework of NTGSBC and the method of problem solving. Fuzzy temporal information is introduced in the model of NTGSBC, it makes generalized inference detection model and method more fitting for practical problems in other fields. Certainly, more detailed studies should be continued in the further.

REFERENCES

[1] Bykova M, Ostermann S, Tjaden B. Detecting network intrusions via a statistical analysis of network packet characteristics. In: Proc. of the 33rd Southeastern Symp. on System Theory. 2001. 309314. http://masaka.cs.ohiou.edu/papers/ssst2001.pdf

[2] Denning DE. An intrusion-detection model. IEEE Trans. on Software Engineering, 1987,13(2):222-232.

[3] Lee W, Stolfo SJ. A framework for constructing features and models for intrusion detection systems. ACM Trans. on Information and System Security, 2000,3(4):227-261.

[4] Valdes A, Skinner K. Adaptive, model-based monitoring for cyber attack detection. In: Debar H, M¨ L, Wu SF, eds. Proc. of the 3rd Int¹Workshop on the Recent Advances in Intrusion Detection (RAID 2000). LNCS 1907, Heidelberg: Springer-Verlag, 2000. 80-93.

[5] Aickelin U, Greensmith J, Twycross J. Immune system approaches to intrusion detectionA review. In: Nicosia G, et al., eds. Proc. of the 3rd Int¹¹onf. on Artificial Immune Systems. LNCS 3239, Heidelberg: Springer-Verlag, 2004. 316-329.

[6] Lee W, Stolfo SJ. A Data mining framework for building intrusion detection models. In: Gong L, Reiter MK, eds. Proc. of the !99 IEEE Symp. on Security and Privacy. Oakland: IEEE Computer Society Press, 1999. 120-132.

[7] Eskin E, Arnold A, Prerau M, Portnoy L, Stolfo SJ. A geometric framework for unsupervised anomaly detection: detecting intrusions in unlabeled data. In: Barbar¨ D, Jajodia S, eds. Applications of Data Mining in Computer Security. Boston: Kluwer Academic Publishers, 2002. 78-99.

[8] Proedru K, Nouretdinov I, Vovk V, Gammerman A. Transductive confidence machine for pattern recognition. In: Elomaa T, et al., eds. Proc. of the 13th European Conf. on Machine Learning. LNAI 2430, Heidelberg: Springer-Verlag, 2002. 381-390.

[9] Barbar¨ D, Domeniconi C, Rogers JP. Detecting outliers using transduction and statistical testing. In: Ungar L, Craven M, Gunopulos D, Eliassi-Rad T, eds. Proc. of the 12th ACM SIGKDD Int¹¹onf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2006. 55-64.

[10] Angiulli F, Pizzuti C. Outlier mining in large high-dimensional data sets. IEEE Trans. on Knowledge and Data Engineering, 2005, 17(2):203-215.

[11] Ghosh AK, Schwartzbard A. A study in using neural networks for anomaly and misuse detection. In: Proc. of the 8th USENIX Security Symp. 1999. 141-151. http://www.usenix.org/events/sec99/full_papers/ghosh/ghosh.ps

[12] Manikopoulos C, Papavassiliou S. Network intrusion and fault detection: A statistical anomaly approach. IEEE Communications Magazine, 2002,40(10):76-82.

[13] Laskov P, Schafer C, Kotenko I. Intrusion detection in unlabeled data with quarter-sphere support vector machines. In: Proc. of the Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA 2004). 2004. 71-82. http://www2.informatik.hu-berlin. de/wm/journalclub/dimva2004.pdf

AUTHORS PROFILE

He Ping is a professor of the Department of Information at Liaoning Police Academy, P.R. China. He is currently Deputy Chairman of the Centre of Information Development at Management Science Academy of China. In 1986 He advance system non-optimum analysis and founded research institute. He has researched analysis of information system for more than 20 years. Since 1990 his work is optimization research on management information system. He has published more than 200 papers and ten books, and is editor of several scientific journals. In 1992 awards Prize for the Outstanding Contribution Recipients of Special Government Allowances P. R. China.

APPENDIX

The proof of theorem 3.1is as follows:

(1) Note

$$Co - ps = \{(a_k, D^+(a_k)) \mid a_k \in A_I^{i+1} - A_I^i, D^+(a_k)$$
$$= D_1^{+(i+1)} - D_1^{+(i)}, 0 \leq i \leq l - 1\}$$,

as n_{i+1} is a successor node of n_i, according to the definition 3.2，it exists $D_j \in D_2^{+(i)}, a_k \in cause(D_j) - A_I^{(i)}$, such that

$n_{i+1} = succ(n_i, a_k)$

$= (A_I^{(i)} \cup \{a_k\}, D_1^{+(i)} \cup some(D_2^{+(i)} \cap effect(a_k)), D_2^{+(i)}$

$- some(D_2^{+(i)} \cap effect(a_k)))$,

a_k is a temporal consistency covering on

$$some(D_2^{+(i)} \cap effect(a_k)) \subseteq D_2^{+(i)} \cap effect(a_k).$$

Due to $a_k \in cause(D_j) - A_I^{(i)}$, it exists

$\{a_k\} \cap A_I^{(i)} = \phi, A_I^{(i+1)} - A_I^{(i)} = A_I^{(i)} \cup \{a_k\} - A_I^{(i)} = \{a_k\}$,

$a_k \in A_I^{(i+1)} - A_I^{i}$.

Based on $D_2^{+(i)} = D^+ - D_1^{+(i)}$, it exists

$D_1^{+(i)} \cap D_2^{+(i)} = \phi, D_1^{+(i)} \cap some(D_2^{+(i)} \cap effect(a_k)) = \phi$,

Then it obtains

$D^+(a_k) = D_1^{+(i+1)} - D_1^{+(i)}$

$= D_1^{+(i)} \cup some(D_2^{+(i)} \cap effect(a_k)) - D_1^{+(i)}$

$= some(D_2^{+(i)} \cap effect(a_k))$

So for $(a_k, D^+(a_k)) \in Co - ps, \{a_k\} = A_I^{i+1} - A_I^i$ is a temporal consistency covering on $D^+(a_k)$.

(2) In order to express itself clearly note $a_k^{i+1} \in A_I^{(i+1)} - A_I^{(i)}, D^{+(i+1)}(a_k^{i+1}) = D_1^{+(i+1)} - D_1^{+(i)}$, then

$$\bigcup_{(a_k, D^+(a_k)) \in co - ps} D^+(a_k) = \bigcup_{0 \le i \le l-1} D^{+(i+1)}(a_k^{+(i+1)}) \quad (A.1) \text{ A}$$

proof by mathematical induction is adopted firstly. For $p \le l-1$, the following expression holds

$$\bigcup_{0 \le i \le p} D^{+(i+1)}(a_k^{+(i+1)}) = D_1^{+(p+1)} \quad (A.2)$$

When $p=0$, noticed that $S_1^{+(0)} = \varphi$ for initial node n_0, it exits

$$\bigcup_{0 \le i \le 0} D^{+(i+1)}(a_k^{(i+1)}) = D^{+(1)}(a_k^{(1)})$$

$= D_1^{+(1)} - D_1^{+(0)} = D_1^{+(0+1)}$

Suppose the expression (A.2) is set up when $p \le t, t \le l-2$, then

$$\bigcup_{0 \le i \le t+1} D^{+(i+1)}(a_k^{(i+1)})$$

$= (\bigcup_{0 \le i \le t} D^{+(i+1)}(a_k^{(i+1)})) \cup D^{+(t+2)}(a_k^{(t+2)})$

$= (D_1^{+(t+1)} \cup (D_1^{+(t+2)} - D_1^{+(t+1)}))$

$= D_1^{+(t+1)} \cup some(D_2^{+(t+1)} \cap effect(a_k^{t+2}))$

$= D_1^{+(t+2)}$

So the expression (A.2) is set up. As well as $n_l = (A_I^{(l)}, D_1^{+(l)}, D_2^{+(l)})$ is a complete end node, $D_1^{+(l)} = D^+, D_2^{+(l)} = \varphi$, hence

$$\bigcup_{0 \le i \le l-1} D^{+(i+1)}(a_k^{(i+1)}) = D_1^{+(l)} = D^+.$$

According to the results of expression (1) and (2), and definition 2.4, the set Co-ps is a complete explanation Co-exp（P）for the problem P.

(3) As Co-ps is a Co-exp（P）of the problem P, according to definition 2.4, a fault set is as follows:

$A_I = \{a_k \mid (a_k, S^+(a_k)) \in Co - ps\}$

$= \bigcup_{0 \le i \le l-1} \{a_k^{(i+1)}\}$

$= \bigcup_{0 \le i \le l-1} A_I^{(i+1)}(A_I^{(i+1)}) - A_I^{(i)})$

Resembling the proof of the expression (A.2) in the step (2), a proof by mathematical induction is adopted. So the set of anomalous behavior explanation $Co - \exp(P)$ is as follow:

$$A_I = \bigcup_{0 \le i \le l-1} (A_I^{(i+1)}) - A_I^{(i)}) = A_I^{(i)}$$

Theorem 3.2 is proved as follows:

$$Pa - ps = \{(a_k, D^+(a_k)) \mid a_k \in (A_I^{(i+1)} - A_I^i),$$

$D^+(a_k) = D_1^{+(i+1)} - D_1^{+(i)}, 0 \le i \le l-1\}$

(1) it is similar to the step (1) in theorem 3.1, it exists $(a_k, D^+(a_k)) \in Pa - ps$. $\{a_k\}$ is a temporal consistency covering of $D^+(a_k)$.

(2) For the set Pa-ps, resembling the proof of step (2) in theorem 3.1, it can be obtained

$$D_1^+ = \bigcup_{(a_k, S^+(a_k)) \in Pa - ps} D^+(a_k) = \bigcup_{0 \le i \le l-1} D^{+(i+1)}(a_k^{+(i+1)}) = D_i^{+(i)} \quad (B.1)$$

Resembling the proof of step (3) in theorem 3.1, for Pa-ps, it exists

$$A_I = \{a_k \mid (a_k, D^+(a_k)) \in Pa - ps\} = \bigcup_{0 \le i \le l-1} \{a_k^{+(i+1)}\} \quad (B.2)$$

$= \Upsilon_{0 \le i \le l-1}(A_I^{(i+1)} - A_I^i) = A_I^i$

As $n_t = (A_I^i, D_1^{+(i)}, D_I^{+(i)})$ is partial end nodes, so

Cause $\{D_2^{+(i)} \subseteq A_I^i, D_2^{+(i)} \ne \varnothing, D_1^{+(i)} = D^+ - D_2^{+(i)} \subset D^+$

Then combining with the expression (B.1) and (B.2), it exists

$$D_1^+ = \bigcup_{(a_k, M^+(a_k)) \in Pa - po} D^+(a_k) = D_I^{+(l)} \subset D^+.$$

For

$D_2^+ = D^+ - D_1^+ = D^+ - D_1^{+(l)} = D_2^{+(l)}$,

$cause(D_2^+) = cause(D_2^+) \subseteq A_I^{(l)} = A_I$,

$cause(D_2^+)k \subseteq A_I$

$= \{a_k \mid (a_k, D^+(a_k) \in Pa - ps\}$.

According to the conclusions in step (1) and (2), and definition 2.5, Pa-ps is a partial explanation of the problem P, namely theorem 3.2 can be established.

E-Participation

Modeling and Developing with Trust for Decision Making Supplement Purpose

Vitri Tundjungsari

YARSI University
Information Technology Faculty (FTI)
Jl. Letjen Suprapto, Jakarta Pusat, Indonesia
1(Student of Department of Computer Science and
Electronics, Gadjah Mada University)

Jazi Eko Istiyanto, Edi Winarko, Retantyo Wardoyo

Department of Computer Science and Electronics
Gadjah Mada University
Jl. Sekip Utara
Yogyakarta, Indonesia

Abstract— **ICT has been employed in various areas, including e-Participation to support citizen participation and achieve democracy ideal. Trust as a social behavior can be used as a method to model preferences and facilitate better participation and interaction in a decision making within a group of decision makers. In this paper we present literatures survey related to e-Participation and trust in computer science; we also proposed a Group Decision Support System model and application, which utilizes trust to support decisions in a collaborative team. In brief, the model is based on the synthesis of group members' preferences following an appropriate aggregation procedure.**

Keywords- e-Participation; trust; group decision making.

I. INTRODUCTION

Information and communication technology (ICT) has become deeply involved in various studies regarding decision making. Advances in ICT facilitate solution on some of the decision making problems such as in politics, economics, and engineering sciences. Nowadays, decision making is not only based on single decision maker perspectives but also involves several decision makers, also called as a group decision making. In politics, especially in governmental concept, this group usually consists of government staffs who also expert in their fields. However, government today with democratic concepts requires much more citizen participation in decision making to achieve democracy ideals stated by Abraham Lincoln's (the government of the people, by the people, and for the people). Furthermore, we believe that ICT could be applied to enhance citizen participation in the policy process [36]. Macintosh [22] mentions the overarching objectives of e-participation are given as: (1) reach a wider audience to enable broader participation; (2) support participation through a range of technologies to cater for the diverse technical and communicative skills of citizens; (3) provide relevant information in a format that is both more accessible and more understandable to the target audience to enable more informed contributions; (4) engage with a wider audience to enable deeper contributions and support deliberative debate.

There is no clear definition of e-Participation from various literatures, however we can conclude that ICT usage (particularly Internet based) in those participation action and mechanisms, with the implication that the technology has the ability to change or transform the communication processes between the participants involving citizens in societal decision making, is also called as e-Participation. It is normally associated with some form of political deliberation or decision-making within the formal political process (e.g. voting), or outside it (e.g. political activism) [30].

Our literatures review concentrates on how participatory research can be applied in the decision-making sense, and also includes e-Participation research's state of the art. Moreover, this paper contributes to the work of defining the emerging research area of e-Participation by: (1) describing and categorizing recent set of relevant and important literatures in area of e-Participation; (2) proposing new model of e-Participation for decision making purpose and developing it into a prototype called Par-GDSS.

The rest of the paper is organized as follows. Section 2 provides a brief review of existing e-Participation research. Section 3 and 4 presents our e-Participation model and application. Finally in section 5 we conclude the paper and state our future work.

II. E-PARTICIPATION LITERATURES SURVEY

E-Participation is a technology-mediated interaction between the civil society sphere and the formal politics sphere, and between the civil society sphere and the administration sphere [30]. The main point of e-Participation is the citizen, i.e., the purpose of e-Participation is to increase citizens' abilities to participate in digital governance, including participation in the political process and the transformation of digital government information and services.

Literatures survey shows that e-Participation is an emerging research area and has been studied and implemented in some developed and developing countries ([11], [26], [30]). However, survey also finds out that e-Participation (related to Decision making) still has limited source both in theory development and its practice. In table 1, we present e-Participation research based on its motivation and reference discipline from 29 paper sources.

E-PARTICIPATION CLASSIFICATION BASED ON RESEARCH MOTIVATION

Research Motivation: Participative Imperative		
Examples	Reference discipline	Reference
Participatory design and ethnographic action research	Information systems (socio-technical)	Watkins & Tacchi (2008)
Political and economic effects on Participatory budgeting	Politics, socio-economics	Spada (2008)
Implementing participative research in German	Socio-economics	Mambrey (2008)
Improving public participation in politics (Obama case)	Information systems, socio-politics	Kes-Erkul & Erkul (2009)
Participatory approach for IS development in developing countries	Information systems	Puri et al. (2004)
Participatory approach in Decision making	Information systems	Byrne & Alexander (2006)
Increasing participation in online communities	Information systems	Bishop (2007)
Research Motivation: Instrumental justification		
Examples	Reference discipline	Reference
Proposed framework to evaluate e-Participation	Information systems	Aichholzer & Westholm (2009)
Assessment of e-Participatory budgeting	Information systems	www.participatorybudgeting.org.uk
Policy-making lifecycle and key dimensions of e-Participation	Public policy	Macintosh (2004)
Proposed framework for mapping, assessing and clustering e-Par projects in EU	Information systems	Charalabidis et al. (2008)
Overview e-Participation projects	Information systems	Ahmed (2006)
Proposed framework ICT tools for e-Par objectives	Computer Science	Phang & Kankanhalli (2008)
e-Governance as dynamic and open socio-technical systems	Information systems (socio-technical)	Dawes (2008)
Overview e-PB in Brazil	Socio-economics	Peixoto (2008)
Overview e-Participation domain	Information systems	Kalampokis et al. (2008)
e-Government usage assessment from socio-technical perspectives	Information systems (socio-technical)	Belanger & Carter (2009)
Proposed framework assessing e-par project and tool	Computer science	Tambouris et al. (2007)
Proposed multidimensional framework of wide participation and common understanding	Information systems	Wang & French (2008)
Improving GDSS and NSS	Computer science	Jinbaek (2008)
Research Motivation: Technology focus		
Examples	Reference discipline	Reference
e-administration and e-politics	Computer science	Grima-Izquierdo &Insua (2006)
Bridging the gap PD & IS methodology	Information systems, computer science	Pekkola et al. (2006)
Searching strategies for citizen-centered e-Government	Information systems	Bertot et al. (2007)
Improving citizen interaction in an e-Deliberation environment (city map online)	Information systems	De Cindio et al. (2008)
Overview e-Voting in the UK	Computer science	Xenakis & Macintosh (2003)
Improving e-Gov interoperability and consistency	Computer science	Davies et al. (2007)
Proposed draft of e-Par ontology	Information systems	Wimmer (2007)
Guide to online participation (social networking software)	Information systems	Sommer & Cullen (2009)
Online participation with social networks	Information systems	Hwang & Mohammed (2008)

Sanford and Rose [30] categorize e-Participation motivation into:

1. The participative imperative. This motivation emphasize in participation principle where stakeholders in society (citizens in various roles and stakeholder groupings) have an intrinsic right to participate in the formation and execution of public policy, especially when it involves their interests. This principle is derived by argumentation from principles in philosophy and political science, and is commonly protected in democratic societies by law, convention, and practice. However, the extent of this intrinsic right, the nature of the participation, and the democratic forms which enable it are open to debate. This research motivation can therefore be associated with a desire to understand, improve or reshape societal participation forms.
2. Instrumental justification. This motivation relate to the study of the effectiveness of government and policy making is, and how to improve it. Stakeholder participation in public affairs can be instrumental in more effective policy making and governance. This is because consultation with societal stakeholders can lead to improved public policies and encourage adoption and implementation of policy and services.
3. Technology focus. This motivation point out the role, effectiveness of information and communications technology to improve participation in the political process through: enhanced reach and range (inclusion); increased storage, analysis, presentation, and dissemination of contributions to the public policy and service debate; better management of scale; and by improvements to the process of organizing the public sphere debate.

However in practical implementation, those motivations often overlap with each other's, thus we present only the dominant motivation of the research paper as described in table 1.

In reference discipline related to e-Participation, Sanford and Rose [30] identify several related fields, i.e.: (1) information systems, (2) political science, (3) social and political philosophy, (4) social and economic philosophy, (5) public policy, (6) computer science. Further explanation of e-Participation can be found in [30]; [22]; [34]; [44]; [28]; [17].

Phang and Kankanhalli [28] in their paper stated that e-Participation initiatives have several objectives, such as: to provide citizens information, to support citizens' participation, to utilise citizen's input in decision making processes, and to investigate citizens' needs. Table 4 shows e-Participation initiatives categorization based on its objectives, i.e.: (1) information exchange, (2) education and support-building, (3) decision-making supplement, and (4) input probing. Our work focuses on how technology can support decision making processes as an e-Participation objective.

Although Non-government organization (NGO) and private sector are also relevant to e-Participation but basically e-Participation is a citizen focused research. Macintosh [22] has identified e-Participation as part of e-Democracy. E-Democracy is defined as ICT usage involving citizens, in order to support a democratic decision making process and to strengthen a representative democracy.

A democratic decision making process is divided into two categories, which are: e-Voting and e-Participation (Macintosh, 2004). Therefore, it can be concluded that technology involved in e-Participation should deliberate citizens' opinion, reform government's citizen and furthermore provide better government's service to its citizens.

In spite of several pro and contras regarding the prospects of ICT towards better governance and democracy, this paper believes that ICT can facilitate participation in decision making processes and also supports collective deliberation.

III. PROPOSED e-PARTICIPATION MODEL

As stated above, the main objective of this study is to support and facilitate wider participation using ICT (e-Participation) as decision-making supplement. For this reason, we propose a model utilize multi-criteria decision analysis (MCDA) aiming to assist decision makers on the entire process following a structured approach (specifically in process 1, figure 1); trust and reputation mechanisms to encourage and assist in knowledge sharing and education process (process 2, figure 1) also group decision making process (process 3, figure 1); and consensus with centrality approach to construct group decision (process 3, figure 1).

Moreover, the model consists of three main processes (figure 1), i.e. [38]:

1. Agenda setting process. This process mainly contains of clarification of the decision context, i.e.: problem and objective definition, criteria and alternatives exploration, and identification of participants.

2. Knowledge sharing and education process. This process involves iterative learning process by providing advice taking service. A decision maker can find and ask advisors using trust and reputation mechanisms, in order to improve her decision's quality. This process also aim to inform participants about the why and how decision is made from other participants' (e.g. experts) perspectives.

3. Group decision making process. This process includes evaluation of decision criteria and alternatives. All alternatives are measured with regard to every decision criterion using a related measurement scale. These evaluations based on subjective judgments by participants themselves. This process also supports participants' decisions iteration and refinement in order to reach consensus, facilitating by supra decision maker role.

Figure 1. The Proposed Model

Related to computer science, trust has been used in various fields, e.g. data/information filtering and collecting strategy [21], [13], [39], [40]; security mechanism [35], [6]; information/ knowledge sharing [19], [4]; and recommender system [42], [24], [46], [32]. Table 2 summarizes several literatures in trust related to computer science application.

As we can see from table 2, trust model has not been utilized in decision making area; while it has been used widely in data/ information collecting and filtering and recommender systems. Our effort is employed trust into our e-Participation development. Reputation and recommendation are the common terminologies used in trust model. Further explanation of how our trust model works within our decision making model can be found in [37] and [38].

For implementing our model into Par-GDSS application, we follow framework of ICT exploitation for e-Participation initiatives described by Phang and Kankanhalli [28]. The framework in table 4 shows there is no single participatory technique and ICT tool that can satisfy all four objectives of e-Participation initiatives. Therefore it is important to first identify the objective to be served by the initiative to improve the probability of success of an e-Participation initiative,

followed by a careful selection of the best-matching techniques and ICT tools for the objective [28].

TRUST MODEL AND APPLICATION USED IN COMPUTER SCIENCE

Aplication	Used terminology			Environment	Author
	Reputation	Recommendation	Others		
Data/ Information collecting and filtering	No	Yes	No	decentralized networks distributed systems	Lin et al. (2007)
	Yes	Yes	No	distributed systems	Umuhoza et al. (2008) O'Donovan & Smyth (2008)
	No	Yes	No		
	No	Yes	No	social networks distributed systems	Walter et al. (2007)
	No	Yes	No		Lifen (2008)
Security mechanisms	Yes	No	penalty evaluation	P2P networks	Tian et al. (2006) Bhanwar & Bawa (2009)
	Yes	Yes	No	grid environment	
Information/ knowledge sharing	Yes	No	risk evaluation	P2P networks	Liang & Shi (2007)
	No	No	No	web-based	Bencina (2007)
Recommender systems	Yes	Yes	No	P2P networks distributed systems ubiquitous environment	Wang & Vassileva (2003)
	No	Yes	No		Montaner et al. (2002)
	No	Yes	No		Yuan et al. (2006)
	No	Yes	No	P2P networks	Song et al. (2004)

Our paper is oriented toward e-Participation efforts with the decision-making supplement objective aim to extract specific information from citizens, such as, citizens' preferences on the use of an empty plot of land in the city. In these e-Participation efforts, citizens are given an opportunity for direct input into the planning process, as mentioned by Phang and Kankanhalli (2008).

Therefore our model provides characteristics, which are: control of participation processes, interactions among planners and participants, mechanisms for data collection. Those characteristics aim to support decision-making supplement as e-Participation objectives.

Related to our e-Participation application (namely as Par-GDSS), we construct and integrate trust model into our proposed decision making model. Table 3 describes trust properties in our model and application. The user as actors in the model consists of four different level, categorized by its trust level (tl) which are:

- Citizen with trust level=1;

- Citizen with trust level=2;
- Citizen with trust level=3;
- Citizen with trust level=4.

TRUST PROPERTIES WITHIN THE PROPOSED MODEL

User/ Actor	Trust Properties	Environment
Citizen, i.e.: • Regular citizen (tl=1 dan tl=2) • Expert citizen (tl=3 dan tl=4)	• Always between two entities (Trustee dan Truster) • Non-symmetrical (e.g.: if A trust B, then doesn't mean B trust A) • Conditionally transitive (e.g.: if A trust B, B trust C, then A trust C under condition that B as a recommender to C, based on C's reputation). • Context-based (a trustee only can be trusted in one/ several specific categories) • Reflexively (a truster has to trust system and his/ her own judgment to determine trustee's trust value) • Dynamically (trust value is dynamically changed).	Trust • Decentral peer • Partial trust algorithm

Environment in this trust model consists of decentral peer and partial trust algorithm. Decentral Peer in this model works for all of the peers within the network can act as end-peer or recommender for other peers.

This is a simple but the least reliable model since there is no need for any recommenders in the system to prove the recognition of its recommendation. While partial trust algorithm applies when a peer uses only some of the recommendations to measure trust value of another peer [25].

For the next step, we then implement our model into an application named Par-GDSS. Par-GDSS is a web based application which can be reached online at www.pargdss.com. We follow a three-step procedure proposed by Phang and Kankanhalli [28] as presented in figure 2.

We define Par-GDSS objective as a decision-making supplement. The participatory technique used is group decision making with consensus achievement. Consensus achievement is produced by the role of Supra decision maker as the leader of the group.

Finally, we then decide Group Decision Support Systems (GDSS) with trust and reputation mechanisms as the ICT tool to support the technique and, in turn, the objective.
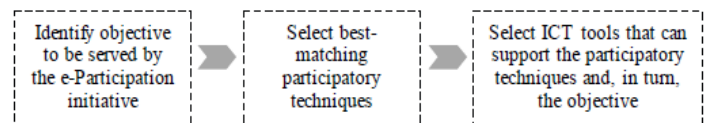
Figure 2.   A three-step procedure to implement e-Participation [28]

TABLE I.   A FRAMEWORK OF ICT EXPLOITATION FOR ICT INITIATIVES [28]

E-Participation Initiatives				
E-Participation Objectives				
	Information exchange	**Education and Support-building**	**Decision-making supplement**	**Input Probing**
Features Desired	*Avenue for citizens to participate freely, Opportunity for two-way open exchange of information*	*Formal selection and engagement of participants from target demographic population, Maintenance of contact with participants over time*	*Control of participation processes, Mechanisms to obtain specific set of information useful for decision making (e.g., ranked preferences) from relevant group*	*Mechanisms for systematic collection and analysis of citizens' input, Collection of citizens' opinion that is neither biased nor influenced by others*
Example(s) of Best-fitting Participatory Techniques	*Drop-in center, Public hearing*	*Citizen advisory committee, Citizen panel*	*Nominal Group Process, Value analysis*	*Citizen survey*
*Characteristics** Restriction on number and target of participants	**No**	**Yes**	Yes	Depends on sample size needed and resources available
Control of participation processes	No	No	**Yes**	No
Maintenance of long-term contact	No	**Yes**	No	No
Interaction among planners and participants	**Yes**	Yes	Limited (mainly one direction)	**No**
Mechanisms for data collection	No	No	**Yes**	**Yes**
Statistical analysis of data	No	No	No	**Yes**
ICT Tools that Provide the Features Desired	*Web portal with Online discussion forum, Online chat*	*Electronic profiling, Online chat, Discussion forum with login feature, Teleconferencing, Videoconferencing, E-mail*	*Group support systems with process restrictiveness feature, Online pair-wise structured survey, Visualization tools*	*Online survey questionnaire, Web comment form, Data analysis tools*
Example of E-Participation Initiatives	*Denmark's Democracy on the Web, Singapore's REACH Portal*	*UK's Askbristol E-Panel, Canadian's Youth Connection Forum*	*Netherlands' Almere Co-Production of Interactive Policy and Technological Policy Solution for Societal Problems*	*Sweden's Kalix Consultation*

IV. PAR-GDSS: E-PARTICIPATION APPLICATION

The architecture of the GDSS (named as Par-GDSS) employs M (Model), V (View), C (Controller) design pattern, which adopted from software engineering approach. MVC separates web application's logic from layer presentation so then it has components to manipulate data, user interface, and application control.

Figure 2 describes the systems architecture of Par-GDSS. At the user side, there are many functionality elements (such as: create account, view event, etc) that could be done by user/ citizen with various role, i.e.: (admin, citizen with tl=1, citizen with tl=2, citizen with tl=3, citizen with tl=4).

There are several modules have been developed for Par-GDSS, i.e.:

- *Citizen*: This module handles all requests related with
(1) *retrieve, add, delete, update Citizen*;

(2) trust and reputation mechanisms and theirs calculation.

- *Event*: This module handles all requests related with
(1) *retrieve, add, delete, update Event*;

(2) individual decision;

(3) trust and reputation mechanisms and theirs calculation;

(4) document and information related with an *Event*;

(5) control and *feedback mechanisms*.

- *Dashboard*: This module handles all information related with an *event* involving a *citizen* (*user*), e.g.: received message (inbox), newest comment, and upload document.

- Modul Forum: This module handles all activities related with Discussion Forum, e.g.: *post topic, post thread, post discussion, view discusssion*.

- Modul *Message*: This module handles all activities related with sending and receiving messages between users (*one to one user*).

- Modul *Session:* This module handles *user session*.

Par-GDSS is a web based application which can be accessed through a login page, where users have to provide appropriate password. The system recognizes two roles: Supra decision maker and citizen.

Supra decision maker works on a fully functional mode of the system, while citizen work on a mode presenting a subset of functionality based on their trust level. Figure 3 shows the functionality for each citizen.
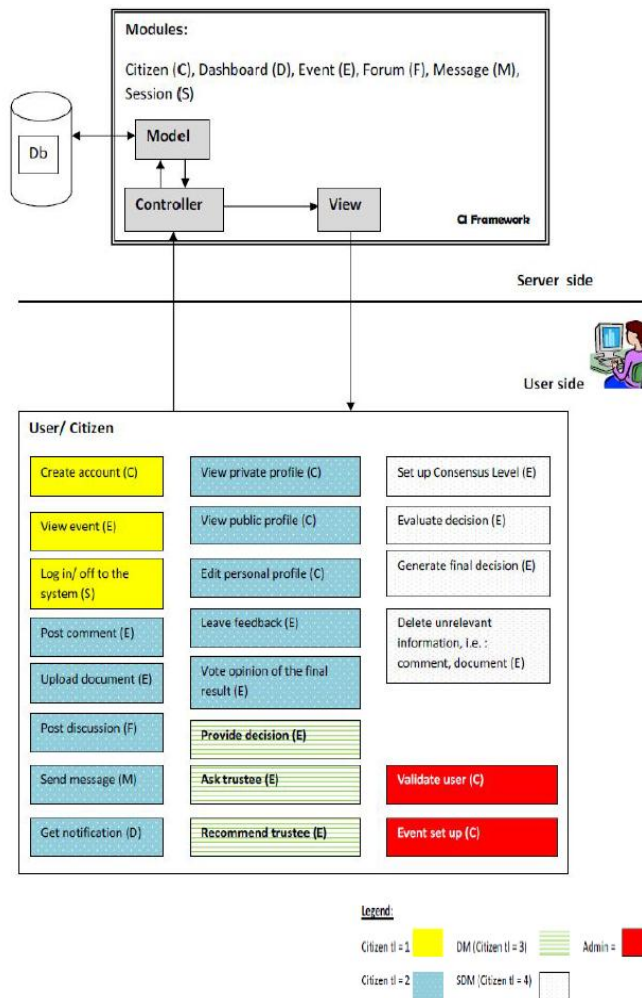
Figure 3.   Par-GDSS Overall Architecture

mechanism utilization is coming from the possibility of any doubtful has been taken by decision maker. Unlike expert decision makers, citizens as decision makers could have no prior information regarding what and how decision should be made of.
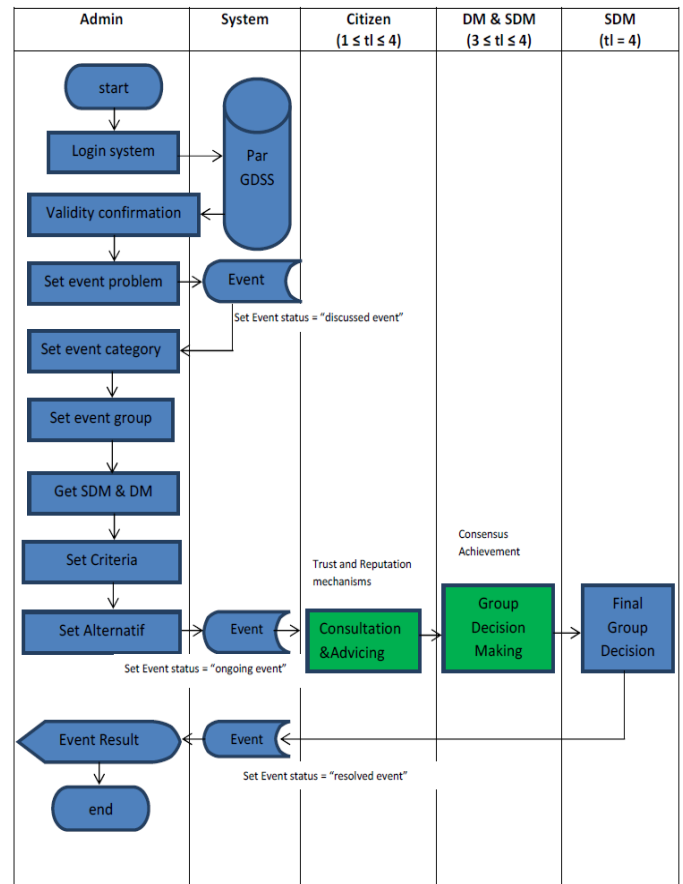
Figure 4.   Par-GDSS Features and Activities Diagram

Figure 4 shows the beginning of the systems, i.e. login to system, until the end of the system, i.e. view the result and provide feedback. The event in Par-GDSS changes several times related to the decision making progress, which consists of:

- Discussed event: this feature is model's interpretation of Agenda Setting process.
- Ongoing event: this feature is model's interpretation of Knowledge Sharing and Education process.
- Resolved event: this feature is model's interpretation of Group Decision Making process.

For every event changes, we provide two kinds of mechanisms, i.e.: trust and reputation mechanisms, consensus achievement mechanism.

V.   CONCLUSION

Our literatures survey finds that e-Participation in decision making area is still lack of research and implementation. Therefore, this paper proposes an e-Participation model using trust to assist participatory decision making, while enhances collective deliberation among citizens. The idea of trust

This paper then proposes to adopt the 'real life' behavior, by providing interaction and consultation between citizens in order to produce better decision's quality. Focusing on the process of making decisions on public matters with citizen participation, this paper also investigates state-of-the art researches related to e-Participation.

The advantages of our proposed approach are:

(1) it provides a transparent control mechanisms by employing trust and reputation mechanisms, so that it can promote democracy aspects and not only works as group decision support systems,

(2) it can support various decision makers with different backgrounds and skills by providing information and knowledge through trust and reputation mechanisms (a beginner decision maker can learn how to decide from other trusted and recommended sources),

(3) it can enhance participation level to some extent by providing wider participation access and resources and also supports better interaction among decision maker (learning and sharing process to achieve final group decision). Our

future works will be concentrated on our proposed model testing and evaluation.

REFERENCES

[1] N. Ahmed, An Overview of e-Participation Models, "United Nations Report (Division for Public Administration and Development Management DPADM/ Department of Economic and Social Affairs UNDESA)", April 2006.

[2] G. Aichholzer, H. Westholm, "Evaluating eParticipation Projects: Practical Examples and Outline of an Evaluation Framework", *European Journal of ePractice*, www.epracticejournal.eu, vol. 1(7), March 2009.

[3] F. Belanger, L. Carter, "The Impact of The Digital Divide on e-Government Use", *Communications of The ACM*, vol. 52 (4), April 2009, pp. 132-135.

[4] J. Bencina, "Web-based Decision Support System for the Public Sector Comprising Linguistic Variables", *Informatica*, vol. 31, 2007, pp. 311–323.

[5] J.C. Bertot, P.T. Jaeger, C.R. McClure, "Citizen-centered E-Government Services: Benefits, Costs, and Research Needs", *Proceeding of the 9th Annual International Digital Government Research Conference 2008,* 2008, pp. 137-141.

[6] S. Bhanwar, S. Bawa, "Establishing and Evaluating Trust in a Grid Environment", in *10th International Symposium on Pervasive Systems, Algorithms, and Networks*, Kaohsiung, Taiwan, December 14-16, 2009.

[7] J. Bishop, 2007, "Increasing participation in online communities: A framework for human–computer Interaction", *Journal of Computers in Human Behavior*, vol. 23(4), July 2007, pp. 1881–1893.

[8] E. Byrne, P.M. Alexander, "Questions of Ethics: Participatory Information Systems Research in Community Settings", *Proceedings of SAICSIT 2006 PDC 2006 - the ninth Participatory Design Conference 2006, Trento, Italy,* 1-5 August, 2006, pp. 117-126.

[9] Y. Charalabidis, L. Kipenis, S. Koussouris, J. Psarras, "Multi-faceted Analysis of Electronic Participation Projects at European Level: A Systemic Approach", www.ep-momentum.eu, *accessed on 12 Oktober 2009.*

[10] J. Davies, T. Janowski, A. Ojo, "Technological Foundations of Electronic Governance", *Proceedings of International Conference on Theory and Practice of Electronic Government (ICEGOV) 2007*, Macao, December 10-13, 2007, pp. 5-11.

[11] S.S. Dawes, "Governance in the information age: a research framework for an uncertain future", *Proceedings of the 9th Annual International Digital Government Research Conference,* Montreal, Canada, May 18, 2008, pp. 290-297.

[12] F. De Cindio, C. Peraboni, L. Sonnante, "Improving citizens' interactions in an e-deliberation Environment", *Proceedings of Advances Visual Interfaces (AVI) 08*, Napoli, Italy, May 28-30, 2008, pp. 486-487, ACM Digital Library.

[13] J. Donovan, B. Smyth, "Trust No One: Evaluating Trust-based Filtering for Recommenders", *Proceedings of the 19th International Joint Conference on Artificial Intelligence IJCAI*, 2005.

[14] C. Grima-Izquierdo, D. Rios Insua, "Designing a General Architecture to Support eGovernment", http://www.publicsphereproject.org/events/diac08/proceedings/10.eGove rnment Architecture.Grima-Izquierdo_and_Rios_Insua.pdf, 2006, accessed on 8 Desember 2009.

[15] J. Hwang, A.B. Mohammed, "Approaching e-democracy: A Case Study Analysis from Jordan", Proceedings of the 3rd 2008 International Conference on Convergence and Hybrid Information Technology, 2008.

[16] K. Jinbaek, "A Model and Case for Supporting Participatory Public Decision Making in e-Democracy", *Journal of Group Decision Negotiation*, vol. 17, pp. 179–193, May 2008, Springer Science & Business Media B.V.

[17] E. Kalampokis, E. Tambouris, K. Tarabanis, "A Domain Model for eParticipation", *Proceedings of The Third International Conference on Internet and Web Applications and Services (ICIW)*, Athens, Greece, June 8-13, 2008, pp, 25-30, IEEE Digital Library.

[18] Kes-Erkul, R.E. Erkul, "Web 2.0 in the Process of e-participation: The Case of Organizing for America and the Obama Administration", *US NCDG (National Center for Digital Government) 2009 Working Paper No. 09-001*, 2009.

[19] Z. Liang, L. Shi, "PET: A Personalized Trust Model with Reputation and Risk Evaluation for P2P Resource Sharing", *Proceedings the 38th Annual Hawaii International Conference on System Sciences*, January 03-06, 2005.

[20] L. Lifen, "Trust Derivation and Transitivity in a Recommendation Trust Model", *Proceedings in International Conference on Computer Science and Software Engineering*, 2008.

[21] Z. Lin, X. Feng, W. Yuan, L. Jian, "A Semantic and Time Related Recommendation-Feedback Trust Model", *Proceedings of Second International Conference on Availaibility, Reliability, and Security (ARES 2007)*, Vienna, Austria, April 10-13, 2007, pp. 765-774, IEEE Digital Library.

[22] A. Macintosh, "Characterizing E-Participation in Policy-Making", *Proceedings of the 37th Hawaii International Conference on System Sciences – 2004 Track 5,* Hawaii USA, January 5-8, 2004, pp. 1-10.

[23] P. Mambrey, "From Participation to e-Participation: The German Case", *Proceedings of the 2nd International Conference on Theory and Practice of Electronic Government (ICEGOV2008)*, Cairo, Egypt, December 1-4, 2008, pp. 355-360.

[24] M. Montaner, B. L´opez, B., J.L. de la Rosa, "Opinion-based Filtering through Trust", *Proceedings of CIA'02 LNAI 2446*, 2002, pp. 164-178.

[25] I. Nurtanti, "Trust Algorithm in Files Sharing P2P Network", Master Thesis, Faculty of System Engineering, Policy Analysis and Management, Delft University of Technology, Delft, the Netherlands, 2008.

[26] T. Peixoto, "*e-Participatory Budgeting: e-Democracy from theory to success'*", e-Working Papers 2008, e-Democracy Centre, Universität Zürich, p. 4. Available at http://www.edemocracycentre.ch/files/Tiago.pdf , accessed on January 19, 2009.

[27] S. Pekkola, N. Kaarilahti, P. Pohjola, "Towards Formalised End-User Participation in Information Systems Development Process: Bridging the Gap between Participatory Design and ISD Methodologies", *Proceedings Participatory Design Conference*, vol. 1, Trento, Italy, August 1-5, 2006, pp. 21-30.

[28] C.W. Phang, A. Kankanhalli, "A Framework of ICT Exploitation for E-Participation Initiatives", *Communications of the ACM*, December 2008, 51(12), pp. 128-132.

[29] S.K. Puri, E. Byrne, J. Nhampossa, Z.B. Leopoldo, Quraishi, "Contextuality of Participation in IS Design: A Developing Country Perspective", *Proceedings of the 8th Participatory Design Conference 2004*, vol. 1, Toronto, Canada, July 27-31, 2004, pp. 42-52, ACM Digital Library.

[30] C.Sanford, J. Rose, J., "Characterizing eParticipation", *International Journal of Information Management*, vol. 27(6), 2007, pp. 406-421.

[31] L. Sommer, R. Cullen, "Participation 2.0: a Case Study of e-Participation within the New Zealand Government", *Proceedings of the 42nd Hawaii International Conference on System Sciences – 2009, Hawaii, USA, January 5-8,* 2009.

[32] W. Song, V.V. Phoha, X. Xin, "An Adaptive Recommendation Trust Model in Multiagent System", *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'04),* 2004.

[33] P. Spada, "The Economic and Political Effects of Participatory Budgeting", Yale University, 2009.

[34] E. Tambouris, N. Liotas, K. Tarabanis, "A Framework for Assessing eParticipation Projects and Tools", *Proceedings of the 40th Hawaii International Conference on System Sciences,* Hawaii, USA, January 3-6, 2007, pp. 1-10.

[35] C.Tian, S. Zhou, W. Wang, S. Cheng, "An Efficient Attack-Resistant Trust Model for P2P Networks", *International Journal of Computer Science and Network Security*, vol. 6(11), 2006.

[36] V. Tundjungsari, J.E. Istiyanto, P. Santoso, "Building Public Trust through Public Participation Using e-Governance", *Proceedings of Rural International Conference on ICT*, 2009, ITB, Bandung, Indonesia, June 17-18, 2009, pp. 155-162.

[37] V. Tundjungsari, J.E. Istiyanto, E. Winarko, R. Wardoyo, "A Trust based Reputation Model to Seek Judgment and Advice in Participatory Decision Making", *Proceedings of the 2nd International Conference on Distributed Framework and Application (DfMA)*, Yogyakarta, Indonesia, August 2-3, 2010, IEEE Digital Library.

[38] V. Tundjungsari, J.E. Istiyanto, E. Winarko, R. Wardoyo, "Enhancing Participation Process in Public Decision Making with MCDA and Trust Modeling", *International Journal of Computer Science Issues*, Vol. 8, Issue 1, January 2011.

[39] D. Umuhoza, J.I. Agbinya, D. Moodley, A. Vahed, "A Reputation based Trust Model for Geospatial Web Services", *Proceedings of 1st WSEAS International Conference on Environmental and Geological Science and Engineering (EG'08)*, Malta, September 11-13, 2008, http://ict4eo.meraka.csir.co.za/ict4eo/publications-and-research-outcomes/documents/A_reputationbased_trustModel_for_Geospatial_WebServices.pdf.

[40] F.E. Walter, S. Battiston, F. Schweitzer, "A Model of a Trust-based Recommendation System on a Social Network", *Journal of Autonomous Agents and Multi-Agent Systems*, vol.16(1), 2007, pp.57-74.

[41] W. Wang, S. French, "A Multi-Dimensional Framework for Facilitating Wide Participation and Common Understanding", *Proceedings of WebScience'08*, Pittsburgh, Pennsylvania, USA, June 19, 2008, http://users.ecs.soton.ac.uk/dem/workshops/webscience08/papers/websci07-wang.pdf.

[42] Y. Wang, J. Vassileva, "Trust and Reputation Model in Peer-to-Peer Networks", *Proceedings of the 3rd International Conference on Peer-to-Peer Computing,* 2003, p.150.

[43] J. Watkins, J.A. Tacchi, "ICT for Creative Engagement : Participatory Design and Ethnographic Action Research". In: *HCI for Community and International Development Workshop*, Florence, Italy, 5-6 April 2008, http://eprints.qut.edu.au/17933/1/c17933.pdf.

[44] M.A. Wimmer, "Ontology for an e-Participation Virtual Resource Centre", *Proceedings of the 1st International Conference on Theory and Practice of Electronic Government (ICEGOV2007)*, Macao, December 10-13, 2007, pp. 89-98.

[45] A. Xenakis, A. Macintosh, "E-electoral Administration: Organizational Lessons Learned from the Deployment of E-voting in the UK", *Proceedings of the 2005 National Conference on Digital Government Research (dg.O),* vol.89, Atlanta, Georgia USA, May 15-18, 2005, pp. 191-197.

[46] W. Yuan, D. Guan, S. Lee, Y. Lee., A. Gavrilov, "Finding Realiable Recommendations for Trust Model", *Lecture Notes in Computer Science*, vol. 4255, 2006, pp. 375-386, Springer-Verlag Publisher.

[47] www.participatorybudgeting.org,uk, accessed on 12 October 2009.

Algorithm design for a supply chain equilibrium management model

Peimin Zhao

Feixian School
Linyi University
Feixian, Shandong, P.R.China

Abstract—**In this paper, we consider a complementary model for the equilibrium management of supply chain. In order to give an optimal decision for the equilibrium management, we propose a new algorithm based on an estimate of the error bound. This algorithm requires neither the existence of a non-degenerate solution nor the non-singularity of the Jacobian matrix at the solution. We also prove the quadratic convergence of the given algorithm. It can be viewed as extensions of previously known results.**

Keywords- Supply chain equilibrium management; complementary model; algorithm; quadratic convergence.

I. INTRODUCTION

Supply chain management involves many aspects, including manufacturing, transportation, logistics and sales markets, etc. Manufacturers produce a commodity and then transport it to sellers. The manufacturers need to pay for the costs of production and transportation. In order to make a profit, they have to determine optimal production and transportation. Sellers need to pay for the costs in stores, and maximize their profits from the commodity price. Commodity price is the price customers are willing to pay. The customers need to determine their optimal level of consumption, which is related to commodity prices at different stages of the supply chain and the commodity handling costs. The supply chain management problem can be abstracted as a "complementary" relationship between the two sets of decision variables, expressed as a complementary model ([1,2,3]).

Let mapping $f : R^n \to R^n$, be continuously differentiable. Find vector $x^* \in R^n$, such that

$$x \geq 0, f(x) \geq 0, x^T f(x) = 0, \qquad (1)$$

We use X^* to denote the solutions of (1)，and assume that X^* is not empty.

The model takes into consideration different behaviors and their interactions of the various participants, and provides a standard to assess the appropriate commodity prices and quantities. It is an equilibrium optimization problem. At its core is the use of mathematical methods, with computer and network as tools, to study various complex systems having this complementary relationship and their corresponding solutions. It provides scientific basis for decision makers, with the goal of ultimately achieving balanced and harmonious development of complex system operation. It is widely used nowadays. In recent years it has become a hot topic of operational research and management studies. Many researchers, especially scholars abroad, are increasingly interested, and great progress has been made in both theoretical research and practical applications. To solve (1), many algorithms have been proposed ([4, 5, 6, 7, 9]). The basic idea of these algorithms is to transform (1) into an unconstrained optimization problem, or into a simple constrained optimization problem ([4, 7, 9]), and then to use the Newton type algorithms or confidence region algorithms. Although the convergence and the corresponding rate of convergence are established, the conditions are mostly quite strong. Most of these algorithms require that the Jacobian matrix at the solution of (1) is non-singular or there exists a non-degenerate solution. As we know, the non-singularity of the Jacobian matrix at the solution of (1) ensures that the application of the well-known Levenberg-Marquardt (L-M) algorithm for solving (1) has the property of quadratic convergence ([7]). Recently, Yamashita, and Fukushima have proved that under certain conditions on the local error bound, the L-M algorithm for solving nonlinear equations also has the property of quadratic convergence. These conditions on local error bound are weaker than the non-singularity of the Jacobian matrix ([8]). This motivates us to transform (1) into a nonlinear equation, and to establish the error bound for (1). With the establishment of an error bound of a residual function from (1), we propose a new algorithm and prove its quadratic convergence, thereby overcoming the drawbacks of the above mentioned algorithms on convergence.

II. PRELIMINARIES

In this section, we mainly give a result on the error bound for (1) ([10]), which play a key role in the establishment of the nonlinear equation and in the proof of the convergence of the proposed algorithm.

Assumption 1 Let $f : R^n \to R^n$ be a uniform P function, i.e., there exists a constant $\mu > 0$ such that

$$\max_{1 \leq i \leq n}\{(x - y)_i [f(x) - f(y)]_i\} \geq m\| x - y \|^2, \forall x, y \in R^n.$$

Theorem 1 Assume that the mapping $f : R^n \to R^n$ in (1) is a uniform P function, and is Lipschitz continuous. Then there exist constant $\eta_1 > 0$, and $x^* \in X^*$ for any $x \in R^n$, such that

$$\| x - x^* \| \leq \eta_1 \| \min\{x, f(x)\} \|. \qquad (2)$$

III. ALGORITHMS AND CONVERGENCE

In this section, we first transform (1) into a nonlinear equation, and then give another type of error bound for (1). Based on this, we give the new algorithm for solving (1) and establish its property of quadratic convergence, overcoming the drawbacks of requiring the non-singularity of the Jacobian matrix at the solution or the existence of a non-degenerate solution.

We transform（1）into a nonlinear equation via Fischer function ([11]) $\phi: R^2 \to R^1$ $\phi(a,b) = \sqrt{a^2+b^2} - a - b, a,b \in R$.

This function has the property $\phi(a,b) = 0 \Leftrightarrow a \geq 0, \ b \geq 0$, $ab = 0$. For any $(a,b) \in R^2$, Fischer function also has the following property ([12])

$$(2-\sqrt{2})|\min\{a,b\}| \leq |\phi(a,b)| \leq (2+\sqrt{2})|\min\{a,b\}|. \quad (3)$$

For any vectors $a,b \in R^n$, we define the following vector-valued function $\Phi(a,b) = (\phi(a_1,b_1), \phi(a_2,b_2), \cdots, \phi(a_n,b_n))^T$, (4)

Where $a = (a_1, a_2, \cdots, a_n)^T, b = (b_1, b_2, \cdots, b_n)^T$. Obviously,

$$\Phi(a,b) = 0 \Leftrightarrow a \geq 0, b \geq 0, a^T b = 0.$$

With (2), (3) and Theorem 1, we can easily prove the following result.

Theorem 2 If the mapping $f: R^n \to R^n$ in (1) is a uniform P function, and is Lipschitz continuous，then there exist constant $\eta_2 > 0$, and $x^* \in X^*$ for any $x \in R^n$, such that $\|x - x^*\| \leq \eta_2 \Phi(x, f(x))$.

Function（4）is not smooth. In order to give a smooth algorithm with the property of quadratic convergence, we give also the following smooth Fisher-Burmeister function $\varphi_\tau: R^2 \to R^1$ $\varphi_\tau(a,b) = \sqrt{a^2 + b^2 + 2\tau^2} - a - b$, where $\tau > 0$ is a smooth parameter. For ease of presentation, let

$$\Theta_\tau(y,z) = (\varphi_\tau(y_1, z_1), \varphi_\tau(y_2, z_2), \cdots, \varphi_\tau(y_n, z_n))^T \in R^n,$$

We define the following vector-valued function

$$F: R^n \times (0,\infty) \to R^n \times (0,\infty), \quad F(x,\tau) := \begin{pmatrix} \Theta_\tau(x, f(x)) \\ \tau \end{pmatrix} \quad (5)$$

and real-valued function $f: R^{n+1} \to R$

$$f(x,\tau) := F(x,\tau)^T F(x,\tau) = \|F(x,\tau)\|^2. \quad (6)$$

Obviously, $x^* \in X^* \Leftrightarrow (x^*, 0)$ is a solution of $F(x,\tau) = 0$. In the following, we give the algorithm for solving

$F(x,\tau) = 0$, and denote the solutions of $F(x,\tau) = 0$ by Ω^*.

First, we give some properties of function $\varphi_\tau(a,b)$ ([13], [14]).

Lemma 1 Function $\varphi_\tau(a,b)$ has the following properties：

(i) Function $\varphi_\tau(a,b)$ is continuously differentiable on $R^2 \times (0,\infty)$；function $\varphi_\tau(a,b)$ is strongly semi-smooth on $(a,b,\tau) \in R^2 \times [0,\infty)$ that is,

$$\varphi_{(\tau+\Delta\tau)}(a+\Delta a, b+\Delta b) - \varphi_\tau(a,b) - V^T(\Delta a, \Delta b, \Delta \tau) = O(\|(\Delta a, \Delta b, \Delta \tau)\|^2)$$

Where $V \in \partial\varphi_{(\tau+\Delta\tau)}(a+\Delta a, b+\Delta b)$, and $(\Delta a, \Delta b, \Delta \tau) \to 0$, $\partial\varphi$ is the Clarke generalized gradient of φ ([15]).

(ii) For any $(a,b,\tau) \in R^2 \times (0,\infty)$, we have

$$|\varphi_0(a,b) - \varphi_\tau(a,b)| \leq \sqrt{2}\tau.$$

For (5), using Lemma 1, we have the following result.

Theorem 3 Function $F(x,\tau) = 0$ has the following properties:

(i)Function $F(x,\tau)$ is continuously differentiable on $R^n \times (0,\infty)$ and is locally Lipschitz continuous and strongly semi-smooth, i.e., for any $(x,\tau) \in R^n \times [0,\infty)$, there exist $L_1 > 0, L_2 > 0$ and $b_1 > 0$ such that

$$\|F(x+\Delta x, \tau+\Delta\tau) - F(x,\tau)\| \leq L_1 \|(\Delta x, \Delta\tau)\|, \quad (7)$$

$$\|F(x+\Delta x, \tau+\Delta\tau) - F(x,\tau) - V^T(\Delta x, \Delta\tau)\| \leq L_2 \|(\Delta x, \Delta\tau)\|^2,$$
$$(8)$$

$$\forall(\Delta x, \Delta\tau) \in N(0,b_1) := \{(\Delta x, \Delta\tau) \in R^n \times [0,\infty) | \|(\Delta x, \Delta\tau)\| \leq b_1, \tau + \Delta\tau \geq 0\},$$

Where $V \in \partial F(x+\Delta x, \tau+\Delta\tau)$, $\partial F(x,\tau)$ is the Clarke generalized gradient of $F(x,\tau)$.

(ii) From Theorem 2, for a solution $(x^*, 0) \in \Omega^*$, there exist a neighbourhood $N((x^*, 0), b_2) \in \Omega^*$ of $(x^*, 0)$ and a constant $c_1 > 0$, such that for any $(x,\tau) \in N((x^*, 0), b_2)$, we have

$$dist((x,\tau), \Omega^*) \leq c_1 \|F(x,\tau)\|, \quad (9)$$

Where $N((x^*, 0), b_2) := \{(x,\tau) | \|(x,\tau) - (x^*, 0)\| \leq b_2, \tau \geq 0\}$.

Proof . (i) is a direct result of Lemma 1.

(ii) From Theorem 2, there exists a constant $1 > b_2 > 0$, such that $dist(x, X^*) \leq \eta_3 \|\Psi(x)\|, \forall x \in N(x^*, b_2)$, where $N(x^*, b_2) := \{x \in R^n | \|x - x^*\| \leq b_2\}$, and there exists $\bar{x} \in X^*$, such that $dist(x, X^*) = \|x - \bar{x}\|$, From Lemma 1(ii), we have

$$\left| \|\Theta_0(x, f(x))\| - \|\Theta_\tau(x, f(x))\| \right| \leq \|\Theta_\tau(x, f(x)) - \Theta_0(x, f(x))\|$$

$$= (\sum_{i=1}^s (\varphi_\tau(x_i, f(x)_i) - \varphi_0(x_i, f(x)_i)^2)^{\frac{1}{2}} \leq \sqrt{2n}\tau.$$

Because of this, for any $(x,\tau) \in N((x^*, 0), b_2)$ we have

$$dist(x, X^*) = \|x - \bar{x}\| \leq \|(x,\tau) - (\bar{x}, 0)\| \leq \|x - \bar{x}\| + \tau \leq \eta_3 \|\Phi(x)\| + \tau$$

$$= \eta_3 \parallel \Theta_0(x, f(x)) \parallel + \tau \; \leq \eta_3(\parallel \Theta_\tau(x, f(x)) \parallel + (\sqrt{2n}\eta_3 + 1)\tau$$

$$\leq (\sqrt{2n}\eta_3 + 1)(\parallel \Theta_\tau(x, f(x)) \parallel + \tau)$$

$$\leq (\sqrt{2n}\eta_3 + 1)(\parallel \Theta_\tau(x, f(x)) \parallel_1 + \tau) \leq (\sqrt{2n}\eta_3 + 1) \parallel F(x, \tau) \parallel_1$$

$$\leq \sqrt{n+1}(\sqrt{2n}\eta_3 + 1) \parallel F(x, \tau) \parallel.$$

In the following we give a smooth Levenberg-Marquardt algorithm for solving (1).

Algorithm：

Step 1: Select an initial point $x^0 \in R^n$, and a parameter

$\tau^0 > 0 \quad \varepsilon > 0$. Let $k = 0$.

Step 2: If $\parallel \nabla f(x^k, \tau^k) \parallel \leq \varepsilon$, stop; otherwise, go to Step 3.

Step 3: Choose H^k, where H^k is the Jacobian matrix of $F(x^k, \tau^k)$, $\mu^k = \parallel F(x^k, \tau^k) \parallel^2$. Let $d^k = (\Delta x^k, \Delta \tau^k) \in R^{n+1}$ be the solution of the following strictly convex quadratic programming

$$\min \quad \parallel F(x^k, \tau^k) + H^k d \parallel^2 + \mu^k \parallel d \parallel^2$$

$$s.t. \qquad |\nabla \tau| \leq \frac{1}{1+\mu^k} \tau^k$$

Step 4: Let $x^{k+1} := x^k + \Delta x^k, \tau^{k+1} := \tau^k + \Delta \tau^k, \; k := k+1$, go to Step 2.

In the following convergence analysis, we assume that Algorithm 1 generates an infinite sequence. From (7)-(9), combined with (14) − (15), and the proof of Theorem 2.1 ([8]), we can obtain the quadratic convergence rate of Algorithm 1.

Theorem 4 Let $\{(x^k, \tau^k)\}$ be an infinite sequence generated by Algorithm 1. If (x^0, τ^0) is close to $(x^*, 0)$ sufficiently, then $dist((x^k, \tau^k), \Omega^*)$ converges to 0 quadratically, and sequence (x^k, τ^k) converges quadratically to $(x^*, 0) \in \Omega^*$ I $N((x^*, 0), b_2 / 2)$.

IV. CONCLUSION AND PROSPECT

In this paper, we present a new algorithm for solving the equilibrium management model of supply chain, and also have showed that method has the quadratic convergence, the conditions guaranteeing the quadratic convergence in this paper are weaker than those in the existing resolving methods in [7], since it does not require the existence of a non-degenerate solution of (1), nor the non-singularity of the Jacobian matrix at the solution of (1),

It is uncertain whether we have that the results in this paper for the functions which contain the P functions discussed in this paper, this is a topic for further research.

REFERENCES

[1] M.C.Ferris and J.S. Pang. Engineering and economic applications of complementarity problems. Society for industrial and applied mathematics, 1997, 39(4), 669-713.

[2] A. Nagurney, J. Dong, D. Zhang. A supply chain network equilibrium model. Transportation Research, 2002, Part E 38, 281.

[3] L.P. Zhang. A nonlinear complementarity model for supply chain network equilibrium. Journal of Industrial and Managment Optimization, 2007, 3(4), 727-737.

[4] R. Andreani, A. Friedlander and S.A. Santos. On the resolution of the generalized nonlinear complementarity problem. SIAM Journal on Optim., 2001, 12, 303-321.

[5] F. Facchinei and J.S. Pang. Finite-Dimensional Variational Inequality and Complementarity Problems. Springer, New York, 2003.

[6] C. Kanzow and M. Fukushima. Equivalence of the generalized complementarity problem to differentiable unconstrained minimization. J. Optim. Theory and Appl., 1996, 90(3),581-603.

[7] Y.J. Wang, F.M. Ma and J.Z. Zhang. A nonsmooth L-M method for solving the generalized nonlinear complementarity problem over a polyhedral cone. Appl. Math. Optim., 2005, 52(1), 73-92.

[8] N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg-Marquardt method. Computing [Suppl]}, 2001, 15, 239-249.

[9] X.Z. Zhang, F.M. Ma and Y.J. Wang. A Newton-type algorithm for generalized linear complementarity problem over a polyhedral cone. Appl. Math. Comput., 2005, vol.169, 388-401.

[10] N.H. Xiu and J.Z. Zhang. Global projection-type error bound for general variational inequlities. J. Optim. Theory Appl., 2002, 112(1), 213-228.

[11] A. Fischer. A special Newton-type optimization method. Optim., 1992, 24, 269-284.

[12] P. Tseng. Growth behavior of a class of merit function for the nonlinear complementarity problem. J. Optim.Theory Appl., 1996, 89, 17-37.

[13] S. Engelke, C. Konzow. Improved smoothing type method for the solution of linear programs. Preprint, Institute of Applied Mathematics, University of Hamburg, Hamburg, March, 2000.

[14] L. Qi, D. Sun. Smoothing function and smoothing Newton method for complmentarity problems and variational inequality problems. Reprot, University of New South Wales, Sydney, Australia, 1998.

[15] F.H. Clarke. Optimization and Nonsmooth Analysis. John Wiley and Sons, New York, NY, 1983.

2D Satellite Image Registration Using Transform Based and Correlation Based Methods

Dr. H.B. Kekre
Senior Professor, Department of
Computer Science,
Mukesh Patel School of Technology
Management and Engineering
Mumbai, India

Dr. Tanuja K. Sarode
Asst. Professor, Computer Engineering
Department,
Thadomal Shahani college Of
engineering
Mumbai, India.

Ms. Ruhina B. Karani
Lecturer, Department of Information
Technology,
D.J. Sanghvi college of
engineering,
Mumbai, India.

Abstract— Image registration is the process of geometrically aligning one image to another image of the same scene taken from different viewpoints or by different sensors. It is a fundamental image processing technique and is very useful in integrating information from different sensors, finding changes in images taken at different times and inferring three-dimensional information from stereo images. Image registration can be done by using two matching method: transform based methods and correlation based methods. When image registration is done using correlation based methods like normalized cross correlation, the results are slow. They are also computationally complex and sensitive to the image intensity changes which are caused by noise and varying illumination. In this paper, an unusual form of image registration is proposed which focuses upon using various transforms for fast and accurate image registration. The data set can be a set of photographs, data from various sensors, from different times, or from different viewpoints. The applications of image registration are in the field of computer vision, medical imaging, military automatic target recognition, and in analyzing images and data from satellites. The proposed technique works on satellite images. It tries to find out area of interest by comparing the unregistered image with source image and finding the part that has highest similarity matching. The paper mainly works on the concept of seeking water or land in the stored image. The proposed technique uses different transforms like Discrete Cosine Transform, Discrete Wavelet Transform, HAAR Transform and Walsh transform to achieve accurate image registration. The paper also focuses upon using normalized cross correlation as an area based technique of image registration for the purpose of comparison. The root mean square error is used as similarity measure. Experimental results show that the proposed algorithm can successfully register the template and can also process local distortion in high-resolution satellite images.

Keywords- Discrete Cosine Transform (DCT); Discrete Wavelet Transform (DWT); HAAR Transform; Walsh transform Normalized Cross Correlation; Interst Point Area Extraction; Image Registration.

I. INTRODUCTION

The accelerated growth in the field of computer vision, image fusion, medical imaging, military automatic target recognition, remote cartography and astrophotography has established the need for the development of good image registration technique for the efficient retrieval of interest point area [10]. Before the development of image registration, there were difficulties in matching the images with angular distortion. As a result interest point matching result was poor [21]. Image registration is the process of transforming different sets of data into one coordinate system. The data set can be a set of photographs, data from various sensors, from different times, or from different viewpoints [1], [2], [5]. In this paper, the use of Discrete Cosine Transform (DCT), Discrete Wavelet transform (DWT), HAAR transform, Walsh transform and normalized cross correlation is investigated.

This paper presents the image registration techniques based on extracting interest point area of satellite images using the transforms mentioned above. Extracting the interest point area is achieved by computing correlation coefficient, DCT, DWT, HAAR and Walsh transforms between target unregistered image of any size and a source image of size NXN [21]. The root mean square error is used as similarity measure. Image registration is achieved by extracting interest point area and highlighting the same on the source image with the help of coordinate extraction.

The first section of this paper contains a brief review of previous relevant work. In the second section, the new algorithm is introduced in detail. Next, we present some experimental results using satellite images. Finally, some concluding remarks are provided.

II. LITERATURE REVIEW

Image registration algorithms can be broadly classified into two categories according to matching method: area based methods (ABM) and feature based methods (FBM) [1], [4], [5]. In ABM algorithms, small window of pixels in the sensed image is compared statistically with window of the same size in the reference image. The most commonly used methods are cross-correlation matching and least-squares matching. The centers of the matched windows are treated as control points, which can be used to solve mapping function parameters between the reference and sensed images [2]. ABM is a classical matching method. Feature-based matching techniques do not use the gray values to describe matching entities, but use image features derived by a feature extraction algorithm [1]. The form of the description as well as the type of feature used for matching depends on the task to be solved.

Interest-point matching is problematic and remains the subject of much research within the communities of photo grammetry, remote sensing, computer vision systems, pattern recognition, and medical image processing [2]. In remote sensing, area-based algorithms are normally suitable for open terrain areas, but the feature-based approaches can provide more accurate results in urban areas. No single technique performs well in both circumstances [4]. Both algorithms have their own unique strengths and weaknesses. Neither type of algorithm can avoid the problem of dealing with ambiguity in smooth (low texture) areas. Feature-based algorithms face the additional problem of the effect of outliers (points with no correspondences) on the results [1], [2], [5], [7].

Because of the large number of feature-based algorithms used in interest-point matching, there are many classification methods for describing these algorithms. Normally, feature based algorithms can be categorized into rigid and non-rigid (according to the transformation between images), global and local (according to the image distortions), or corrected and uncorrected (according to the image variations) [1].

In addition, most of the feature-based algorithms search for correspondences and also address the refinement of a transformation function. Therefore, feature-based algorithms can also be grouped into three additional categories [3]. They either solve the correspondence only, solve the transformation only, or solve both the correspondence and the transformation. Although numerous feature-based algorithms have been developed, there is no general algorithm which is suitable for a variety of different applications. Every method must take into account the specific geometric image deformation [6].

For non-rigid local distortions, more complicated transformations are developed. TPS was proposed initially for global transformations, but it was improved for smooth local distortions for medical image registration [1], [2], [5].

III. METHODOLOGY

The proposed algorithm aims upon using DCT, DWT, HAAR, WALSH transforms and normalized cross correlation for the purpose of image registration. The algorithm works by dividing the source image into overlapping blocks of predefined size having some horizontal and vertical overlap. The interest point area extraction is achieved by finding DCT, HAAR and WALSH of unregistered target image and of each block of source image.

The minimum root mean square error criterion is used as a similarity measure. For applying DWT, first level decomposition of unregistered target image and each block of source image is taken. The similarity measure is used between unregistered image and each block of source image. The second approach used for interest point area extraction deals with computing normalized cross correlation between each block of source image and the unregistered target image.

A. Correlation Based Approach for Image Registration

This section of the paper focuses upon using normalized cross correlation to achieve image registration. Correlation based methods, sometimes called as template matching merge the feature detection step with the matching part [2].

1) Normalized Cross Correlation Approach

Windows of predefined size or even entire images are used for the correspondence estimation during the registration step. Area-based methods put emphasis on the feature matching step rather than on their detection. No features are detected in these approaches so the first step of image registration is omitted. Area-based methods, sometimes called correlation-like methods or template matching merge the feature detection step with the matching part. These methods deal with the images without attempting to detect salient objects [2]. Correlation coefficient can be calculated using equation (1).

$$r = \frac{Syx}{\sqrt{SyySyx}}$$

$$= \sum_{i=1}^{n} (Yi - \sim Y)(Xi - \sim x)/(\sqrt{(\sum_{i=1}^{n}(Yi - \sim Y)^2 * \sum_{i=1}^{n}(Xi - \sim x)^2)}) \quad (1)$$

a) Algorithm

Step 1. Resize the target image template to M X N pixels.

Step 2. Partition source image into overlapping blocks of size MX N with some horizontal and vertical overlap.

Step 3. Compute the correlation coefficient between each block of source image and that of target image template.

Step 4. Extract the region having maximum correlation from source image.

Step 5. Set the threshold by trial and error.

Step 6. Extract the coordinates of all blocks correlation greater than the threshold.

Step 7. Draw a region boundary enclosed by the coordinates.

- *Advantages*
  1. Simple to implement.
  2. Works best for open terrain areas.
  3. Suitable for images with little distortion.
- *Drawbacks*
  1. Slow because of computational complexity.
  2. Cannot deal with smooth areas.
  3. Sensitive to image intensity changes caused by noise or varying illumination.

B. Transform Based Approach for Image Registration

This section of the paper focuses upon making use of different transforms mentioned below to achieve image registration. The new proposed algorithm is discussed for each of these transforms along with some of the required equations for implementation of the same.

1) Discrete Cosine Transform Approach

The discrete cosine transform (DCT) is closely related to the discrete Fourier transform. It is a separable linear transformation; that is, the two-dimensional transform is equivalent to a one-dimensional DCT performed along a single dimension followed by a one-dimensional DCT in the other dimension [11]. Two dimensional DCT transform can be calculated using equation (2).

$$F(u,v) = \frac{Cu}{2} * \frac{Cv}{2} \sum_{y=0}^{T} \sum_{x=0}^{T} f(x,y) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right] \quad (2)$$

With

$$Cu = \frac{1}{\sqrt{2}} \quad if\ u = 0 \qquad Cv = \frac{1}{\sqrt{2}} \quad if\ v = 0$$
$$Cu = 1 \quad if\ u > 0 \qquad Cv = 1 \quad if\ v > 0$$

The basic operation of the DCT is as follows:

- The input image is N by M

- f(i,j) is the intensity of the pixel in row i and column j

- F(u,v) is the DCT coefficient in row 'u' and column 'v' of the DCT matrix.

a) Algorithm

Step 1.   Resize the target image to M X N pixels.

Step 2.   Partition source image into overlapping blocks of size M X N by taking block size manually according to the size of source image and some horizontal and vertical overlap.

Step 3.   Compute the two dimensional DCT of target image and each block of source image starting from the leftmost top block.

Step 4.   Calculate root mean square error between DCT of target image and each block of source image respectively.

Step 5.   Compute the root mean square error.

Step 6.   Set a threshold i.e. .minimum allowable error.

Step 7.   The block having minimum RMSE is extracted from source image.

Step 8.   Extract the coordinates of all blocks having error value less than the threshold and draw a region boundary enclosed by the coordinates.

2) Haar and Walsh Transform Approach

When k=0, the Haar function is defined as a constant,

$$ho(t) = \frac{1}{\sqrt{N}} \qquad\qquad (3)$$

When k>0, the Haar function is defined by,

$$hk(t) = \frac{1}{\sqrt{N}} * \left(2^{\frac{p}{2}}\right) \qquad \frac{q-1}{2^p} \le t < \frac{q-0.5}{2^p}$$

$$hk(t) = \frac{1}{\sqrt{N}} * \left(-2^{\frac{p}{2}}\right) \qquad \frac{q-0.5}{2^p} \le t < \frac{q}{2^p}$$

$$hk(t) = 0 \quad otherwise \qquad (4)$$

Walsh transform matrix is defined using a Hadamard matrix of order N. The Walsh transform matrix row is the row of the Hadamard matrix specified by the Walsh code index, which must be an integer in the range [0, ..., N - 1]. Walsh transform matrix is defined as a set of N rows, denoted W_j, for

j   =   0,1….,N-1,   which   has   the   following properties:

- W_j takes on the values +1 and -1.
- $W_j[0] = 1$ for all j.
- $W_j \times [W_k]^T = 0$ for $j \ne k$ and $W_j \times [W_k]^T = N$ for $j = k$
- W_j has exactly j zero crossings, for j = 0, 1, …., N-1.
- Each row W_j is either even or odd with respect to its midpoint

a) Algorithm

Step 1.   Resize the target image to M X N pixels.

Step 2.   Partition source image into overlapping blocks of size M X N by taking block size manually according to the size of source image and some horizontal and vertical overlap.

Step 3.   Design HAAR/WALSH kernel of size MXN.

Step 4.   Apply the kernel on each block as well as on target image individually.

Step 5.   Compute the average of HAAR/WALSH result.

Step 6.   Calculate root mean square error between result of target image and each block.

Step 7.   The block having minimum RMSE is extracted from source image.

Step 8.   Extract the coordinates of all blocks having error value less than the threshold and draw a region boundary enclosed by the coordinates.

3) Discrete Wavelet Transform Approach

Wavelet analysis is similar to Fourier analysis which allows a target function over an interval to be represented in terms of an orthonormal function.

Figure 1 shows the single level decomposition diagram using discrete wavelet transfrom.

LL	HL
LH	HH
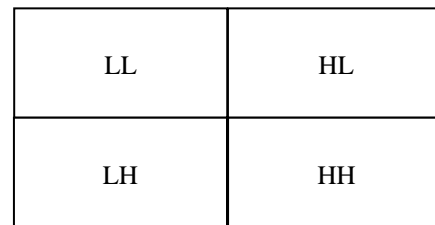
Figure 1. Single level decomposition using descrete wavelet transform

a) Algorithm

Step 1.   Resize the target image to M X N pixels.

Step 2.   Partition source image into overlapping blocks of size M X N by taking block size manually according to the size of source image and some horizontal and vertical overlap.

Step 3.   Perform single level decomposition of each block and that of target image.

Step 4. Calculate root mean square error between result of target image and each block

Step 5. The block having minimum RMSE is extracted from source image.

Step 6. Extract the coordinates of all blocks having error value less than the threshold and draw a region boundary enclosed by the coordinates.

- *Advantages*

1. Faster than area based methods.
2. Provides good result.
3. Works best for urban areas.
- *Drawbacks*
    1. Features cannot be exactly matched because of noise.
    2. Involves human intervention for manually selecting control points.
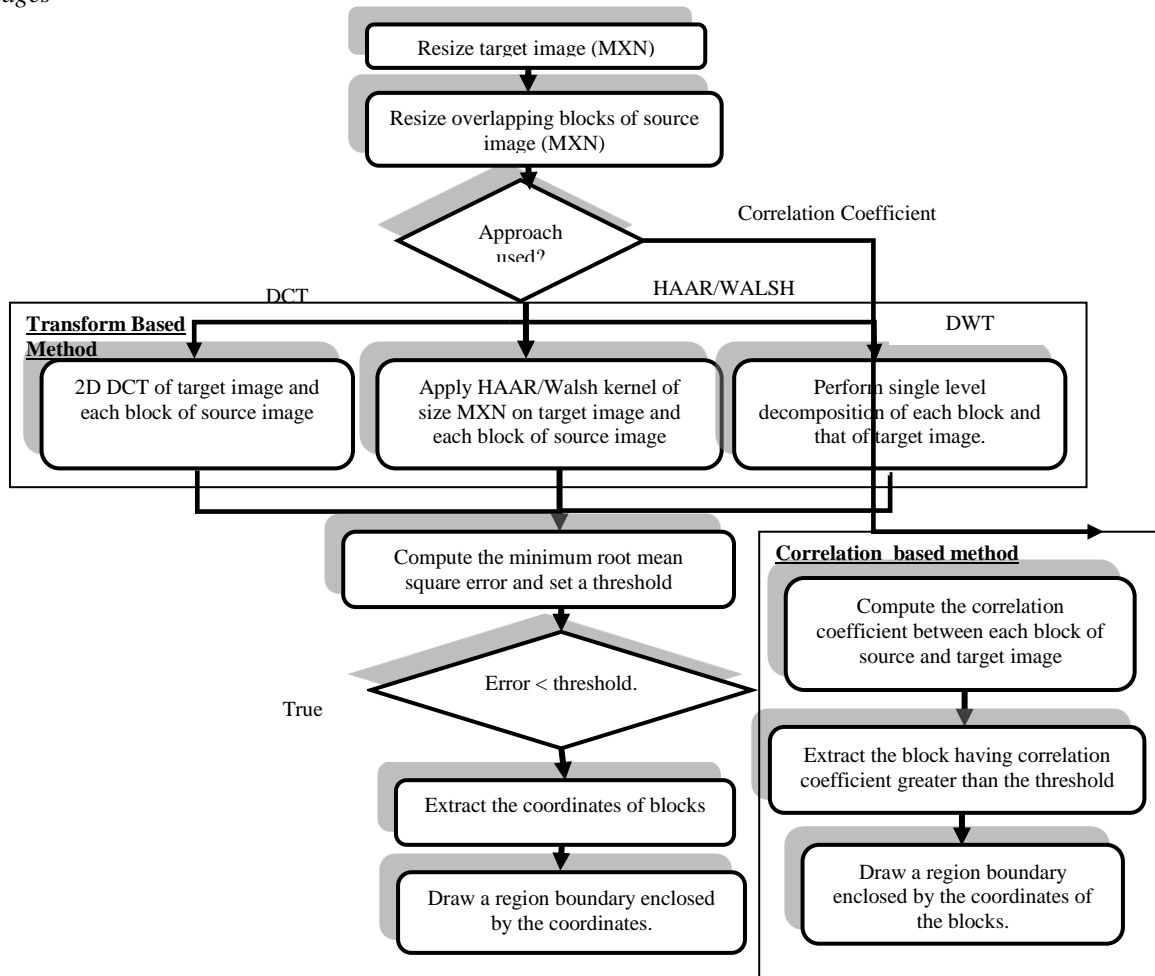
Figure 2. Flowchart for complete implementation

Figure 2 shows the complete implementation scheme of proposed technique. Initially the target image template is resized to the predefined size MXN and the stored source image is divided into overlapping blocks of same size as that of template by providing some horizontal and vertical overlap. Based on the approach used for image registration (i.e. transform based approach and correlation based approach), the corresponding algorithm mentioned for that is applied. The transform based approach applies DCT, HAAR, WALSH and wavelet transforms on the resized target image template and each block of source image. The wavelet transform uses first level decomposition technique for the same. The root mean square error is computed as a similarity measure between transformation result of target image and each block of source image. Threshold is established for minimum root mean square values. The coordinates of the corresponding blocks of source image satisfying the threshold criterion (i.e. error<threshold) are extracted and a region boundary is drawn connecting these coordinates. Correlation based approach tries to compute normalized cross correlation between the resized target image and each block of source image. The maximum value of correlation coefficient is 1. The threshold is established for computed correlation coefficient. The coordinates of the block of source image for which correlation coefficient is greater than the threshold value are extracted and the region boundary is drawn connecting these coordinate pixels.

IV. RESULTS AND DISCUSSIONS

The implementation of the proposed algorithm is done in MATLAB R2009a using a computer with Intel Core2 Duo Processor (2.20GHz) and 2 GB RAM.

The proposed technique works on the idea of applying image registration algorithm explained in the paper on satellite images. To achieve these results, the algorithm is executed with various satellite images of different sizes. The transform based and correlation based approach is tested on each of them. The results obtained are shown in the following figures.

A. Results Obtained For Satellite Image

This section of the paper focuses on satellite images having land and water regions. The technique takes any water image template as an input and a source satellite image is stored in a database. The goal of the technique is to find out where exaclty the water lies in the source satellite image by applying image registration.

Figure 3 shows a satellite image on which image registration has to be performed. Figure 4 shows unregistered water image temaplate taken randomly from any satellite image. Results included are obtained both by applying correlation based method as well as transform based methods. Figure 5 shows the result of image registration using normalized cross correlation approach from which it is clear that correlation based method does not detect entire water region from the source image.

Figure 6 shows the result of applying transform based approach. It shows that all the transforms i.e. DCT, HAAR, WALSH and WAVELET transfroms produce the same results. Figure 6 also shows that the result obtained by transform based approach cover most of the water region except some. Results obtained in this section imply that transform based approach performs better than correlation based approach for image registration.

Figure 3. Source Image

Figure 4. Unregistered Image of Water

Figure 5. Registered image using normalized cross correlation

Figure 6. Registered image using DCT/DWT/WALSH/HAAR transforms

Figure 7. Source image taken from different angle

B. Results Obtained For Angularly Distorted Images

This section of the paper focuses on satellite images captured from different camera angles or by different viewpoints. The goal of this section is to register the image template taken out from the image taken from one angle to the image taken from some different angle thus satisfying the basic definition of image registration explained in section I. Figure 7 shows image of some region captured from one angle whereas Figure 8 shows the same image captured from different angle. The image registration technique mentioned in the paper works by extracting image template of size MXN as shown in figure 9 and figure 10 from the source image taken from different angle.

The source image 1 is then divided into overlapping blocks of same size as the template by providing some horizontal and vertical overlap. Both the techniques (i.e. Correlation Based Approach and Transform Based Approach) are applied on the template and each block of source image. Figure 11 shows the result obtained by applying both the methods for registration of the image template shown in figure 9.

Figure 12 shows the result obtained by applying both the methods for registration of the image template shown in figure 10. Experimental results imply that both the technique produce same result. Both the techniques correctly outline the interest point area. Experimental results also show that Correlation based approach consumes more time to produce the result as compared to Transform based approach.

Figure 8.   Source image 1

Figure 8.   Unregistered image template 1

Figure 9.   Unregistered image template 2

Figure 10.   Registered image 1 using correlation based and transfrom based methods

Figure 11.   Registered image 2 using correlation based and transfrom based methods

V.   CONCLUSION

Image registration plays an important role in the alignment of two or more images of the same scene. In this paper, a simple but effective algorithm of Image Registration is presented which uses transform based approach as well as area based approach. The algorithm works with the idea of dividing the source image into overlapping blocks of size same as that of unregistered target image.

The image registration method of this paper is for the field of satellite imaging. Image registration is achieved by extracting the coordinates of blocks satisfying the criteria of threshold value thus making entire procedure automatic for interest point area detection. This paper also takes angularly distorted images into consideration. It tries to locate the exact position of given template of the image taken from different angle or by different viewpoint thus achieving basic definition of image registration.

Experimental results on various images in image registration shows that the interest point area computed using transform based approach provides more accurate results as compared to the registration performed using normalized cross correlation.

REFERENCES

[1]   Xiong and Yun Zhang "A Novel Interest-Point-Matching Algorithm for High-Resolution Satellite Images" ieee transactions on geoscience and remote sensing, vol. 47, no. 12, december 2009.

[2]   Gang Hong, Yun Zhang "Combination of Feature-based and Area- based Image Registration Technique for High Resolution Remote Sensing Image" ieee international conference on Geoscience and Remote Sensing Symposium, 2007.

[3]   Ronald W. K. So and Albert C. S. Chung "Multi-level non-rigid image registration using graph-cuts" ieee international conference on Acoustics, Speech and Signal Processing, 2009.

[4]   M. Auer, P. Regitnig, and G. A. Holzapfel, "An automatic non rigid registration for stained histological sections," IEEE Trans. Image Process.,vol. 14, no. 4, pp. 475–486, Apr. 2005.

[5]   Goshtasby, A., "Piecewise linear mapping functions for image registration, Pattern Recognition", vol.19, no.6, pp.459-466,1986.

[6]   Dongjin Kwon, "Rolled Fingerprint Construction Using MRF-Based Nonrigid Image Registration", ieee transactions on image processing, vol. 19, no. 12, december 2010

[7]   Boffy, Y. Tsin, and Y. Genc, "Real-time feature matching using adaptive and spatially distributed classification trees," Brit. Mach. Vis. Conf., Jul. 2006.

[8]   Ming-Sui Lee, Meiyin Shen, Akio Yoneyama and C. -C. Jay Kuo "DCT-Domain Image Registration Techniques for Compressed Video" ieee international conference on circuits and systems, 2005.

[9]   L. G. Brown, "A survey of image registration techniques," ACM Comput.Surv., vol. 24, no. 4, pp. 325–376, Dec. 1992.

[10] Dr. H.B.Kekre, Sudeep D. Thepade, Akshay M aloo, "Image Retrieval using Fractional Coefficients of Transformed Image using DCT and Walsh Transform", International Journal of Engineering Science and Technology , Vol. 2(4), pp .362-371, 2010.

[11] H. Chui and A. Rangarajan, "A new point matching algorithm for nonrigid registration," Comput. Vis. Image Underst., vol. 89, no. 2/3, pp. 114– 141, Feb. 2003.

[12] C. Harris and M. Stephens, "A combined corner and edge detector," in Proc. Alvey Vis. Conf., 1988, pp. 147–151.

[13] G. K. Wallace, "Overview of the JPEG still Image Compression standard," SPIE 1244 (1990) 220-233.

[14] J. Rexilius, S. K. Warfield, C. R. G. Guttmann, X. Wei, R. Benson,L. Wolfson, M. Shenton, H. Handels, and R. Kikinis,"A novel

nonrigidregistration algorithm and applications," in *Proc.MICCAI*, W. Niessen and M. Viergever, Eds., 2001, vol. 2208, pp. 923–931.

[15] J. Williams and M. Bennamoun, "Simultaneous registration of multiple corresponding point sets," *Comput. Vis. Image Underst.*, vol. 81, no. 1, pp. 117–142, Jan. 2001.

[16] Sang-M i Lee, Hee_Jung Bae, and Sung-Hwan Jung, "Efficient Content-Based Image Retrieval M ethods Using Color and Texture", *ETRI Journal* 20 (1998) 272-283.

[17] Sahil Suri and Peter Reinartz, "Mutual-Information-Based Registration of TerraSAR-X and Ikonos Imagery in Urban Areas" ieee transactions on geoscience and remote sensing, vol. 48, no. 2, february 2010

[18] Mingchao Sun, Bao Zhang, linghong Liu, Yongyang Wang and Quan Yang," The Registration of Aerial Infrared and Visible Images" International Coriference on Educational and Information Technology 2010.

[19] L. Zagorchev, and A. Goshtasby, "A comparative study of transformation functions for nonrigid image registration," *IEEE Transactions on Image Processing,* vol.15, pp. 529-538, 2006.

[20] Ruhina Karani, Dr. Tanuja K. Sarode, "Image Registration using Discrete Cosine Transform and Normalized Cross Correlation", International Journal of Computer Application, ISBN No- 973-93-80864-48-5, 2012

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engineering. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970 He has worked as Faculty of Electrical Engg. and then HOD Computer Science and Engg. at IIT Bombay. For 13 years he was working as a professor and head in the Department of Computer Engg. at Thadomal Shahani Engineering. College, Mumbai. Now he is Senior Professor at MPSTME, SVKM's NMIMS University. He has guided 17 Ph.Ds, more than 100 M.E./M.Tech and several B.E./B.Tech projects.

His areas of interest are Digital Signal processing, Image Processing and Computer Networking. He has more than 350 papers in National / International Conferences and Journals to his credit. He was Senior Member of IEEE. Presently He is Fellow of IETE and Life Member of ISTE Recently ten students working under his guidance have received best paper awards and two have been conferred Ph.D. degree of SVKM's NMIMS University.Currently 10 research scholars are pursuing Ph.D. program under his guidance.

Dr. Tanuja K. Sarode has Received Bsc.(Mathematics) from Mumbai University in 1996, Bsc.Tech.(Computer Technology) from Mumbai University in 1999, M.E. (Computer Engineering) degree from Mumbai University in 2004, Ph.D. from Mukesh Patel School of Technology, Management and Engineering, SVKM"s NMIMS University, Vile-Parle (W), Mumbai, INDIA. She has more than 11 years of experience in teaching. Currently working as Assistant Professor in Dept. of Computer Engineering at Thadomal Shahani Engineering College, Mumbai. She is life member of IETE, member of International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT), Singapore. Her areas of interest are Image Processing, Signal Processing and Computer Graphics. She has more than 100 papers in National /International Conferences/journal to her credit.

Ms. Ruhina B. Karani has received B.E. (Computer Engineering) from Mumbai University in 2007. Currently pursuing her M.E. (Computer Engineering) from Thadomal Shahani Engineering College, Mumbai. She has 4 years of experience in teaching. She was working as a lecturer in K.J. Somaiya College of engineering and is currently working as a lecturer in D.J. Sanghvi college of engineering, Mumbai. She is a life time member of ISTE. Her areas of interest are Image processing, Microprocessor and microcontroller. She has guided two B.E. projects. She has one research paper in international journal to her credit.

Validation of the IS Impact Model for Measuring the Impact of e-Learning Systems in KSA Universities: Student Perspective

Salem Alkhalaf, Steve Drew, Anne Nguyen

School of ICT, Griffith University,

Gold Coast, Australia

Abstract—**The IS-Impact Measurement Model, developed by Gable, Sedera and Chan in 2008, represents the to-date and expected stream of net profits from a given information system (IS), as perceived by all major user classes. Although this model has been stringently validated in previous studies, its generalizability and verified effectiveness are enhanced through this new application in e-learning. This paper focuses on the re-validation of the findings of the IS-Impact Model in two universities in the Kingdom of Saudi Arabia (KSA). Among the users of 2 universities e-learning systems, 528 students were recruited. A formative validation measurement with SmartPLS, a graphical structural equation modeling tool was used to analyse the collected data. On the basis of the SmartPLS results, as well as with the aid of data-supported IS impact measurements and dimensions, we confirmed the validity of the IS-Impact Model for assessing the effect of e-learning systems in KSA universities. The newly constructed model is more understandable, its use was proved as robust and applicable to various circumstances.**

Keywords – IS-Impact Model; e-learning systems; Saudi Arabia

I. INTRODUCTION

According to the Communications and Information Technology Commission (CITC), the Kingdom of Saudi Arabia (KSA) is one of the fastest growing countries in the world in terms of e-learning [1]. CITC data shows an explosive growth in the number of Internet users generally, from a mere 200,000 in 2000 to 4.8 million in 2006. The number of students enrolled in Saudi institutions of higher education has also increased significantly in 2011 to 11.4 million [1]. As a result, many of these institutions have turned to e-learning systems as a means to help broaden and enhance access to their courses and subjects [2].

Reflecting this trend, a growing number of research studies have been conducted on e-learning in the Kingdom of Saudi Arabia [3-5]. Movement and development in e-learning seem to be fast and strong, especially in the universities in KSA [3].

Many of these studies have focused on identifying the key factors that differentiate online education from face-to-face learning, analysing the in-principle advantages and disadvantages of online courses, or developing strategies to achieve a suitable online learning environment [6]. To date, however, little attention has been paid to the issue of assessing the e-learning environments that have been set up in the country. Indeed, it would appear that relatively little research

has been done regarding the evaluation of e-Learning systems in general [7, 8].
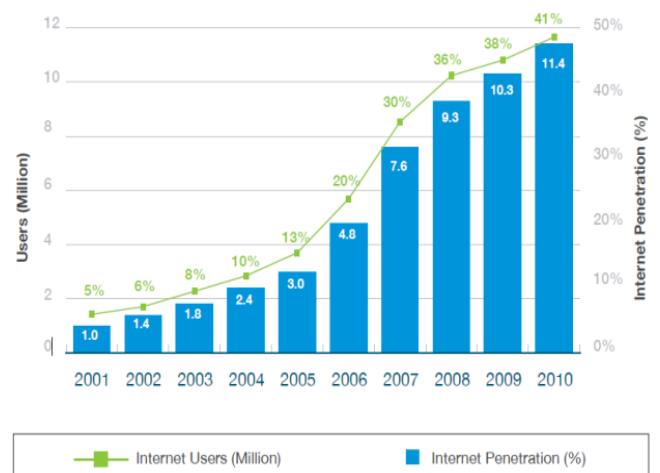
Figure 1. Internet Market Evolution for KSA (2001-2010) (adopted from [1])

II. BACKGROUND

Interest in e-learning has grown rapidly during the past decade or so in KSA, for a number of reasons [9]. First, the demand for higher education has far outstripped supply, such that institutions are faced with overcrowding and insufficiency of facilities and human resources for the delivery of traditional-style education to all of the nation's qualified applicants for admission. E-learning has been suggested as a means to overcome these limitations.

Second, KSA is a large country in terms of geographical area, with a significant number of communities being isolated from major population centres. E-learning offers the potential to deliver educational services to remote locations, thereby reducing disparities across the various regions and areas [5, 10].

Third, in KSA's higher education, men and women receive their instruction in separate classes not mixed "ikhtilat" classes, for cultural and religious reasons [11]. So, Male instructors can only teach female students through eLearning tools or distance

learning tools [10]. This puts further strains on the limited facilities and human resources available. It has been observed, accordingly, that women are often among the strongest supporters of e-learning, which potentially facilitates their access to higher education [12-14].

In 2008 the KSA Ministry of Higher Education established a National Centre of E-learning & Distance Learning to promote and facilitate the spread of e-learning systems in Saudi universities [15]. It has was estimated that in 2008, annual turnover of the e-learning industry in KSA had already reached US$ 125 million, with further expected growth of about 33 per cent per year over the following 5 years [16]. It is timely, therefore, to investigate the issue of assessing the success or impact of the e-learning systems that have been set up to date[5].

III. CONCEPTUAL FRAMEWORK AND MEASUREMENT DIMENSIONS

The IS-Success/Impact Measurement framework was selected because it comprehensively takes into account four dimensions of success/impact in the context of IT systems [17]. The IS-Success Model [18] is one of the most cited models in IS research [19]. More recently it has been supplemented by the IS-Impact Measurement Model [20]. As illustrated in Figure II, within the IS-Success/Impact framework, the success and impact of an IS system can be measured in terms of

- the quality of the information produced (information quality),

- the performance of the system from a technical perspective (system quality),

- the impact on individual users (individual impact), and

- the impact on the relevant organisation (organisational impact).

For e-learning systems, the third and fourth dimensions are the most important ones, as they represent the end-goals of the system [5].
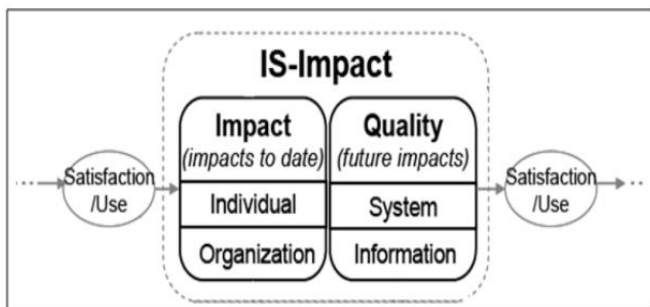
Figure 2. IS-Impact Model (adapted from [20])

Gable, Sedera, & Chan [20] stated that on behalf of an organisation, Individual Impact measures the extent to which an IS influences the capabilities and effectiveness of key stakeholders or users [21]. Organisational Impact is a measure of the degree to which the IS advances the enhancement of organisational outcomes and abilities. Information Quality is a measure of the quality of (the IS) output (e.g., the quality of the

information that the system produces in reports and onscreen) [20]. System Quality refers to a measure of (the IS) performance from a technical and design perspective [20].

The IS-Impact Model is considered the theoretical foundation for this study. We designed the model in such a way that it is robust and generalisable, and produces simple results that are highly comparable across stakeholders, time, and various types of systems [21]. We employ a continuous measurement scheme, thereby creating a general instrument that is easy to understand and respond to, by all relevant stakeholder groups. This measurement scheme enables the comparison of stakeholder perspectives.
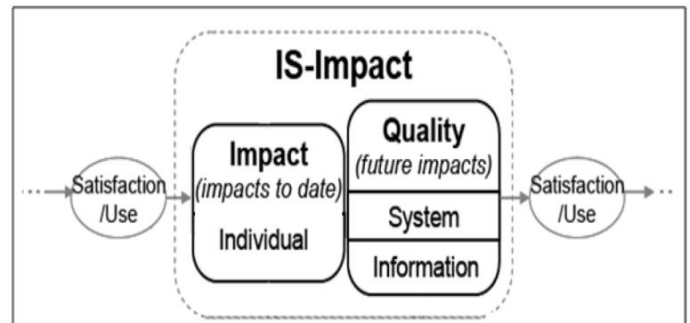
Figure 3. New measurement model for students using e-learning systems

The present study identifies relevant new measures that were not determined by previous researchers. In order to concentrate on the impact that e-learning systems have on students, as the key stakeholders, for the purposes of this paper we exclude the measure of "Organizational Impact". Thus, we use only 26 measures (the a priori model developed in [20]) to ensure model completeness.

IV. RESEARCH METHODOLOGY

To ensure the suitability of the model for KSA universities, we modified the original survey instrument. It was translated into Arabic for students who are less conversant in English. Following the suggestions of Brislin and McGorry [21, 22], the instrument was translated using both back translation and decentering approaches. Sentence structure and word choice were slightly modified, ensuring that no changes in meaning occurred [21]. The Arabic version was then examined for face validity. Given that the Organisational Impact dimension is more strongly related to staff and faculty member issues than to student attributes, we deducted the items that make up this dimension. For staff members, other models such as fourth dimensions were introduced.

The questionnaire comprises two main sections. The first is intended to collect demographic information on the respondents, while the second features the 26 measures of the IS-Impact Model and many dependent variables used to test construct validity. A five-point Likert scale (with "strongly agree" and "strongly disagree" as the end points) was used to rank the responses of the participants regarding the model items [5]. Figure IV shows the IS-Impact Model and its 26 measures.
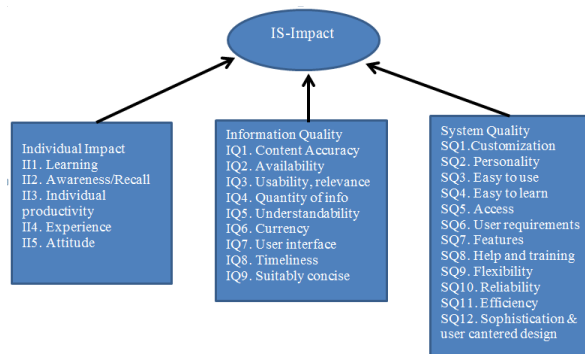
Figure 4.   IS-Impact Model with 26 measures

The participants were required to answer all the questions, which include those related to the descriptive items, IS-Impact measures, and dependent variables. This instruction is written at the beginning of the questionnaire and repeated in the introductory page.

Qassim University and King AbdulAziz University in KSA were chosen as the study sites for data collection. A mixture of cluster, convenience, and snowball sampling methods were used to select the respondents. The respondents belong to e-learning classes or use the e-learning systems of the universities. A hard copy of the questionnaire was distributed to the sample. Out of the 800 questionnaires distributed, 560 were returned; 32 were rejected because of incomplete answers. The final sample comprised 528 students (328 males, 200 females).

V.   RESULTS

A clear definition and understanding of formative construct and its difference with reflective construct were obtained; in specifying formative constructs, few studies provided exemplary interpretations of formative measurement results, but we were able to find credible ones from [23] and [24]. An illustrative example of formative construct validation was found in [25], [26], and [27]. The authors demonstrated the method for estimating and assessing constructs with partial least squares (PLS) software. One of the principal advantages of PLS is that it supports both formative and reflective measures—a highly beneficial feature because employing covariance-based SEM techniques, such as LISREL or EQS [28, 29], is difficult to accomplish. Assessing the validity of a formative measurement model involves four steps, shown in the Table I below.

Validity test for the formative measurement model (adopted from [21])

Test of	Description
Multicollinearity	The presence of multicollinearity of the items is identified by conducting a test. Higher collinearity among items shows conceptual redundancy.
External validity	External validity is used to evaluate the validity by examining how well the formative items capture the construct by relating these measures with a reflective variable of the same construct.
Nomological validity	Nomological validity is used to

(Nomological net)	evaluate validity by connecting the items to other constructs that have important and strong relationships known through previous research; that is, according to previous, research linking the formative measurement model with the antecedents and/or consequence constructs to which a structural path exists.
Significance of weights	Formative measurement models are observed by significant weights.

A.   Multicollinearity

Collinearity diagnostics is a regression test on the items in a formative construct (the independent variables) with a dependent variable. The results show the presence of collinearity. The presence of multicollinearity was also confirmed by the tolerance and variation inflation factor (VIF) value shown in the "Coefficients" table. The results demonstrate that all the 26 measures were below the general VIF cutoff point of 10 [23], [30], [31], with 1.9 as the largest VIF observed.

Diamantopolous & Winklhofer [23] stated that one approach to testing item quality is observing the correlation of items with a different variable that may be external to an index. The items that show a strong correlation with the variable should be retained. Further, Diamantopolous & Winklhofer [23] suggested that the relationships among the items with the proposed dependent variable (the global item) at each dimension can be examined by using four global items that "summarise the essence of the construct that the index purports to measure".

B.   Assessing the Validity of the IS-Impact Model through Structural Relationships

The IS-Impact Model was tested by determining the structural relationship between unobserved variables (inner model), also known as nomological (net) validity [21]. The tests were carried out using SmartPLS, a software application for (graphical) path modelling with latent variables, in which the PLS approach is used to analyse latent variables, or both latent variables and observed or manifest variables (outer model) [21, 32]. Figure V shows the results of the structural relationship evaluation.

Two reflective measures given in the "IS-Impact summary" were used to evaluate the structural relationship of the measurement model (outer model). The analysis shows adjusted R-squares of 0.622 and 0.941, indicating that 94.1% of the variance in the IS-Impact Model is explained by Individual Impact, Information Quality, and System Quality. Chin [33] and Henseler, Ringle, & Sinkovics [27] suggested R-square values of 0.67 (substantial), 0.33 (moderate), and 0.19 (weak). The IS-Impact Model is almost substantial, as shown by the path analysis. At $p < 0.05$ or better (estimated by a bootstrapping procedure with 528 bootstrapping samples), all the structural paths were significant and the IS impact was considerably influenced by System Quality and Information Quality.
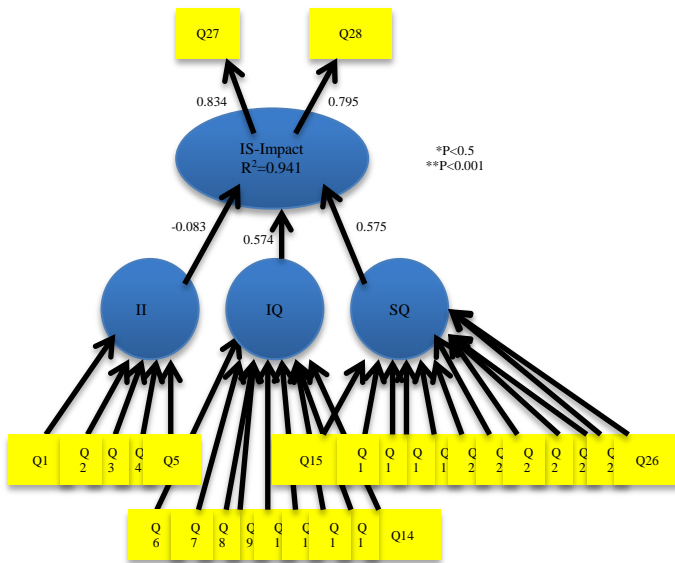
Figure 5.   Structural model

C. Explanatory Power of the Model

After the PLS test, the R-square was varied to determine the effect of Individual Impact, Information Quality, and System Quality on the overarching IS-Impact construct. This approach involves the repeated estimation of PLS and evaluation of effect size, in which each of the PLS is run with one dimension omitted. The effect size was derived using the following equation:

$$\text{Effect size, } f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}} \qquad (1)$$

Effect size (adopted from [21])

Table II shows that Individual Impact, Information Quality, and System Quality exhibited an average effect of greater than 0.15 on IS impact. Cohen [34] denoted a value of 0.02 as small, 0.15 as medium, and 0.35 as large. The effect of System Quality on IS impact was larger at 0.35. Demonstrating the addition of the three dimensions as a complete measurement model is the main objective of this analysis. The results show that mixing all the dimensions in a model enabled a high incremental change in the R-square variations.

TABLE I.   EFFECT SIZE

R-square included		0.941		
Run	**Removed**	**R-square Excluded**	**Effect Size**	**Interpretation**
1	Individual impact (II)	0.937	0.068	Small effect
2	Information quality (IQ)	0.919	0.322	Large effect
3	System quality (SQ)	0.922	0.373	Medium effect

D. Variance Inflation Factor

The VIF factor scores for all the three dimension were low (all VIF < 2), strongly indicating that most of the predictor variables in the model were confounded [35]. Because the values (VIF) less than 10 are the critical values for this

factor, no overlap was observed between these factors; they are independent of determinants   with a significant overlap. This result is unimportant and can be considered similar to the unexplained change in the follower, but it also indicates no multicollinearity problem [36].

TABLE II.   VIF

Category	VIF
Individual impact (II)	1.9
Information quality (IQ)	1.5
System quality (SQ)	1.4

VI. DISCUSSION AND CONCLUSION

The IS-Impact Model can be used to determine the effect of e-learning information systems used to date, as well as to forecast its future effects on KSA universities. IS impact and all the pathways between the construct and higher constructs were significant, suggesting that. The hypotheses below are supported by the Model, Therefore, the students were definitely satisfied because of the considerable impact of the IS.

TABLE III.   HYPOTHESES

No	Hypotheses	Results
H1	Individual Impact is positively affected by the use of collaborative eLearning	Supported.
H2	System Quality is positively associated with the use of collaborative eLearning system.	Supported.
H3	Information Quality is positively affected by employing a collaborative eLearning system.	Supported.

We conducted many path estimate tests to exclude the measures with low path coefficients and observe the incremental change in R-square values. When these measures were excluded, we noticed a decrease in R-square values at the end of the tests, indicating that the explanatory power of the model diminishes. Therefore, although some measures may not be strong predictors of the construct, they are still considered to have important correlations in this study. The IS-Impact construct was considerably influenced by all the 26 measures, and none should be excluded.

We addressed the generalisability of the IS-Impact Model and expanded its application to new settings; that is, we used varied languages, cultures, and types of systems. In the context of KSA universities, the Model proves to be a valid tool. Adding or excluding a measure from the original set resulted in a more understandable model for assessing the effect of IS on KSA universities. Both the Arabic and English instruments were also suitable for determining the validity of the Model. Therefore, the IS-Impact Model is rich and can be applied to various circumstances (custom package, KSA universities, and English or Arabic language).

The limitations of this study can serve as opportunities for future research. Although contexts differ from one another, the differences can be minimised by using the Model. The e-learning system was used as a unit of analysis, similar to the original work on the IS-Impact Model. The generalisability of the Model was tested and the validity of the IS evaluation

system was expanded to the KSA context. Future studies can widen this scope by testing validity under different contexts. For instance, the assessment of the effect of IS on organisations can be carried out by developing better standardised measurement systems.

Finally, future work would be to developing the IS-Impact Model for assessing the effect of e-learning systems in other educational environments of eLearning. Also focusing on supporting processes of eLearning, the research will be conducted taking into account the general requirements of eLearning system.

REFERENCES

[1] CITC. (2010). *Communication and Information Technology Commission* Available: http://www.citc.gov.sa/citcportal/Homepage/tabid/106/cmspid/%7B611C6ED D-85C5-4800-A0DA-A997A624D0D0%7D/Default.aspx

[2] A. AlSaif, "The motivating and inhibiting factors affecting the use of Web-Based Instruction at the University of Qassim in Saudi Arabia," *ETD Collection for Wayne State University,* 2005.

[3] A. M. Al-Shehri, "E-learning in Saudi Arabia:'To E or not to E, that is the question'," *Journal of family and community medicine,* vol. 17, p. 147, 2010.

[4] A. Alwalidi and P. Lefrere, "Making E-Learning Invisible: Experience at King Khalid University, Saudi Arabia," *Educational Technology,* vol. 50, p. 4, 2010.

[5] S. Alkhalaf, A. Nguyen, and S. Drew, "Assessing eLearning Systems in the Kingdom of Saudi Arabia's Higher Education Sector:An exploratory analysis," in *2010 International Conference on Intelligent Network and Computing (ICINC 2010)*, Kuala Lumpur, Malaysia, 2010, p. 4.

[6] A. Alshehri, "Assessing faculty attitudes toward the significant factors for facilitating the implementation of online courses at the Institute of Public Administration in Saudi Arabia," *Mississippi State University* 2005.

[7] S. Aceto, C. Delrio, C. Dondi, T. Fischer, N. Kastis, R. Klein, W. Kugemann, F. Nascimbeni, M. P. Garcia, N. Rabemiafara, and A. Szûcs, "e-Learning for Innovation: Executive Summary of the Helios Yearly Report 2007," Brussels: MENON Network EEIG2007.

[8] Y. Wang, H. Wang, and D. Shee, "Measuring e-learning systems success in an organizational context: Scale development and validation," *Computers in Human Behavior,* vol. 23, pp. 1792-1808, 2007.

[9] M. Albalawi, "Critical factors related to the implementation of web-based instruction by higher-education faculty at three universities in the kingdom of saudi arabia," The University of West Florida, 2007.

[10] H. S. Al-Khalifa, "eLEARNING IN SAUDI ARABIA," *CASES ON CHALLENGES FACING E-LEARNING AND NATIONAL DEVELOPMENT: Institutional Studies and Practices,* vol. 2, pp. 745-772, 2010.

[11] E. A. Doumato, "Education in Saudi Arabia: gender, jobs, and the price of religion," *Women and Globalization in the Arab Middle East,* pp. 239-258, 2003.

[12] T. Bates. (2009, 30 July 2010). *A personal view of e-learning in Saudi Arabia* Available: http://www.tonybates.ca/2009/11/05/a-personal-view-of-e-learning-in-saudi-arabia/

[13] A. Mirza, "Students'Perceived Barriers to In-Class Participation in a Distributed and Gender Segregated Educational Environment," *The Electronic Journal of Information Systems in Developing Countries,* vol. 35, pp. 1-17, 2008 2008.

[14] A. Mirza, "Utilizing Distance Learning Technologies to Deliver Courses in a Segregated Educational Environment," in *World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, Vancouver, Canada, 2007, pp. 126-129.

[15] UWN. (2008, 01 June 2008). SAUDI ARABIA: E-learning education shake-up. *(0030).* Available: http://www.universityworldnews.com/article.php?story=20080529145753433

[16] AMEinfo.com. (2008, 30 July 2010). *Saudi Arabia's eLearning industry to touch $125m in 2008.* Available: http://www.ameinfo.com/154480.html

[17] L. Cao and N. F. Elias, "Validating the IS-Impact model : two exploratory case studies in China and Malaysia," in *Association for Information Systems (AIS)*, Hotel Novotel, Hyderabad, India, 2009, pp. 1-14.

[18] W. DeLone and E. McLean, "Information systems success: the quest for the dependent variable," *Information systems research,* vol. 3, pp. 60-95, 1992.

[19] B. Myers, L. Kappelman, and R. Victor, "Prybutok, A comprehensive model for assessing the quality and productivity of the information systems function: toward a theory for information systems assessment," *Information Resources Management Journal,* vol. 10, pp. 6-25, 1997.

[20] G. Gable, D. Sedera, and T. Chan, "Re-conceptualizing information system success : the IS-Impact Measurement Model," *Journal of the Association for Information Systems,* vol. 9, pp. 377-408, 2008.

[21] N. F. Elias, "Measuring the Impact of Information Systems in Malaysia," *International Journal on Advanced Science, Engineering and Information Technology,* vol. 1, pp. 206-211, 2011.

[22] S. Y. McGorry, "Measurement in a cross-cultural environment: survey translation issues," *Qualitative Market Research: An International Journal,* vol. 3, pp. 74-81, 2000.

[23] A. Diamantopoulos and H. M. Winklhofer, "Index construction with formative indicators: An alternative to scale development," *Journal of Marketing research,* pp. 269-277, 2001.

[24] S. Petter, D. W. Straub, and A. Rai, "Specifying formative constructs in information systems research," *Management Information Systems Quarterly,* vol. 31, pp. 623-656, 2007.

[25] P. Andreev, T. Heart, H. Maoz, and N. Pliskin, "Validating formative partial least squares (PLS) models: methodological review and empirical illustration," 2009.

[26] R. T. Cenfetelli and G. Bassellier, "Interpretation of formative measurement in information systems research," *Management Information Systems Quarterly,* vol. 33, p. 7, 2009.

[27] J. Henseler, C. M. Ringle, and R. R. Sinkovics, "The use of partial least squares path modeling in international marketing," *Advances in International Marketing,* vol. 20, pp. 277-319, 2009.

[28] D. Goodhue, W. Lewis, and R. Thompson, "PLS, Small Sample Size, and Statistical Power in MIS Research," in *System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference on*, 2006, pp. 202b-202b.

[29] W. W. Chin, "Commentary: Issues and Opinion on Structural Equation Modeling," *MIS quarterly,* vol. 22, pp. vii-xvi, 1998.

[30]   K. Mathieson, E. Peacock, and W. W. Chin, "Extending the technology acceptance model: the influence of perceived user resources," *ACM SIGMIS Database,* vol. 32, pp. 86-112, 2001.

[31]   J. Pallant, *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows Version 15*: Open University Press, 2007.

[32]   C. M. Ringle, S. Wende, and A. Will, "SmartPLS (Version 2.0 beta)," *Hamburg, Germany: University of Hamburg,* 2005.

[33]   W. W. Chin, "Issues and opinion on structural equation modeling," *Management Information Systems Quarterly,* vol. 22, p. 3, 1998.

[34]   J. Cohen, *Statistical power analysis for the behavioral sciences*: Lawrence Erlbaum, 1988.

[35]   D. Janetzko, "Objectivity, reliability, and validity of search engine count estimates," *International journal of internet science,* vol. 3, pp. 7-33, 2008.

[36]   R. M. O'brien, "A caution regarding rules of thumb for variance inflation factors," *Quality & Quantity,* vol. 41, pp. 673-690, 2007.

AUTHORS' PROFILES

Salem Alkhalaf graduated Bachelor of Education degree in Computer Since from the Department of Computer, Teachers College (Riyadh) in 2003, Also he graduated with Honors degree. And he graduated the Master of ICT from Griffith University (Brisbane, Australia) in 2008. Currently he is working toward the PhD in ICT from Griffith University (Gold Coast, Australia). Finally, he is interest developing e-learning environments for universities especially in KSA.

Steve Drew is a Senior Lecturer in Higher Education at Griffith University and is an adjunct research fellow in the School of Information and Communication Technology. His research interests include acceptance and adoption of information systems, innovative learning environments, and enhanced student engagement with learning. Steve holds a PhD in Computer Science (QUT), and a Masters in Higher Education (Griffith).

Anne Nguyen is a Senior Lecturer in the School of Information and Communication Technology at Griffith University and conducts research into a range of e-Systems technologies. Her research interests include e-Commerce, e-Banking, Human-Computer Interactions and User Interface Design.

Effect of a Video-Based Laboratory on the High School Pupils' Understanding of Constant Speed Motion

Louis Trudel

Faculté d'éducation
Université d'Ottawa
Ottawa, Canada

Abdeljalil Métioui

Département de didactique, Faculté des sciences de
l'éducation, Université du Québec à Montréal,
Montréal, Canada

Abstract—**Among the physical phenomena studied in high school, the kinematical concepts are important because they constitute a precondition for the study of subsequent concepts of mechanics. Our research aims at studying the effect of a computer-assisted scientific investigation on high school pupils' understanding of the constant speed motion. Experimentation took place in a high school physics classroom. A repeated measures analysis of variance shows that, during the implementation of this strategy, the pupils' understanding of kinematical concepts increased in a significant way. In conclusion, we specify advantages and limits of the study and give future research directions concerning the design of a computer-assisted laboratory in high school physics.**

Keywords-speed; computer-assisted laboratory; understanding; high school; Rasch measurement

I. INTRODUCTION

In reaction to what some people called traditional physics education, constituted by lectures followed by sessions of exercises and verification of theories in the laboratory, Anderson [1] recommends that the pupil plays a more active role by constructing his knowledge under the supervision of the teacher. In this approach, called guided discovery, learning results from activities of putting to the test by experiments the pupils' ideas about motion phenomena. However, experiments offered to the pupils in physics lessons do not usually take into account their alternative schemas in relation to the speed so that they can experience difficulties in identifying the appropriate factors and expressing them quantitatively in the form of equations. Therefore, our first research objective aims at conceiving and identifying the conditions of implementation of a strategy of scientific investigation which takes into account alternative schemas of the pupils and allow them to prove their hypotheses in the laboratory. Our second research objective consists in assessing the effect of such a strategy on the high school pupils' understanding of the constant speed concept. In the second section, we present a review of the literature concerning difficulties of the pupils in kinematics and the use of the computer-assisted laboratory to palliate these difficulties. In the third section, we describe a somewhat specialized version of the computer-assisted laboratory, which is called the video-based laboratory (VBL), and specify the conditions in which it was implemented in a high school. In the fourth and fifth sections, we present the methodology used and the results obtained with respect to the evaluation of the effect of the VBL

on the high school pupils' understanding of motion at constant speed. A discussion will follow in the sixth section on the conditions of implementation of the VBL and its effects on the pupils' understanding of kinematics. In conclusion, we draw the limits of our study and suggest directions for future researches.

II. REVIEW OF LITERATURE

If there is a domain which causes a lot of difficulties to the pupils, it is kinematics, defined as the study of the motion of objects without being concerned about its causes [2]. There are two main reasons put forward by the researchers: alternative schemas which the pupils already have on the properties of motion and the emphasis put on the mathematization of its properties in the teaching of kinematics. Firstly, the pupils have, before arriving in the physics course, a broad experience about the properties of motion which they have acquired in their interactions with daily events. These experiences allowed them to construct schemas with which they can interpret the phenomena of motion [3]. Particularly, these schemas resemble those developed by historical figures such as Aristote [4]. These schemas are completely adapted to the common life tasks: drive a bike, catch an object, etc. However, these schemas may differ from scientific concepts. In certain cases, these schemas may even interfere with learning, especially if the teacher does not take them into account. In that case, there is great danger that the pupils differentiate school knowledge, which works in the school (for instance, in the laboratory), of daily knowledge, who allows them to react with effectiveness to events of the common life [2]. Secondly, during laboratory activities, kinematics is often approached with the aid of a mathematization to which the pupils are not accustomed [2]. For instance, a common pedagogic technique consists in bringing the pupils, at the beginning of the study of kinematics, to the laboratory where they measure different properties of motion which they then put in graphs. Back in class, they analyze their results and perform calculations with the aid of mathematical expressions to get the values of the speed and acceleration. And yet, it appears that the pupils perform these various operations without a real understanding of what they are doing [2,5]. In this respect, during the laboratory activities, it seems that the pupils do not have enough opportunity to propose their own hypotheses [2,6,7]. Indeed, a study of protocols offered by the science laboratory textbooks in

Quebec shows that the pupils have seldom the opportunity to get involved in genuine research, since textbooks laboratory protocols put generally emphasis on the procedures of data collecting and analysis [8].

To overcome these difficulties and propose more genuine investigation of kinematical phenomena, the use of technology would make easier the data collecting and analysis while supporting the pupil in his investigation of scientific phenomena [9]. In this approach, called 'video-based laboratory' or VBL, the motion of objects are recorded as videos, treated by softwares allowing at the same time the measure of the positions of objects according to time and the organization of these data in tables and graphs. Such an approach has several advantages: 1) it allows the pupil to focus on the generation of hypotheses and the interpretation of results, two skills not much developed in traditional laboratories [10]; 2) it allows the pupil to generate and to prove several hypotheses much faster, by making easier strategies of variation of parameters necessary for the formulation of hypotheses regarding the properties of phenomena [11,12]; 3) in physical situations where it is necessary to come back on the results of an experience to check the accuracy of the results obtained or possibly to change the original hypothesis, VBL can allow the traditional laboratory to become iterative in spite of the school constraints with respect to time or equipment. In effect, it is often necessary for the pupil to come back on the results of an experience to study the reasons of the gap between his ideas and his experimental results, thus favoring conceptual change in sciences [13,14].

III. CONCEPTION, IMPLEMENTATION AND EVALUATION OF THE VIDEO-BASED LABORATORY

As regards the activities of conceptual change of kinematics phenomena, we conceived them in order to study the characteristics of constant speed motion. In the first section, we describe how we conceive the video-based laboratory. In the second section, we describe how the planned activities went ahead in the classroom. In the third section, we present the comments of both the pupils and their teacher about how they evaluate the activities. Finally, in the last section, we draw limits and propose improvements to the VBL to envision its future scope.

A. Conception of the video-based laboratory

To allow the pupils to work in small groups of four or five persons, we conceived a guide of activities allowing supervising the steps of the pupils. The guide introduces two cases of constant speed motion to study different aspects of this type of motion. These two cases consist of activities (questions, graphics to draw, etc.) which guide the modelling process of the pupils. The process of conceptual change is structured as a POE task (Prediction> Observation> Explanation) [15]. Every POE task takes place in the following way. A physical situation represented under a concrete form by a physical set-up is explained to the pupils in the guide. Questions linked to this case ask the pupil to predict what is going to arrive if experience is to be performed and to write them in their notebook. Once written, they compare their predictions with their peers in small group discussions to reach a general agreement. Upon agreement, each group send their

representative to present their predictions to the whole class. In this step, teacher acts as a facilitator, asking questions to get pupils clarify their ideas. Then, when all teams have presented their ideas, the teacher demonstrates the phenomenon in front of the pupils and records it under video form with the help of pupils' volunteers. These videos are then transferred to USB keys and distributed to every team. Having inserted these sequences of pictures in the REGAVI1 software, the pupils of every team can then, with the aid of a cursor, take measures of the successive positions of the ball according to time. These measures are automatically put in tables by REGAVI. Later, these tables can be transferred for analysis to the REGRESSI2 software [16]. This last software possesses functions allowing the pupil to produce different graphs of position and speed according to time. As such, REGRESSI makes easier the discovery of relations between variables by providing means to compare the adjustment of different curves (linear, quadratic, exponential, etc.) in gathered data. At last, the pupils try then to explain the gaps, if need be, between their predictions and their results.

The role of the teacher in the VBL is to introduce activities to the pupils, to allocate roles to the pupils during small group discussions, to perform the demonstration of every case of motion in front of the pupils, to record these movements under video form and to distribute them to the pupils and, finally, to make easier exchanges between the pupils during whole class discussion [7]. In order to do so, we planned a training period of two hours duration, conducted by the main researcher, where the teacher could master the elements of the VBL and practice the skills to conduct discussions efficiently [7]. As regards the adaptation of the approach to the schoolroom, both the main researcher and the teacher met, throughout research, to undertake adjustments requested according to the evolution of pupils' understanding. Besides, two supplementary meetings prior to the implementation allowed the teacher to gain knowledge of the approach and the main researcher to add modifications in order to adapt the activities to the context of the school and to the characteristics of the pupils.

Regarding uniform straight motion, the first case subjected to the pupils is represented in the guide in the following way: « A ball is thrown on a horizontal rail. The circle in grey points out its initial position at the time of launching. The circle with symbol 1 inside points out the position of the ball after 1 second (see fig. 1) ».

---

[1] The REGAVI software allows the collection of data from a video of an object in motion in form of AVI file. This software contains functions allowing the measure of the successive positions of this object which it organizes in form of tables. It is possible later to transfer these data in the file REGRESSI to aims of analysis (see the following site: www.micrelec.fr/equipelabo/pics_art/pdf/M0314G26.pdf · PDF file).

[2] The REGRESSI software accomplishes Cartesian graphs of data transferred to a data collection software such as REGAVI. REGRESSI also contains functions allowing to calculate new variables (speed, acceleration) from the measurements of position and of time, to find the best curve of group of experimental points, etc.
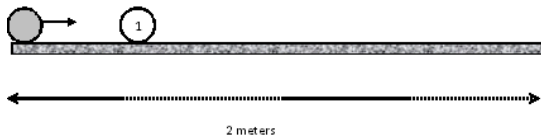
Figure 1.   Motion of a ball traveling a horizontal rail.

The POE task consists then in predicting what will be the successive positions of the ball in the following seconds, knowing the distance gone through in the first second. From these predictions, the guide asks the pupils to draw a graph of what would be the curve of position according to time. Later, the pupils observe the motion of the ball when their teacher demontrates the phenomenon in front of them. After doing it several times, the teacher records the motion of the ball with the digital camera as a video. The video is then distributed to the different teams for data collection and analysis. These different steps allow the pupils to produce graphs of position and speed according to time in order to compare their predictions with their results and to change their hypotheses, if need be.

Regarding the second case of uniform straight motion, it is presented to the pupils in the guide in the following way: A light impulse is given to the ball (A) so as to make it move from an end of a horizontal rail of 2 metres to the other. The position of the ball (A) after 1 second is pointed out in figure 2. Some seconds later, a second ball (B) is pushed with a bigger impulse than the ball A on a parallel rail to the first (fig. 2). With respect to the prediction, the guide asks the pupils to draw by hand on the figure 2, the positions he expected the balls (A) and (B) will be during the seconds following the impulsion. They also have to answer the following question: Do you think that the ball (B) will catch the ball (A)? If yes, they have to point out by one X the place where the two balls may go side by side and explain their answer. The steps of experience and explanation take place as in the first case described before.
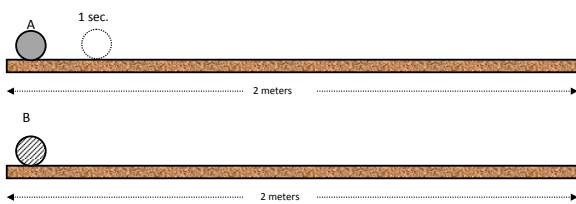
Figure 2.   Motion of two balls moving at different speeds on two horizontal parallel rails

B. Unfolding of activities

Our pilot study consisted of a class of 32 French-speaking pupils attending a physics course in a high school of the province of Ontario in Canada. Research took place at the beginning of the second semester. It is to note that the school where took place the research had adopted a calendar where lessons, which normally stretch over all school year, were condensed in one semester. Consequently, the lessons of the first semester differed from the lessons of the second

semester3. To study the implementation of the VBL, the main researcher held a research diary where he recorded his observations on the sequence of events, the critical details regarding the introduction of the VBL by the teacher, comments of the teacher in meetings with the main researcher, and links that the main researcher could establish between his observations and the theoretical framework of the present research [17]. During the experimentation, the main researcher or one of his research assistants were present at each of the periods to observe the upfolding of the events and take the measures of pupils' understanding. The main researcher or one of his assistants also played the role of monitor of laboratory, to solve the difficulties which may arise with the experimental set up or the use of data and analysis software by the pupils.

Activities took place in a physics laboratory distinct from the schoolroom. This place was suitable for our research, because it had several mobile tables that could be regrouped in small islets where the pupils could work in small teams. A unique set up allowing the demonstration of the cases of motion was located at the front of the class at short distance from the islets. The VBL described in the previous paragraph was implemented in the 11th grade physics course during five successive periods of one hour and quarter at the very beginning of the second semester. During the period immediately preceding the beginning of research, fifteen minutes had been dedicated at the end of the period to introduce the research to pupils.

As a matter of fact, during these last fifteen minutes, the teacher introduced the pupils to the research objectives, briefly described the activities and the time it would take and answered their questions. Furthermore, the teacher mentioned that these activities were part of the curriculum, that he was going to animate them with the materials provided by the main researcher and as a result all pupils had to participate in it. The teacher then divided the pupils in eight groups of three to four pupils of various skills. Moreover, the main researcher distributed to the pupils a letter of consent which their parents had to sign and returned to the main researcher at the end of the five periods dedicated to the research. At the end of the prior period, five minutes were dedicated to the completion of a quiz where the pupils had to answer a question testing their initial understanding of kinematics.

During the first period, the teacher present to the pupils the physical set up of the first case (see fig. 1), without performing the demonstration nevertheless. He asks them to write individually in their guide their predictions, i.e. to draw the successive positions of the ball on the horizontal rail as well as to draw what they expect would be the position-time and speed-time graphs. Then in small teams, the pupils discuss their answers to reach a general agreement which they write on large paper sheets. Once students have ended this last step, the teacher asks a representative of every team to present their predictions to the whole class. During their presentation, the teacher asks them questions to clarify their ideas. Because of time constraints, only the first six teams could make their

---

[3] In Canada, the school year is divided into two semesters each of a length of four months.

presentations. During the last five minutes, the pupils answer a question of understanding.

During the second period, the teacher invites the last two teams to present their predictions to the class. After these presentations, the teacher then calls two pupils to volunteer for doing the demonstration of the phenomenon. All groups remain attentive except for a group located at the back of the class. The pupils point out very quickly that the rail is not perfectly horizontal and issue hypotheses on the reason for which it would not be so; in this respect, the pupils call into question the precision of measurements and the horizontality of the table. The ball is then lightly thrown by the volunteers so that the class can better notice the motion of the ball. The teacher mentions then to the pupils that it is necessary, besides observing the motion of the ball, to take measures of positions according to time if they want to prove their hypotheses and to compare them with their predictions. In order to do so, the teacher, helped by two volunteers, records a video of the motion, which is then copied on USB memories by one of the research assistants and later given to every team so that the pupils can download the video on their computer. The teacher announces to the pupils that they are now going to analyze data with the aid of the software REGAVI and REGRESSI to prove their ideas on the motion. In this respect, it was planned to broadcast with a projector the tutorials to the whole class. Because of technical problems, we asked the pupils to listen to the videos in group of two4. Our observations point out that the pupils are engaged in the listening of tutorials, that some of them take notes, that they explain software functions to each other and that they begin at the same time studying the motion recorded on video. Once the REGAVI software is understood, the pupils undertake the measurement of the positions of the ball according to time to constitute a table. They export this table towards the REGRESSI software where they produce both the graphs of position and speed according to time. In this respect, most of the teams could obtain their position-time and speed-time graphs which they record on their USB key and give it back to the research assistants. Upon reception of these, the assistants print for every member of the team both graphs which will be analyzed in the next period. Since this step is critically important for the activities of the third period, the research group and the teacher decides to report the measure of the understanding planned at the end of period to the beginning of the next period.

As a consequence, at the beginning of the third period, the pupils answer a question of the quiz. Then the teacher distributes the graphs of position-time and speed-time produced the period before to the pupils. He mentions them that they have to answer the questions of the explanation part of the first case. He passes from one team to the other to supervise their work. With the aid of a projector linked up with a computer, the teacher makes a review with the pupils of their answers by choosing the graphs produced by the pupils. The pupils answer the questions of the teacher with respect to these graphs. In the second half of this period, the teacher present the set-up of case 2 (see fig. 2) to the pupils. The sequence of the activities of the

second case is similar to the first case. The study of case 2 extended from the second half of the third period to the middle of the fifth period (that is during two periods as the first case). During the second half of the fifth period, the research assistant animates a discussion with the various teams, asking pupils prepared questions about benefits and difficulties they have experienced with the activities and gathers their comments with a recorder.

C. Evaluation of the implementation of the VBL in the classroom

The implementation of the VBL was assessed by the pupils and by the teacher with the aid of semi-directed discussions envisaged at the end of research. As regards the evaluation by the pupils, their comments and suggestions were gathered during discussion in small groups by one of the research assistants. As regards the evaluation of the activities by the teacher, his comments and suggestions were gathered during an interview conducted by the main researcher at the end of activities. We shall analyze the comments of the pupils in the first subsection and those of the teacher in the second.

1)   *Evaluation of the activities by the pupils:* The pupils made comments about the following topics: the benefits of the activities, the difficulties they had to face, the easiness of use of technology (digital camera, soft wares of collection and analysis of information), the pertinence of activities in relation to the study of the concepts of motion, their suggestions or recommendations to improve the activities, their criticisms and the limits they saw with respect to the activities. As regards the benefits of activities, a pupil mentioned that it was good for visual persons and those who liked technology. Several pupils mentioned that they acquired new knowledge about motion, for example that the ball, contrary to their expectations, rolled at constant speed on the horizontal rail. Besides, some pupils could make links with what they had already known, notably with the notions of graph and slope which they had seen in the mathematics lessons. Another pupil mentioned the concrete character of activities which he preferred to written exercises because he could manipulate and touch. Another pupil found activities more interesting than to be in class. Regarding activities, a pupil mentioned that he had made physics without realizing it. Finally, a pupil mentioned that he had learnt to use computer programs and to insert them into physics. As regards the pertinence of activities, almost all pupils mentioned that activities were appropriate to learn motion, that they had learnt a lot and that this was going to be of use for them in their physics course.

2)   As regards encountered difficulties, a pupil mentioned that he would have preferred making more experiments rather than to waste his time explaining hypotheses. In the same vein, another pupil lamented that he had not sufficient knowledge to compare with laboratory results. Several pupils mentioned difficulties in the collection of data: 1) the use of the paving rather than the mouse made the measurement of the various positions of the ball long and difficult; 2) balls being the same colour as the rail, it was sometimes difficult to differentiate between them, and that generated measurement errors. For the same reason, it was also difficult in the second case to differentiate between both balls when they were in the same

---

[4] To increase participation of all pupils in the data collecting and analysis, two computers were allocated to each team of four pupils so that they could work by subteams of two.

position. In this respect, some pupils suggested to use balls of different colour to make measurement easier. The majority of the pupils mentioned problems associated with the tutorials to which they reproached their lacking suppleness. Hence, while using them, the pupils could not make a pause or come back to listen again to an explanation. According to pupils, these tutorials contained too many steps to follow, so they had to listen again to the whole band if they had not understood the first time. Moreover, some pupils experienced difficulties in producing graphs with the REGRESSI software.

3) *Evaluation of activities by the teacher:*

During the interview at the end of research, the main researcher and the teacher discussed about the same topics as during small group discussions with the pupils (see previous subsection). Above all, the teacher spotted indications that the pupils learned kinematics because, for example, they could express themselves in their own words and they were clear regarding what they meant about motion. From day to day, the teacher could see the progress of his pupils because their questions were more definite, their comments more appropriate. According to the teacher, the pupils were more advanced than ordinarily in the study of kinematics, a sign that they progressed tremendously. Certain pupils spoke about the force of gravity. A pupil even linked up the force of gravity with friction. According to the teacher, the pupils used force notions which were not necessarily what they needed for the study of motion, but that were linked up in fact, but they went farther than when we teach in a lecture. Therefore, according to him, the activities have the pupils question themselves, discuss so that they get a broader overview in fact than what they have habitually in kinematics. Consequently, during the second case, the teacher noticed that several teams had succeeded in predicting the outcome of the phenomenon (see previous subsect.).

As regards the experimental method, the teacher noticed that the pupils discussed even during pauses or before the lessons of their predictions and their hypotheses concerning motion. They exchanged on their predictions in form of hypotheses and on the manner of proving them. According to the teacher, it is not necessarily what the pupils do in the physics course, partly because of the lack of equipment, but also because of the emphasis put on problem solving. However, during the VBL activities, the pupils could apply the scientific method. Besides, the teacher is of the opinion that the pupils gained knowledge of the use of technology in the laboratory, taken into account the graphs which they produced and that were well made. As regards the curriculum, the teacher considers that the activities were in line with the objectives associated with uniform motion. In effect, according to him, the pupils did graphic analysis, approached the mathematical notions of slope and studied notions to understand uniform movement, as well as factors which influence it, and as a result they covered practically about half of the unit in motion. With respect to the follow up to be given to the VBL activities, the teacher plans now to make a quick flyover because he considers that the pupils had covered the majority of the notions of uniform straight motion.

According to the teacher, the difficulties met by the pupils consisted especially in gaining knowledge of the software of data collection and analysis. But once this objective was achieved, the teacher realized that, especially during the second case, the pupils could perform the same tasks (collection and analysis of data) as in the first case in only fifteen to twenty minutes. As regards the large number of pupils for a science class (32), the break up in small teams helped a lot as well as the arrangement of tables in the form of islets. This disposition allowed the pupils to work well all the more so as they were divided in team of different skills, some pupils having more skills in sciences, others in communication. This sharing out in different skill groups was preferable according to him to groups of pupils urged just to calculate. As regards technical difficulties, they were minimized by the focus placed upon the study of only two cases, which simplified the task of the teacher to help the pupils use the software.

D. *Future scope*

Since this part is concerned mostly with envisioning what would be the limits and possible range of the VBL, the comments gathered are mainly the teacher say. As such, the pupils mainly comments on the concrete aspects of the VBL, their difficulties encountered with the softwares, with the data collection and analysis, etc. Regarding improvements, the teacher suggests introducing the pupils to the data collecting and analysis softwares with the aid of already available experiments store into a data bank. The pupils could so gain knowledge of the softwares before the experimentation really begins. Thus, it could save time during the first activities and help the class keep a good rhythm with respect to content coverage. This training would not necessarily be long, at the very most thirty minutes. Moreover, he considers that tutorials representing and explaining the functioning of software helped a lot the pupils. With respect to the guide, he noticed that the pupils have the impression that the questions repeat themselves. However, the teacher is not sure that his pupils did not understand either the sense of the questions or the concepts itself. Taken into account this situation, the teacher explains that, for some of his pupils, French is not their mother tongue so that the shape of sentences could put them off. Besides, the cultural aspect could influence their understanding of questions because certain pupils do not understand certain terms which bear contextual, cultural or social references.

As regards the limits of activities, the teacher mentions that the time constitute the main limit. Indeed, he thinks that to dedicate five periods to this type of activities is really appropriate, but that if it had to persist, he fears losing the interest of the pupils, as besides for any other type of activities. He considers that after some time, it becomes repetitive, the set-up, the equipment. However, he noted that even so the pupils participated till the end. As regards the incorporation of activities in the curriculum, he considers that they are well integrated with the curriculum. However, the teacher pointed out a problem concerning the perception of the pupils who don't consider these activities being real learning. This is in line with some pupils comments about how they had learnt a lot and that it would be beneficial to them in their physics course. In this respect, the teacher wonders if to persuade them, they should be assessed midway so they could see their progress. According to the teacher, it is probable that the pupils are not accustomed to this method and it is only later they are

going to realize that they understand all this. However, noticing the manner the pupils discussed with him, the way they answered questions in the guide, how they worked, he could point out the progress they made. Perhaps according to him, the activities worked so well that the pupils had not the impression that by making less effort they could understand the concepts better and better.

Even if at the beginning, the teacher had concerns about possible disturbances research could have to the normal sequence of activities and about what this research could bring him as a teacher. However, he considers that because of the adaptation, the activities took place in a natural manner. As a matter of fact, the teacher considers that activities were advantageous to him as well as to his pupils and constitute a very good introduction to kinematics. Moreover, our observations confirm that the approach was well suitable for the style of the teacher, notably he demonstrated the faculty to ask good questions to the pupils, to grab in flight their comments and that in general he controlled and managed activities well.

IV. METHODOLOGY

Our pilot study consisted of a class of 32 French-speaking pupils, that is 18 girls and 14 boys, of a 11th grade physics course given in a high school of the province of Ontario in Canada. The pupils of this class had chosen a science orientation program of studies given by the school so that it is assumed they are interested by sciences. The teacher involved in this research had a teaching experience of 20 years and already tested a similar strategy the previous year [18]. The approach described in the previous section took place during five successive periods of one hour and quarter at the very beginning of the second semester. Besides, two periods were added to the implementation of the VBL, a prior period to introduce activities to the pupils and a subsequent period to take an additional measure of pupils' understanding of constant speed motion. In each of these periods, the pupils had to answer a question taken haphazardly in a bank of problem during five minutes of the period. This disposition allowed us to follow the evolution of the understanding of every pupil throughout the experimentation of the strategy of the VBL.

To measure the evolution of understanding according to the number of periods dedicated to the use of the VBL, we conceived a test of understanding of the constant speed [19]. Since every pupil is measured at several occasions, conditions in which these measurements are taken may vary, that it is in the day of the week, the hour during day, etc. So, a pupil may not get the same result in answering questions of identical difficulty because he is tired or irritated during a particular occasion. In such a case, where the temporal dimension is important, the choice of an item response model must take into account the variations in the course of the time of the answers of the pupils. As such, the model with facets developed by Linacre [20] allows considering the influence of these different factors or 'facets' on the measure of understanding. The calculation of the different parameters linked to these factors (or facets) is made by the model equation which links up the values of parameters with observations [20]:

$$\text{Log } [(P_{nijk}) / (P_{nijk-1})] = (B_n + T_i) - (D_j + F_k) \qquad (1)$$

where:

P_{nijk} is likelihood that the pupil 'n' sees itself granted a level 'k' at occasion 'i' when he answers question 'j'.

P_{nijk-1} is likelihood that the pupil 'n' sees itself granted a level 'k-1' at occasion 'i' when he answers question 'j'.

B_n is the skill of the pupil 'n'.

T_i is the more or less easiness with which the pupils answer at occasion 'i'.

F_k is the difficulty linked to the jump of level k-1 at level k.

The various properties of the Facets model make it an appropriate tool for our analysis [20]. Firstly, the various parameters calculated by the model from observations have the properties of an interval scale. Secondly, the model allows the calculation of values for all pupils during all occasions of measure, including missing data, which increases the power of the statistical tests used. In order to do so, the model calculates for every pupil a value of the logarithm of its likelihood to produce a correct answer to the question subjected to each of the periods of the experimentation. To determine if the VBL had an effect on the understanding of the pupils, a repeated measures analysis of the variance was performed of the computed values. This analysis allows us to compare the results of the pupils between the different occasions of measure and to determine if one of these results differs significantly from the others. This comparison can take a specific form called contrast [21]. To prove our research hypothesis, it is necessary to determine if the increase of understanding according to the number of periods of the implementation of the VBL is linear. As such, it is possible, with the aid of orthogonal polynomials to separate the contributions from the linear tendencies and higher polynomials (quadratic, cubic, etc). Besides, these elements of variance being independent to each other, they can be separately tested [21]. Nevertheless, the comparison of every pupil in comparison with himself may be correlated. This possible dependency between repeated measurements of the same pupil means that the repeated measures analysis of variance cannot be used unless the condition of sphericity is satisfied [22]. This test, called also Mauchly's test, imposes certain conditions on the matrix of variances and covariance [23]. When the result of the test of Mauchly exceeds the threshold alpha of 0.05, we can consider that the condition of sphericity is satisfied and as a result that the analysis of variance by repeated measurements can be used.

V. RESULTS

The figure 3 introduces the average values of understanding acquired by the pupils at every occasion of measurement, each value being calculated with the help of equation (1). By inspecting figure 3, we note that, in spite of certain variations, there is a positive tendency in the increase of understanding. This linear tendency is significant at the threshold alpha 0.05 ($p < 0.02$) according to the results of the analysis of the variance of the table 1.

Moreover, we also notice the presence of significant tendencies of order 4 and 5. Let us note that the test of Mauchly ($p = 0.98$) points out that the condition of sphericity is satisfied

and as a result that the analysis of variance by repeated measurements can be used. Finally, the omnibus test of the variance of repeated measurements points out that there is at least one of the averages which differs significantly from the others (p <0.000) at the threshold alpha 0.05.
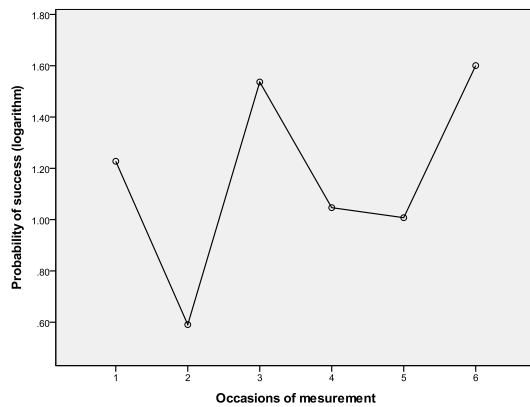
Figure 3.   Average likelihood of success (logarithm) according to the number of periods

TABLE 1
Analysis of the variance of repeated measurements

Source	Trend	Sum of squares	D.f.	Mean Squares	F	Sig.
Factor	Linear	3.153	1	3.153	6.03	.020
	Quadratic	1.861	1	1.861	2.79	.105
	Cubic	.146	1	.146	.31	.579
	Order 4	11.703	1	11.703	14.98	.001
	Order 5	5.550	1	5.550	13.04	.001
Error	Linear	16.218	31	.523		
	Quadratic	20.691	31	.667		
	Cubic	14.460	31	.466		
	Order 4	24.212	31	.781		
	Order 5	13.194	31	.426		

VI.   DISCUSSION

Our research is inspired by the constructivist approach, where the pupil constructs his knowledge by interacting with his environment. Given the difficulties pupils encountered in the physics course to gain conceptual understanding of kinematics concepts, we conceived a video-based laboratory (VBL), a specialized version of the usual computer-assisted laboratory, allowing the pupils to generate and prove hypotheses with the help of data collection and analysis softwares while discussing in small groups about videos of objects in motion.

Concerning the implementation of the VBL, the comments gathered during interviews with the pupils and the teacher show that these activities were advantageous to the pupils in their understanding of the different aspects of motion at constant speed. Some pupils appreciated the concrete character of activities which they prefered to lectures. A pupil even mentioned that he had made physics without realizing it. In the same vein, they could make links with mathematical notions such as the production and the interpretation of Cartesian coordinates as well as the calculus of the slope of a line. Moreover they familiarize themselves with the use of data collection and analysis softwares in the physics laboratory. Besides, it seems that the presentation format of motion phenomena as demonstrations did not prevent the pupils from considering these activities as laboratories [24].

However, improvements remain to be made with respect to the tutorials aimed to familiarize pupils to the collection and analysis software, notably by increasing their easiness of use (including pause, step and backwards functions) and by shortening the number of steps covered. In this respect, the teacher suggested to introduce the pupils in the handling of the softwares by giving them a prior training of around thirty minutes with the aid of a bank of activities already recorded under video. Perhaps one of the main obstacles to be overcome is the perception of the pupils regarding these activities. According to the teacher, the pupils did not have the impression to accomplish learning activities but rather consider them more as preparation to the physics course. This impression draws its source perhaps in the conceptions pupils may have of what constitutes learning in a physics course, constituted mostly by problem solving activities, as explained by the teacher. Besides, the emphasis put on the discussion of hypotheses between pupils runs opposite to pupils' conceptions of the role of laboratory in physics courses that mainly attempt to verify theories [25]. Moreover, according to the teacher, the time dedicated to this type of activities should not exceed five periods, owing to the lassitude which could manifest itself after some time. Finally, since discussion is one of the important elements of this approach, a particular attention must be turned to the linguistic and cultural aspects that can influence pupils' learning [26].

Concerning the effect of the VBL on the understanding of concepts of constant motion, our results point out that the curve of understanding according to the number of periods dedicated to the laboratory shows a positive and significant linear tendency. These results can point out that the VBL such as conceived has an effect on the understanding of the pupils. Indeed, observations noted in the research diary by the main researcher and the comments of the pupils and the teacher expressed during post-research interviews tend to corroborate these results. As such, the teacher mentioned that he noticed a constant progress in the understanding of the pupils of concepts of motion by the type of question which the pupils put down, by the quality of their explanation on the properties of motion phenomena, by their interpretation of the graphs, etc. However, the presence of significant tendencies of higher degree polynomials (of order 4 and 5) may point out to important fluctuations in results from one occasion to the other. These fluctuations can be caused by imprecision in measurement, variations in the interest and attention of the pupils according to the instant of the day when the measure was taken. It is also possible that these fluctuations are the result of swings linked up to transitions between levels of understanding [27].

VII. CONCLUSION

This study undertaken with a single group of pupils cannot pretend to formulate conclusions that can be generalized to all high school pupils. As a result, these conclusions have a speculative character and are to be considered in the light of the exploratory aim of our study. This research adopts the perspective that the usage of computer science in the physics laboratory is revolutionizing the education of this discipline. However, computer-assisted experimentation is too often dedicated to the technical side of the data collecting and organization in form of tables and graphs. This emphasis on the technical precision of measurements, in spite of his rigor, risks of making us forget that it is often necessary for the pupils to develop their qualitative reasoning as well as their quantitative counterpart. However, it is not a question of leaving out the mathematization of the properties of phenomena but to consider it only when the essential elements of problem are qualitatively understood by the pupils. The approach introduced here use the capacities of the computer so that the pupil can, from common sense reasoning of a qualitative nature about properties of motion phenomena, make the transition to a mathematical representation in form of graphs of position-time and speed-time. Researches undertaken with several groups of pupils may help to confirm results obtained in the present study [28].

REFERENCES

[1] R.E. Anderson, "Reforming science teaching: What research says about inquiry". Journal of Science Teacher Education, vol. 13, no 1, pp. 1-12, 2002.

[2] A.B. Arons, A guide to introductory physics teaching, 2nd ed. Toronto: John Wiley & Sons, 1997.

[3] R.D. Knight, Five Easy Lessons: Strategies for Successful Physics Teaching. San Francisco: Addison Wesley, 2004.

[4] F. Espinoza, "An analysis of the historical development of ideas about motion and its implications for teaching", Physics Education, vol. 40, no 2, pp. 139-146, 2005.

[5] G. De Vecchi, Enseigner l'expérimental en classe: Pour une véritable éducation scientifique. Paris : Hachette, 2006.

[6] P. Nonnon, and A. Métioui, "L'appropriation du processus de vérification expérimentale par des étudiants en formation des maîtres au secondaire : exemple de la diffusion de la chaleur dans un liquide", RES (Revue de l'Enseignement Supérieur), Academica, vol. 21, no 1, pp. 39-61, 2003.

[7] L. Trudel and A. Métioui, "Influence d'une discussion préalable sur la participation des élèves dans un laboratoire de physique du secondaire". Conférence Internationale Éducation, Économie et Société, Novotel Paris Tour Eiffel, 17-19 Juillet 2008, CD-ROM, pp. 410-424, 2008.

[8] A. Métioui and L. Trudel, "Analyse critique des expériences proposées dans les manuels destinés aux jeunes de 8 à 12 ans: Magnétisme, électrostatique et circuits électriques" . CD-ROM : Critical Analysis of School Science Textbooks, P. Clément (ed.), IOSTE International Meeting Tunisia Hammamet, 7 to 10 February, 12 pages, 2007.

[9] D. Jonassen, J Strobel and J. Gottdenker, "Model building for conceptual change", Interactive Learning Environments, vol. 13, no 1-2, pp. 15-37, 2005.

[10] C. Gianino, "Microcomputer-based laboratory for Archimedes' principle and density of liquids", The Physics Teacher, vol. 46, pp. 52-54, 2008.

[11] E. Koleza and J. Pappas, "The effect of motion analysis activities in a video-based laboratory in students' understanding of position, velocity and frames of reference", International Journal of Mathematical Education, vol. 39, no 6, pp. 701-723, 2006.

[12] M. Riopel, Conception et mise à l'essai d'un environnement d'apprentissage intégrant l'expérimentation assistée par ordinateur et la simulation assistée par ordinateur. Thèse de doctorat, Montréal, Université de Montréal, 2005.

[13] L. Trudel, Impact d'une méthode de discussion sur la compréhension des concepts de la cinématique chez les élèves de cinquième secondaire. Thèse de doctorat, Montréal, Université du Québec à Montréal, 2005.

[14] J.Y. Lin, "Responses to anomalous data obtained from repeatable experiments in the laboratory", Journal of Research in Science Teaching, vol. 44, no 3, pp. 506-528, 2007.

[15] D.W. Russell, K.B. Lucas and C.J. McRobbie, "Role of the Microcomputer-Based Laboratory Display in Supporting the Construction of New Understandings in Thermal Physics", Journal of Research in Science Teaching, vol. 41, no 2, pp. 165-85, 2004.

[16] G. Durliat and J.M. Millet, "L'informatisation des dosages phmétriques avec l'interface Orphy et le logiciel Regressi, EPI, vol. 64, pp. 163-172, 1991.

[17] H. Altrichter and M.L. Hollly, "Research Diaries", In Research Methods in the Social Sciences, chap. 2, B. Somekh et C. Lewin, Eds. Thousand Oaks (Californie) : Éditions SAGE, 2005.

[18] L. Trudel and A. Métioui, "A Second Order Cybernetic Model of Scientific Conceptual Understanding : The Case of Kinematics", Journal of Systemics, Cybernetics and Informatics, vol. 10, no 1, pp. 7-17, 2012.

[19] L. Trudel, C. Parent and R. Auger, "Développement et validation d'un test mesurant la compréhension des concepts cinématiques en physique au secondaire". Mesure et évaluation en éducation, vol. 31, no 1, pp. 93-120, 2008.

[20] T. G. Bond and C. M. Fox, Applying the Rasch model : Fundamental measurement in the human sciences, 2nd ed. Mahwah (New Jersey) : Lawrence Erlbaum Associates, 2007.

[21] D.C. Howell, Méthodes statistiques en sciences humaines. Bruxelles : De Boeck, 2008.

[22] L. K. Edwards, Applied analysis of variance in behavioral science. New York : Marcel Dekkers Inc., 1993.

[23] K. Leik, Experimental design and the analysis of variance. Thousands Oak (Californie) : Pine Forge Press, 1997.

[24] Roth, Wolff-Michael, Campbell J. McRobbie, Keith B. Lucas and Sylvie Boutonné, "Why may students fail to learn from demonstrations? A social practice perspective on learning in physics", Journal of Research in Science Teaching, vol. 34, no 5, pp. 509-533, 1997.

[25] M. Larochelle and J. Désautels, "Of course, it's just obvious': adolescents' ideas of scientific knowledge". International Journal of Science Education, vol. 13, no 4, pp. 373-389, 2007.

[26] L. Rivard, "La langue : Outil pour construire son savoir en sciences", Spectre, vol. 28, no 1, pp. 32-36, 1998.

[27] N. Granott, "How microdevelopment creates macrodevelopment: Reiterated sequences, backward transitions, and the Zone of Current development". In Microdevelopment: Transition Processes in Development and Learning, N. Granott and J. Parziale, Eds. Cambridge: Cambridge University Press, 2002, pp. 213-242.

[28] R.E. Slavin, Educational research in the age of accountability. Boston: Pearson, 2007.

AUTHORS PROFILE

Dr. Louis Trudel is assistant professor at the faculty of education at University of Ottawa, Ottawa, Canada. He was awarded a Ph.D. in education at the Université du Québec à Montréal, Montreal, Canada. Dr. Trudel has published several papers on science education and the pedagogical applications of computer science to the field of education.

A. Métioui received his Ph.D. in didactics from Laval University, Canada, in 1987. He is currently a professor at the Département de Didactique, Université du Québec à Montréal, Montréal, Quebec. Dr. Métioui has several publications to his credit. His research interests involve teachers' and student's representations in science and technology, and the design of the interactive computer environments that take into account the students and scientific representations related to physics.

QoS Parameters Investigations and Load Intensity Analysis, (A Case for Reengineered DCN)

[1]Udeze Chidiebele. C, [2]Prof. H. C Inyiama

[1,3] R & D Department, Electronics Development Institute (FMST-NASENI), Awka, Nigeria.

3 Okafor Kennedy .C, [4]Dr C. C. Okezie,

[2,4] Electronics and Computer Engineering Department, Nnamdi Azikiwe University, Awka, Nigeria.

Abstract—**This paper presents the simulation results on Reengineered DCN model considering Quality of Service (QoS) parameters in a homogeneous network for enterprise web application support. To make it feasible for the computer aided design (CAD) tool, a complete and robust real performance assessment of simulation scenarios is considered. This paper deals with a performance measurement and evaluation of a number of homogeneous end-to-end (e2e) DCN subnets taking into account a wide range of statistics. An active measurement approach of QoS parameters has been adopted for the studying of proprieties we called concise process model QoS. This paper validates the performance of the Reengineered DCN through simulations with MATLAB SimEvent which presents efficient performance metrics for its deployment. Considering the model hierarchy in context, the results shows that with the design model, a highly scalable DCN will be achieved.**

Keywords-QoS; DCN; Model; CAD; Homogenous; Process component.

I. INTRODUCTION

Contemporarily, data centre networks (DCNs) have continued to attract a lot of interest in the enterprise networking market segments [1]. According to [2], a Data Centre is the consolidation point for provisioning multiple services that drive an enterprise business processes. A discrete event based process model of a reengineered data center communication network comprising of four data center LAN nodes with an inclusion of a virtualization server for creating virtual instances of resources and applications for DCN nodes was analyzed and presented in [3]. The perceived quality of TCP and UDP services on the reengineered DCN is primarily determined by parameters like packet loss, delay, jitter, throughput, and available bandwidth. For this purpose, several reference documents containing constraints regarding these parameters have been defined [4], [5], [6], and research have focused on the effects of these parameters on real time traffic (i.e. telephony) [7]. Understanding the statistics of QoS parameters is important for the appropriate design of DCN algorithms (routing, flow control, streaming), for the evaluation of network capability to support new value-added services (i.e. telephony, multimedia), for the study of network performance, for developing algorithms to detect security vulnerabilities and, finally, for the definition of Service Level Agreements (SLAs).

1. Why do packets drop in switched DCNs? And what are the effects in mission critical applications like: ERPs, Application Servers, E-commerce servers in DCN networks, etc.

2. How do we control traffic flow from access layer sources into switched data center networks without affecting network performance?

3. What is the behaviour of switch buffers at the congestion points, and reaction points under network oversubscription considering the various performance indices?

4. How do we achieve Agility in a Congested switched links for high performance computing?

A lot of works exists in literature that focused on the analysis of QoS parameters over different types of networks and protocols that runs on them [8], [9], [10], [11], [12]. Basically, by expanding distribution core and subnet backbone networks, performance parameters such as delay, utilization, bandwidth, and alongside the Carrier Sense Multiple Access with Collision Detection and Traffic Arbitration Message Protocol (CSMA/CD+TMAP) will create a stable and efficient DCN. This paper shall focus on metrics that will enhance performance in the reengineered DCN setup as shown in our system model in Fig.1.

II. SYSTEM DESIGN AND ASSUMPTIONS

A. System Model

The high performance Reengineered DCN model configuration is shown in Fig. 1. It is assumed that the virtualization subnet is a composite powerful system for resource provisioning in the DCN. As discussed in [3], this paper develops a discrete event based process model and simulates a reengineered data center communication network comprising of four data center LAN nodes with an inclusion of a virtualization server for creating virtual instances of resources and applications for DCN nodes. A T-junction link modeled for 10GPs bandwidth was used. The MAC Controller block implements the CSMA/CD + TAMP (Carrier Sense Multiple Access with Collision Detection and Traffic Arbitration Message Protocol) Mechanism. This access technique of MAC bus system makes for an efficient and flexible data throughput while optimizing bandwidth in the reengineered DCN model. MATLAB SimEvent was employed in the model. Also, the CSMA/CD+ TAMP was modeled into the DCN MAC controllers for full duplex traffic flow while processing queues (frames) in such a manner as to avoid packet losses and congestion in the DCN when scalability consideration arises. Fig. 2 shows the process model architecture developed in this paper. This paper practically details the implementation

methodology as well as discusses the System architecture mechanism in the section IV.

B. Design Goals

In designing our proposed DCN model, the main goals are to maintain throughput, low latency, fast convergence network and to create a cost effective network infrastructure that is based on robust power management infrastructure (SMART GRIDS), hence scalable, flexible with less administrative overhead.

The network should be capable of supporting high performance workloads (HPW) besides conventional ethernet traffic like cloud computing services. Meanwhile, multiple flows from various subnets are localized with VLAN strategies, but still shares link capacity fairly (buffer capacities). In the following, each of these goals is explained in details. The performance of today's DCN's (relevant to multimedia applications) is measured by QoS parameters such as:

1. Throughput or bandwidth
2. Delay or latency
3. Delay variation (delay jitter)
4. Loss or error rate
5. Convergence/Stability

From the application perspective, throughput refers to the data rate (bits per second) generated by the application. Throughput, measured in the number of bits per second, sometimes is called bit rate or bandwidth. Bandwidth is considered to be the network resource that needs to be properly managed and allocated to applications.

The throughput required by an application depends on the application characteristics. Fig. 1 shows the proposed system model developed in [3]. As shown in Fig 2, The process model architecture was implemented with MATLAB SimEvent. The top level model which includes four subnets DCN LAN nodes, each consisting of the following:

- An Application data packet block that models the DCN source data with its interface to a switch and router gateway.
- A MAC (medium access control) controller that governs the DCN's use of the shared channel and link resources.
- A T-junction that connects a LAN node to other DCN subnets networks.
- Switching Path Combiners, Gateway IP Clouds and Sinks.

The packet generation rate and packet size range are specified at the Application data packet blocks, the transmission buffer size specified at the MAC Controller blocks, and the length of the cable specified at the Cable blocks. The terminator, T-junction, and cable blocks at the bottom of the model represent physical components of the network.

C. Experimental Testbed

In this paper, to measure the system performance of the proposed Reengineered DCN infrastructure, MATLAB SimEvent [13] was used to achieve the set objective. We performed our experiments over the real test-bed described in Fig. 2. It is composed of a number of homogenous DCN networks which simplified the abstraction as depicted in Table 1. Over such test-bed, several configurations have been taken into account; indeed, the experiments have been performed by varying a number of configuration parameters like process model blocks, operating system, end user device, access network, transport protocol, and traffic condition. By combining all these variables we obtained about 20 different network conditions. The measurement stage has been performed in the time period between December 2012 and January 2012, in the simulation environment. For our proposed test bed, the infrastructure components include:

i. 4 DCN MAC controller subnets (80 DCN nodes) encapsulated in 4 subnets

ii. DCN Switch/Load balancer

iii. 4 Client Gateway for connection to the IP cloud

iv. Management Server with Virtualization support

v. Power management with SMART GRIDS

vi. Back office Applications (Network Monitoring, Configuration, Authentication, Authorization and Accounting) : Virtualization Management server for virtual instances, central network configuration & Monitoring (Server 2008).

In Fig. 1, the Reengineered DCN model utilizes the infrastructure components outline above. A management server in subnet 3 has the role authenticating and monitoring the overall network for efficient service delivery as well as creating virtual instances for applications in the DCN. The DCN switch enables VLAN segmentation and propagates frames across the various subnets. Also, it interfaces the management virtual server, and DCN nodes, as shown in Fig. 2. The DCN LAN nodes were setup in their various subnets. The bridge switch is a speed redundancy layer supporting virtualization and load balancing. Access to the IP internet cloud is completed by the gateway and modem. After setting up the model, a simulation run was carried out to generate our graphical plots shown in this work (Fig. 3). Also, a consistency test carried out shows that the design model is stable and consistent before the simulation execution.

III. DISCUSSION

As for the active measurement procedure adopted in this work, we used the MATLAB SimEvent tool. By combining different pairs of process block set parameters, we generated a multitude of data for traffic patterns. In this way we generated controlled synthetic data that is realistic in the context of our proposal (Reengineered DCN).
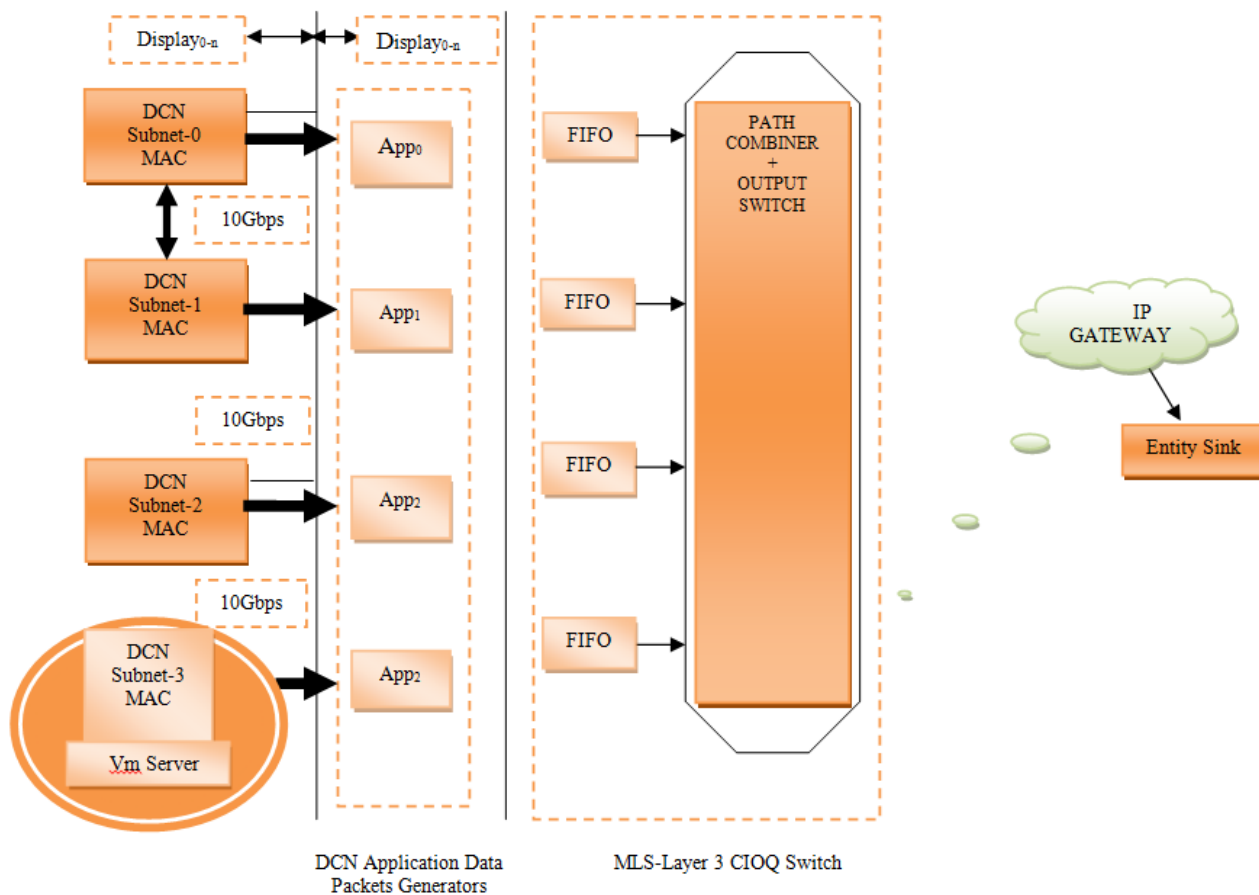
Figure 1. The Process Model Architecture [3]
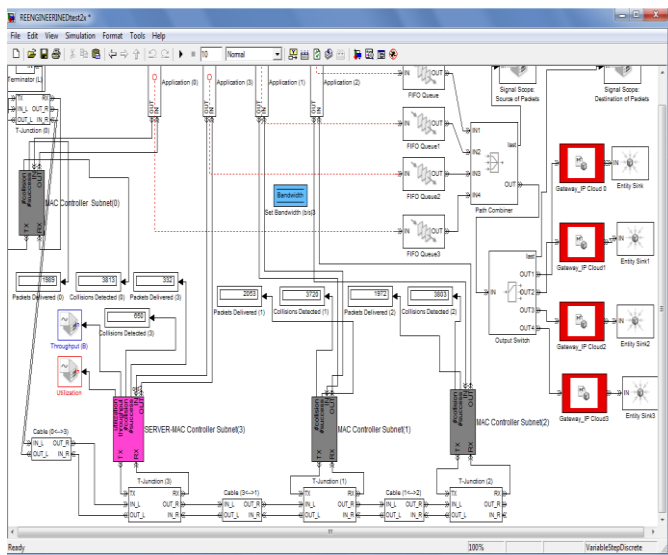
Figure 2. The SimEvent model

Also, in order to draw a reference curve for the parameter outlined in this paper, we considered only data traffic profile generated within our simulation environment. We would like to underline that our experiments have been carried out by using three traffic conditions namely Low, Medium, and High Traffic. For each of them a corresponding QoS is obtained, but interestingly our observation confirmed our model convergence and stability. Due to the nominal bandwidth used in our connections, we considered here only both high and low traffic condition for the communication pattern. We showed the behavior of throughput, latency, utilization, losses for our reengineered model measured over TCP connections.

Finally, it is worth noting that the presented results have been averaged on several tests in order to minimize the effect of random error on measures. Since there no changes in the QoS parameters with respect to variation in other parameters, we argue that our model is very stable.

The Reengineered DCN supports the following functionalities:

i. Parallel LAN and SAN infrastructure.

ii. Service oriented Architecture.

iii. Software as a service (SaaS), Platform as a service (PaaS).

iv. Server consolidation and Virtualization.

v. Dynamic Host Configuration Strategy (DHCS).

vi. Automatic Configuration.

vii. Bandwidth Control/Optimization.

viii. Multiple Services.

ix. Enterprise Security Layout.

IV. SIMULATION PARAMETERS

In this section, we provide simulation results that support section III. We used MATLAB SimEvent [3] to generate the parameters for various case scenarios in the simulations. The model was configured for simulation using the traffic attributes as listed in Table 1.

TABLE I. PROCESS MODEL ATTRIBUTES

Configurations	Values
Simulation Duration for Each Scenario	1Sec-15Secs
Link Propagation Delays	0.5usecs
Switch Output Buffer	100-1400 packets
Simulation Seed	128
Update Interval	50000 Events
Simulation Kernel	OS Optimized

By integrating established and well-known simulation strategies found in different works, we have set up a methodology and provided simulation result data for analysis from our validation tool. The data were captured via the SimEvent to "To Workspace block" which inputs a signal and writes the signal data to the MATLAB workspace [14]. The block writes the data to an array or structure that has the name specified by the block's variable name parameter. Table 2 shows the Re-DCN parameters used for the various load intensities in the process and network models.

A. PERFORMANCE EVALUATION

The Re-engineered DCN is a high performance model that is robust and scalable which is deemed fit for the enterprise domains. Having run the simulation in normal mode and collected the statistics for various cases, the figures below show their analysis. This paper used an aggregate scenario approach for homogeneous load intensities just by varying the key model parameters viz: Time, buffer sizes, and packet generation rates as depicted in Table 2a.To demonstrate the effects of these; we employed the MAC controller DCN simulation scenarios with the combinations of values for the above listed parameters. After running 20 simulations alongside varying the process model block parameters for our stipulated load intensities, the QoS parameters viz: Average throughput, latency, utilization stability and losses were computed for analysis. Table 2b shows the packet delivered data and collision detected data without TAMP.

B. SIMULATION RESULTS

The process model developed with MATLAB Simevent was used to generate Fig. 3 to Fig. 15. Traffic characterization for the reengineered DCN gave the observed plot trends. From the process model, the subnet MAC Controller and FIFO Queues holds packets based on service policy of the DCN design hierarchy , hence the DCN switch is capable of determining the flow sequence in each of the incoming buffers at the beginning of service of each packet/frame arrival. The model behavior shows a workload support for enterprise wide deployments viz: cloud computing ERPs, content delivery network systems, and other services. Considering the various process model subnets, this work presents, a graphical plot analysis justifying our selected QoS.

In Fig. 3, packet flow in subnet 0,1,2, and 3 was observed to be greatly influenced by some forms of collisions which tend to affect data throughput in the regeneered DCN model. This work introduced an enhance layer 2 protocol to suppress and enhance packet delivery in the reengineered DCN (subnet 0-3). Since the conventional ethernet networks works with CSMA/CD, traces of packet collisions in the subnet broadcast domains definitely affect the throughput response. Essentially, the Traffic arbitration message protocol (TAMP) strategy in the subnet MAC Controller block regulates efficient throughput by suppressing collision effects. Using linear regression concept in Table 2b, the DCN model residual value R (R= 0.999), gives a significant index for the reengineered subnet 0, 1, 2 and 3, hence guaranteed packet delivery. Figures 4, 5, 6 show that with the CSMA/CD+TAMP, the model will scale gracefully offering better performance than the conventional DCNs.
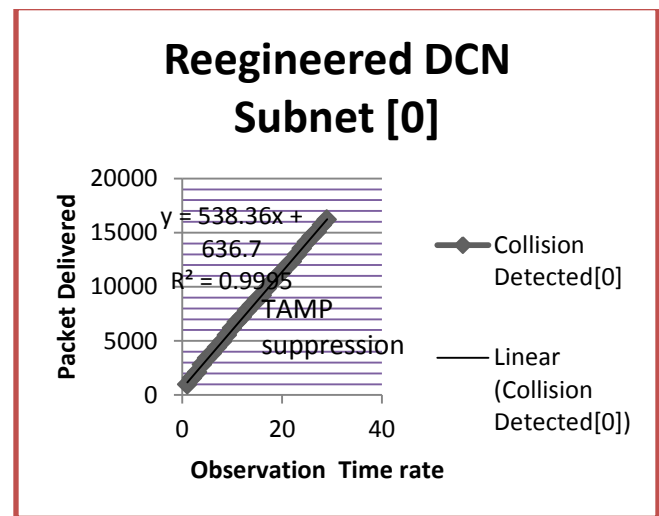
Figure 3. A plot of TAMP Suppression in Subnet [0]

TABLE II.   A: R-DCN SIMULATION PARAMETERS

Time (Secs)	Buffer Sizes	MTUmax	Packet Generation rate	Average. Throughput	Average. Latency	Average. Utilization	Average. Stability	Average. Losses
1	25	1500	100	111.4	107.4286	110.001818	109	110.83
1.5	50	1500	150	130.2	100.125	105.1667	110.5	110.29
2.0	100	1500	200	256	100.025	105.25	112	109.75
2.5	150	1500	250	256.25	100.375	131.6888889	113.5	109.21
3.0	200	1500	300	320	100.5	105.3333	115	108.67
3.5	250	1500	350	320.6	100.625	141.6168889	116.5	108.13
4.0	300	1500	400	355.5	100.75	105.4167	118	107.59
4.5	350	1500	450	424	100.875	105.5	119.5	107.05
5.0	400	1500	500	500	101	101.4866667	121	106.51
5.5	450	1500	550	2179.6	101.125	105.5833	122.5	105.97
6.0	500	1500	600	6081.8	94.3333	105.6667	124	105.43
6.5	550	1500	650	6145.8	94.4444	110.1616889	125.5	104.89
7.0	600	1500	700	6158.9	94.5556	105.75	127	104.35
7.5	650	1500	750	6200.11	94.6667	100.6923	128.5	103.81
8.0	700	1500	800	6240.5	94.7778	97.0714	130	103.27
8.5	750	1500	850	6260	94.8889	100.8462	131.5	102.73
9.0	800	1500	900	6239	95	100.9231	133	102.19
9.5	850	1500	950	6300.5	95.1111	101	134.5	101.65
10.0	900	1500	1000	6350.98	95.2222	101.0769	136	101.11
10.5	950	1500	1050	6500	95.3333	101.1538	137.5	100.57
11.0	1000	1500	1100	6800	94.4444	101.2308	139	100.03
11.5	1050	1500	1150	7800	94.5556	101.3077	140.5	99.49
12.0	1100	1500	1200	7950	94.6667	101.3846	142	98.95
12.5	1150	1500	1250	7960	94.7778	100.7692	143.5	98.41
13.0	1200	1500	1300	8000	94.8889	100.8462	145	97.87
13.5	1250	1500	1350	8020	95	100.9231	146.5	97.33
14.0	1300	1500	1400	8030	95.1111	101	148	96.79
14.5	1350	1500	1450	8037	95.2222	101.0769	149.5	96.25
15.0	1400	1500	1500	8046	95.3333	101.1538	151	95.71

TABLE II.   B: R-DCN TRAFFIC CHARACTERIZATION

Time (Secs)	Packet Delivered[0]	Collision Detected[0]	Packet Delivered[1]	Collision Detected[1]	Packet Delivered[2]	Collision Detected[2]	Packet Delivered[3]	Collision Detected[3]
1	546	970	546	948	511	984	86	165
1.5	882	1559	890	1524	862	1551	147	283
2.0	1103	2086	1115	2043	1089	2070	183	387
2.5	1488	2794	1496	2730	1455	2796	246	483
3.0	1759	3402	1789	3317	1728	3386	292	592
3.5	2017	3843	2077	3756	1996	3835	341	660
4.0	2325	4389	2383	4289	2283	4379	392	761
4.5	2685	4928	2757	4816	2627	4931	441	845
5.0	2999	5525	3094	5393	2926	5524	487	948
5.5	3429	6182	3519	6020	3347	6181	531	1024
6.0	3793	6742	3913	6566	3741	6750	601	1089
6.5	4108	7217	4230	7036	4047	7229	666	1154
7.0	4493	7802	4605	7626	4421	7815	717	1221
7.5	4756	8279	4877	8102	4685	8295	762	1288
8.0	5079	8809	5212	8619	5020	8844	799	1354
8.5	5332	9194	5460	8996	5279	9220	885	1424
9.0	5752	9851	5911	9640	5708	9896	913	1509
9.5	6181	1.043e+004	6300	1.02e+004	6085	1.048e+004	913	1509
10.0	6469	1.089e+004	6587	1.066e+004	6321	1.096e+004	1011	1625
10.5	6821	1.135e+004	6931	1.112e+004	6679	1.142e+004	1048	1674
11.0	7138	1.189e+004	7256	1.164e+004	7001	1.194e+004	1090	1744
11.5	7480	1.233e+004	7615	1.206e+004	7351	1.238e+004	1141	1782
12.0	7893	1.294e+004	8004	1.268e+004	7729	1.302e+004	1191	1862
12.5	8326	1.354e+004	8428	1.327e+004	8140	1.361e+004	1238	1910
13.0	8707	1.412e+004	8808	1.383e+004	8524	1.419e+004	1295	1986
13.5	8977	1.463e+004	9113	1.432e+004	8785	1.47e+004	1342	2066
14.0	9386	1.507e+004	9515	1.475e+004	9142	1.516e+004	1405	2122
14.5	9803	1.569e+004	9958	1.536e+004	9589	1.579e+004	1466	2188
15.0	1.019+004	1.62e+004	1.036e+004	1.586e+004	9959	1.629e+004	1518	2249

Figure 4.   A plot of TAMP Suppression in Subnet [1]

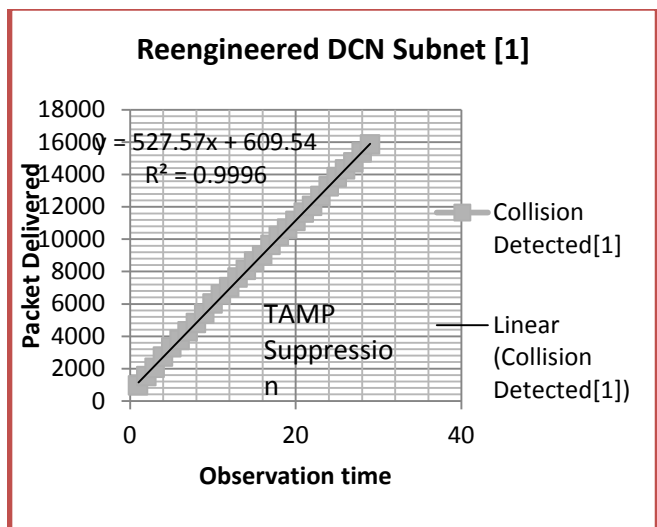Figure 5.   A plot of TAMP Suppression in Subnet [2]

Figure 6.   A plot of TAMP Suppression in Subnet [3]

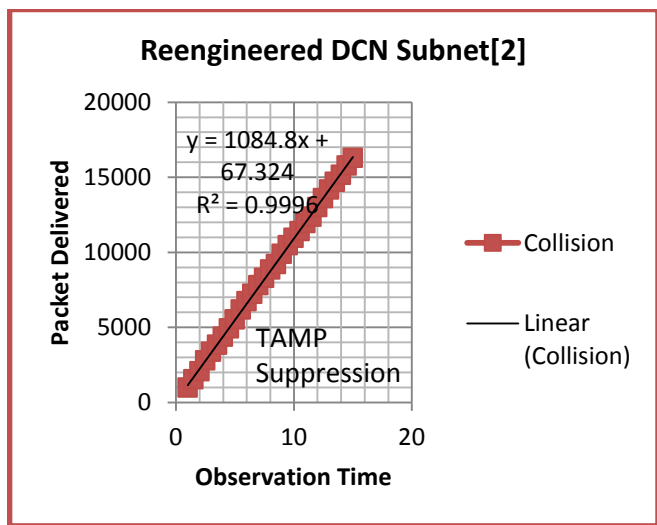Figure 7.   A plot of TAMP Suppression for packet and broadcast traffic
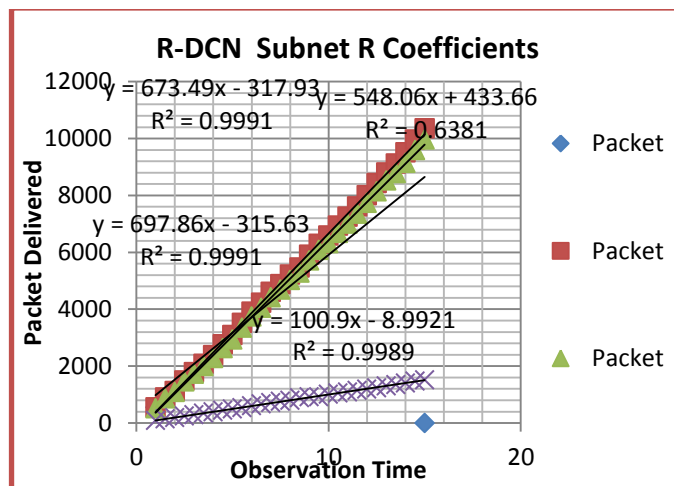

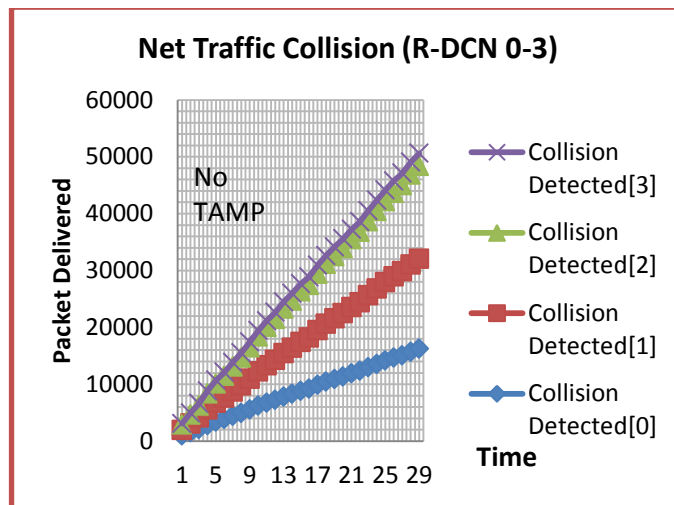Figure 8.   A Plot of Suppression R coefficients for High Scale Model

Figure 9.   A Plot of Net Collision for subnets 0-3

From Fig. 7, It is known that packet delivered in a data center network may not be stable at the connection level under some unfair bandwidth allocation policies, even when the normal offered load condition is satisfied, i.e., the average traffic load at each link is less than its capacity. From the plot, under the normal offered load condition in the reengineered DCN, a data network is stable when the bandwidth of the network is allocated so as to maximize a class of general utility functions. In context, packet delivery and collision occurrence are optimized throughout the simulation run time majorly because of the suppression strategy adopted in the model.

Regression analysis as a statistical tool for investigating into performance index and suppression index is used in this work. From Fig. 8, the four R depicts a positive correlation thereby justifying the model validity. The observation time is the simulation time during which the model statistics are gathered. The packet delivery mode is seen to be reliable, and efficient.

In general, this work assumes that the network should be stable under the normal load condition following the effects of CSMA/CD, but with no TAMP, it shows possible collision cases detected. Though layer 2 switches represent collision domains, yet the full duplex functionality literally does not eradicate collision possibility on the interface ports. For a high end application running on such networks, it is pertinent that all possibilities of collisions be addressed. In this work, Fig. 9 presents the net collision effects for subnet 0-3. As discussed earlier, the approach adopted in this work suppresses interferences and collision possibilities.

Fig. 10 shows a buffer size distribution in a pie chart and presented average throughput. The variation is based on an increment of 50 packets after a previous buffer size selection. This distribution optimally reflects efficient switch traffic carrying capacities with respect to the average throughput in the DCN model. It ranges from 25 packets to 1400 packets for the core switch in the model R-DCN setup. The result of the observation shows that the higher the R-DCN switch buffer, the higher and more effective the throughput with minimal losses. Also, wireless access points with considerable buffer size play significant role in good throughput propagation.

Fig. 11 shows the average network utilization effect for the R-DCN while Fig. 12 shows latency throughput response in the reengineered DCN design. In the context of enterprise DCN, latency, throughput, utilization, and bandwidth usage are vital resources considered in design phases. With the traffic load sources in the hierarchical topology used in this research (Access, Aggregate and Core layers), an initial gradient rise was established before resource allocations were fairly distributed. With a connection request, feasible regions of resource allocation are first established comfortably with CSMA/CD+TAMP. From the test bed, enterprise servers are dumped in Subnet 3 only for server consolidation and virtualization. All the load sources from the model polls resources from this server. Hence, resource utilization in these regions is quite high as shown by the peaks from 6000 and above load times. These are the regions with high VLAN priority tags and as such the computational power of the DCN servers are high. Thus, it was observed that in model, the

subnet 3 region have the highest resource utilization cycles. Also, at zero load time from the plot, the connection requests have been established in the model as the buffer capacity of the switches presents a geometric expansion. This literally depicts a high performance network since server virtualization in this region presents a cost effective infrastructure management and maintains high performance traffic workload computing. Hence, the cloud server is hosted in this domain.

According to [15], network throughput is defined as the average rate of successful message delivery over a communication network (access, distribution and core layers in our context). From Fig. 13, the throughput response was seen to be greatly influenced by the CSMA/CD+TAMP. From a packet generation rate of 500 and above, the throughput response was seen to exponentially rise to higher values with very little latency effect, hence depicting a fast convergence network model. It might take a long time for each packet to reach its destination, because it gets held up in long queues, or takes a less direct route to avoid congestion. This is different from throughput, as the delay can build up over time, even if the throughput is almost normal. In some cases, excessive latency can render an application such as VoIP or online gaming unusable. The model in this context shows low latency-high throughput responses for all applications on this network.

According to [16], the quality of service (QoS) refers to several related aspects of telephony and computer networks that allow the transport of traffic with special requirements. In particular, traffic delay is a considered metric for multimedia conversations in the reengineered DCN mode, as well as supporting new applications with even more strict service demands. Under the influence of the adopted protocol, after initial startup, at within 4secs, the network delay collapses and maintains a stable load traffic state. Average delay, in the reengineered DCN model is highly optimized hence giving rise to a stable and faster network model. Essentially, Fig. 13 and Fig. 14 depict typically reengineered traffic characterization. Considering the CSMA/CD+TAMP, this work via its generated data in table 2a, 2b shows that QoS parameters chosen reflects the desired model behavior.

Fig. 15 shows a fast convergence and effective resource utilization in the proposed model. Application network baseline which is a functional description of application software, how it is used on the proposed network, and the transmission characteristics of the application were all considered in this context. Critical information demanding good utilization and fast convergence include the following: Application server platform, Server location(s), Client locations (number on each LAN) and Application data transmission detail for discrete application transactions.

Conclusively, in all cases, owing to our implementation technologies like virtualization and consolidations, effective service provisioning during the simulation runs shows a high performance reengineered DCN model.
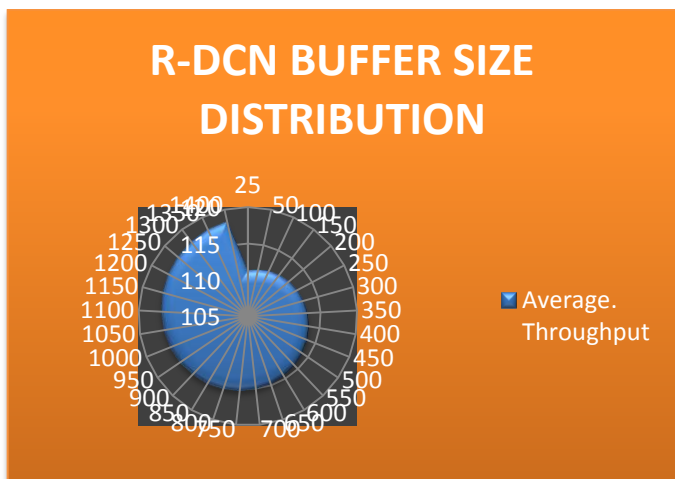
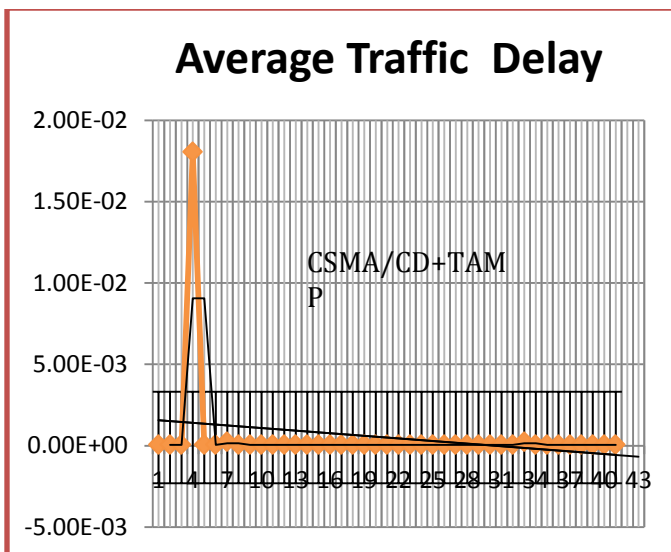Figure 10. A plot of reengineered DCN switch buffer size distribution

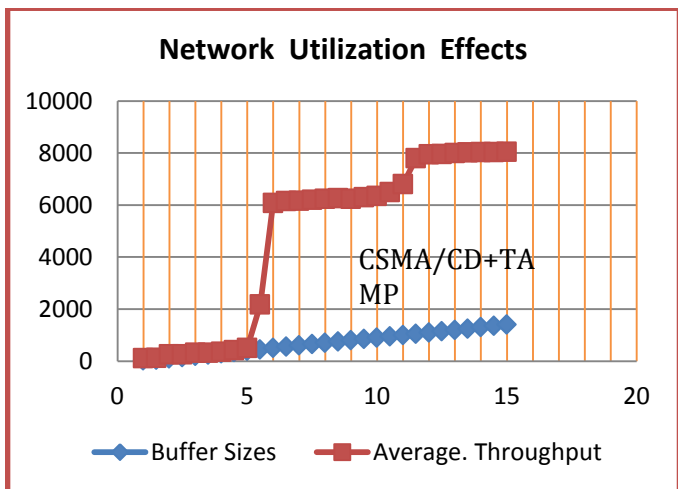Figure 13. A plot of Reengineered DCN Average Traffic Delay

Figure 11. A plot of Reengineered DCN Network Utilization Effects.
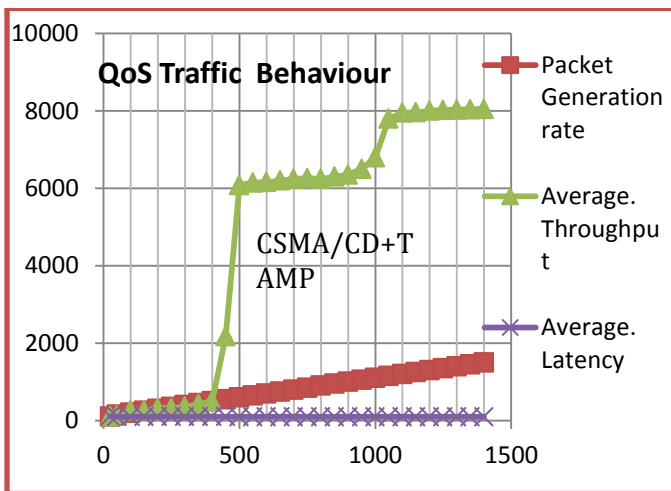
Figure 14. A plot of Reengineered DCN QoS Traffic Behaviour
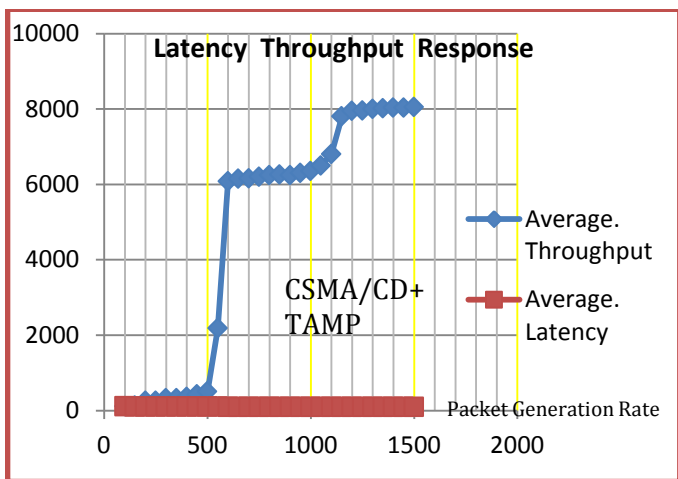
Figure 12. A plot of Reengineered DCN latency Throughput Response
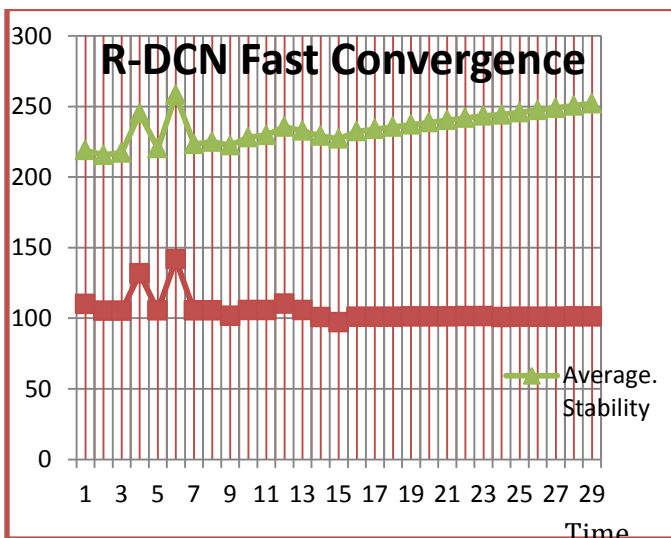
Figure 15. A plot of Reengineered DCN Fast Convergence

REFERENCES

[1] Okafor K.C., "SVLAN Approach to Congestion Management in Data Center Networks", M.Eng Thesis, University of Nigeria, Nsukka," 2010.

[2] http://ieee802.org/802_tutorials/Data-Center-Bridging-Tutorial-Nov-2007-v2.pdfGigabit Ethernet-Wikipedia.

[3] O. Kennedy, U. Chidiebele. C, C, Prof. H. C. Inyiama, Dr. C.C Okezie, "MATLAB Simevent: A Process Model Approach for Event-Based Communication Network Design, (A Case for Reengineered DCN)" (Unpublished).

[4] ITU-T SG16, "One-way Transmission Time", Recommendation G.114

[5] ITU-T SG13, "Network Performance Objectives for IP-Based Services", Y.1541

[6] ITU-T SG12, "The E-model, a computational model for use in transmission planning", Recommendation G.107.

[7] L. Zheng, L. Zhang, D. Xu, "Characteristics of network delay and delay jitter and its effect on voice over IP (VoIP)", ICC 2001, V. 1, June 01 pp.122–126.

[8] Rozalina Dimora, Georgi Georgiev, Zlatko Stanchev. "Performance Analysis of QoS parameters for Voice over IP Applications in a LAN Segment" International Scientific Conference on Computer Science 2008. Pp. 232-237.

[9] Nalin Gahlaut et al. "Comparative Analysis & Study of Different QoS parameter of wireless AD-HOC Network2". International Journal of Advances in Engineering & Technology,. Vol. 1 Issue 2, pp. 176-182.

[10] Vijayalaskhmi M, et al "QoS Parameter Analysis on AODV and DSDV Protocols in a wireless Network" International Journal of Communication Network & Security, Volume1, Issue 1, 2011 pp. 62-70.

[11] I. Miloucleva, A, Nassri, AArizaloni. "Automated Analysis of Network QoS parameters for Voice over IP Applications". IPS 2004, BUDAPEST HUNGARY pp. 184-193.

[12] S. R. Biradar, et al. "Analysis of Qos Parameters for MANETs Routing Protocols" International Journal on Computer Science and Engineering. Vol. 02, No. 03, 2010 pp. 593-599.

[13] Mathworks. 2007. http://www.mathworks.com

[14] http://www.mathworks.in/help/toolbox/simulink/slref/toworkspace.html

[15] http://en.wikipedia.org/wiki/ Throughput

[16] http://en.wikipedia.org/wiki/ Quality_of_service

AUTHORS PROFILE

Udeze Chidiebele C. received his B.Eng and M.Sc in Electrical Electronics Engineering with Nnamdi Azikwe University, Awka, Nigeria .He is a Senior R & D Engineer with Electronics development Institute Awka, Nigeria and also a member of NSE. He is currently running his Ph. D programme in computer and control systems engineering. His current research interest is on data center networks, wireless sensor networks and control systems engineering.

Prof H.C Inyiama is a senior lecturer in the electronic and computer department of Nnamdi Azikiwe University Awka, Nigeria. He is an expert consultant in computer and control systems engineering. He has supervised many masters and Ph.D works in electronics engineering, computer science and engineering, and has many publications to his credit.

Okafor Kennedy. C. is a Senior R & D Engineer. He has B.Eng & M.Engr in Digital Electronics & Computer Engineering from the university of Nigeria Nsukka, (UNN). He a Cisco expert and currently works with the ICT department of Electronics Development Institute, ELDI. He is a member of IEEE, IAENG and NCS.

C.C Okezie is a senior lecturer in the electronic and computer department of Nnamdi Azikiwe University Awka. She received her B. Eng, M. Eng, and Ph.D from Enugu State University of Science and Technology, Enugu, Nigeria. She majors in digital systems and control engineering and has many publications to her credit.

Test Case Generation For Concurrent Object-Oriented Systems Using Combinational Uml Models

Swagatika Dalai
School of Computer Engineering
KIIT University
Bhubaneswar

Arup Abhinna Acharya
School of Computer Engineering
KIIT University
Bhubaneswar

Durga Prasad Mohapatra
Department of Computer Science
Engineering
National Institute of Technology
Rourkela

Abstract—**Software testing is an important phase of software development to ensure the quality and reliability of the software. Due to some limitations of code based testing method, the researcher has been taken a new method to work upon UML model based testing. It is found that different UML model is having different coverage and capable of detecting different kinds of faults. Here we have taken combinational UML models to have better coverage and fault detection capability. Testing concurrent system is difficult task because due to concurrent interaction among the threads and the system results in test case explosion. In this paper we have presented an approach of generating test cases for concurrent systems using combinational UML models i.e. sequence diagram and activity diagram .Then a Sequence-Activity Graph (SAG) is constructed from these two diagrams. Then that graph is traversed to generate test cases which are able to minimize test case explosion.**

Keywords-Software Maintenance; Regression Testing; Test case Prioritization.

I. INTRODUCTION

Software testing plays a vital role in System Development Life cycle. It is an investigation which is to be conducted to provide information about the quality of the product. According to IEEE testing is "the process of exercising or evaluating a system or system components by manual or automated means to verify that it satisfies specified requirements". In other words testing is a process of gathering information by making observation and comparing them to expectations. The facts like the primary objective and the risks of software implementation are provided by software testing. Testing is not only meant for finding the bug in the code, but also it checks whether the program is behaving according to the given specifications and testing strategies [1]. If the software does not perform the required and expected result then a software failure is occurred. Therefore maximum of software development effort is being spent on testing. Generally the Testing process consists of three things: i) test case generation ii) test case execution iii) test case evaluation. The test case generation process plays a vital role among the three cases. The test case consists of three things i.e. input to the system, state of the system and the expected output from the system. If the test case detects maximum number of faults with minimum number of test cases then it is said to be having good coverage. During the software development process software testing can be implemented at any time. But the testing is implemented after all the requirements are defined and the coding process is over. But Code based testing having certain disadvantages which are as follows:

- Code based testing is not capable of extracting the behavioural aspects of the system.

- Code based software testing is not suitable for component based software development, because the source code may not be available to the developer.

An alternative approach is to generate test cases from the models which represent the software. It has the specific feature that the testing techniques can be applied throughout the development process depending on the requirement specification and design models. Another advantage of model based testing is that the generated test data is independent of any particular implementation of design. The model based testing reduces the testing time and effort.

Therefore now the researchers have used the analysis and design models like Unified Modelling Language (UML) for test case generation. UML models are very popular because UML is a solution of standardization and utilization of design methodologies. Another advantage of UML models is that it provide different diagram for representing different view of system models.

Concurrent computing is a property of systems in which several computations are executing simultaneously and also interacting with each other. Testing a concurrent system is a very difficult task because this type of system can reveal different responses depending upon different concurrency condition.

A concurrent system may be implemented via processes and/or threads. Due to concurrency a major problem arises known as explosion of test case because of the possibility for arbitrary interference of concurrent threads. These threads are executing relatively independently. However, since they are acting towards some goal, they must need to communicate and coordinate. So, issues arise while testing the concurrent system. In case of object oriented system objects interact with each other to accomplish task. While objects are interacting with each other there may be arbitrary interference of concurrent threads. Each thread may have more than one activity which

may be dependent on each other; that may be inter thread dependency or intra thread dependency. Due to these problems there issues like Communication deadlock and synchronization may arise.

In this paper, we propose an approach for generating test cases using combinational UML diagram. In our approach we have taken the combination of sequence and activity diagram which are then traversed to generate the optimized test suite.

The rest of the paper is organized as follows: Section II describe the related work. The Basic concepts are described in section III. Analysis of our proposed methodology is discussed in Section IV. Section V contains the Conclusion and Future work.

II.   RELATED WORK

Sarma et al. [2] proposed a method for generating Test cases from UML Sequence Diagram (SD). In this technique the author first derive different operation scenarios from the Sequence Diagram. Here operation scenario means the set of messages that are flows between different objects. Based on this Scenario an intermediate format is constructed known as Sequence Diagram Graph (SDG). The state of the object changes when a message flows from one object to other. Here the author represents each state as a node and also assign an edge between the nodes. Here the graph has two ends i.e. one for true condition and other for false condition. Then the graph is traversed using graph traversal algorithms i.e. BFS and DFS. By applying All Sequence Message Path Criterion the author has find out all the possible message paths from the starting node to the end node. After traversal some test case are generated which are able to detect the operational faults as well as interaction faults.

Kim et al. [3] proposed a method to generate test cases from Activity Diagram (AD). The author first convert the Activity Diagram into an intermediate format known as called I/O explicit Activity Diagram (IOAD). In this diagram the author has suppressed the non-external input and output. He has only represented the external input and output as the internal activities are less important than the external activities to avoid test case explosion. Then an directed graph is being constructed using the basic path coverage criterion. After that the graph is traversed using DFS algorithm. Finally a set of basic paths are derived to generate the test cases.

Samuel et al. [4] proposed a method to generate the Test Sequences from UML 2.0 Sequence Diagram. In this approach the author first find out the different types of relationship like indirect message dependency, direct message dependency, simple indirect message dependency, simple direct message dependency that exists between the messages. Depending on the relationship different message sequences are generated and a graph is constructed known as Sequence Dependency Graph (SDG).Here each node in the SDG represent a message or a set of messages. Here the author associate node with the message number. Finally the SDG is traversed to generate the test cases.

The author Khandai et al. [5] proposed a method to generate test cases from UML 2.0 Sequence Diagram. To represent the complex scenarios she has used the Combined Fragment (CF). By applying some mapping rule In this approach first the Sequence Diagram (SD) is transformed into an intermediate form called Concurrent Composite Graph (CCG) to represent different scenario and their flow. Then the CCG is traversed using Message Sequence Paths Criteria (MSPC) to generate the test cases. The test cases are useful for detecting scenario, interaction as well as operational faults.

Sarma et al. [6] proposed a method for generating test cases from combination of UML models i.e. Sequence diagram (SD) and Use case Diagram (UD). In this approach first of all SD is converted into Sequence Diagram Graph (SDG). Then the UD is converted into Use case Diagram Graph (UDG.)In the UDG each actor is represented as node in the graph and assigns some edges between the nodes. Then the SDG and the UDG are combined to form a graph called as System Testing Graph (STG). Then STG is being traversed to generate the test cases.

The author Sun et al. [7] proposed a transformation-based approach to generate scenario oriented Test cases from UML Activity Diagram. The approach consists of three basic steps: First the UML Activity Diagram is transformed into an intermediate representation known as Binary Extended AND OR Tree (BET) via a set of transformation rule. The transformation rule is applied on different types of nodes like Fork node, Join node, Branch node and Merge node. Then the author applied an algorithm on the intermediate format to generate the Extended AND OR Tree. After that the tree is traversed using DFS algorithm to generate   a set of test scenario.   Finally from the test scenario a set of test cases are derived. The proposed this method for testing the concurrent system.

Kundu et al.[8] proposed an approach to generate test cases from UML Activity Diagram(AD). Then the AD is converted into an intermediate format known as Activity Graph (AG) by applying some transformation rules. The AG is traversed using BFS and DFS algo. BFS algorithm is used traverse all the concurrent activities and the rest are traversed by using DFS. Here the author uses the Activity Path Coverage Criterion to generate different Activity Paths.

Fan et al. [9] proposes a method for generating test cases from sub activity diagram to compound activity diagram in a hierarchical manner. They introduce the idea of this method by taking the thought of functional decomposition, bottom-up integration testing strategy and round-robin strategy.

Khandai et al. [10] proposed a method to generate test cases from combinational UML models such as Sequence Diagram (SD) and Activity Diagram (AD). In her approach AD is converted into an intermediate format known as Activity Graph (AG). After that test sequences are generated from AG by applying Activity Path Coverage Criteria. Then SD is converted into Sequence Graph (SG) and the test sequences are generated by applying All Message Path Coverage Criterion. For having better coverage and high fault detection capability the author constructed a Activity Sequence Graph (ASG) which has the combine features of AG and SG. Finally the ASG is traversed to generate the test cases.

III. Basic Concepts

UML Models

a) Sequence Diagram

The Sequence Diagram is a type of interaction diagram that is used for dynamic modelling which focuses on identifying the behaviour within the system. It represents object interaction arranged in a time sequence. These diagrams are used to represent or model the flow of messages, events and actions between the objects or components of a system. Sequence diagrams are typically used to describe the object-oriented system. It describes the objects and classes involved in the scenario and also the sequence of messages exchanged between the objects to carry out the functionality of the scenario.

b) Activity Diagram

An UML Activity Diagram is suitable for representing concurrent interaction among multiple threads. An Activity Diagram describes how multiple objects collaborate to do a specific set of operation. The basic elements of Activity Diagram are activity and transition. Activity is used as a state for doing something and the transition is represented as a directed line which connects different activities. A transition can be message flow, object flow or control flow. Activity Diagram is used for both conditional and parallel behaviour. Conditional behaviour can be denoted as a branch and a merge, and parallel behaviour is denoted by a fork and a join.

c) Concurrent System

In a Concurrent System different programs or threads are represented as collections of interacting computational processes that may be executed in parallel. The execution of threads begins from fork node and ends at join node. The main limitations in designing concurrent program to ensure the correct sequencing of the interactions or communications between different computational processes and coordinate the access to shared resources.

d) Activity path coverage criteria

This coverage criterion is used for both loop testing and concurrency among activities of activity diagrams. Here in this coverage criterion precedence relationship is maintained. An activity path is a path that allows a loop maximum two times and also maintains precedence relationship between different concurrent and non-concurrent activities.

IV. Proposed Methodology

We have proposed a methodology whose model diagram is shown in Fig. 1. The model has been proposed for generating test cases for concurrent system. In our approach we have presented an approach to generate test cases for concurrent system using combinational UML models. Here we have used the combination of Sequence diagram and Activity diagram. In a concurrent system several computations are processed simultaneously and also potentially interacting with each other.
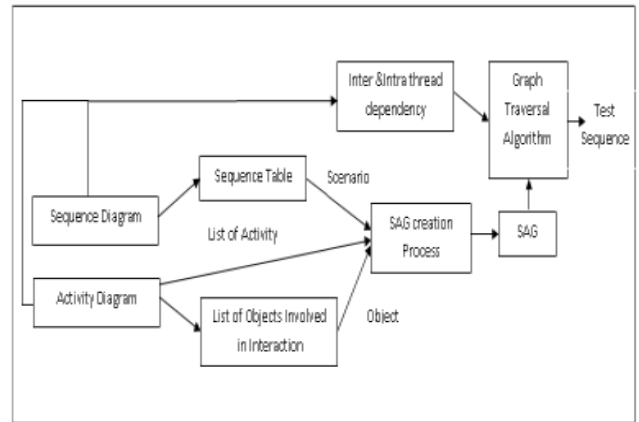
Figure 1. A Frame work of our proposed methodology

In our approach we have taken a sequence diagram and an activity diagram. From these two diagrams we have constructed a graph named as Sequence-Activity graph (SAG) by using an algorithm named as sequence-activity graph. Then the SAG is traversed using graph traversal algorithm i.e. BFS and DFS to generate test cases.

The resultant shows that the generated test cases are being capable of addressing the issue like test case explosion as we are dealing with concurrent system.

Concurrent computing is a form of computational process in which processes are designed as a collection of interacting computational processes that are executed in parallel. Here we have taken the sequence diagram because by using sequence diagram we can provide a dynamic view of the system in a graphical manner to display how the messages are passed between the objects at run time in order to perform tasks. Here we are dealing with concurrent system. Software testing is especially very much difficult when a system contains concurrently executing objects.

For this we have taken the activity diagram as the activity diagram is very much suitable for concurrent system as the activity diagram is capable of showing the parallel execution of different activities in a concurrent system.

Testing a concurrent system is a difficult task due to the arbitrary interference of different threads. Each thread may have more than one activity. Activities present in same thread maintain partial order relationship which shows dependency called as Intra thread dependency. Also activities of different thread may be dependent on each other called as Inter thread dependency.

Due to these types of dependency some critical situation arises during message passing which leads to a communication deadlock and also it causes test case explosion. So our objective is to minimize the test case explosion while generating the test cases.

Our approach consists of the following steps:

1) Construction of a Sequence Diagram (SD) and an Activity Diagram (AD).

2) Maintaining a Sequence Table (ST) with different schema as Source object(SC), Destination object(DO), Message ID(MI) and Message Content(MC) by taking the information from SD.

3) Then construction of Sequence-Activity Graph (SAG) by combining the features from SD and AD.

4) Finally traversal of SAG to generate test cases.

Example of Sequence diagram and Activity diagram

In this section we have presented the dummy examples of sequence and activity diagrams. Here we have used the Activity Diagram and Sequence diagram due to the following reasons.

Activity diagram is suitable for representing the concurrent activities as there is no state explosion of objects a in state chart diagram. Activity diagram shows the sequence of activity flows and also it represents parallel activities taken place in fork and join node.

Sequence Diagram (SD), also known as Interaction diagram represent the sequence of messages passed between the objects to specify some task without leading to test case explosion.
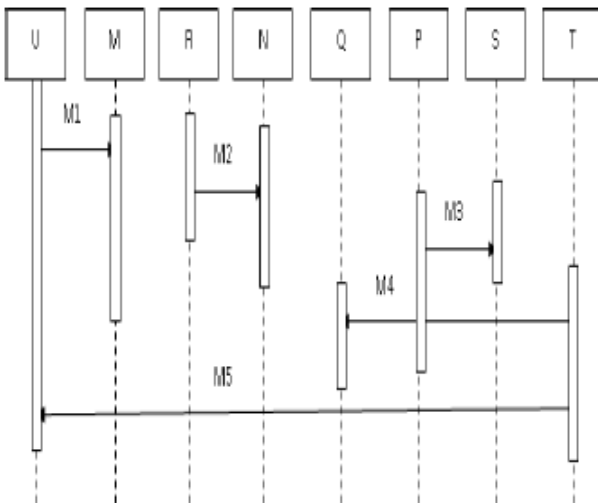
Figure 2. Sequence Diagram

U, M, R, N, Q, P, S, T are different objects of Sequence Diagram shown in Fig 2 . M1, M2 , M3,M4 and M5 are different messages that are flows between the objects.

The activity diagram represents the sequence of activity taken place. In this Activity diagram the Inter and Intra thread dependency are present. In above diagram different objects are like U,M,N,P,T,R,S,V are created after or before completion of different activities.
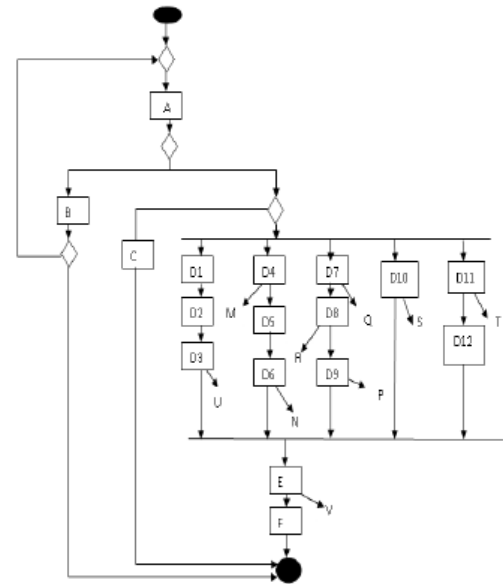
Technique for constructing Sequence Table (ST)

We have constructed a Sequence Table(ST) by taking the data from the Sequence Diagram as described in the section1.1.The table is constructed with the schema as Source object(SC), Destination object(DO), Message ID(MI) and Message Content(MC).

TABLE I. SEQUENCE TABLE

SO	DO	MI	MC
U	M	M1	aaa
R	N	M2	bbb
P	S	M3	ccc
T	Q	M4	ddd
T	U	M5	eee

Construction of SAG (Sequence-Activity Graph)

In this approach we propose a technique for generating SAG by combining both the features of Sequence diagram and Activity diagram. Whenever a transition is there we then take the two nodes in SAG and assign an edge between them. The nodes shown in the graph may be the Object node or Activity node. More no of threads are present in the Activity diagram in concurrent system. The activities present in different threads may be dependent on each other which are called as Inter thread dependency. The objects are created after activities are being taken place and communicating with each other by passing message among them. In the graph Inter thread dependency is represented by dashed arrows.

Node Representation of SAG

To represent a node in the graph two types of lists are required i.e. node list and edge list.

Node List

Status Bit	NodeID	Address of next node	Adjacent Node

<p style="text-align:center;">Figure 3. Node Structure</p>

Here the status field shows whether the node is a object node or activity node. 0 means object node and 1 means the node is an activity node. Node ID represents the unique ID of the node and the third field represents the address of the next node. and address of next edge shows the next adjacent node.

Edge List

Dest	Dependency Bit	Link

<p style="text-align:center;">Figure 4. Edge Structure</p>

Here the Dest field describes the destination link of the node and the Dependency Bit field describes the edge connecting to the nodes are Intra thread or Inter thread. I is denoted as Inter thread and O is denoted as intra thread dependency. Link field denotes the address of the next adjacent node of the Edge structure.

Traversal of SAG

Here we are using the graph traversal algorithm to traverse the graph and generate the test sequences by using the activity path coverage criteria.

Algorithm: SAG Traversal Algorithm
  Input: Sequence-Activity Graph
  Output: Set of test Sequence
  1. Start.
  2. Traverse the SAG using DFS.
  3. While(Ncurr !=Nend)   //Ncurr=Current Node
     Nend=End Node
  4. If(Ncurr= =Nf)   //Nf=Fork Node
     Then traverse the sub tree rooted at node Nf using BFS.

  5. If Node List[Status bit= =1]
     Then traverse the edge list.
     Else
     Ncurr=Ncurr->next
  6. If Edge List[Dependency bit= =0]
  Then there is inter thread dependency and traverse   that activity.

  7. Go to step4.
  8. Else Ncurr=Ncurr->next
  9. End If
  10. End while
  11. Exit.
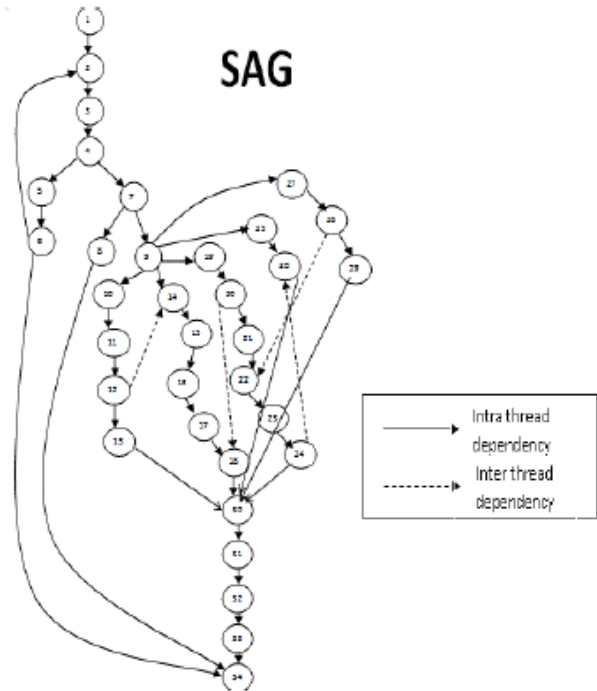
<p style="text-align:center;">Figure 5. Sequence-Activity Graph (SAG)</p>

After the construction of SAG we have traversed the SAG using Graph traversal algorithm. While traversing the nodes whenever a fork node is encountered, we will apply BFS (Breadth first Search) algorithm and for the rest nodes we have applied DFS (Depth First Search) algorithm in order to generate test cases. In this approach we have used the Activity path coverage criteria to generate test cases. After traversing the graph some test sequences are generated which are given below.

T1=1-2-3-4-5-6-2-3-4-5-6-34
T2=1-2-3-4-7-8-34
T3=1-2-3-4-7-9-10-11-14-12-13-15-16-17-18-19-20-21-22-23-26-24-25-27-28-29-30-31-32-33-34

V. CONCLUSION

In this paper we have proposed an approach for generating test cases using combinational UML diagram. In our approach we have taken the combination of sequence and activity diagram to construct a graph known as Sequence-Activity Graph(SAG).The partial ordering relationship that exists due to inter thread and intra thread communication is also included in the approach. The SAG is then traversed to generate the optimized test suite which minimizes the test case explosion.

VI. FUTURE WORK

The scalability of the approach is yet to be tested. A suitable optimization technique like Genetic Algorithm or Particle Swarm Optimization can be used to further reduce the volume of test data to be generated.

REFERENCES

[1] Rajib Mall, Fundamentals of Software Engineering, Prentice-Hall of India Comm, New Delhi, 2007.

[2] M. Sharma, D. Kundu, and R. Mal, Automatic test case generation from uml sequence diagram, In 15th International Conference on Advance Computing and Communication, IEEE, pages 60-65, 2007.

[3] H. Kim, S. Kang, J. Baik, and I. Ko, Test cases generation from UML activity diagram

[4] P. Samuel and A. T. Joseph, Test sequence generation from uml sequence diagrams, In Ninth ACIS International Conference on Software Engineering, Artificial intelligence, Networking, and parallel/distributed computing, IEEE, pages 879 - 887, 2008.

[5] M. Khandai, A. A. Acharya, D. P. Mohapatro, A Novel Approach of Test Case Generation for Concurrent Systems Using UML Sequence Diagram, 3rd International Conference on Electronics Computer Technology, ICECT, 2011, pages 157-161.

[6] M. Sharma and R. Mal, Automatic test case generation from uml models, In 10th International conference on Information Technology, IEEE, pages 196-201, 2007.

[7] C. Sun,A transformation-based approach to generating scenario-oriented test cases from uml activity diagram for concurrent applications, In Annual IEEE International Computer Software and Applications Conference, IEEE, pages 160- 167, 2008.

[8] D. Kundu and D.Samanta, A novel approach to generate test cases from UML activity diagrams, volume 8 no 3, pages 65 - 83, May-June 2009.

[9] X. Fan, J. Shu, L. Liu, and Q. Liango,Test case generation from uml subactivity and activity diagram ,In Second International Symposium on Electronic Commerce and Security, IEEE, pages 244 - 248, 2009.

[10] M. Khandai, A. A. Acharya, D. P. Mohapatro, Test Case Generation for Concurrent System using UML Combinational Diagram, International Journal of Computer Science and Information Technologies, IJCSIT,2011, Vol.2, Issue.

AUTHOR PROFILE

Swagatika Dalai is a faculty in Kalinga Polytechnique, KIIT University, Bhubaneswar, Odisha, INDIA. Her research areas include Object-Oriented System Testing. Now she is a M.tech student in KIIT University. She has a teaching experience of 5 years.

Arup Abhinna Acharya is an Assistant Professor and research scholar in the School of Computer Engineering, KIIT University, Bhubaneswar, Odisha, INDIA. He received his Masters degree from KIIT University Bhubaneswar. His research areas include Object Oriented Software Testing, Software Cost Estimation, and Data mining. Many publications are there to his credit in many International and National level journal and proceedings. He is having eight years of teaching experience. He is a member of ISTE.

Durga Prasad Mohapatra received his Masters degree from National Institute of Technology, Rourkela, India. He has received his Ph.D. from Indian Institute of Technology, Kharagpur, India. He is currently working as an Associate Professor at National Institute of Technology, Rourkela. His special fields of interest include Software Engineering, Discrete Mathematical Structure, Program Slicing and Distributed Computing. Many publications are there to his credit in many International and National level journal and proceedings.

Prioritizing Test Cases Using Business Criticality Test Value

Sonali Khandai
School of Computer Engineering
KIIT University
Bhubaneswar

Arup Abhinna Acharya
School of Computer Engineering
KIIT University
Bhubaneswar

Durga Prasad Mohapatra
Department of Computer Science
Engineering
National Institute of Technology
Rourkela

Abstract—**Software maintenance is an important and costly activity of the software development lifecycle. Regression testing is the process of validating modifications introduced in a system during software maintenance. It is very inefficient to re-execute every test case in regression testing for small changes. This issue of retesting of software systems can be handled using a good test case prioritization technique. A prioritization technique schedules the test cases for execution so that the test cases with higher priority executed before lower priority. The objective of test case prioritization is to detect fault as early as possible. Early fault detection can provide a faster feedback generating a scope for debuggers to carry out their task at an early stage. Model Based Prioritization has an edge over Code Based Prioritization techniques. The issue of dynamic changes that occur during the maintenance phase of software development can only be addressed by maintaining statistical data for system models, change models and fault models. In this paper we present a novel approach for test case prioritization by evaluating the Business Criticality Value (BCV) of the various functions present in the software using the statistical data. Then according to the business criticality value of various functions present in the change and fault model we prioritize the test cases are prioritized.**

Keywords- Software Maintenance; Regression Testing; Test case Prioritization; Business Criticality.

I. INTRODUCTION

Software developers often save the test suites they develop for their software, so that they can reuse those suites later as the software evolves. Such test suites are reused of regression testing [1]. Regression testing is the re-execution of some subset of test that has already been conducted. So regression testing can be defined as follows: Let P be a program and P' be a modified version of P with T be a test suite developed for P, then regression testing is concerned with validating P'. Integration testing occurs in regression testing, so number of regression tests increases and it is impractical and inefficient to reexecute every test for the software when some changes occur. For this reason, researchers have considered various techniques for reducing the cost of regression testing, like regression test selection, and test suite minimization [2, 3] etc. Regression test selection technique attempt to reduce the time required to retest a modified program by selecting some subset of the exiting test suite. Test suite minimization technique reduces testing costs by permanently eliminating redundant test cases from test suites in terms of codes or functionalities exercised. However

Regression test selection and test suite minimization techniques have some drawbacks. Although some empirical evidence indicates that, in certain cases, there is little or no loss in the ability of a minimized test suite to reveal faults in comparison to the unminimized one, other empirical evidence shows that the fault detection capabilities of test suites can be severely compromised by minimization [4]. Similarly, although there is safe regression test selection techniques that can ensure that the selected subset of a test suite has the same fault detection capabilities as the original test suite, the conditions under which safety can be achieved do not always hold. So for these reasons testers may want to order their test cases or reschedule the test cases [5]. A prioritization technique schedules the test cases for execution so that the test cases with higher priority executed before lower priority, according to some criterion:

Rothermal et al. [6] defines the test case problem as follows, where:

Problem: Find T' belongs to PT such that (for all T") (T" belongs to PT) (T" ≠ T')

$[f(T') \geq f(T")]$.

T: a test suite,

PT: the set of permutations of T,

f: a function from PT to the real numbers.

Here, PT represents the set of all possible prioritization of T and f is a function that, applied to any such ordering, yields an award value for that ordering.

The objective of test case prioritization is fault detection rate, which is a measure of how quickly faults are detected during the testing process. There are two categories of test case prioritization: general test case prioritization and version specific test case prioritization [6]. In general test case prioritization, given a program P and test suite T, we prioritize the test cases in T with the intent of finding an ordering of test cases that will be useful over a succession of subsequent modified versions of P.

Thus, general test case prioritization can be performed following the release of some version of the program during off-peak hours, and the cost of performing the prioritization is amortized over the subsequent releases. In contrast, in version-specific test case prioritization, given program P and test suite

T, we prioritize the test cases in T with the intent of finding an ordering that will be useful on a specific version P' of P. Version-specific prioritization is performed after a set of changes have been made to P and prior to regression testing P'. Because this prioritization is accomplished after P' is available. General test case prioritization does not use information about specific modified versions of P, whereas version-specific prioritization does use such information.

Prioritization can again be categorized as Code Based Test Case Prioritization and Model Based Test Case Prioritization [7]. Most of the test case prioritization methods are code based. Code Based Test Case Prioritization methods are based on the source code of the system. Code based test case prioritization techniques are dependent on information relating the tests of test suite to various elements of a system's code of the original system i.e. it can utilize the information about the number of statements executed or the number of blocks of code executed. Model Based Test Case Prioritization methods are based on the system models. System modeling is widely used to model state-based system. System models are used to capture some aspects of the system behavior. Several modeling language have been developed such as Extended Finite State Machine (EFSM) and Specification Description Language (SDL).

Test case prioritization is ordering the test cases in a test suite to improve the efficiency of regression testing. Regression testing is a process of retesting the modified system using the old test suite to have confidence that the system does not have faults. During retesting of the system developers face the issue of ordering the tests for execution, which can be addressed using a good prioritization technique. One of the objectives of test case prioritization is 'early fault detection rate', which is a measure of how quickly faults are detected during testing process. Here a metric is used which calculates the average faults found per minute and also with the help of APFD (Average Percentage of Faults Detected) metric the effectiveness of the prioritized and non-prioritized case is compared [8,9]. The APFD is calculated by taking the weighted average of the number of faults detected during the run of the test suite.

APFD can be calculated using the following notations:

Let T = the test suite under evaluation,

m = the number of faults contained in the program under test P,

n = the total number of test cases

And TF_i= the position of the first test in T that exposes fault i.

$$APFD = 1 - \frac{TF1 + TF2 + \cdots + TFm}{mn} + \frac{1}{2n} \qquad (1)$$

But calculating APFD is only possible when prior knowledge of faults is available. Various experiment were conducted in which the rate of fault detection for each test case is calculated and order of test suite is evaluated in decreasing order of the value of rate of fault detection. Then the APFD value is determined for both the prioritized and non-prioritized test suite and it is found that the APFD value of prioritized test suite is higher than the non-prioritized test suite.

The rest of the paper is organized as follows: Section 2 summarizes the related works. Discussions and the analysis of our proposed methodology are given in Section 3. Section 4 presents a case study: Shopping Mall Automation System. Comparison with the related work is discussed in section 5. The paper concludes in Section 6 and Future works are highlighted in section 7.

II. RELATED WORK

Srivastava et al.[11] proposed a prioritization technique and also used a metric called APFD (Average Percentage of Faults Detected) for calculating the effectiveness of the test case prioritization methods. The disadvantage of the method proposed is that calculation of APFD is only possible when prior knowledge of faults is available. APFD calculations therefore are only used for evaluation of effectiveness of various prioritization techniques.

Rothermel et al. [6, 8, 10] presented 21 different techniques for code based test case prioritization, which are classified into three different groups i.e. comparator group, statement level group and function level group. To measure the effectiveness of these techniques, an experiment was conducted where 7 different programs were taken. Here several dimensions like granularity were taken for test case prioritization. The main disadvantages of code based test case prioritization are it is very expensive as its execution is slow because of the execution of the actual code and code based test case prioritization may not be sensitive to the correct or incorrect information provided by the testers or the developer.

Srikanth et al. [12] presented a technique that extend the code-coverage TCP techniques and apply test case prioritization at a system-level for both new and regression tests. Here the advantage is that the author uses a system level test case prioritization techniques which is called the Prioritization of Requirements for Test (PORT) based techniques.

The PORT technique prioritizes system test cases based upon four factors: requirements volatility, customer priority, implementation complexity, and fault proneness of the requirements. System level test case prioritization techniques are very beneficial because it improve the rate of fault detection of severe faults. PORT technique requires the team to conduct system analysis and write concrete test cases. The act of writing concrete test cases immediately after requirements specification can lead to the identification of ambiguous and unclear requirements, allowing requirements errors to be identified and rectified earlier.

The PORT technique allows the engineering team to monitor the requirements covered in system test; the ability to monitor requirements covered in system test is believed to be one of the challenges faced by the industry. But here only the experiment is done for four projects developed by students in advanced graduate software testing class. So the authors have to test the scalability of the PORT method.

Korel et al.[13, 14] present a model based test case prioritization method which can be used for any modification of the EFSM (Extended Finite State Machine) system model.

Here an experimental study is done which is used to compare the early fault detection of the various test case prioritization techniques presented in this paper i.e. random prioritization, two version of selective prioritization (version I & II) and model dependence-based test case prioritization. Here the author used RP (d), the most likely relative position of the first failed test that detects fault d to measure the early fault detection. The experimental result show that the version II of selective prioritization and model-dependence based prioritization may improve the effectiveness of the test case prioritization. But here in this paper the author used very small model for the test case prioritization, so the effectiveness of the model based test case prioritization method can't not understand properly. Also in this paper a fixed model is used i.e here the model is not changed in the system. In the further work Korel et al. [14] present model-based test case prioritization methods in which information about the system model and its behavior is used to prioritize the test suite for system retesting. There are several model based test prioritization methods are present in this paper. Such as Selective test prioritization, Heuristic #1 test prioritization, Heuristic #2 test prioritization, Heuristic #3 test prioritization and Model dependence-based test prioritization. In selective test prioritization techniques high priority is assigned to those test cases that execute modified transitions in the modified model. A low priority is assigned to those test cases that do not execute any modified transition. And Heuristics #1, #2 and #3 have been developed for modifications with multiple marked transitions. The idea of model dependence-based test prioritization is to use model dependence analysis [15] to identify different ways in which added and deleted transitions interact with the remaining parts of the model and use this information to prioritize high priority tests. Here the authors have done a experimental study and from the experimental study it indicate that model based test prioritization techniques may improve on average the effectiveness of early fault detection as compared to random prioritization techniques.

Korel et al.[7] presented a comparison between codebased and model-based test case prioritization. The results from the experimental study indicate that model-based test prioritization detects early fault as compare to code-based test prioritization. However due to the sensitiveness property the early fault detection of model-based prioritization may be deteriorate if incorrect response is given by the tester or the developer. The model-based test case prioritization is less expensive than the code-based test case prioritization because execution of the model is faster than the execution of the whole code.

Acharya et al.[16] presented a method for prioritize the test cases for testing component dependency in a Component Based Software Development (CBSD) environment using Greedy Approach. Here the author first convert the system model i.e. sequence diagram to An Object Interaction Graph (OIG) using an algorithm. Then the OIG is traversed to calculate the total number of inter component object interactions and intra component object interactions. Depending upon the number of interactions between the object an objective function is calculated and then the test cases are ordered accordingly.

Swain et al.[17] proposed an approach to generate test cases and prioritize those test cases based on a test case prioritization metric. Here the author has used UML sequence and activity diagrams for their purpose. The sequence and activity diagrams are converted into testing flow graph (TFG) from which test cases are generated. Then the TFG is converted to a model dependency graph (MDG). Next, he calculated various weights for nodes (message-method/activity) as well as edge (condition) of the MDG based on a rational criterion. Weight of the node is calculated by using the number of nodes in Forward Slice (NFS) of node of MDG and weight of the edge is calculated by multiplication of the number of incoming control dependencies (edges) of node Ni and the number of outgoing control dependencies (edges) of node Nj. After calculating the weights of the node and edge the he calculated the weight of the basic path by adding the weight of the node and edge. Then he prioritized the test case in descending order of the weight.

Kumar et al.[18] have proposed a new approach which considers the severity of faults based on requirement prioritization. They considered four different factors to assign the weights to the requirements: Business Value Measure (BVM), Project Change Volatility (PCV), Development Complexity (DC), and Fault Proneness of Requirements. To calculate the Total percentage of fault detected (TSFD), the author used severity measure(SM) of each fault. Once the fault has been detected then they assign some severity measure to each fault according to requirement weights, to which it is mapped. Total Severity of Faults Detected (TSFD) is the summation of severity measures of all faults identified for a product.

Mall et al.[19] presented a method for model based approach to prioritize regression test cases for object oriented programs. Here the author represents all relevant object-oriented features such as inheritance, polymorphism, association, aggregation and exception. Here the authors also included dynamic aspects such as message path sequences from UML sequence diagrams. The author also considered the dependencies among test cases for test case prioritization. Here the author named their proposed model as Extended Object oriented

System Dependence Graph (EOSDG). This model extends LH-SDG and includes exceptions and message path sequencing information. The author named their prioritized techniques as Model-based Regression Test Case Prioritization technique. This approach involves two activity diagrams one activity diagram represent the activities that are performed before the testing process and the second activity diagram represent the activities that are performed each time a software is modified. The author constructed a backward slicing of the EOSDG n then constructed a backward slicing of the EOSDG and the collect the model elements in both the slicing and then he prioritized the test cases in descending order of the coverage of the model elements.

III. PROPOSED METHODOLOGY

In this section we discuss our proposed approach to generate a prioritize test cases. Our approach consists of the following three steps.

a) Maintaining a repository.

b) Matching the project type of the new projects on which regression test has to be performing with the existing projects contained in the repository and identifying the affected functions. Assigning business criticality values to the affected functions using statistical data stored in the repository.

c) Prioritizing the test cases according to the Business Criticality Test Value (BCTV) of the test cases in descending order.

The framework of proposed methodology is shown in Fig. 1. The input to our proposed approach is activity diagram of the new project.
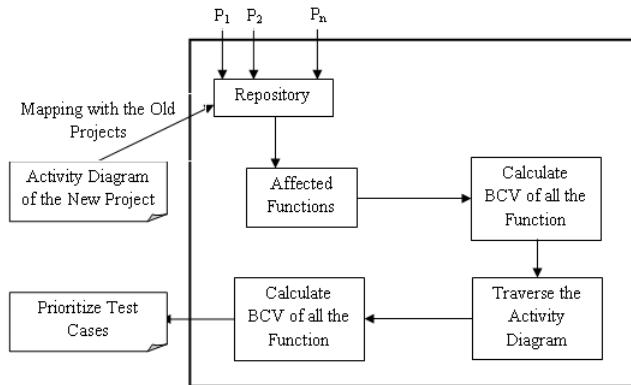

Fig. 1. Framework of the Proposed Methodology

As shown in Fig. 1 we first map our new project with the repository and then we found out the affected functions due to the change in the new project. After that we found out the BCV (Business Criticality Value) of each function. Then we traverse the activity graph of the new project. Using the BCV, we calculate the BCTV (Business Criticality Test Value) of each test case are found out. Finally, we prioritize the generated test cases according to BCTV.

B. Maintaining Repository

A repository is maintained for various historical projects of different category. Before maintaining the repository, a historical search is perform for finding various existing projects of unrelated categories such as application projects, networking projects, database projects, etc. for satisfying different needs of the end user. After finding out various projects belonging to different category, a table is update for each project which keeping track of following information.

- First the types of changes that have occurred during the maintenance of the project satisfying the end user's requirements are identified.

- Secondly different functions that are being affected due to the changes occurred in the projects have also been maintained, which will help us in finding out the prioritized test case. The affected functions can be identified for each change of the project with the help of foreword slicing method and expertise judgments. Foreword slicing method is applied to the particular node which are being added or changed according the end user's requirements.

The affected functions have been calculated with the help of foreward slicing algorithm which is shown in the Algorithm 1.

ALGORITHM1: FOREWARD SLICING ALGORITHM:

Input: An activity Diagram that has a single start node and an empty set of node identifiers associated with each node.

Output: Forward slice of each node.

1) **Initialization:** Set $S_i = \phi$ and $V_i = 0$ for $\forall i$. where S_i is the set associated with node N_i and V denotes the visited status of node N_i.
2) Call *ForwardSlice(start node)*
3) ForwardSlice (node N_i)
4) Begin
5) if $V_i=1$
6) exit(0);
7) else
8) begin $V_i=1$ /* Mark node as visited */
9) Find $F_i = N_i / N_{(i+1)}$ depends on N_i
10) Set $S_i = S_i \cup F_i$
11) for(each node $N_i \in F_i$)
12) ForwardSlice (N_i);   /*Function called recursively*/
13) End
14) end if
15) end

For example different projects is having C1, C2, C3...Cn types of changes which are affecting functions from the set of function. F1, F2, F3...Fn types of function. The types of changes and the list of affected functions are given in Table 1.

The Table 1 maintains three attributes such as Project ID, Types of change and affected functions. From the Table II it is found that Project-1 under gone C1, C2 and C3 types of changes. C1 is affecting F1, F2, F4, F6, F7, F8, F10, F12, F13, F15 functions, C2 is affecting F1, F5, F6, F7, F8, F9, F11, F12, F14, F15 types of functionalities and C3 is affecting F2, F3, F4, F8, F9, F10, F11, F12, F13 functionalities. In this similar manner information for other projects undergone regression test have also been maintained in the repository table.

C. Evaluating Business Criticality Value

In this section whenever regression testing has to carry on a new project, the business criticality value of the various affected functions have been calculated in the following manner. Suppose a new project has encountered along with the information with us about the subsequent changes that the project has undergone according to the customer requirement. Then we will first match the new project with the existing projects that are maintained in the repository. After finding out the matching project during the matching process we will then find out the function that are being affected due to the changes occurred in the new project from the repository.

For example we encountered a new project P NEW1 having changes C2, C7, C9 which match with the project PID_2 from the repository then we will find that function F3, F4, F5, F6,

F7, F8, F10, F11, F12, F13, F15 that have been affected by the change C2, factors F2, F3, F4, F5, F8, F9, F10, F11, F12, F14, F15 have been affected by the change C7 and factors F2, F4, F6, F8, F10, F12, F14, F15 have been affected by the change C9. Since the new project P NEW1 with changes C2, C7, C9 matches with the PID 2 type. The affected functions are listed in Table 2.

Table 1. Repository Table

Project ID	Types of changes	Affected functions
PID_1	C1	F1,F2,F4,F6,F7,F8,F10,F12,F13,F15
	C2	F1,F5,F6,F7,F8,F9,F11,F12,F14,F15
	C3	F2,F3,F4,F8,F9,F10,F11,F12,F13
	C4	F1,F3,F5,F6,F7,F8,F11,F12,F13,F16
	C5	F3,F4,F6,F7,F9,F10,F12,F14,F15,F16
	C9	F6,F7,F8,F9,F10,F11,F12,F13,F14,F16
	C10	F3,F4,F5,F7,F8,F9,F13,F15
PID_2	C2	F3,F4,F5,F6,F7,F8,F10,F11,F12,F13,F15
	C3	F1,F2,F3,F6,F8,F9,F10,F12,F13,F14
	C6	F1,F2,F3,F4,F5,F6,F8,F9,10
	C7	F2,F3,F4,F5,F8,F9,F10,F11,F12,F14,F15
	C8	F1,F2,F3,F4,F5,F6,F7,F9,F10,F11
	C9	F2,F4,F6,F8,F10,F12,F14,F15
	C10	F1,F3,F5,F7,F9,F11,F13,F14,F15
	.	.
	.	.
PID_n	C1	F1,F3,F5,F7,F9,F10,F11,F12,F13,F15
	C3	F1,F4,F5,F6,F7,F9,F10,F11,F13,F14
	C5	F2,F4,F6,F7,F8,F9
	C7	F3,F4,F5,F6,F7,F8,F9,F10,F11,F14,F15
	C8	F2,F3,F4,F5,F6,F7,F8,F9,F10,F11,F15
	C9	F4,F5,F6,F7,F8,F9,F10,F11,F15
	C10	F2,F3,F4,F7,F8,F9,F10,F12,F13,F14

Table. 2. Change_Detail_New

Changes in P_NEW1	Affected Factors
C2	F3, F4, F5, F6, F7, F8, F10, F11, F12, F13, F15
C7	F2, F3, F4, F5, F8, F9, F10, F11, F12, F14, F15
C9	F2, F4, F6, F8, F10, F12, F14, F15

The Table 2 stores the information about the changes that are being made in the new project along with the functions that are being affected.

Further the Business Criticality Value (BCV) of the various functions is found out. A Business Criticality Value (BCV) is defined as the "amount of the function contribution towards the success of the project implementation." For example in a Banking Automation Software, there are two activities such as to do transaction and collect feedback. The money transaction activity will be having higher BCV than the feedback collection activity.

The Business Critical Test Value is calculated as follows:

- First find out those functions that are being affected due to the changes made in the project.
- Find out the average interaction of each function within that project.

The Business Critical Value is calculated by the following formula 2.

$$\text{BCV (Fn)} = \frac{\text{No.of times Fn encounter}}{\text{Total no of factor being affected}} \qquad (2)$$

Now the BCV of each functionality have been calculated according to the formula and the value has been store in the Table 3. BCV table i.e. Table 3 stores certain kind of

information such as the function name along with the average interaction and the BCV value of each function.

D. 3.3 Prioritizing Test Cases

In this section the test cases are prioritized according to the Business Criticality Values of the different factors. Every testcase executes different factors of a project according to the Depth First Search (DFS) of individual test cases are identified known as Business Criticality Test Value (BCTV) of that test case the project. And every factor is having different Business Critical Values.

Initially different test cases are identified and BCTV of the function that is encounter during the traversal process. This is shown in Table 4. Then we ordered our test cases in descending order of the Business Criticality Values.

Table.3. Business Criticality Value (BCV) Table

Factors	No. of times each factor encounter	BCV
F1	0	0
F2	2	0.067
F3	2	0.067
F4	3	0.1
F5	2	0.067
F6	2	0.067
F7	1	0.033
F8	3	0.1
F9	1	0.033
F10	3	0.1
F11	2	0.067
F12	3	0.1
F13	1	0.033
F14	2	0.067
F15	2	0.067

PRIORITIZATION ALGORITHM:

Our proposed technique to prioritize regression test cases is algorithmically represented in algorithm 2.

Algorithm 2:
1. Maintain a repository which contains different types of projects, no. of changes and the affected functionality due to the changes.
2. Matching the new project with the repository and identifying the no. of changes and the affected functions respectively.
3. Calculate the business criticality value of each function according to the equation 2.
4. Then traverse the activity diagram of the new coming project with the help of DFS with individual test case.
5. Find the BCTV of each test case by adding the BCV value of each factor.
6. Then prioritize the test cases according to the descending order of BCTV for each test case.

Table. 4. Prioritization Table

Test Case	Traversing Factors	BCTV of each Test Case
t_1	F1,F2,F3,F4,F6,F8,F9,F10,F11,F12	0.701
t_2	F1,F3,F4,F5,F7,F8,F9,F10,F15,F16,F17	0.567
t_3	F2,F3,F4,F5,F6,F7,F8,F10,F11,F12,F13,F15,F18	1.371
t_4	F3,F5,F6,F7,F8,F9,F10	0.467
t_5	F1,F3,F5,F7,F8,F9,F10,F11,F12,F14,F15	0.701

The prioritize table stores the information about the test cases that are obtained after the traversal process. It also stores the information about the functions that are being encountered during the traversal process. Finally we have found out the BCTV of each test case and prioritized the test suite according to the descending order of the BCTV values.

Hence the prioritize test suite is:

t3, t1, t5, t2, t4 or t3, t5, t1, t2, t4.

IV. A CASE STUDY: SHOPPING MALL AUTOMATION

In this section our proposed approach is described with a case study of a Shopping Mall Automation System. Since in our proposed method we are matching the new projects goes through change with the repository to find out the affected functions. So that whenever a new project is encounter we can find out the affected function.

After finding out the affected function from the repository the BCV value of each function is calculated. Then the total BCTV value of each test case is calculated by adding the BCTV value of those functions that are visited during the DFS traversal of the Activity Diagram of the new encounter project. Finally the test cases are prioritizing according to descending order of their BCTV value. Suppose in the past there was a need for a big bazaar project.
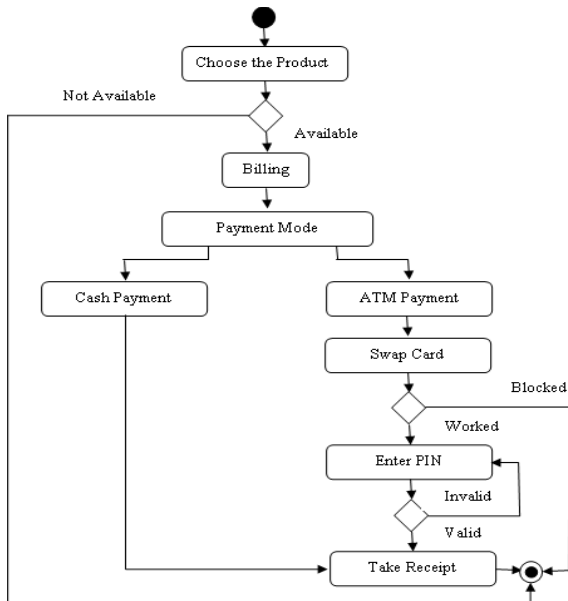
Fig. 2. Activity Diagram of Shopping Mall Automation System

So projects have been designed whose activity diagram is shown in Fig. 2.

After the project have been submitted, it was found that there was a need of some additional functionality for big bazaar application so the necessary changes have been made to the submitted project, which activity diagram is shown in Fig. 3. A functionality detail table has been maintained which store the different types of functions present in the activity diagram of the Big Bazaar Automation System and their function id as shown in Table 5.
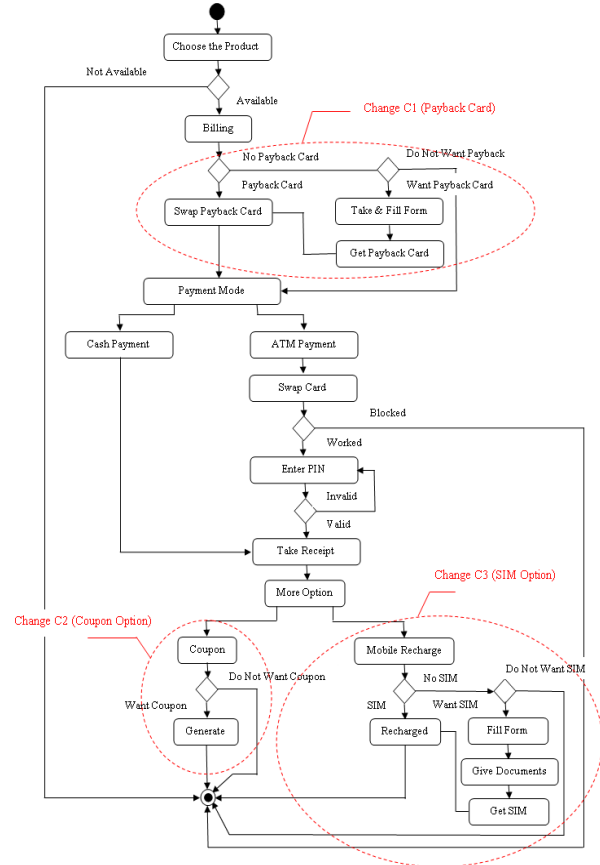
Fig. 3. Activity Diagram for A Big Bazaar Automation System

The types of changes that have been made in the project with the functionalities that have been affected by the changes have been maintained in a repository which is shown in Table 6.

In our project there are three numbers of changes have been made. C1 change is for payback card. C2 change is for coupon option and C3 change is for SIM card option. The changes made are highlighted through dotted line in the activity diagram of the project which is shown in Fig. 3.

A. Matching a project with the repository

Suppose we now encounter a new shopping mall project which Activity Diagram is shown in Fig. 4. We found that the new project matches to the big bazaar project and having C1 and C2 types of changes those are highlighted in through dotted line in Fig. 4. Now we match it with the repository and find the affected functionalities due to the changes C1 and C2. This is shown in Table 7.

Now the BCV of each functionality have been calculated according to the formula 2 and the value has been store in the Table 8.

Table.5. Functionality Detail Table

FUNCTION ID	FUCTIONALITY	FUNCTION ID	FUCTIONALITY
F1	START	F16	Check PIN
F2	Choose the Product	F17	Take Receipt
F3	Check Available	F18	More Option
F4	Billing	F19	Mobile Recharge
F5	Have Payback Card	F20	Have SIM
F6	Want Payback Card	F21	Want SIM
F7	Swap Payback Card	F22	Fill Form
F8	Take and Fill Form	F23	Give Documents
F9	Get Payback Card	F24	Get SIM
F10	Payment Mode	F25	Recharged
F11	Cash Payment	F26	Coupon
F12	ATM Payment	F27	Want Coupon
F13	Swap Card	F28	Generate
F14	Check Working	F29	FINISH
F15	Enter PIN		

Fig. 4. Activity Diagram of New Shopping

Table. 6. Repository (A Shopping Mall Project)

No. Of Changes	Affected Functionality
C1	F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16, F17, F18, F19, F20, F21, F22, F23, F24, F25, F26, F27
C2	F18, F26, F27, F28, F29
C3	F18, F19, F20, F21, F22, F23, F24, F29

Table. 7. Affected Function due to the Change in the New Project

No. Of Changes	Affected Functionality
C1	F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16, F17, F18, F19, F20, F21, F22, F23, F24, F25, F26, F27
C2	F18, F26, F27, F28, F29

Table. 8. Business Criticality Value (BCTV) Table

FUCTIONALITY(Fn)	No. of times each Function Encounter	BCV	FUCTIONALITY(Fn)	No. of times each Function Encounter	BCV
F1	0	0	F16	1	0.04
F2	0	0	F17	1	0.04
F3	0	0	F18	2	0.08
F4	0	0	F19	1	0.04
F5	1	0.04	F20	1	0.04
F6	1	0.04	F21	1	0.04
F7	1	0.04	F22	1	0.04
F8	1	0.04	F23	1	0.04
F9	1	0.04	F24	1	0.04
F10	1	0.04	F25	1	0.04
F11	1	0.04	F26	1	0.04
F12	1	0.04	F27	1	0.04
F13	1	0.04	F28	1	0.04
F14	1	0.04	F29	1	0.04
F15	1	0.04			

Now the activity diagram shown in Fig.4 has been traversed according to the Depth First Search (DFS) and the traversing functionalities of each test case are found out. After that the BCTV of each test cases are calculated by adding the BCV value of the functions and all these information have been maintained in table 9.

Table 9: Prioritization Table

Test Case	Traversing Functionality	BCTV of each Test Case
t_1	F1, F2, F3, F29	0.04
t_2	F1, F2, F3, F4, F5, F6, F10, F11, F17, F18, F26, F27, F29	0.40
t_3	F1, F2, F3, F4, F5, F6, F10, F11, F17, F18, F26, F27, F28, F29	0.44
t_4	F1, F2, F3, F4, F5, F6, F10, F12, F13, F14, F29	0.28
t_5	F1, F2, F3, F4, F5, F6, F10, F12, F13, F14, F15, F16, F17, F18, F26, F27, F29	0.56
t_6	F1, F2, F3, F4, F5, F6, F10, F12, F13, F14, F15, F16, F17, F18, F26, F27, F28, F29	0.6
t_7	F1, F2, F3, F4, F5, F9, F10, F11, F17, F18, F26, F27, F29	0.40
t_8	F1, F2, F3, F4, F5, F9, F10, F11, F17, F18, F26, F27, F28, F29	0.44
t_9	F1, F2, F3, F4, F5, F9, F10, F12, F13, F14, F29	0.28
t_{10}	F1, F2, F3, F4, F5, F9, F10, F12, F13, F14, F15, F16, F17, F18, F26, F27, F29	0.56
t_{11}	F1, F2, F3, F4, F5, F9, F10, F12, F13, F14, F15, F16, F17, F18, F26, F27, F28, F29	0.6
t_{12}	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F11, F17, F18, F26, F27, F29	0.52
t_{13}	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F11, F17, F18, F26, F27, F28, F29	0.56
t_{14}	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F12, F13, F14, F29	0.4
t_{15}	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F12, F13, F14, F15, F16, F17, F18, F26, F27, F29	0.68
t_{16}	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F12, F13, F14, F15, F16, F17, F18, F26, F27, F28, F29	0.72

After calculating the BCTV value of all the test cases we than prioritized the test suite according to the descending order of the BCTV values of the test cases.

Hence The Prioritize Test Sequence is:

t16, t15, t6, t11, t5, t10, t13, t12, t3, t8, t2, t7, t14, t4, t9, t1

Hence, our approach generates effective prioritized test cases because it more priority to the test cases those are having more Business Criticality Value.

V. COMPARISON WITH THE RELATED WORK

Several test case prioritization techniques are discussed in section 2. Code based test case prioritization techniques are discussed in [6, 8, 10] but as these are based on code so execution of code is very slow. Model based test case prioritization techniques are discussed in [12, 13, 14] and these are based on the no of mark transition executed by the test cases. Our approach is based on model and it considers the business criticality value of the function. As our approach is based on business criticality value so we gave more importance to that functionality whose business criticality value is more. So our approach is detecting fault as earlier to the other approach.

VI. CONCLUSION

In this paper we proposed a model based test case prioritization technique using the business criticality value of each functions. Business Criticality Value (BCV) is defining "as the amount of contribution towards the business of the project." The BCV of each factors are calculated based on the affected functionality of the project due to the subsequent changes of the project for satisfying the requirement of the customers. So the generated prioritization sequence is more efficient because it is generated based on the requirement of the customers. So the proposed prioritization method is more effective and efficient. This gives an early change to the debuggers to work with the most critical function first.

VII. FUTURE WORK

Our approach is a model-based test case prioritization which is specifically contains the functional features of the project. We can also extend our approach by adding the non-functional features of the project. In future we also implement our proposed approach with the help of IBM Quality Seed: Functional Tester for Regression Testing.

REFERENCES

[1] K. Onoma, W-T. Tsai, M. Poonawala, and H. Suganuma, "Regression Testing in an Industrial Environment", Comm. ACM, vol. 41, no. 5, pp.81-86, May 1988.

[2] D. Binkley, "Semantics Guided Regression Test Cost Reduction", IEEE Trans. Software Eng., vol. 23, no. 8, pp. 498-516, Aug. 1997.

[3] T.Y. Chen and M.F. Lau, "Dividing Strategies for the Optimization of a Test Suite", Information Processing Letters, vol. 60, no. 3, pp. 135-141, Mar.1996.

[4] W.E. Wong, J.R. Horgan, S. London, and A.P. Mathur, "Effect of Test Set Minimization on Fault Detection Effectiveness", Software Practice and Experience, vol. 28, no. 4, pp. 347-369, Apr. 1998.

[5] G. Rothermel, M.J. Harrold, J. Ostrin, and C. Hong ,"An Empirical Study of the Effects of Minimization on the Fault Detection Capabilities of Test Suites", Proc. Int'l Conf. Software Maintenance, pp. 34-43, Nov. 1998.

[6] G. Rothermel, R.H. Untch, C. Chu, and M.J. Harrold, "Prioritizing Test Cases For Regression Testing", IEEE Transactions on Software Engineering, vol. 27, No. 10, pp 929-948, October 2001.

[7] Bogdan Korel, George Koutsogiannakis, "Experimental Comparison of Code-Based and Model-Based Test Prioritization", IEEE International Conference on Software Testing Verification and Validation Workshops, pp.77-84.

[8] G. Rothermel, R.H. Untch, C. Chu, and M.J. Harrold, "Test Case Prioritization: An Empirical Study", Proc. Int'l Conf. Software Maintenance, pp. 179-188, Aug. 1999.

[9] Alexey G. Malishevsky, Joseph R. Ruthruff, Gregg Rothermel, Sebastian Elbaum, "Costcognizant Test Case Prioritization", Proc. IEEE International Conference on Software Maintenance,2006.

[10] S. Elbaum, A. Malishevsky, G. Rothermel, "Test Case Prioritization: A Family of Empirical Studies", IEEE Transactions on Software Engineering, vol. 28, No. 2, pp 159-182, 2002.

[11] Praveen Ranjan Srivastava, "Test Case Prioritization", Journal of Theoretical and Applied Information Technology, 2005 - 2008 JATIT, pp. 178-181.

[12] H. Srikanth, L. Williams, J. Osborne, "System Test Case Prioritization of New and Regression Test Cases", IEEE, 2005, pp.64-73.

[13] B. Korel, L. Tahat, M. Harman, "Test Prioritization Using System Models", Proc. 21st IEEE International Conference Software Maintenance (ICSM'05), pp. 559-568, 2005.

[14] B. Korel, G. Koutsogiannakis, L. Tahat, "Application of System Models in Regression Test Suite Prioritization", Proc. 24st IEEE International Conference Software Maintenance (ICSM '08), pp. 247-256, 2008.

[15] B. Korel, L. Tahat, B. Vaysburg, "Model Based Regression Test Reduction Using Dependence Analysis", Proc. IEEE International Conference on Software Maintenance, pp. 214-223, 2002.

[16] A. Acharya, D. P. Mohapatra and N. Panda, "Model based test case prioritization for testing component dependency in cbsd using uml sequence diagram", IJACSA, vol. 1, no. 3, pp. 108-113, December. 2010.

[17] S. K. Swain, "Test Case Prioritization Based on UML Sequence and Activity Diagrams", PhD thesis, KIIT University, 2010.

[18] Dr. V. Kumar, Sujata and M. Kumar, "Test Case Prioritization Using Fault Severity", IJCST, vol. 1, Issue 1, pp. 67-71, September. 2010.

[19] R. Mall and C. R. Panigrahi, "Test case prioritization of object oriented Program", In SETLabs Brieng, Infosys, vol. 9, pp. 31-40, 2011.

AUTHOR PROFILE

Sonali Khandai is a M.tech student of KIIT University, Bhubaneswar, Odisha, INDIA. Her research areas include Software Testing. She has completed her B.tech in Computer Science and Engineering from BPUT in the year 2010. She can be reached at khandaisonali@gmail.com

Arup Abhinna Acharya is an Assistant Professor and research scholar in the School of Computer Engineering, KIIT University, Bhubaneswar, Odisha, INDIA. He received his Master's degree from KIIT University Bhubaneswar. His research areas include Object Oriented Software Testing, Software Cost Estimation, and Data mining. Many publications are there to his credit in many International and National level journal and proceedings. He is having eight years of teaching experience. He is a member of ISTE. He can be reached at aacharyafcs@kiit.ac.in.

Durga Prasad Mohapatra received his Master's degree from National Institute of Technology, Rourkela, India. He has received his Ph.D. from Indian Institute of Technology, Kharagpur, India. He is currently working as an Associate Professor at National Institute of Technology, Rourkela. His special fields of interest include Software Engineering, Discrete Mathematical Structure, Program Slicing and Distributed Computing. Many publications are there to his credit in many International and National level journal and proceedings. He is a member of IEEE. He can be reached at durga@nitrkl.ac.in.

Troubleshooting Microprocessor Based System using An Object Oriented Expert System

D.V. Kodavade

Associate Professor in Computer Sc. & Engineering, DKTE Textile & Engineering Institute Ichalkaranji, INDIA.

S.D.Apte

Professor in Electronics Engineering,
Rajshree Shau College of Engineering, Pune.INDIA.

*Abstract -*The paper presents an object oriented fault diagnostic expert system framework which analyses observations from the unit under test when fault occurs and infers the causes of failures. The frame work is characterized by two basic features. The first includes a fault diagnostic strategy   which utilizes the fault classification and checks knowledge about unit under test. The fault classification knowledge reduces the complexity in fault diagnosis by partitioning the fault section. The second characteristic is object oriented inference mechanism using backward chaining with message passing   within objects. The refractoriness and recency property   of inference mechanism improve efficiency in fault diagnosis. The developed framework demonstrates its effectiveness and superiority   compared to earlier approaches.

Keywords- Inference mechanism; object; Fault diagnosis.

I. INTRODUCTION

Expert Systems have traditionally been built using large collection of rules based on empirical associations; Interest has built up recently in use of Artificial Intelligence techniques that reason from first principles i.e. from an understanding of causality of the device being diagnosed. Randall Davis [1] discussed causal interaction model for fault diagnosis. Expert system that reason based on understanding of the structure and function of the unit under test has been explored in number of domains, including medicine [2-3], computer fault diagnosis [4], automobile engine fault diagnosis [5], and electronics equipment fault diagnosis [6]. Proposed work focuses on the last domain.

Fault diagnosis methodology operates on observed erroneous behavior and   hardware structure of the unit under test. The erroneous behavior consists of responses of different components on the output lines on specific input values. Development of a methodology which determines possible sources of causes in minimum time for a specified fault is basic aim of the research. In digital components, there is fixed deterministic flow of signals from input to output. The signals are binary in nature and flow through various components.

In keeping with the notion of reasoning from first principles, a scheme is proposed herewith to develop a system capable of reasoning in a fashion similar to an experienced electronic engineer. In particular, the system is built by capturing skill exhibited by an engineer who can diagnose faults   from schematic even though he may never have seen that particular unit before. However, the average person who does not possess the experience has to check all components that may be faulty. It leads to low efficiency in troubleshooting process and is not acceptable for large & complex devices. Diagnosing a faulty component from an electronic circuit board is challenging problem. Applying artificial intelligence approach to solve this problem is true motivation behind this research. Fault diagnosis requires expertise and knowledge in specific domain. Initial focus of research is to develop a rule based expert system for fault diagnosis in microprocessor system and then explore it to object oriented framework to evaluate the correctness compared to other approaches.

II. BRIEF REVIEW ON MAJOR EXISTING METHODS FOR FAULT DIAGNOSIS

A. Expert Systems (ES)

Many expert systems have been developed for fault diagnosis in different domains. C. Angeli [7] discussed diagnostic expert system for real time application using functional reasoning. To handle   online diagnostic constraints, model based approach was proposed for fault diagnosis in real time application.

Jinyu Qu, Liyan Liary [8] proposed a production rule based expert system for electronic control automatic transmission fault diagnosis. Here every fault and cause of fault (fault reason) has been assigned aunique codes and both are stored in database. Rule base is designed for mapping relationship between fault reasons and fault types using AND/OR form of forward reasoning.

Ioan Borlea, Adrain Buta [9] devised an expert system for fault diagnosis in Timisoara 220 KV Substation. Fault diagnosis method uses reasoning based on rules inferred from operation of substation's primary equipment and main bus bar and auto transformer protections.

Chen Jingie, Chen Xia xia [10] presented the traditional airborne electronic equipment fault diagnostic system. It executes the dynamic processing by subsystems, then summaries   information and makes the integrated diagnosis by the expert system which is embedded in flash memory. It uses forward extract rule base approach for inference mechanism.

Ting Han,Bo Li, Linei Xu [11] has proposed a universal fault diagnosis expert system based on Bayesian network, it utilizes expert knowledge to diagnose the possible root causes and the corresponding probabilities for maintenance decision making support. Bayesian network is used as an inference

engine for raw data analysis. Authors has tested the system on production line of a chipset factory and obtained satisfactory results.

Sebastien Gebus,Kauko Leivisa[12] dicussed how defect related knowledge on an electronic assembly line can be integrated in decision making process at an operational and organizational level. It focuses particularly on the efficient acquisition of shallow knowledge concerned with production. Authors concluded that, the effective decision support system is essential to provide workers with information necessary to identify the causes of problems and takes appropriate action to solve it.

John W.Coffey, Alberto J. Canas, et al.[13] has discussed EI-Tech expert system to provide performance support and training for electronic technicians in troubleshooting RD-379A(V)/UNH, a redundant, fault tolerant, air traffic control recorder system.

Ning Yang,Shaocheng Zhang Xu et al.[14] built an expert system for vibration fault diagnosis in large steam turbine generator set. Knowledge base is constructed using production rules and inference engine is based on confidence factors, a mathematical model is proposed by authors to calculate Confidence Factor (CF)during reasoning process. Diagnostic system consist of two parts: data acquisition system and fault diagnostic expert system. Data acquisition system is responsible for collecting vibration signals and the diagnostic expert system analyses it.

B. Artificial Neural Netwoks (ANNs) :

Many neural network models were suggested for fault diagnosis and prediction problems. Yong Chun Liang, Xiao-Yun Sun et al.[15] proposed a combinatorial probabilistic neural network(PNN) model for fault diagnosis of power transformers. PNN model is based on Bayesian classification. Four PNN models for fault classification are proposed to classify normal heat fault, partial discharge fault, general over heating faults and severe overheating faults. Authors obtained better accuracy compared to other approaches.

An adaptive neural network based fault detection and diagnosis using unmeasured states is proposed by C.S.Liu,S-J Zhang,S.-S. Hu.[16]. Authors built a fault diagnostic architecture for unknown nonlinear systems with unmeasured states. A radial basis function (RBF) and adaptive RBF neural network approaches are used to approximate the model of unknown systems and for on line updates respectively.

Yanghong Tan, Yigang He, Chun Cui et al.[17] has proposed a neural network and genetic algorithm based approach for analog fault diagnosis . By understanding the measurable voltage deviation in the deviation space the unified fault vectors for single, double and triple faults are characterized. The classification of faults is done using artificial neural network.

C. Hybrid intelligent Systems (HISs) :

The combination of neural network and rule based expert system is proposed by Rye Senjen, Muriel De Beler [18]. The reasoning mechanism is implemented using neural networks. The hybrid system is developed for performance monitoring and fault diagnosis in telecommunication networks. Here performance monitoring is carried out using neural network and fault diagnosis is carried out using rule based expert systems.

Damian Grzechca,Jerzy Rutkowski [19] discussed Neuro-Fuzzy approach to time domain electronic circuits fault diagnosis. Proposed method belongs to Simulation Before Test (SBT) technique , a simple step input is given to unit under test and response is analyzed. The information acquired such as a rise time ,input output delay, overshoot are fuzzyfied and fuzzy neural dictionary is created. Feed forward back propagation network classifier algorithm is demonstrated with analog filter circuit.

D. Petri–Nets (PN) :

Petri Nets are used for multiprocessing and on line system modeling. Antonio Ramfrez-Trevino,Elvia Ruiz-Beltran et al.[20] proposed an online model-based for fault diagnosis of discrete event systems. Model of the system is built using the interpreted Petri Nets (IPN).Model includes all system states as well as all possible faulty states.IPN modeling methodology follows a modular bottom-up strategy. A diagnostic algorithm is used to diagnose the faulty component.

A fuzzy petri-nets approach for fault diagnosis for electro mechanical equipment is discussed by Qunming Li,Ling Zhu,Zhen Xu [21] . The information flow in fuzzy petri net model (FFDPN) is driven inversely, and the production rules are defined backwards. The author has demonstrated how this proposed model can be used for other domains as well.

Chunlai Zhou, Zhongcheng Jiang [22] devised a fault diagnosis approach for TV transmitters based on Fuzzy PetriNets. All the knowledge of fault diagnosis is summarized into fuzzy rules, these fuzzy rules then translated into fuzzy petri nets by using an algorithm. A parallel reasoning algorithm is proposed for reasoning in fault diagnosis

E. Fuzzy Logic (FL) :

To handle incomplete and linguistic knowledge fuzzy logic is used. As per survey fuzzy logic is applied to may fields for handling inexact situations. Yan Qu et al. [23] discussed fuzzy diagnostic expert system for electric control engine. Comix fuzzy reasoning method is used in inference engine. Proposed expert system includes knowledge base, reasoning machine, explain system, management system and human machine interface modules.

An intelligent fault diagnosis framework based on fuzzy integrals is built by M. Karakose I et al.[24]. The method consist two frameworks. The first framework used to identify the relations between features and a specified fault and the second framework integrates different diagnostic algorithms to improve accuracy rate. Approach is experimented on 0.37 KW induction motor, where broken rotor bar and stator faults were evaluated to validate the model.

An electronic equipment fault diagnosis in air crafts using fuzzy fault tree is described by Lians Xiao- lin et al. [25]. The complexity, ambiguity and uncertainty in fault diagnosis process for equipment fault diagnosis is modeled using fuzzy fault tree. The list of the most suspected faults is given by the system with fuzzy measures.

III. ARCHITECTURE OF RULE BASED EXPERT SYSTEM

Rule based expert systems have a wide range of applications for diagnostic tasks where expertise with deep knowledge is rarely available. The architecture of basic expert system is shown in Figure 1. The system consists of knowledge base, inference engine and user interface. The expert knowledge is elicited by knowledge engineer and represented in suitable knowledge structures. The knowledge base is constructed using production rules. It describes the action that should be taken if a symptom is observed. The main feature of rule based system is empirical association between premises and conclusions in the knowledge base. These associations further describe cause effect relationship to establish logical reasoning.
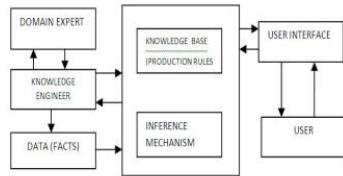

Figure 1.   Basic Structure of Rule-Based Expert System.

For fault diagnosis in a typical microprocessor system board, 60 different faults have been considered. Diagnostic rules are typically of the form if <X is true> Then < add Y to the suspect list>.

An example from the rule base, is

IF (1) Pin_30 of 8085 is low continuously AND

(2) +5 v present at pin no.40 of 8085 AND

(3) CLOCK pin 1 is pulsating

THEN The IC 8085 faulty.

This rule "fire" if conditions (1), (2) and (3) were found to be true then component listed in the "then" statement get added to the list of suspect faulty components.

Inference engine uses depth first search technique with back ward reasoning. The expert system   diagnoses the fault by interacting with the user. The results obtained using this system for typical faults are listed in Table 1.

TABLE I.   RESULTS OBTAINED USING RULE BASED EXPERT SYSTEM

FaultQuery	List of possible Faults
1.System is dead	Power supply faulty Power supply connector loose connections Power cable defective
2. Keyboard not working	8279 faulty Keyboard faulty
3. 8253 port A not working in Mode 0	8253 faulty
4. Memory read from C000 H not working	Ram _6116 faulty 74ls138_U6 faulty

IV. ARCHITECTURE OF THE PROPOSED OBJECT ORIENTED EXPERT SYSTEM

The overall architecture of the object oriented expert system is shown in Fig. 2. The system consists of knowledge base, inference mechanism, user interface. Knowledge base consists of declarative knowledge and procedural knowledge. Inference mechanism uses backward chaining & message passing technique. To get observations from the unit under test user interface is provided. The object attribute values obtained during diagnosis are stored in working memory.
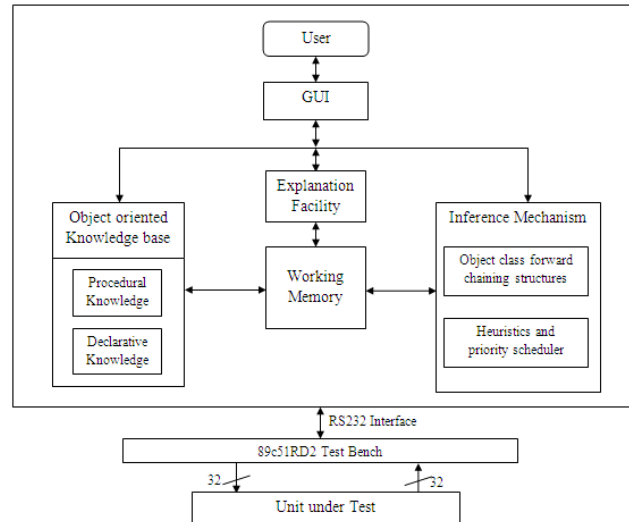

Figure 2.   Architecture of the Object Oriented Expert System.

For off line testing of the components like 74Ls138 decoder a test bench interface is provided using RS232 port. Test bench uses 89c51RD2 microcontroller with 32 input /output test probes.

V. OBJECT ORIENTED KNOWLEDGE REPRESENTATION

The object oriented knowledge base is lumped by many objects, and is a modular, uniform and structured paradigm. It can be uploaded easily which increases generality of the system. Again, by using inheritance property, knowledge reusage increases as compared to structured representation. The complex data types like heuristics can be implemented as attributes in object oriented knowledge base. Under the microprocessor super class there are many sub classes like CLASS_8255, CLASS_8279, CLASS_8253, CLASS_6116, CLASS_2764. The super class inherits all common attributes of sub classes. The pin no. of ICs acts as objects. The status of the pin can be obtained by passing the object attributes novice technician and the attribute values get stored in working memory.

The connectivity between the components is described using frame structures and the functional behavior of the components is described using object-rules. The system uses fault classification and checks knowledge. The fault classification knowledge isolates the object space as per the selected fault query. It is represented as a FAULT_ISOLATE class. This class uses metaknowledge, which determines valid classes for the selected fault query with priority. After the fault

query is selected the FAULT_ISOLATE class analyses the fault query by getting the attribute values form the user

For example following part of metaknowledge is applicable for the microprocessor system.

1. If there is fault in power supply, checking the behavior of circuit components which depends on it not necessary.
2. If microprocessor chip is faulty, then check knowledge about the program is invalid.

Checks knowledge describes functional behavior of components. It uses object-rules associations within the respective sub class. Interconnection between the components is represented using frames and slots. Declarative knowledge assists inference mechanism and novice technician in fault diagnosis by providing information about component connectivity. Pseudo code for frame representation in Visual Prolog is shown in Fig.3. Here, frame is constructed to represent pin connections of 8255 IC with FRC connector on circuit board. Frame name is 8255, slot 1 represents pin no.1 of 8255 IC, and the description of the pin name is given in facet. Another facet is used to represent its connection with FRC connector pin no.1.Next again, slot2 describe pin no.2. of 8255 and its connection to FRC connector pin no.2. like this all pins and their connections   is represented. For addition of new component assertz predicate is used to update knowledge base.

assertz(frame("8255",[slot("PIN_01",[facet("NAME","P1.0"),facet("CONNECT_TO","01FRC1")]), slot("PIN_02",[facet("NAME","P1.1"),facet("CONNECT_TO","02FRC1")]), slot("PIN_03",[facet("NAME","P1.2"),facet("CONNECT_TO","03FRC1")]), slot("PIN_04",[facet("NAME","P1.3"),facet("CONNECT_TO","04FRC1")]), slot("PIN_05",[facet("NAME","P1.4"),facet("CONNECT_TO","05FRC1")]), slot("PIN_06",[facet("NAME","P1.5"),facet("CONNECT_TO","06FRC1")]), slot("PIN_07",[facet("NAME","P1.6"),facet("CONNECT_TO","07FRC1")]),

Figure 3.   Pseudo Code For Frame Representation

VI.   INFERENCE MECHANISM

Inference mechanism uses backward chaining mechanism to find causes of fault. It uses conflict resolution strategy by selecting one object to fire from conflict set. Using object oriented structure the most recent observation will be treated as most promising and will be inherited from other objects like human expert does. For the present diagnostic approach only one fault query is selected at a time. On selection of the query from the user the system first isolates fault and applies the objects from isolated class, every node act as an object under sub class. System applies Depth First Search (DFS) strategy for searching the object tree. As the search starts from left side the most promising object is kept at left side as per expert judgment. Search proceeds by passing attributes and getting positive responses from user till fault is diagnosed. If attribute value is not positive the tree traversing stops at that sub node and backtracks till next node. As fault diagnosis is complex and probabilistic process, there is likelihood that other probable faults also get diagnosed under the same observed symptoms. To identify most promising one, confidence Value (CV) is computed. Every sub class is associated with confidence value

attribute. The confidence value of the conclusion   is obtained by taking minimum of attribute values associated with objects under that class and multiplying it with class attribute value. A threshold is kept to limit the suspected fault list.   Conclusions whose confidence values are greater than threshold get displayed as most suspected faults.

As an example, the fault query "Display not working" is selected by the user, fault isolate class isolates the fault area after getting attribute value, and activates DISPLAY_CLASS, under   this   class   Object (pin_3_of_FND_507_+5v,<attribute><attribute value>) passes the query to user to get status of pin no.3 of display. The attribute value obtained is supposing not positive and   get stored in the working memory. As shown in Fig. 4. As per DFS next object from the tree get fired. To diagnose the track open fault there is necessary of checking the status of   pin _3_U11_74ls145 and   FND 507 pin no.3 again suppose it is positive but   since status of FND 507 pin no3 checked previously and inherited it skips it and fires next object. Since pin_1_74245and pin 32-35_8279 are checked in the previous diagnosis it skips it and returns with track open fault
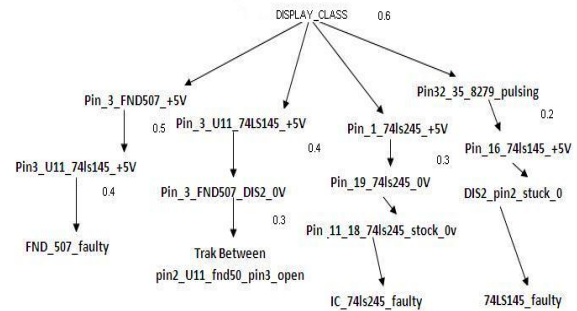

Figure 4.   Object Search Tree

VII.   RESULTS

Table1. Shows the results obtained for some typical faults in 8085 microprocessor board using two approaches. For the first fault query "system not getting started" rule based approach uses five rules to diagnose four faults while object oriented approach   has fired only   two objects and diagnosed fault correctly. The results obtained by two approaches are compared and validated by industrial experts. Fault diagnosis using Object oriented approach is more accurate and uses less object search space and hence memory. Similarly 80 different faults are considered and validated by industrial experts and found correct.

VIII.   CONCLUSION

This work is an attempt to speed up the fault diagnosis process using expert knowledge base and inference mechanism. Using object oriented approach problem domain divides naturally   and diagnosis is carried out as expert troubleshooter as predicted. Using   inheritance property the inference mechanism efficiency is increased and becomes more flexible and modular. As discussed in results, object oriented approach takes less time to diagnose the fault compared to other approaches.   As per validation report, results obtained using object oriented approach is 85% accurate.

TABLE II.   RESULT OBTAINED FOR TYPICAL FAULT

S No	Fault Query	Diagnosed Faults	
		Using Rule Based Approach	**Using Object Oriented Approach**
1.	System not getting started	8085 processor 74ls373 latch Crystal Diode in reset logic (No. of Rules fired = 5)	8085 faulty (No. of objects fired -02) CV = 0.8
2.	No display power is ON	EPROM 2764 6162 RAM No. Of rules 03	2764 No of objects -01 CV =0.7
3.	Data Not getting written to C100 onwards	74HS32 U18 8085 Strapping P6,P17 Open Rules - 03	Strapping P6 open Object- 01 Cv =0.6
4	Input Port Of 8255 is not working in input mode	74ls138U7 8255, 8085 Rules -03	8255 Objects -02 Cv =0.8
5.	8253 is not working in mode 0	74LS138 U7 8253 STRAPPING P18 Rules -04	8253 Objects 01 Cv =0.8

REFERENCES

[1] R.Davis "Reasoning from First Principles in Electronic Troubleshooting ," International Journal of Man Machine Studies,pp. 403-423,Nov.1983.

[2] "Rule Based Expert" *Systems, M.Sasikumar,S.Ramani* et al.

 a.   NarosaPublication 2007.

[3] B.G. Buchannan and R.O. Duda, "Principles of Rule-based Expert Systems," *Advances In Computers,* Vol.22, 1983, pp.163-21

[4] Roger T Hartley "CRIB: Computer Fault –finding ThroughKnowledge Engineering" *IEEE Computer* March 1984.   PP 76-81.

[5] Suk I.Yoo, Il Kon Kim," DIAS1 : An Expert System for Diagnosing Automobiles with Electronic Control Units" International Journal on Expert Systems With Applications, Vol 4. PP 69-78,1992.

[6] F. Pipitone,"The FIS Electronics Troubleshooting System,"IEEE Computers,pp.68-76,July 1986.

[7] C.Angeli., "Diagnostic Expert Systems: From Expert's

 a.   Knowledge to Real –Time Systems. Advanced Knowledge Based Systems : Model ,Applications & Research. Vol.1. PP 50-

 b.   73,2010.

[8] Jinyu Qu, Liyan Liang, " A Production Rule Based Expert System   for Electronic Control Automatic Transmission Fault Diagnosis" Procc. Of 2009 IEEE International Conference on Test & Automation.

[9] Ioan Borlea ,Adrian Buta" DIASE –Expert System Fault Diagnosis for Timisoara 22kv Substation" Proc of Eurocon 2005. PP 251 255.

[10] Chen Jingjie,Chen Xiaxia," Research on Embedded Airborne

 a.   Electronic Fault Diagnosis Expert System" Proceedings of

 b.   International Conference on Tests & Automation, 2010.

[11] Ting Han,Bo Li, Limei Xu ," A Universal Fault Diagnostic Expert System Based on Bayesian Network" Proceedings of 2008 IEEE International Conference on Computer Sc. & Software Engineering.PP 260-265.

[12] Sebastien Gebus, Kauko Leiviska," Knowledge acquisition for

 a.   decision support systems on an electronic assembly line"

 b.   International Journal on Expert Systems with Applications, Elsevier, 2007 PP94-101

[13] John W.Coffey, Alberto J. Canas et al. "Knowledge modeling and the creation of EI-Tech: a performance support and training system for electronic technicians", *International Journal on Expert Systems with Applications,* Elsevier Vol 25.2003, PP 483-492.

[14] Ning Yang,Shaocheng Zhang et al. " An Expert System for

 a.   Vibration fault Diagnosis of Large Steam Turbine Generator Set", Proc. IEEE International Conference on Test Automation, 2011.PP 217-221.

[15] Yong- Chun LiANG, Xiao- Yun Sun,Dong –Hui Liu et al. "Applications Of Combinatorial Probabilistic Nural Network In Fault Diagnosis Of Power Transformer" Procc. Of Sixth IEEE

 a.   International Conference on Machine Learning and Cybernetics,

 b.   2007.1115-111.

[16] C.-S. Liu,S.-J. Zhang, S.-S. Hu ," Adaptive neural –network –based fault detection and diagnosis using unmeasured states", International Journal on IET Control Theory Applications 2008 Vol2. No.12. PP 1066-1076.

[17] Yanghong Tan,Yigang He,Chun Cui and Guanyuan Qiu," A Novel Method for Analog Fault Diagnosis Based on Neural Networks and Genetic Algorithms" IEEE Transaction on Instrumentation and Measurement. Vol.57, No.11 Nov.2008. PP 2631-2635.

Rye Senjen,Muriel de Beler,Chris Leckie et al. " Hybrid Expert Systems for Monitoring and Fault Diagnosis" IEEE Computer 1993. PP 235-236.

[18] Damian Grzechca,Jerzy Rutkowski," Use of Neuro-Fuzzy System to Time Domain Electronic Circuit Fault Diagnosis", Proc. Of IEEE International Conference on Test and Automation, 2005. PP 1-4.

[19] Antonio Ramfrez et al. "Online Fault Diagnosis of Discrete Event Systems. A Petri Net- Based Approach", IEEE *Transaction on Automation Science and Engineering,* Vol.4.No.1. Jan 2007.pp 31

[20] Qunming Li,Ling Zhu,Zhen Xu," Fuzzy Petri-Nets based Fault Diangosis for Mechanical –electrical Equipment" Procc. of 2007 IEEE International Conference on Control and Automation, 2007 PP 2539 -2543.

[21] Chunlai Zhou and Zhongcheng Jiang, " Fault Diagnosis of TV Transmitters Based on Fuzzy Petri Nets" Procc. of IMACS Multiconference on " Computational Engineering in Systems Applications , 2006. PP 2003-2007.

[22] Yan Qu et al. "A Fuzzy Expert System Framework Using

a. ObjectOriented Technique" *Proceeding of 2008 IEEE Pacific –*

b. *Asia workshop on Computational Intelligence and Industrial*

c. *Applications* .PP474-478.

[23] M. Karakose, I.Aydin, ,E Akin "The Intelligent Fault Diagnosis Framework Based on Fuzzy Integral" *SPEEDAM 2010 international Symposium on Power Electronics, Electrical Drives, Automation and Motion.* PP 1634-1639.

[24] Liang Xiao-lin, Zhao Yan-xia .et al. "Research on Applications of Fuzzy Fault Analysis in the Electronics Equipment Fault Diagnosis" IEEE System Engineering & Electronics,Vol 2, 2010.

[25] Hruday: A Heart Disease Advisor Expert System, Vidwan- An Expert System Shell: User Manual", *National Centre for Software Technology,Mumbai,* 1993.PP18519

AUTHORS PROFILE

D.V.Kodavade, is PhD Scholar form Shivaji University, Kolhapur. He has completed B.E. Electronics in 1990 and M.E. Electronics in 1995 from the same university.Since 1992,he has been working as Assistant Professor in Electronics Engineering. Currently, he is an Associate Professor in Computer Sc. & Engineering at D.K.T.E Society's Textile & EngineerinInstitute, Ichalkaranji,Kolhapur, INDIA.

S.D.Apte received the PhD in Electronics Engg degree in 2001from Shivaji University, Kolhapur. She has worked as Professor in Electronics Engineering Department at Walchand College Of Engineering,Sangli form 1980-2007.Currently,she is working as Professor in Electronics Engineering at Rajshree Shau College of Engineering , Pune.She has published 28 research papers in international journals and obtained one patent on "Generation of Mother Wavelet".

Countermeasure for Round Trip Delay Which Occurs in Between Satellite and Ground with Software Network Accelerator

Kohei Arai

Graduate School of Science and Engineering

Saga University

Saga City, Japan

Abstract—**Countermeasure for round trip delay which occurs in between satellite and ground with network accelerator is investigated together with operating system dependency on effectiveness of accelerator. Also disaster relief data transmission experiments are conducted for mitigation of disaster together with acceleration of disaster related data transmission between local government and disaster prevention center. Disaster relief information including remote sensing satellite images and information from the disaster occurred areas to local government for creation of evacuation information is accelerated so that it becomes possible to send them to the residents in the suffered areas due to disaster through data broadcasting in the digital TV channel.**

Keywords-IP comunication through Internet Satellite; WINDS; software accerelator.

I. INTRODUCTION

Wideband Inter-Networking engineering test and Demonstration Satellite: WINDS (KIZUNA[1] in Japanese) was launched on February 23, 2008 (JST) from the Tanegashima Space Center (one of the launch sites in Japan) to establish the world's most advanced information and telecommunications network. It is expected that this information and telecommunications network's speed and capacity will be much higher than anything achieved previously.

KIZUNA satellite communication system aims for a maximum speed of 155Mbps (receiving) / 6Mbps (transmitting) for households with 45-centimetre aperture antennas (the same size as existing Communications Satellite antennas), and ultra-high speed 1.2Gbps communication for offices with five-meter antennas. In addition to establishing a domestic ultra high speed Internet network, the project also aims to construct ultra high speed international Internet access, especially with Asian Pacific countries and regions that are more closely related to Japan.

KIZUNA project is responsible for the demonstration of the validity and usefulness of technologies related to large-capacity data communications in our space infrastructure project, "i-Space,"[2] the purpose of which is to promote the use of satellites

in such fields as Internet communications, education, medicine, disaster measures and Intelligent Transport Systems.

One of the problems on KIZUNA is around 0.8 sec of the delay time of round trip delay which occurs in between satellite and ground (0.3 sec by the factor of two) and the delay time in Asynchronous Transfer Mode: ATM switch [3] onboard KIZUNA satellite (0.2 sec). Because KIZUNA provides IP communications based on TCP/IP protocol[4] so that shake-hand protocol including acknowledge confirmation is required results in effective network speed (data rate) is degraded due to the delay time.

In order to overcome such problem, hardware type of accelerator such as Sky-X[5], Blue coat[6] is developed which is a little bit expensive though. In this paper, software type of accelerator[7] is proposed with some evidences of experimental results which are obtained with KIZUNA satellite. Through experiments with KIZUNA satellite, performance of the proposed software TCP accelerator is confirmed [1]-[9]. Such accelerator is useful for not only satellite communication but also the other surface communications with TCP/IP protocol communications.

The following section describes the proposed software accelerator followed by some experimental results. Then concluding remarks with some discussions is followed by.

II. PROPOSED SOFTWARE ACCERELATOR

A. Problem Due to Delay

Under the TCP/IP protocol, sender transmits a shake-hand message first with header information including packet size, sender address, etc., and then receiver replies with ACK[8]: Acknowledgement. After that sender send packets as shown in Figure 1.

---

[1] http://www.jaxa.jp/countdown/f14/overview/kizuna_e.html
[2] http://ja.wikipedia.org/wiki/I-Space

[3] http://en.wikipedia.org/wiki/Asynchronous_Transfer_Mode
[4] http://www.comptechdoc.org/independent/networking/protocol/protnet.html
[5] http://www.1st-computer-networks.co.uk/packeteerSkyXAccelerator-250.php
[6] http://www.bluecoat.com/
[7] http://www.viasat.com/government-communications/information-assurance/compact-xpep
http://www.bitpipe.com/tlist/TCP/IP-Acceleration.html
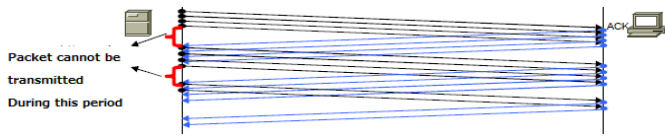[8] http://en.wikipedia.org/wiki/Transmission_Control_Protocol

Figure 1. Communication procedure of TCP/IP protocol

If there is some delay in the network links, then packet cannot be sent during the red portions of period in Figure 1. Thus effective data rate is somewhat degraded depending on window size and delay time. The typical round trip time (Delay time) is 5ms for Local Area Network: LAN, 10ms for Metropolitan Area Network: MAN, 100ms for Wide Area Network: WAN and 600ms for geostationary satellites as well as 1200ms for geostationary satellite double hop. Throughput for 5ms delay of LAN with 8KB of window size is 12.8Mbps while that with 32 KB window size is 51.2Mbps. Meanwhile, throughput for 10ms delay of MAN with 8KB window size is 64Mbps while that with 32KB window size is 25.6Mbps. On the other hands, throughput for 100ms delay of WAN with 8KB window size is 640Kbps while that with 32KB window size is 2.6Mbps. Furthermore, throughput for 600ms delay of geostationary satellite link with 8KB window size is 107Kbps while that with 32KB is 427Kbps and throughput for 1200ms delay of geostationary double hop link with 8KB window size is 53Kbps while that with 32KB window size is 213Kbps. In accordance with window size, throughput is increased. Meanwhile throughput is decreased with increasing of delay time.

B. Hardware Accelerator

TCP protocol has three major merits, error corrections, end-to-end flow control and collision avoidance control so that the most of IP communication adopts TCP protocol. Remaining such these merits, TCP is converted to XTP (Xpress Transport Protocol) which is appropriate protocol for high speed of long distance communications. Sky-X [9] adopts XTP [10] and TCP accelerations. XTP includes (1) Dynamic window sizing, (2) Rate control and flow control, (3) Selective retransmission mechanism, (4) Data compressions, and (5) HTTP acceleration [11]. Meanwhile, TCP acceleration is based on SCPS [12]: Space Communication Protocol Standard of MIL-STD-2045-4000 [13].

C. Software Accelerator

Packets on the network flow have to be handled by software accelerator. LAN driver software has to be modified to accelerate together with some modifications of LAN board. Free open source software of virtual LAN driver is used for the proposed software accelerator. It is known that all the functions of TCP protocol have not been realized with the virtual LAN driver. Therefore, the proposed software accelerator may choose functionalities TCP using the control panel shown in

[9] http://www.netone.co.jp/seminar/tfa9q100000031eb-att/06DSE_SkyX.pdf

[10] http://www.sophia-it.com/content/Xpress+Transport+Protocol

[11] http://en.wikipedia.org/wiki/Varnish_(software)

[12] http://www.scps.org/

[13] http://www.everyspec.com/MIL-STD/MIL-STD+(2000+-+2999)/MIL-STD-2045_47001C_25099/
http://www.gwg.nga.mil/ntb/baseline/docs/44500/index.html

Figure 2. Thus efficiency or effective data rate, throughput, etc. can be measured with the selected functions, in particular for status of the buffer memory.

The most important function of the proposed software accelerator is an acknowledge handling. Delay time can be shortened because network performance is degraded due to delay time. Although it is not possible that physical distance is shortened, the required time for acknowledge sending can be shortened. Actually, acknowledge is replied immediately after receiving TCP packet on the virtual LAN driver [14]. The actual data packet can be notified to the receiver with the different protocol. Thus acknowledge can be replied immediately after receiving TCP packet so that delay time can be shortened remarkable (ideally, delay time is going to be zero). Although this method may avoid the delay time induced problems, the actual data packets have to be sent to the receiver certainly.

Maximum buffer size for receiving has to be noticed. TCP protocol of automatic tuning of receiving buffer size based on RFC1323 [15] and Windows Vista may control buffer size for increasing effective data rate (maximum buffer size for sending is also to be notified for Windows Vista [16] of automatic tuning). The receiving buffer size does not include data packet which are not yet received by application layer. In case of IP communication through KIZUNA, delay time is 0.8 sec so that all the data packets which are not yet received by application layer can be processed with the current performance of personal computers. That is the same thing for the other communication satellite in the geostationary orbit (Delay time is 0.6sec). Therefore, available receiving buffer size cannot be sent properly. The proposed software accelerator rewrites the available receiving buffer size is replaced to the maximum receiving buffer size.
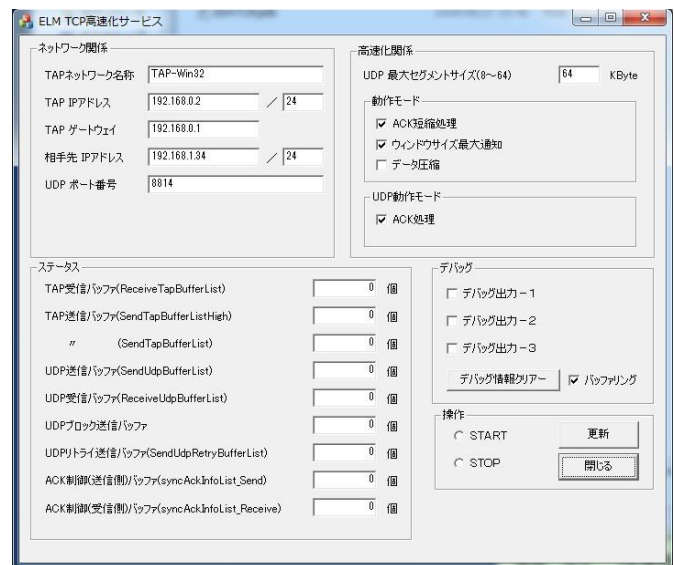

Figure 2. Control panel for selection of functionalities of TCP protocol.

[14] http://ja.broadcom.com/support/ethernet_nic/netxtreme_desktop.php

[15] http://www.ietf.org/rfc/rfc1323.txt

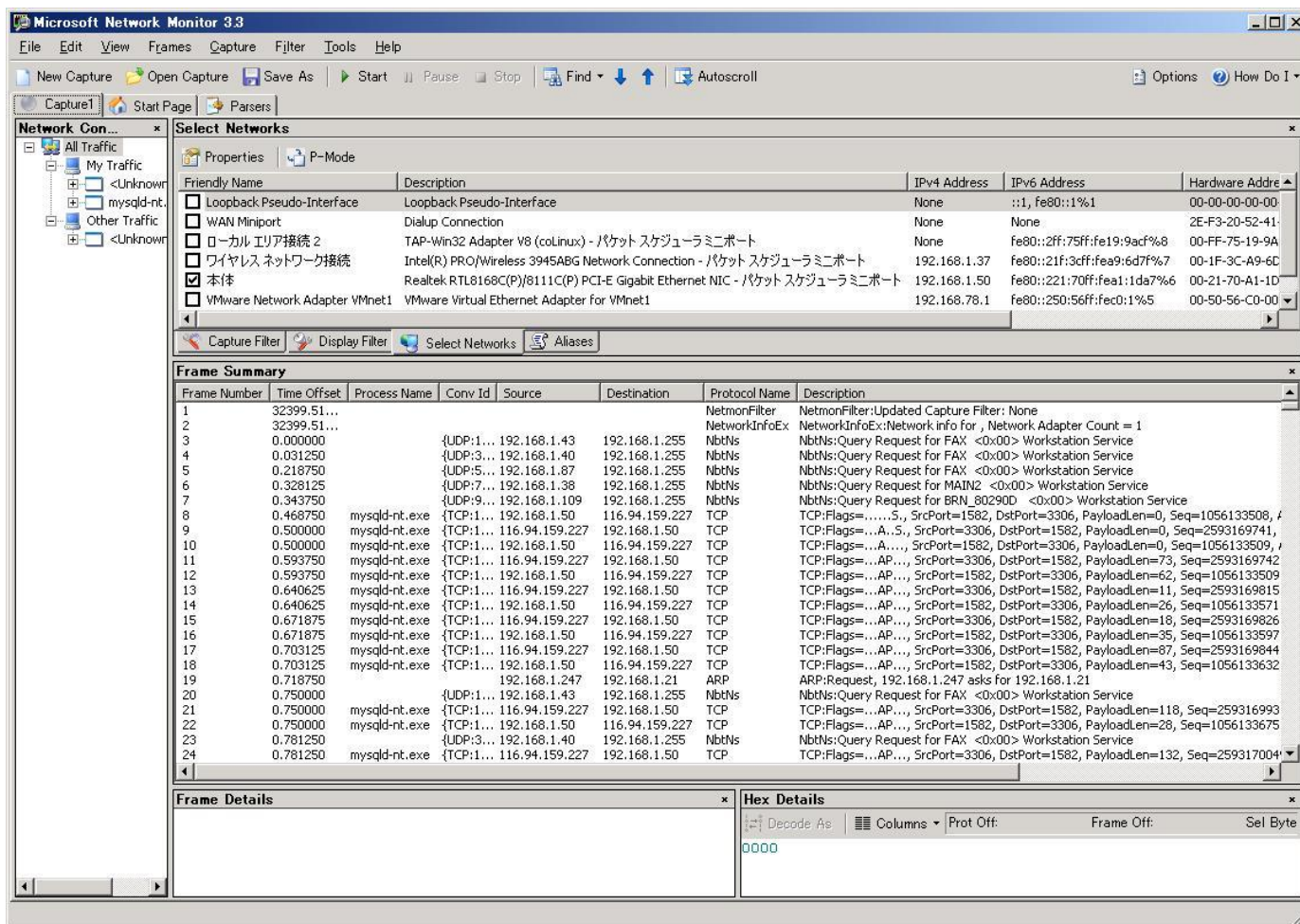[16] http://ja.wikipedia.org/wiki/Microsoft_Windows_Vista

Figure 3. Screen shot image of Microsoft Network Monitor 3.3

The proposed software accelerator takes into account data compression. Namely, data packets with TCP are converted to that with User Datagram Protocol: UPD protocol[17] and then transmit to the sender. Data compression can be applied to the data packets converted to UDP protocol so that the effective data rate can be improved.

III. EXPERIMENTS

A. Experiemnt Method

The effectiveness of the proposed software accelerator is evaluated with file transfer based on ftp protocol. Packet data flow and capacity is measured on the packet monitor with the well-known free open source software of Microsoft Network Monitor 3.3 [18]. Figure 3 shows the screen shot image of Microsoft Network Monitor 3.3. It is confirmed that all the data packets are transmitted perfectly. By using TCP segment flag of SYN and FIN[19], the start and the end time are measured. Therefore, the time required for transmission of data packets can be measured. The measured time duration is shorter than

that is measured on DOS window[20]. Therefore, it is taken into account such duration time differences when the effective data rate is evaluated.
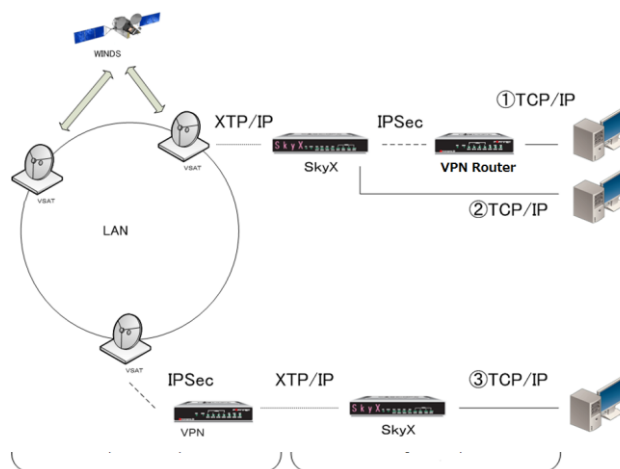
Figure 4. Network configuration

---

[17] http://ja.wikipedia.org/wiki/User_Datagram_Protocol

[18] http://www.microsoft.com/download/en/details.aspx?id=4865

[19] http://en.wikipedia.org/wiki/Transmission_Control_Protocol

[20] http://www.ne.jp/asahi/hishidama/home/tech/windows/command.html

Network configuration is shown in Figure 4 while experimental system configuration is shown in Figure 5, respectively. Two major terminals, Saga University and Kyushu University are connected through KIZUNA satellite. Internet connection is also confirmed as shown in Figure 6. In Figure 4, IPsec denotes Security Architecture for Internet Protocol[21], VPN denotes Virtual Private Network[22]
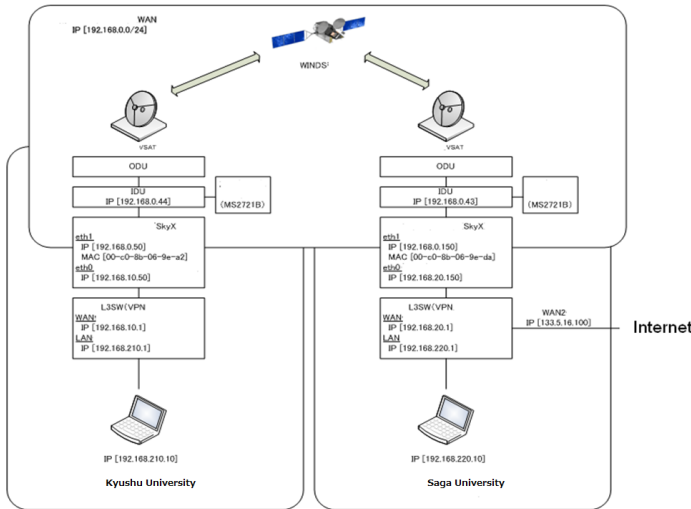

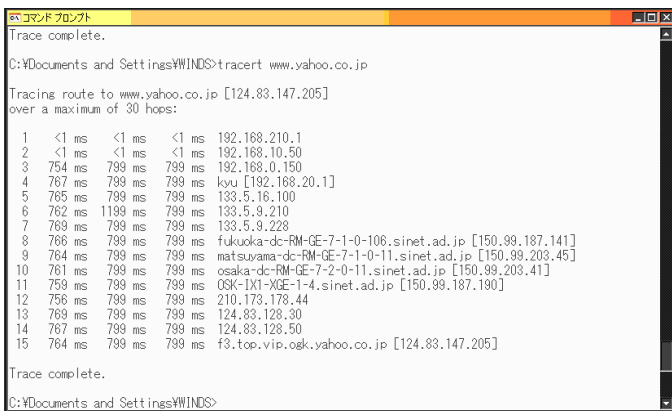Figure 5. Experimental system configurations

Figure 6. Internet connection confirmations

The experiments are conducted on 27 August 2008. Designated up-link data rate is 51Mbps between both universities. There are two datasets for the experiments, small size of (1079064Byte) and large size of (310677846Byte). These datasets are disaster relief satellite imagery data derived from Moderate resolution Imaging Spectrometer: MODIS[23] and Advanced Spaceborne Thermal Emission and Reflection radiometer: ASTER[24] (both mission instruments are onboard Terra satellite) as shown in Figure 7. ASTER and MODIS imagery data of Ureshino city[25], Saga, Japan (33°8′N 130°4′E) is acquired on February 19 2009.

---

[21] http://ja.wikipedia.org/wiki/IPsec

[22] http://ja.wikipedia.org/wiki/Virtual_Private_Network

[23] http://modis.gsfc.nasa.gov/

[24] http://asterweb.jpl.nasa.gov/

[25] http://en.wikipedia.org/wiki/Ureshino,_Saga
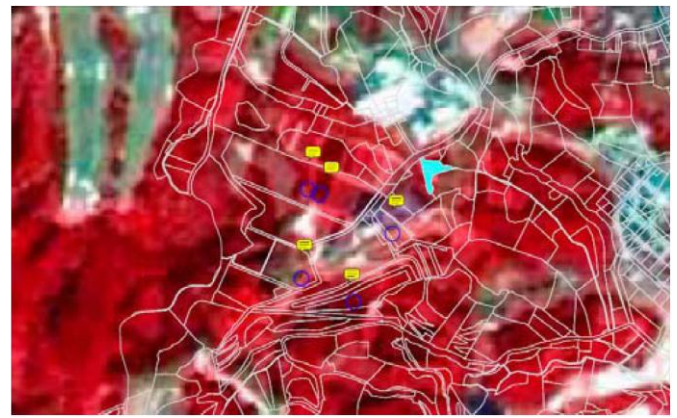
(a)ASTER imagery data of Ureshino: large size of (310677846Byte)
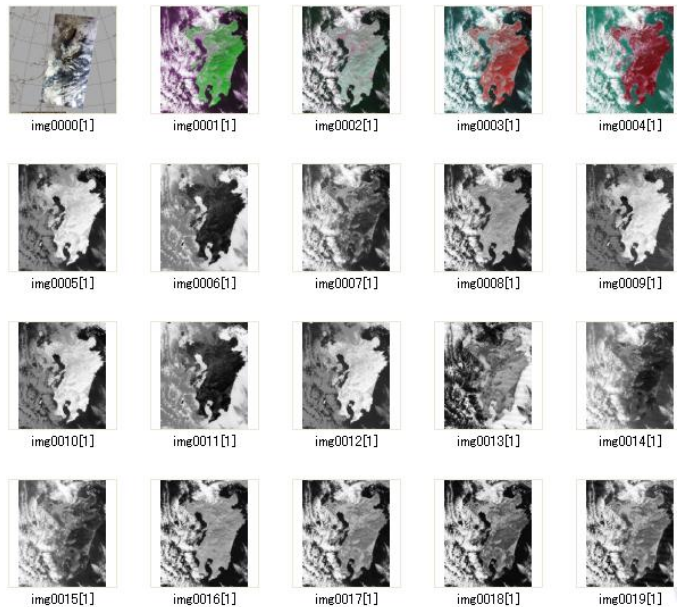
(b)MODIS imagery data of Kyushu: small size of (1079064Byte)

(c)Location of Ureshino and Kyushu in Japan

Figure 7. Disaster relief of ASTER and MODIS imagery data used for data transmission experiments through KIZUNA satellite

B. Experiemntal Results

There are three major features of the proposed software accelerator, 1) acknowledge handling, 2) Maximum receiving buffer size notification, and 3) data compression. The experiments are conducted with combinations of 1), 2), and 3).

(1) Combination of 1), 2), and 3) with small dataset:

1.426757sec

(2) Only 1) with small dataset: 1.422575sec

(3) Combination of 1) and 2) with large dataset: 64.688477sec

(4) Sky-X with small dataset: 1.08sec

(5) Sky-X with large dataset: 75.94sec

(6) Without Sky-X (large dataset): 146.63sec, 7.36MBps

(7) Without Sky-X (small dataset): 1.33sec, 812.55MBps

In comparison between case (1) and (2), contributions of 2) maximum receiving buffer size notification and 3) data compression are not so significant. Also it may say that 1) acknowledge handling is the most effective method for the proposed software accelerator. In comparison between case (2) and (4), the proposed software accelerator does work so as Sky-X does. Approximately, 24% much longer time is required for the proposed software accelerator comparing to the Sky-X based hardware accelerator. It, however, depends on the data volume as well as characteristic of the dataset used for experiments. In fact, 14.8% much shorter time would be enough to send the large size of dataset for the proposed software accelerator in comparison to the Sky-X based hardware accelerator. Therefore it may concluded that the proposed software accelerator with acknowledge handling has almost same performance as Sky-X based hardware acceleration does on acceleration of data transmission even when network link includes delay of 0.8sec.

Figure 8 shows network performance measurement data. Horizontal axis shows the different MODIS data. Figure 8 (a) shows the network performance of TCP data transmission from Saga University to Kyushu University for the case that Sky-X is off while Figure 8 (b) is that for the case of Sky-X is on. Meanwhile, Figure 8 (c) shows the network performance of UDP data transmission from Saga University to Kyushu University for the case that Sky-X is off while Figure 8 (d) is that for the case of Sky-X is on. Figure 8 (e) shows a comparison between UDP and TCP networks performances when the Sky-X is off.
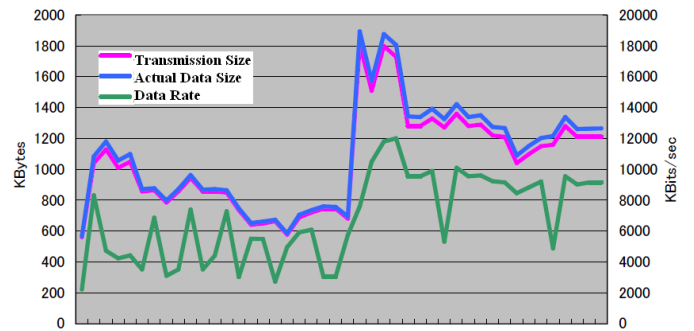
(a) TCP, Sky-X is off

(b) TCP, Sky-X is on
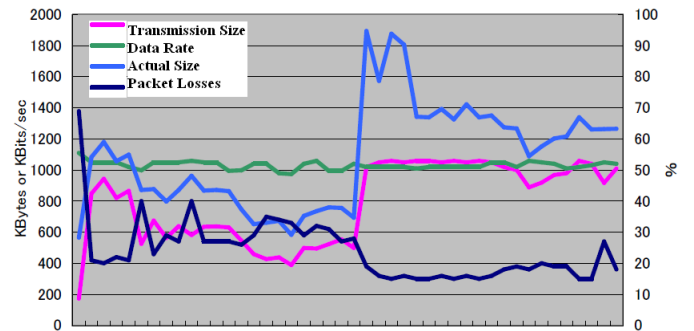
(c) UDP, Sky-X is off
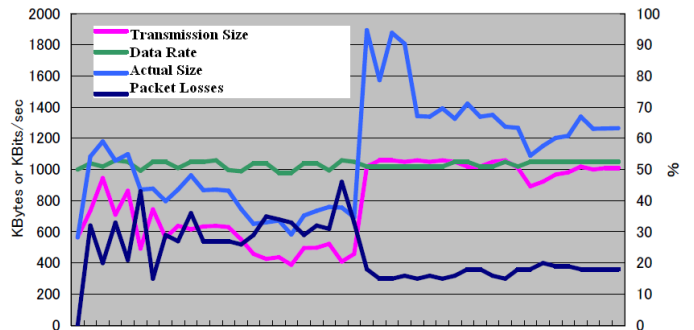
(d) UDP, Sky-X is on
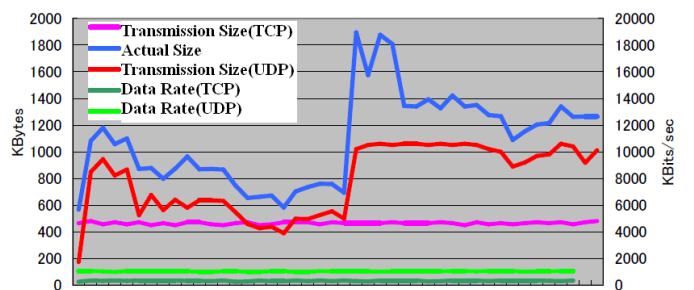
(e) Comparison between TCP and UDP when Sky-X is off
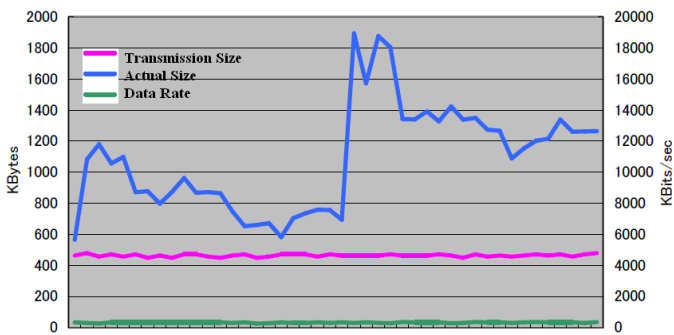
Figure 8.   Network performance measured data

As a result, it is found that data rate for Sky-X is on is around 100 times much larger than that for Sky-X is off.

IV.   CONCLUDING REMARKS AND SOME DISCUSSIONS

Through the experiments with University-to-University communication link through KIZUNA satellite, it is found that the proposed acknowledge handling based software accelerator is effective and is almost same as that of the widely used

hardware accelerator. The effectiveness is much better than that of maximum receiving buffer size notification as well as data compression (conversion from TCP to UDP and application of data compression method).

The proposed acknowledge handling does not ensure TCP based data transmission so that the data transmission has to be done with the other protocols. Actually, Sky-X accelerates data transmissions based on the conversion from TCP to XTP.

Maximum receiving buffer size notification does not work in the experiments. Furthermore, RFC1323 of window size adjustment does also not work so well. Turns out, automatic window size tuning based on Windows Vista does work because it adjusts not only receiving buffer size but also sending buffer size is adjusted dynamically. The effectiveness of the automatic tuning is less than that of the proposed acknowledge handling.

Data compression does not work so well because the packet size of ftp transmission is not so large (around 1500 byte). Although data compression rate depends on the data compression method used, it cannot be compressed in some cases with a consideration of header information treatments.

Also the time required for data conversion from TCP to UDP would be a bottle neck. This method might be useful for a narrow band networks.

The delay time induced problem is affecting to not only TCP based data transmission but also the other ground based surface network links with some delay and even for mobile phone communication links. Demands on the proposed software network accelerator are getting increased in the near future.

REFERENCES

[1] Kohei Arai, Kiyotaka Fujisaki, Hiroaki Ikemi, Masato Masuya, Terumasa Miyahara, Backup communication routing through Internet Satellite, WINDS, for transmission of disaster relief data, Proceedings of the International Symposium on WINDS Application Experiments, 2010.

[2] Kohei Arai, Back-up communication link with WINDS satellite in disaster and its application ot disaster mitigation information network system, Proceedings of the SCOPE Research Results Workshop, 2010.

[3] Kohei Arai, Research and developments of a back-up communication link with WINDS satellite in disaster and its application ot disaster mitigation information network system, Proceedings of the SCOPE Research Results Workshop in 2010, 2010.

[4] Kohei Arai, Kiyotaka Fujisaki, Hiroaki Ikemi, Masato Masuya, Terumasa Miyahara, Backup communication network through WINDS satellite for disaster mitigation, Journal of the Science and Engineering Faculty, Saga University, 38, 2, 7-12, 2009.

[5] Kohei Arai, Backup communication routing through Internet satellite WINDS for transmission of disaster relief data, Proceedings of the International Symposium on WINDS, 2010

[6] Kohei Arai and Kiyotaka Fujisaki, Countermeasure for round trip delay which occurs in between satellite and ground with network accelerator, Proceedings of the Expert Meeting and the 5th Plenary Workshop of the Association for Application Experiments of WINDS, 2010.

[7] Kohei Arai, Kiyotaka Fujisaki, Hiroaki Ikemi, Ibrahim Djawaluddin, Masato Ikemi, Terumasa Miyahara, Software accelerator for high speed TCP data transmission and GIS client-server model through WINDS satellite, Journal of the Science and Engineering Faculty, Saga University, 38, 2, 13-18, 2009.

[8] Kohei Arai, Back-up communication routing through Internet satellite WINDS for transmitting of disaster relief data, International Journal of Advanced Computer Science and Applications, 2, 9, 21-26, 2011.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a counselor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission "A" of ICSU/COSPAR since 2008. He wrote 30 books and published 322 journal papers.

Bridging content's quality between the participative web and the physical world

Anouar Abtoy[1], Noura Aknin[2], Boubker Sbihi[1]

Ahmed El Moussaoui[2] & Kamal Eddine El Kadiri[1].

[1] LIROSA, [2] LaSIT

Abdelmalek Essaadi University

Tétouan, Morroco

**Abstract— web 2.0 and its applications spread widely among internet users. Taking advantage of the easy, the friendly and the rich user interfaces. As consequence, the creation and the production of content become available to anyone. The ordinary user has step forward toward being a producer rather than remaining passive consumer. With this usage shifting, a new concern emerged: the quality of the User-Generated Content UGC or the User-Created Content UCC. Our team has developed a new concept of a Framework for managing the quality of the validated content in the participative web based on the evaluation of content's quality during its lifetime on the web. As continuation of our work, we will present a concept of extending the quality assessment of UGC or UCC into the real world by creating a bridge between digital content and its homologue in the physical world (e.g. printed version …). We will combine existing technologies such as QR codes or RFID in order to perform the linking between the digital and the physical content. This approach well offers the possibility of following the quality of UGC or UCC; and eventually evaluates it; even in hard copies. The evaluation and the assessment of physical content originated from digital content generated through web 2.0 applications will be done in real-time. The proposed approach is implemented to our framework by integrating the features into the UML diagrams i*n the Blog case.*

Keywords-Collaborative web; content management; digital content; physical content.

I. INTRODUCTION

Since its introduction Tim O'Reilly in 2004, web 2.0 has become an essential source of information for every single need in our era; this is mainly due the revolution of the concept of the web and its applications [1] [2]. The internet user has passed from being a simple consumer of content on the net to an active producer.

According to Tim O'Reilly [1], Web 2.0 brings new features to the Web thanks to the seven concepts which are:

- The Web is a services platform.
- The power of collective intelligence should be exploited efficiently.
- Data is the next Intel inside.
- End of the software release cycle.
- Lightweight programming models
- Software above the level of a single device.
- Rich user experiences.

The technological and the usage revolution of web 2.0 enabled various applications that took the web to a whole new level of surfing experience (Fig.1).

Figure 1. Web 2.0 applications

This change in the philosophy of content production and publication increases exponentially the mass of information available on the web. As a direct impact, new issues rise on the net. The lack of the quality and the relevance of both information and content are emerging as a major need for research in the area of the participative web [3].

At first we will introduce the new concept of content validation, and the validated Content Management Framework. Then, we will present our new vision content management for both digital and physical content in real-time.

II. VALIDATED CONTENT MANAGEMENT FRAMEWORK

The framework proposed by our team [4] is based on two major components: the categorization of information into types of quality, and the categorization of users into groups [3]:

- Categorization of information: unlike the earlier web, the information is subdivided into categories that represent a certain degree of information quality.
- Categorization of users: users are grouped into categories: users (eventually a producer), a validator and an expert. Each one of them has certain roles, responsibilities and privileges.

To ensure a certain quality of content, it is submitted to two validation processes: static and dynamic.

- Static validation: When a user produces content, it is submitted to an expert who assigns two validators whom

will evaluate its relevance. Depending on their decision, the content will be published with an initial quality rating given by the combination of their note or it will be rejected (Fig.2).
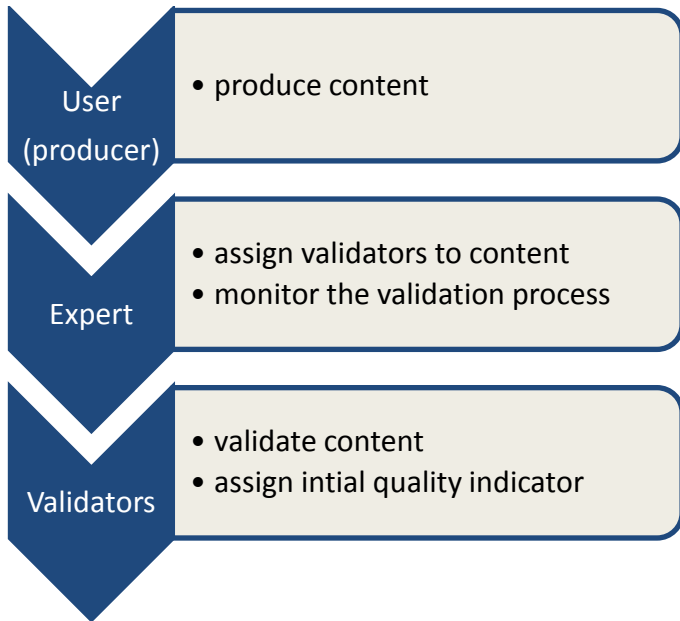

Figure 2. Static validation process [4]

- Dynamic validation: when content is validated and published through of a static validation on the net, the Internet community takes care of the ongoing validation of this content during its life on the web. The degradation of the quality of content till certain threshold causes the elimination or archiving of this content (Fig.3).
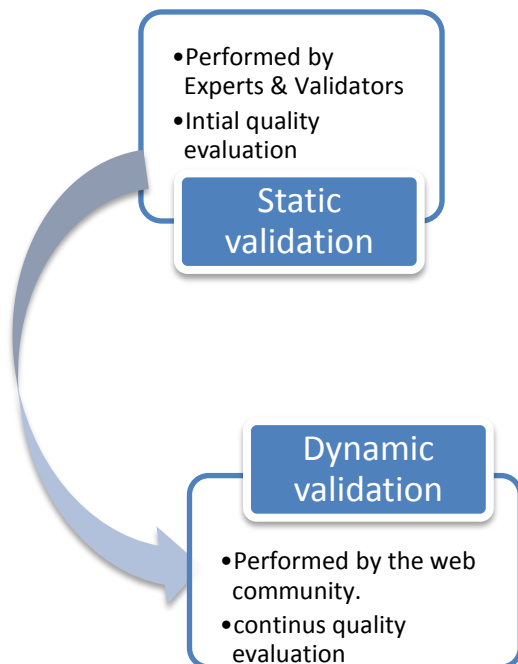

Figure 3. Static and dynamic validation complementary processes [4].

III. BRIDIGING CONTENT QUALITY BETWEEN DIGITAL AND PYSICAL WORLD

With easiness of accessibility of web 2.0 applications such as blogs, wikis and social networks, everyone can produce or generate a content which are called *User-Generated Content UGC* or *User-Created Content UCC.* Based on our Framework, the content's quality can be monitored and evaluated form the moment of its creation to its elimination or archiving. This approach is valid as long as content remains in its digital form.

But usually, users tend to print digital content in order to keep printed version for offline reading. For multimedia content, users also tend to download it and put it on portable media player or burn it on CD, DVD, B-RD ….

The form shifting of content from digital to physical looses the connection between content and its quality indicator evaluated with our Framework.

The problem of linking the same content with different form was already discussed by researchers in other topics. An example this problem is the case of digital rights management for content distribution [5]. The difficulty was how to track and preserve digital rights of contents with continues changing form of those goods or assets (CD, DVD,Blu-ray, flash memory…) during distribution.

Our research is focused on the quality of User-Generated Content UGC or the User-Created Content UCC produced by ordinary users in the context of web 2.0 applications. The Framework that we presented [4] can give a way to evaluate UGC and UCC over the web. But once those content shift from digital to physical media, we need a new vision to assess and link the quality of these content with their original digital content.

We propose to add identification item to content in its physical form with certain tag that can be used to reach the original digital content on the net and its quality details. Our study focuses on two technologies: RFID and QR codes.

- RFID: *Radio Frequency IDentification*, it's a technology that uses tags in order to identify an object from certain distance unlike bar codes. It is used more and more on various application such as handling manufactured goods and materials [6]. This promising technology is tending to be combined with web services to offer presence-aware infrastructure for diverse application scenarios such as committee meeting scenario [7].

The RFID technology has been used in various researches to track digital content. It was proposed in the context of a Framework to mange digital rights and tracks it in its physical form such as digital mediums (Fig.4): CD, DVD or memory card [8]
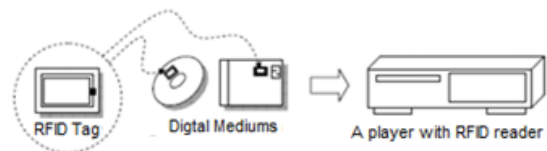

Figure 4. Tracking digital rights based on digital mediums and RFID Tags[8]

The RFID Tags can contain various information related to content, including URL that points to the content on the web.

- QR code: *Quick Response code* is 2 dimensional code or matrix barcode that was initially designed for automotive industry. QR codes offer more potential than the 1D barcodes.

Figure 5.   QR Code (2D) and EAN-13 (1D) barcodes encoding the same data (3254565174110)[6].

Some researches already used QR codes to link content with a web service by scanning the public QR code to obtain a URL in order to reach other information (Private part). The retrieved data are combined with the scanned to adapt contextual messages [9].

Linking print and digital content with QR codes is not a new concept [10], especially that any Smartphone become a QR code scanner. In our approach, we propose to link physical content (print or digital medium) via QR code with its correspondent digital content in the web. With this process, the user can know the current quality of the physical content that he possesses and eventually he can give his feedback and evaluate it.

Conceptually, the two proposed technologies are quite similar; both are intended to provide rapid and reliable item identification-and-tracking capabilities. The key difference between the two technologies is that QR code scans a printed label with imaging technology.

On the other hand, RFID scans or interrogates a tag using radio frequency signals. The table below summarizes the advantages and the inconveniences of both QR code and RFID:

	QR Code	RFID
Advantages	• Simple • Economic • Reader available in a simple phone	• Tag can be hidden • Read/write • Information update • Lifecycle
Inconveniences	• Read only • Labels can be erased by time.	• sensitivity to the environment • cost • reader implementation

Table 1.   The link between the quality of digital and physical content

Our approach relies on generating and integrating tags or labels (RFID or QR code) with physical version of the UGC or UCC. The evaluation of content's quality based on our approach it's a sustainable process. So when the content is printed or burned at certain moment t_n, the user can't know the state of this content's quality at t_{n+1} based on its physical version.

Nowadays basic Smartphone offers the functionality to scan QR code labels, others can even be extended with RFID reader. Via the Tag in the content, the Smartphone can read it and then point to the unique digital version of the initial content on the web. Then, user can know the actual content's quality at t_{n+1} and participate in its evaluation process.
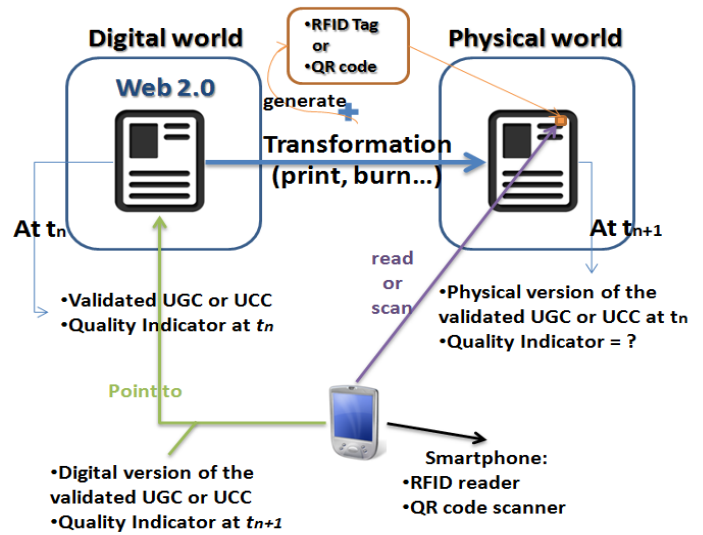
Figure 6.   The link between the quality of digital and physical content [11]

IV. APPROACH IMPLANTATION: BLOG EXTENSTION CASE STUDY

The framework of the validated content proposed by our team was implemented in a blog system to publish and mange validated blog posts [4, 11]. The presented approach in this paper will bring new functionalities by extending the framework; therefore the modeling done in the previous work needs to be adjusted in order to reach the new requirements.
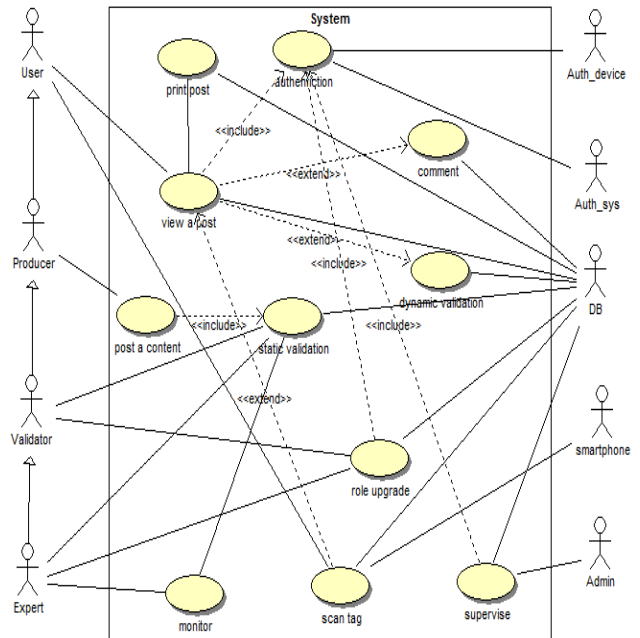
Figure 7.   Use case diagram

The new use cases diagram includes new actor that was identified by analyzing the new requirements of the system. We modeled in this situation the Smartphone because in one hand, it becomes a popular device among users. On the other hand, the mobile technologies have known a breakthrough with technological development made on both levels: software (android, windows 7) and hardware (mobile processors architecture, high imaging capabilities)

The uses case diagram previous use case diagram contains several use cases. Based on the new vision of validating both digital and physical content, the use case diagram (see figure below) will contain two new use cases: "scan tag" and "print post" use cases.
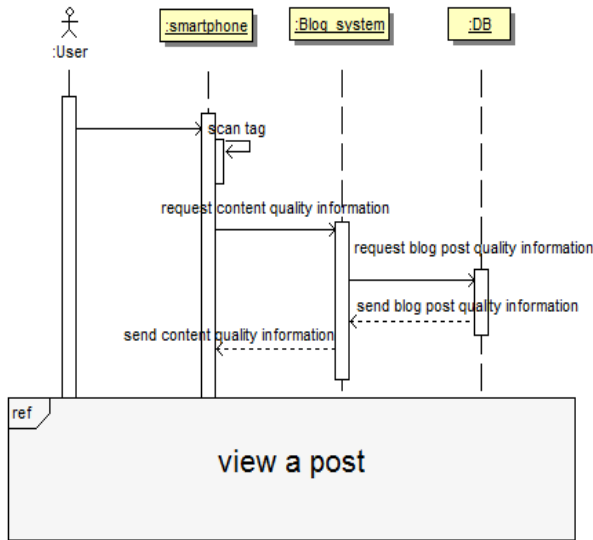
Figure 8. Scan tag sequence diagram

Each use case represent the new features integrated with the framework so that the system can deliver the quality information with physical version of the blog post, at the same time users can know the state of information quality including its validation state form the physical version.

In "scan tag" sequence diagram, the user uses his mobile device or Smartphone equipped by digital scanner for QR code to scan the tag of the physical version of blog post. Then, the Smartphone requests the blog post quality information from the system. As result, the system will send back this information to the user through his device. After accomplishing this phase, users have always the right to view the actual blog post content with its current quality information. Eventually, after authenticating, users can participate in the validation by dynamically validating the blog post.

Another possibility added to the system, is the preparation of the version that will be used to print the blog post or save it in a digital medium. The portable version is considered a pre-physical version for the digital blog post. In our case, the suitable version is PDF (Portable Document Format) files.
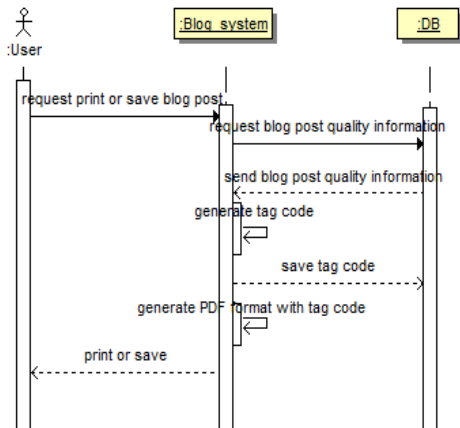
Figure 9. Print post sequence diagram

In the "print post" sequence diagram described above, the user requests from the blog system to save or print the current blog post. The Blog system retrieves the needed information from the database. The next step includes the generation of the tag by the system. In the case of QR codes technology, the system generates the necessary tag that points to the blog post and its quality information state. After saving the generated tag in the data base, the system precedes the generation of the blog post in its pre-physical version that includes the tag as a header. The finale phase is sending the final documents to the user so that he can print it or save in a digital medium for future purpose.
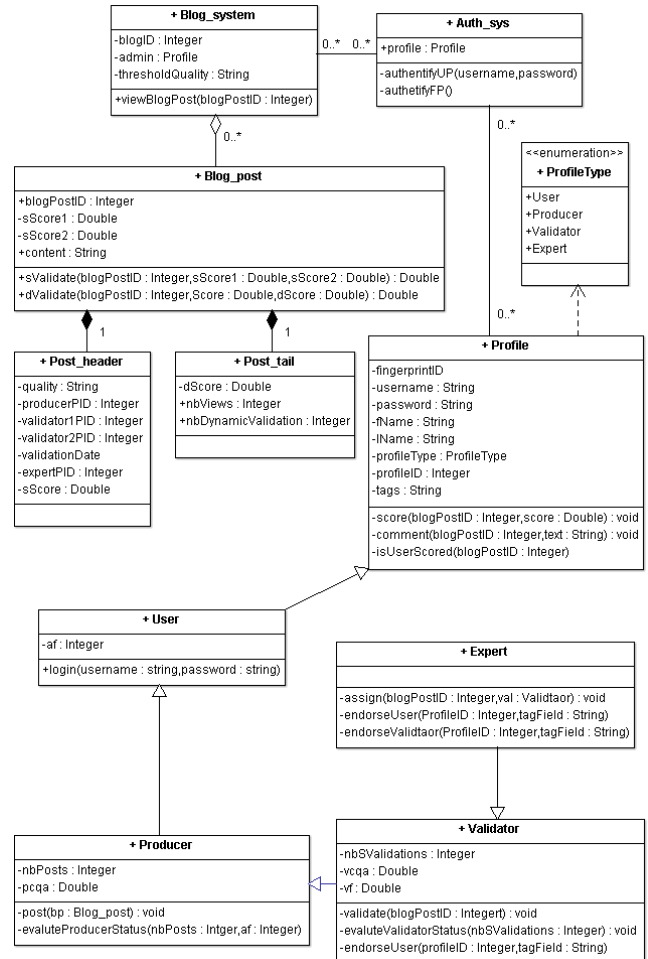
Figure 10. Class diagram of the enhenced blog system

The enhancements done in the system don't have implications on the class diagram molding (see figure 10). Mainly the generated QR code will point to the blog post based on its *blogID* attribute; therefore all the needed information will be retrieved based on this attribute form the system.

The enhanced blog system offers the possibility to create not only a validated blog posts with known quality, but also to create validated versions of the same content published within the system in their physical versions.

After integrating the approach with the validated content framework, it can be applied to other web 2.0 applications. The priority must be given to wiki systems because of their crucial role in the actual knowledge of users on the web.

The results obtained from our case study proved that bridging the concept of content's quality assessment between

two different worlds (digital and physical) is possible. Also, the results in the case of Blogs encourage us to extend this approach in the case of other web 2.0 applications (Wikis, Social networks…)

V. CONCLUSION

The web 2.0 with its collaborative and participative feature encourages users to create and generate content easily. As a result, the web becomes a giant container for UGC and UCC. Our Framework for the management of the validated content offers the ability of to evaluate this information and content's quality over the web. But when it comes to the physical version of this content, it remains unable to assess and evaluates its quality. In this paper, we presented a new concept to extend our Framework to enclose even the physical content originated form participative web. We give an overview for the two possible technologies candidates for this process that are RFID and QR code. The implementation of the new approach into our framework is given in the case of blogs through UML diagrams.

The aim of this approach is bridging the physical and the digital world through simple technologies available to anyone. The potential is very promising and may lead to create a complete validated web that contains the UGC and UCC in their various formats and versions.

As perspective of our work, we propose to develop an approach of a system capable of tracking and classifying physical version of UGC/UCC. The purpose of this approach is to integrate the content into the new vision of internet of things.

REFERENCE

[1] T. O'Reilly, "what Is Web 2.0", 2005, available at: http://oreilly.com/web2/archive/what-is-web-20.html.

[2] J.F Gervais, Web 2.0: les internautes au pouvoir, Edition, Dunod, 2007.

[3] B. Sbihi, and K.E. El Kadiri, "Web 2.2: Toward classified information on the Web", International Journal of Web Applications, Vol. 1, No. 2, Juin 2009, pp. 102- 109.

[4] A. Abtoy, N. Aknin, B. Sbihi, A. El Moussaoui and K.E. El Kadiri, "Towards a Framework for a validated content management on the collaborative Web: Blogs case", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, pp. 96-104.

[5] Q. Lui, R. Safavi-Naini, N.P. Sheppard, "Digital rights management for content distribution ",Proceeding ACSW Frontiers '03 Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003 - Volume 21, 2003.

[6] R. Want, "An introduction to RFID technology", IEEE Pervasive Computing, Volume 5 Issue 1 , Jan.-March 2006.

[7] C. Kerer, S. Dustdar, M. Jazayeri, D. Gomes, A. Szego, J. A. Burgos Caja, "Presence-Aware Infrastructure using Web services and RFID technologies", Proceedings of the 2nd European Workshop on Object Orientation and Web Services technologies, 2004 – Oslo, Norway.

[8] Suh, J.H. and S.C. Park, 2006. U-DRM: A Unified Framework of digital rights management based on RFID and Application of its usage data. International Journal of Computer and Network Security, 6 (8B): 151-155.

[9] J. Rouillard, "Contextual QR Codes", Proceedings of the Third nternational Multi-Conference on Computing in the Global Information Technology. ICCGI 2008. July 27 - August 1, 2008 - Athens, Greece.

[10] D. A. MacRae, "Introducing QR Codes: Linking Print and Digital Content Via Smartphone", Neurosurgery, Volume 68 , Issue 4 , pp 854-855, April 2011.

[11] A. Abtoy, Z. Itahriouan, N. Aknin, B. Sbihi, A. El Moussaoui and K.E. El Kadiri. "Towards quality assessment for both digital and physical Content." Proceedings of the third Mediterranean Symposium on Telecommunications, Fez – Morocco ,March 22,23 & 24, 2012.

AUTHORS PROFILE

Anouar ABTOY received the Master degree in electronics and telecommunications in 2008 from Abdelmalek Essaadi University in Tetuan, Morocco. Currently, he is a PhD Student in Computer Science. He is also member of the Internet Society (ISOC) and student member of IEEE Computer Society. Ongoing research interests: Web 2.0, collaborative and collective intelligence, online identity, evaluation and assessment of information and content.

Noura AKNIN received the the PhD degree in Electrical Engineering in 1998 from Abdelmalek Essaadi University in Tetuan, Morocco. She is a Professor of Telecommunications and Computer Engineering in Abdelmalek Essaadi University since 2000. She is the Co-founder of the IEEE Morocco Section since November 2004 and she is the Women in Engineering Coordinator. She has been a member of the Organizing and the Scientific Committees of several symposia and conferences dealing with RF, Mobile Networks, Social Web and information technologies.

Boubker SBIHI received the PhD degree in computer science. Professor of computer science at the School of Information Science in Morocco. He is also the head of Information management department. He has published many articles on E-learning and Web 2.0. He is part of many boards of international journals and international conferences.

Ahmed EL MOUSSAOUI received the PhD degree in electronics at the University of BRADFORD in 1990. In 2007 he received the international master in E-learning in the Curt Bosh institute in Switzerland. He has been a member of the Organizing and the Scientific Committees of several symposia and conferences dealing with RF, Mobile Networks and information technologies. He has participated in several projects with France and Spain. Currently, his is the vice-president of Abdelmalek Essaadi University in Tetuan - Morocco.

Kamal Eddine EL KADIRI received the "Thèse de troisième cycles" degree in Data analysis at Paris VI University in 1984. He received "Thèse d'état" degree in Computer Science from Granada in 1994. . In 2007 he received the international master in E-learning in the Curt Bosh institute in Switzerland. He is a professor of Mathematics and Computer Science at Faculty of Sciences of Tetuan in Morocco. Currently, he is the director of the National School of Applied Sciences (ENSA) of Tetuan and also the director of LIROSA laboratory. He has published several articles on E-learning and Web 2.0. He is also part of many boards of international journals and international conferences.

Energy Efficient Clustering and Cluster Head Rotation Scheme for Wireless Sensor Networks

Ashok Kumar, Vinod Kumar

Department of E&CE
National Institute of Technology
Hamirpur (HP) - INDIA

Narottam Chand

Department of CS&E
National Institute of Technology
Hamirpur (HP) - INDIA

Abstract— **Wireless sensor nodes are highly energy constrained devices. They have limited battery life due to various constraints of sensor nodes such as size and cost, etc. Moreover, most of the Wireless Sensor Network (WSN) applications render it impossible to charge or replace the battery of sensor nodes. Therefore, optimal use of node energy is a major challenge in wireless sensor networks. Clustering of sensor nodes is an effective method to use the node energy optimally and prolong the lifetime of energy constrained wireless sensor network. In this paper, we propose a location based protocol for WSN- supporting an energy efficient clustering, cluster head selection/rotation and data routing method to prolong the lifetime of sensor network. Proposed clustering protocol ensures balanced size cluster formation within the sensing field with least number of transmit-receive operations. Cluster head rotation protocol ensures balanced dissipation of node energy despite the non-uniform energy requirements of cluster head and sensor nodes in a cluster. The cluster head rotation protocol has been proposed to achieve the balanced energy consumption among the nodes within the cluster thus prolonging the lifetime of network Simulation results demonstrate that proposed protocol prolongs network lifetime due to the use of efficient clustering, cluster head selection/rotation and data routing.**

Keywords- Corona; clusters;cluster head; sink; network lifetime.

I. INTRODUCTION

Recent advancements in large scale integration and wireless communication technologies have enabled the development of small size, low cost, and multi-functional devices known as sensor nodes. Sensor nodes are capable of sensing the desired environmental parameters within their vicinity, such as temperature, pressure, moisture, and pollutants, etc., converting the sensed variable to electrical signal and transmitting the sensed data to the desired destination. To achieve these objectives, sensor nodes are equipped with sensors, microcontrollers/microprocessors, and wireless trans-receivers. When these sensor nodes are deployed in large number to monitor an area, they form self-organizing cooperative wireless ad-hoc network, known as wireless sensor network. The wireless sensor networks are mostly deployed in remote and hazardous locations, where manual monitoring is very difficult or almost impossible. Due to deployment of wireless sensors in unattended harsh environment, it is not possible to charge or replace their batteries. Therefore, energy efficient operation of wireless sensors to prolong the lifetime of overall wireless sensor network is of utmost importance [1, 2]. Due to their low

power radio, wireless sensor nodes cannot transmit the data to large distance in single hop, which makes multi-hop communication essential in case of real life deployment. However, in multi-hop cases, if the energy consumption of sensor nodes is not managed properly it may create energy-hole problem in the network [3]. In literature, a number of protocols have been proposed to manage and reduce the energy consumption of sensor nodes [1-8]. Grouping sensor nodes into clusters has been widely used to achieve this objective. In clustered networks, one of the sensor nodes is elected as cluster head for each cluster. Sensor nodes in each cluster transmit data to their respective cluster head and the cluster head in turn forwards the data after aggregation/fusion to sink node through single/multi-hop transmission.

LEACH is one of the most popular distributed single-hop clustering protocols [5]. In this protocol, the clusters are formed, based on received signal strength. The role of cluster head is periodically rotated amongst the sensor nodes present in the cluster to ensure balanced energy consumption of sensor nodes. This algorithm becomes very inefficient in case of large area sensor networks due to single hop communication of cluster heads to the sink. A number of improvements have been proposed in literature to overcome the shortcomings of LEACH [6, 15-17]. Some of them are LEACH-C [6], PEGASIS [16], TEEN [15], HEED [17], etc. Hausdorff [10] and ERP-SCDS (Energy Efficient Routing Protocol for wireless networks with Static Clustering and Dynamic Structure) [11] are recent clustering algorithms. In this paper, we propose an energy efficient protocol consisting of clustering, cluster head selection/rotation and data routing method to prolong the lifetime of sensor network. In proposed protocol, clusters are formed only once during the lifetime of sensor network, which results in substantial saving of energy.

The simulation experiments demonstrate that the proposed protocol substantially enhances the network lifetime of wireless sensor network. The rest of the paper is organized as follows: Section II describes the related work. Section III describes the system model, which includes network model and energy dissipation model for wireless sensor networks. Section IV describes proposed protocol, including formation of clusters, cluster head selection. Section V describes, energy cost calculation of cluster for its normal operation, and energy cost calculation for optimal cluster head selection/rotation process. Results of simulations for various performance metrics are

given in Section VI. Concluding remarks have been incorporated in Section VII.

II. RELATED WORK

A number of clustering algorithms have been proposed for mobile ad hoc and wireless sensor networks [5-18]. Linked cluster algorithm (LCA) is one of the earliest works aimed towards prolonging network lifetime in mobile ad hoc networks [18, 19]. LCA basically was focused towards forming an efficient network topology to handle the mobility of nodes. It was therefore geared for maximizing network connectivity. The LCA resulted in large number of clusters. Therefore, the algorithm was refined in [20]. Lin and Gerla proposed the efficient support of multimedia applications in multi-hop mobile networks using CDMA. Random competition based clustering (RCC), designed for mobile ad hoc networks can also be used for wireless sensor networks [21]. It mainly focuses at cluster stability in order to support mobile nodes. RCC is based on first declare first win rule. In this approach the first node declaring it as cluster head, forms a cluster within its radio range and rest of nodes in its radio coverage give up their right to become the cluster head and joins its cluster as members. CLUBS is an algorithm that forms cluster through local broadcast and coverage in a time proportional to the local density of nodes [22].

A multi-tier hierarchical control clustering algorithm was proposed by Banerjee and Khullar [23]. In this process, any node can initiate the process of cluster formation and clustering process is carried out in hierarchical order. Low Energy Adaptive Clustering Hierarchy (LEACH) is one of the most popular clustering algorithms for wireless sensor networks [5]. In LEACH clusters are formed based on the received signal strength. The sensor nodes make autonomous decision about cluster formation without any centralized control and cluster head nodes are used to route the data to the base station. The algorithm ensures the balanced energy usage of nodes by random rotation of cluster head role amongst the nodes inside cluster. LEACH forms one-hop intra and inter cluster topology, where each node transmits directly to the cluster head and after aggregation, the cluster head transmits the data to base station in single-hop. This is a serious drawback of LEACH, which makes it less efficient in real life for network deployed in large area. It is also less efficient if the cluster heads are far away from the sink. A number of algorithms have been proposed to overcome the shortcomings of LEACH, such as LEACH-C [6], TEEN [15], APTEEN [24], and PEGASIS [16], etc. Fast Local Clustering Service (FLOC) was proposed as a distributed technique that produces approximately balanced clusters with minimum overlap [25]. In Energy Efficient Clustering Scheme (EECS) the cluster head candidates compete amongst themselves based on the residual energy, for a given round and cluster head candidate with maximum residual energy get selected as cluster head [8]. EECS extends the LEACH algorithm by dynamic sizing of clusters based on the distance of cluster from the sink node/base station. This improves the balanced energy distribution throughout the network resulting in the extended lifetime of network. Hybrid Energy Distributed Clustering (HEED) is multi-hop clustering algorithm, where cluster heads are elected based on the residual energy and intra-cluster communication cost [17]. Intra-cluster communication cost is proportional to the node's proximity to its neighbor and is used by the node in deciding to join the network. It provide uniform cluster head distribution throughout the network and better load balancing. Energy-efficient Unequal Clustering (EEUC) addresses the hop-spot problem that exists due to the fact that cluster heads close to the base station die faster because they need to relay more data traffic [9]. EEUC solves this problem by balancing the energy consumption for forming unequal clusters, where cluster near the sink node are smaller in size than the clusters away from the sink node in order to save intra-cluster energy. Power Efficient and Adaptive Clustering Hierarchy (PEACH) cluster formation is based on overhearing characteristics of wireless communication to support adaptive multi-level clustering [26].

Hausdorff clustering algorithm is a static clustering method, in which the Hausdorff distance between two node set is used as clustering metric [10]. It carries out cluster formation based on minimum Hausdorff distance between nodes of same cluster and between the nodes of neighboring clusters, to ensure connectivity within the network. ERP-SCDS (Energy efficient routing protocol for wireless networks with static clustering and dynamic structure) utilizes virtual points in a corona based wireless sensor network, forming static clusters and dynamic structure [11]. This is a location based clustering protocol, where clustering process is carried out by virtual points around the sink. The location of virtual points is calculated by sink after receiving the location information from all sensor nodes in the network. The protocol involves the transmission of location information by all sensor nodes to the sink for initiation of clustering process. It increases the energy cost of clustering; specially, in case of large area sensor networks due to multi hop communication between nodes and sink.

III. SYSTEM MODEL

Wireless sensor network model based on circular monitoring area A of radius Z with uniform node distribution density ρ having only one sink node at the center of monitoring area has been considered. There are T sensor nodes deployed in the sensing area, designated as $N_1, N_2, \ldots N_T$. Such model has been widely used in literature [6].

It is assumed that the transmitter electronics of all sensor nodes is capable of multi range transmission of the data with two types of radio ranges for intra cluster and inter cluster communication: low power broadcast range $R_1=R/2$ meters and high power range of $R_2=R$ meters, where R is the maximum transmission range of a node. Low power broadcast range R_1 is used for intra cluster communication between a sensor node and cluster head. The maximum diameter of any cluster is assumed to be R/2 meters, so that even if a node is located at boundary of a cluster, it may be selected as cluster head, and data connectivity is still maintained between the sensor nodes and cluster head within the cluster.

High power broadcast rang R_2 is used by cluster head node to transmit data either to next hop cluster head or to the sink node, in case the sink node is at a distance of one hop from the cluster head. The proposed model ensures that two neighboring cluster head are always at maximum distance of R from each other, so that connectivity is always maintained between such nodes. The sensor nodes use radio range $R_3=R/4$ during the

clustering process in order to form clusters of balanced size. As the data from various sensor nodes arrive at the cluster head, it performs data aggregation or data fusion over the arrived data as per the nature of sensed data to reduce the transmission overhead. The cluster head transmits the aggregated/fused data to its next hop cluster head towards sink node. To avoid the collision of intra and inter cluster data; multiple access technique of TDMA and direct sequence spread spectrum (DS-SS) respectively is assumed [10]. All sensing nodes deployed in the sensing area are assumed to be static and have the knowledge of their location. The sensing field is divided into m numbers of concentric circles of equal width, known as corona [14]. In proposed model, the width of each corona is assumed to be R/2 meters, as illustrated in the Fig. 1.
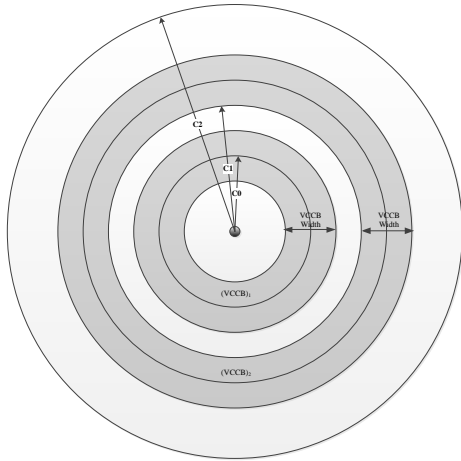

Figure 1. Corona based wireless sensor network and Assignment of concentric circles band (VCCB) to the sensor nodes.

A. Energy Model

Generalized energy consumption model based on first order radio energy consumption is used for calculation of energy consumption for sensor nodes within the sensing area [5, 6, 12, 13].The energy consumption of a sensor node for transmitting k bits of data over a distance d can be expressed as [6, 13]:

$$E_{Tx}(k,d) = E_{elect-Tx}(k) + E_{amp-Tx}(k,d) \qquad (1)$$

$$E_{Tx}(k,d) = \begin{cases} kE_{elect} + k\grave{o}_{fs}d^2, & d < d_0 \\ kE_{elect} + k\grave{o}_{mp}d^4, & d \geq d_0 \end{cases} \qquad (2)$$

$E_{elect-Tx}$ is transmission electronics energy; which is energy consumed by the sensor node for modulation, coding, spreading schemes, filtering operations, etc. E_{amp-Tx} (k,d) is the power amplifier stage energy consumption of sensor node to transmit k bits of data over a distance of d meter with acceptable signal to noise ratio (SNR). E_{elect} (nJ/bit) is energy dissipation per bit to run transmitter and receiver electronic circuitry. ε_{fs} (pJ/(bit-m^{-2})) is energy coefficient of power amplifier stage of sensor node for free space energy dissipation model, when transmission distance is less than threshold i.e. d < d_0. ε_{mp} (pJ/(bit-m^{-4})) is energy coefficient of power amplifier stage of sensor node for multipath energy dissipation model,

when transmission distance is greater than threshold i.e. d ≥ d_0. The energy consumption of sensor node to receiver k bits of data is given by:

$$E_{Rx} = kE_{elect} \qquad (3)$$

IV. PROPOSED CLUSTERING PROTOCOL

The proposed protocol, Virtual Concentric Circle Band Based Clustering (VCCBC) is divided into several phases. After the deployment of sensor nodes in sensing area, the clustering and cluster head selection phase starts with transmission of beacon by sink node. The beacon contains information regarding the location of sink node (x_s, y_s) within the network. After receiving the sink node location, all the sensor nodes calculate their respective Euclidian distance d_{si}, from the sink node. Once, all sensor nodes calculate their respective distance from the sink node, the formation of concentric circles around the sink node starts. In this process the sensor nodes are assigned concentric circle index C_j (j = 1, 2, 3, ….. m), each of width R/2. The assignment of concentric circle to the sensor node N_i is carried out with the formula:

$$C_j = \frac{2|d_{si}|}{R} \qquad (4)$$

A. Cluster Formation and Cluster Head Selection

To achieve energy balanced clustering in the network, in proposed scheme, virtual concentric circles are designated as V_1, V_2, V_3…. A virtual concentric circle band lies at the midway between two concentric circles and has width ± δ i.e. 2δ. The index V_j of virtual concentric bands can be calculated as:

$$V_j = \left[\left\{ \left(\frac{R}{2} * C_i \right) + \frac{R}{4} \right\} \pm \delta \right] \qquad (5)$$

$$If \quad x_i = \left\{ \left(\frac{R}{2} * C_i \right) + \frac{R}{4} \right\} \quad Then \ V_j = [x_i \pm \delta] \qquad (6)$$

The information regarding the sensor node radio range R is known apriori to all nodes and value of δ is made known to all sensor nodes at the time of node deployment. The value of δ depends upon the node density in the network. All sensor nodes in the network calculate their respective VCCB index V_j and compare it with their respective distance from the sink node d_{si}. If the distance d_{si}, falls within the VCCB index [x_i ± δ]; the sensor node is declared as a probable candidate for the election of cluster head (CH) in the first round. All other sensor nodes opt out from the process of cluster head selection for the first round and wait for the declaration of first round cluster head, to associate themselves with one of such declared cluster head. On completion of the process of selection of probable candidates for first round cluster head, the process of first round cluster head election starts. The best possible candidate for the election of first round cluster head would be the one located exactly midway between the two concentric circles. For the first round cluster head election, it has been assumed that the energy of all probable cluster head candidates is same. Therefore, only location of cluster head candidate has been

used as sole criteria for first round cluster head selection. All the probable cluster head candidates calculate their distance from center of their respective VCCB as:

$$d_{i(vc)} = \left[\left[\left\{ \left(\frac{R}{2} * C_i \right) + \frac{R}{4} \right\} - d_{si} \right] \right] \qquad (7)$$

Initially back-off timer value of node N_i is set to be t_i, and is given by:

$$t_i = \frac{d_{i(vc)}}{\delta} * T_{ch} \qquad (8)$$

Where T_{ch} is the time allocated for the first round cluster head election. The sink node assists the cluster head formation by sending START message, which directs all the probable cluster head candidates in the network to start their back-off timer at same time. As the back-off timer value is directly proportional to the distance of cluster head candidate from center of VCCB, the back-off timer of the node near to the center will expire first. As soon as the back-off timer of particular cluster head candidate expires, it sends advertising message CH_ADVT, declaring itself as first round cluster head. This message will be transmitted within the radio range $R_3 = (R/4)$ of sensor node. Any other probable cluster head candidate within the radio range $(R/4)$ of the node which has declared itself as cluster head, will stop its back-off timer and associate itself with the declared cluster head for formation of cluster. Similarly all non-cluster head candidates, falling within the radio range $(R/4)$ of declared cluster head, will also associate themselves with the cluster head and thus form a uniform cluster. Same process takes place in all the concentric circles throughout the sensing area and simultaneous cluster head selection and cluster formation takes place within the network.

V. CLUSTER HEAD ROTATION

The role of cluster head in a cluster must be rotated regularly amongst the sensor nodes to prolong the life time of sensor network by balancing the energy consumption of various sensor nodes. Since, cluster head is required to perform extra task of data gathering and data relaying, compared to the regular sensor nodes, its energy drains out faster. Therefore, some mechanism must be adopted to rotate the role of cluster head. The rotation mechanism must ensure balanced energy consumption of all the sensor nodes in cluster. A number of methods for cluster head rotation have been discussed in the literature [13-19]. Most of these methods are based on residual energy of cluster head or distance of prospective cluster head from the center of the cluster.

If uniform transmission power is assumed for intra cluster communication and cluster is assumed to be static, the distance parameter of elected cluster head from center of cluster becomes insignificant, since power consumption for communication remains same irrespective of the location of cluster head inside the cluster. For such cases, the residual energy of cluster head is important parameter to initiate the process of cluster head rotation. For our proposed model, the cluster head rotation process has been performed based on the residual energy of the cluster head, as explained in the next

section. The frequency and timing of cluster head rotation process, aimed at optimum life time of wireless sensor network, is decided by calculating the energy consumption for regular sensor nodes and cluster head nodes for various tasks performed by these nodes including data relaying and cluster
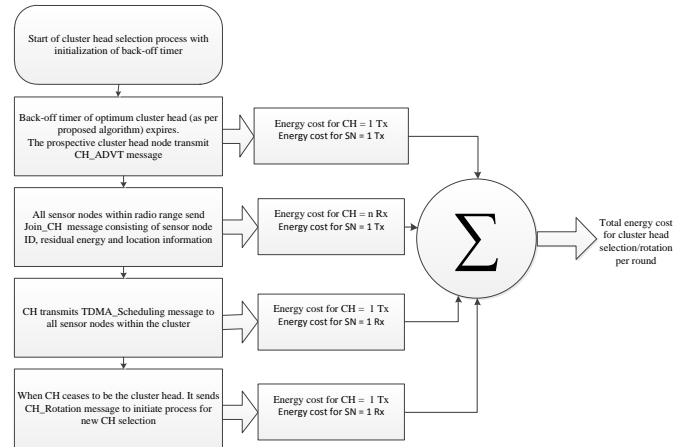

Figure 2. Total energy cost calculation for cluster head selection/rotation per round.

head selection/rotation. After assessing the actual energy consumption of sensor nodes, the balanced energy consumption is ensured by optimal rotation of cluster head node at regular intervals.

A. Energy cost of sensor nodes for normal operation

For normal operation of regular sensor node, energy consumption is basically attributed to the task of sensing, processing and communication of sensed data to the cluster head. Energy consumption of regular sensor node for normal operation can be expressed as:

$$E_{SN} = \left[E_{elect} + E_{sense} + \in d_{SN-CH}^{n} \right] k \qquad (9)$$

Where, E_{sense} is the energy consumed by sensor node for sensing single bit of data and d_{SN-CH} is the distance of sensor node from its cluster head. If, there are n numbers of regular sensor nodes present in a cluster. Then, total energy consumed by all regular nodes in a cluster can be expressed as:

$$E_{SN} = \left[E_{elect} + E_{sense} + \in_{fs} \left(\frac{R}{2} \right)^2 \right] nk \qquad (10)$$

The cluster head receives k bits data sensed by each node present in the cluster during one round of sensing. Apart from receiving the sensed data, cluster head also acts as a regular node to sense the area in its vicinity. It is also responsible for carrying out data aggregation or data fusion operation over the data received from all regular sensor nodes present in the cluster. Moreover, the cluster head is required to transmit data to next hop cluster head or sink node with maximum radio power which is good enough to transmit data over a distance R. Total energy consumption within the cluster for sensing k bits

of data and transmitting this data towards sink node through single/multi-hop communication can be expressed as:

$$E_{total(normal)} = E_{SN} + E_{CH} \tag{11}$$

$$E_{total(normal)} = \left[E_{elect} + E_{sense} + \in_{fs} \left(\frac{R}{2} \right)^2 \right] nk + nk E_{elect}$$
$$+ k E_{sense} (n+1) k E_{aggr} + \left[E_{elect} + \in_{fs} R^2 \right] (n+1)\alpha k \tag{12}$$

Where, E_{aggr} is energy consumed for aggregating the data and α is the data aggregation coefficient. The sequence of messaging is shown in Fig. 2. Extra energy cost for normal sensor node during cluster head selection/rotation process has been calculated as per data transmission and reception carried out during one round of cluster head selection/rotation is given by:

$$E_{SN(extra)} = 1 Tx_{(R/2)} + 3Rx \tag{13}$$

$$E_{SN(extra)} = k \left[E_{elect} + \in_{fs} \left(\frac{R}{2} \right)^2 \right] + 3k E_{elect} \tag{14}$$

$$E_{SN(extra)} = k \left[4 E_{elect} + \in_{fs} \left(\frac{R}{2} \right)^2 \right] \tag{15}$$

Since, there are n normal sensor nodes in the cluster, the total energy consumption of normal sensor nodes, per cluster head rotation process can be expressed as:

$$E_{SN(extra)} = nk \left[4 E_{elect} + \in_{fs} \left(\frac{R}{2} \right)^2 \right] \tag{16}$$

B. Optimal Rotation of Cluster Head

Optimal time/energy threshold to initiate the cluster head rotation process is calculated by analyzing the energy consumption of sensor node for regular sensing and cluster head selection/rotation process. Sensor energy calculation for regular and cluster head selection/rotation process has been analyzed in Section V A. In order to prolong the life time of the network, energy consumption of sensor nodes within the cluster has to be balanced. This balance can only be achieved if the role of a sensor node as cluster head is rotated at appropriate time/energy threshold. This threshold is calculated by assuming n+1 number of nodes in a cluster, out of which, n nodes are assumed to act as regular sensing nodes and one sensor node in the cluster acts as cluster head. If energy of N_i sensor node is assumed to be E_i, total energy of the cluster can be given as:

$$E_{total-cluster} = \sum_{i=1}^{n+1} E_i \quad Joules / cluster \tag{17}$$

Ideally, in a cluster based sensor network, the life of cluster can be prolonged when each sensor node in the network plays the role of cluster head at the most once during the whole life time of the network [9, 15]. This ensures minimum energy consumption during the cluster head rotation and uniform

energy drainage of sensor nodes. We have assumed n+1 number of sensor nodes in a cluster. Therefore, if every sensor node is given a chance to become cluster once during the life time of the sensor network; total number of cluster head rotations will be n. Energy cost of cluster for n cluster head rotations is given as:

$$E_{total(extra-cluster)} = n \left[3k \left\{ E_{elect} + \in_{fs} \left(\frac{R}{2} \right)^2 \right\} + \left(\begin{array}{c} E_{elect} + \\ \in_{fs} R^2 \end{array} \right) + \\ kn E_{elect} + nk \left\{ 4E_{elect} + \in_{fs} \left(\frac{R}{2} \right)^2 \right\} \right] \tag{18}$$

Now, maximum number of possible rounds of data sensing and transmission (taking into consideration the additional energy cost of cluster head rotation/selection) will depend upon the total initial energy of cluster can be given by:

$$N_{optimal} = \frac{E_{total-cluster} - E_{total(extra-cluster)}}{E_{total(normal)}} \tag{19}$$

$$N_{optimal} = \frac{\left[(n+1)E - n \left[3k \left\{ E_{elect} + \in_{fs} \left(\frac{R}{2} \right)^2 \right\} + \left(\begin{array}{c} E_{elect} + \\ \in_{fs} R^2 \end{array} \right) + kn E_{elect} + nk \left\{ 4E_{elect} + \in_{fs} \left(\frac{R}{2} \right)^2 \right\} \right] \right]}{\left[E_{elect} + E_{sense} + \in_{fs} \left(\frac{R}{2} \right)^2 \right] nk + nk E_{elect} + k E_{sense} (n+1) k E_{aggr} + \left[E_{elect} + \in_{fs} R^2 \right] (n+1)\alpha k} \tag{20}$$

For minimal energy cost of cluster head rotation, the cluster head rotation must take place uniformly after $[(N_{optimal})/n]$ rounds. In the proposed model, cluster head rotation process is carried out based on the number of sensing rounds for cluster head. The cluster head in each cluster keeps on counting the number of sensing rounds, it had performed, and after each $[(N_{optimal})/n]$ rounds, it initiates the process of cluster head rotation and after completion of the cluster head rotation process next sensor node start acting as cluster head in the cluster. As $[(N_{optimal})/n]$ rounds are calculated based on the optimized energy threshold; proper energy balance is achieved to prolong the life of sensor network. However, in our model, the cluster head, instead of tracking its residual energy only uses simple counter to count the number of sensing rounds.

VI. PERFORMANCE EVALUATION

In this section the performance of VCCBC protocol is evaluated through simulation experiments. We have implemented the simulator in MATLAB. The performance of VCCBC protocol is compared with Hausdorff [10] and ERP-SCDS [11] protocols.

TABLE I. SIMULATION PARAMETERS

Parameter	Default Value	Range
Area	(100×100) m^2	(50×50)~(400×400) m^2
Number of nodes	400	100~500
Initial Energy of node	2 Joule	
Data packet size (k)	100 byte	
E_{elect}	50 nJ/bit	
ε_{fs}	10 pJ/bit/m^2	
E_{mp}	0.00134 pJ/bit/m^4	
E_{aggr}	5 nJ/bit/signal	
Threshold distance (d$_0$)	87 meters	

The performance metrics include clustering energy dissipation, and life-time of network. The probability of signal collision and signal interference is ignored, assuming TDMA scheduling for intra cluster and DS-SS for inter cluster communication. The simulation parameters are listed in Table I. For each parameter, simulation has been run many times and average result of all runs has been taken for evaluation.
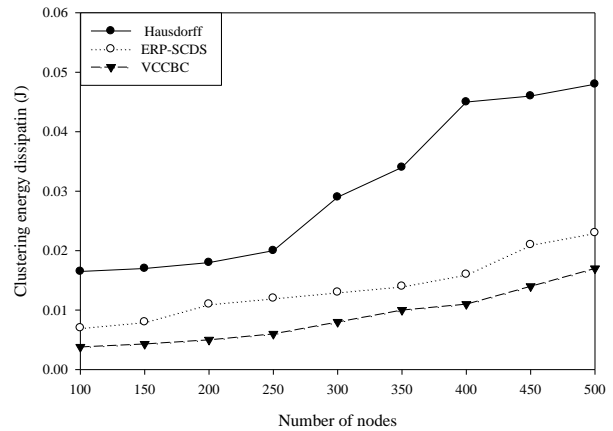
A. Clustering Energy Dissipation

Minimum clustering energy requirement can prolong the network lifetime and can be used as parameter to demonstrate the efficiency of any clustering and routing protocol. Fig. 3 shows the variation in clustering energy dissipation with variation in number of nodes and area of sensing field. Fig. 3(a) shows the simulation results of variation in clustering energy dissipation, in case of Hausdorff, ERP-SCDS and VCCBC protocol with variation in number of nodes from 100 to 500, when nodes are randomly deployed in a sensing field of 100×100 m^2. The clustering energy dissipation is highest in case of Hausdorff, because it uses more control messages for the formation of clusters. ERP-SCDS requires less energy for clustering than Hausdorff protocol, yet its clustering energy requirement is more than VCCBC protocol. Since, it requires the location information transmission of every node to the sink and after receiving this information, the sink initiates the clustering process, dissipating more energy than our proposed protocol VCCBC. This process involves large number of transmit/receive operations, increasing the clustering energy cost. In case of VCCBC protocol, first round cluster head selection and clustering is carried out simultaneously throughout the network. It uses less number of transmit/receive operations since, less number of control messages are exchanged for clustering. Fig. 3(b) shows the simulation results of variation in the clustering energy dissipation with increase in sensing area side length from 50 to 400 meters, when 400 numbers of nodes are deployed. The simulation results show the increase in clustering energy dissipation for Hausdorff and ERP-SCDS. This increase can be attributed to the exchange of more control messages in case of Hausdorff and more transmission of location information from all sensor nodes to the sink with increase in distance of nodes from the sink, in case of ERP-SCDS.
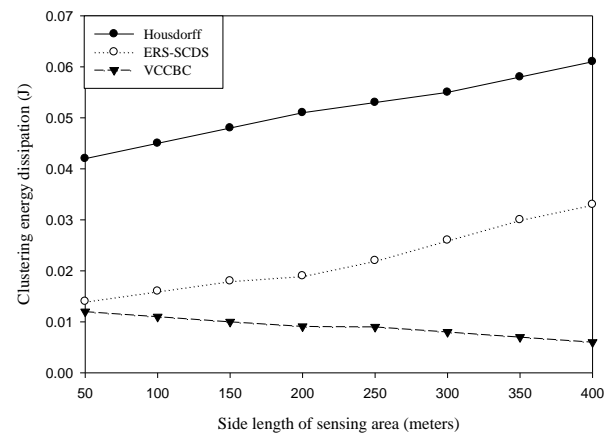
B. Network lifetime

Network lifetime of wireless sensor network is the time span from the deployment to the instant the network ceases to achieve objectives of its deployment. We have considered number of rounds till first node dies (FND), as parameter for the evaluation of VCCBC protocol and existing protocols. Fig. 3 shows simulation results of variation in network life time, i.e. total number of rounds from the deployment to the instant when first node dies in the network. In Fig. 4(a) the number of rounds has been plotted against number of sensor nodes varying from 100 to 500 deployed in sensing field of 100×100 m^2. The simulation results show that VCCBC protocol exhibits better lifetime than Hausdorff and ERP-SCDS protocols. In case of Hausdorff, the unbalanced clustering and more overhead messages for clustering are responsible for the lower lifetime. ERP-SCDS has better network lifetime than Hausdorff due to formation of balanced clusters, yet clustering

requires more energy, causing less lifetime of network in comparison to proposed protocol. Proposed protocol shows

(a)

(b)

Figure 3. Variation in Clustering energy dissipation with (a) number of nodes deployed in sensing field, and (b) side length of sensing field.

highest network lifetime among these protocols, because, it require less energy for clustering, the cluster formed are well balanced and cluster head rotation is carried out considering all kinds of energy consumption within the network, providing more balanced energy depletion of sensor nodes in the cluster. Fig. 4(b) shows the simulation results of variation in network lifetime with increase in side length of sensing field from 50 to 400, keeping the node deployment fixed to 400. Proposed VCCBC protocol exhibits better network life time in comparison to ERP-SCDS and Hausdorff protocols, despite increase in network area. This is due to the fact that in case of VCCBC protocol, the percentage increase in nodes acting as cluster head, decrease in number of nodes per cluster, and increase in percentage of single node clusters is smaller in comparison to these protocols.

VII. CONCLUSIONS

An energy efficient clustering, cluster head selection/rotation and data routing protocol is proposed in this

paper. The proposed virtual concentric circle band based clustering protocol (VCCBC) ensures the formation of uniform clusters in
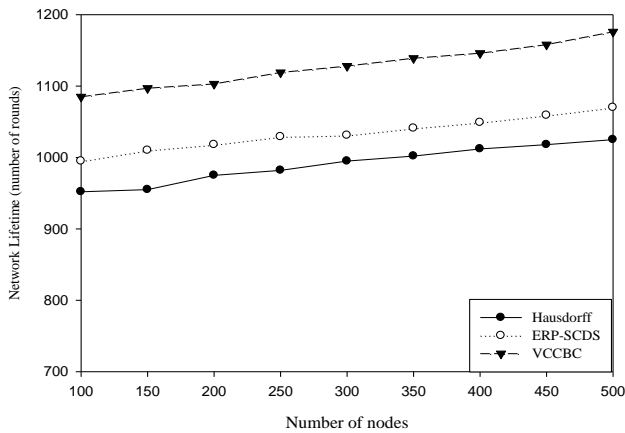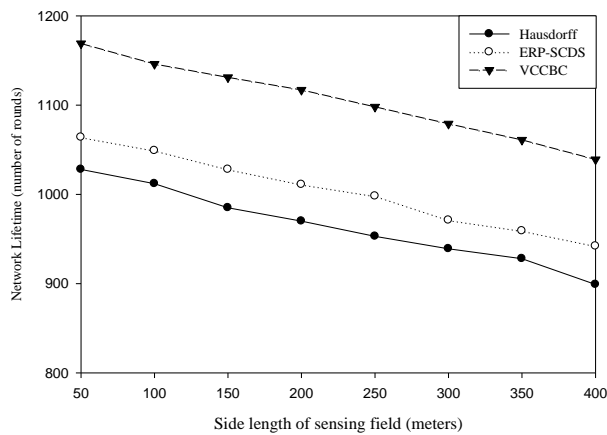

(a)

(b)

Figure 4.   Variation in network lifetime with (a) number of nodes deployed in sensing field, and (b) side length of sensing field.

virtual concentric circular bands around the sink. In proposed method the cluster formation takes place only once in network lifetime, thus avoiding energy wastage associated with re-clustering process. The cluster head rotation process is based on energy calculation of various tasks performed by the sensor nodes and cluster head in a cluster. The timing and frequency of cluster head rotation is carried out by balancing the energy consumption. This results in balanced energy drainage of the node in network, improving the network lifetime. The simulation results demonstrate the effectiveness of proposed protocol in term of clustering energy dissipation.

REFERENCES

[1] C.Y. Chong and S.P. Kumar, "Sensor networks: Evolution, Opportunities, and Challenges," Proceedings of the IEEE, vol. 91, no. 8, pp. 1247–1256, August 2003.

[2] Ameer Ahmed Abbasi and Mohamed Younis, "A Survey on Clustering Algorithms for Wireless Sensor Networks," Elsevier Journal of Computer Communications, vol. 30, pp. 2826–2841, October 2007.

[3] Xiaobing Wu, Guihai Chen and S.K. Das, "Avoiding Energy Holes in Wireless Sensor Networks with Nonuniform Node Distribution," IEEE Transactions on Parallel and Distributed Systems, vol. 19, no. 5, pp. 710-720, May 2008.

[4] M. Younis, M. Youssef and K. Arisha, "Energy-Aware Management in cluster-Based Sensor Networks," International Journal on Computer Networks, vol.43, no. 5, pp. 649–668, December 2003.

[5] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Micro Sensor Networks" Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00), vol.2, pp. 234-243, January 2000.

[6] Wendi B. Heinzelman, Anantha P. Chandrakasan and Hari Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks", Transactions on Wireless Communications, Vol. 4, pp.660 –670, October 2002.

[7] An Amir Sepasi Zahmati, Bahman Abolhassani, Ali Asghar Beheshti Shirazi, and Ali Shojaee Bakhtiari, "Energy-Efficient Protocol with Static Clustering for Wireless Sensor Networks," World Academy of Science, Engineering and Technology vol. 28, pp. 69-72, 2007.

[8] M. Ye, C. Li, G. Chen and J. Wu, "An Energy Efficient Clustering Scheme in Wireless Sensor Networks," Journal of Ad Hoc & Sensor Wireless Networks, Vol.1, pp. 1–21, 2006.

[9] C. Li, M. Ye, G. Chen and J. Wu, "An Energy-Efficient Unequal Clustering Mechanism for Wireless Sensor Networks," Proceedings of IEEE International Conference on Mobile Ad hoc and Sensor Systems Conference(MASS05), Washington, D.C., pp. 604-611, November 2005.

[10] Xiaorong Zhu, Lianfeng Shen and Tak Shing Peter Yum, "Hausdorff Clustering and Minimum Energy Routing for Wireless Sensor Networks," IEEE Transaction on Vehicular Technology, vol. 58, no.2, pp. 990-997, February 2009.

[11] Huei Wen Ferng, Robby Tendean and Arief Kurniawan, "Energy Efficient Routing Protocol for Wireless Networks with Static Clustering and Dynamic Structure," Springer Journal of wireless personal communication (published online), March 2001.

[12] D. Goldenberg, A. Krishnamurthy, W. Maness, R. Yang, A. Young, A.S. Morse, A. Savvides and B.D.O. Anderson, "Network localization in partially localizable networks," IEEE INFOCOM, vol. 1, pp. 313-326, 2005.

[13] I. Rekleitis, D. Meger and G. Dudek, "Simultaneous Planning, Localization, and Mapping in a Camera Sensor Network," Robotics and Autonomous Systems, vol. 54, no.11, pp.921–932, 2006.

[14] Haibo Zhang and Hong Shen, "Balancing Energy Consumption to Maximize Network Lifetime in Data-Gathering Sensor Networks," IEEE Transactions on Parallel and Distributed Systems, vol. 20, no.10, pp.1526-1539, October 2009.

[15] A. Manjeshwar and D. Agrawal, "TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks," Proceedings of 15th International Parallel and Distributed Processing Symposium (IPDPS'01), San Francisco, CA, pp. 2009-2015, April 2001.

[16] S. Lindsey and C. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems," Proceedings of IEEE Aerospace Conference, Big Sky, Montana, USA, vol.3, pp.1125-1130, March 2002.

[17] O. Younis and S. Fahmy, "HEED: A Hybrid Energy-Efficient Distributed Clustering Approach for Ad Hoc Sensor Networks," IEEE Transactions on Mobile Computing, vol. 3, no.4, 2004.

[18] D.J Baker and A. Ephremides, "The Architectural Organization of a Mobile Radio Network via a Distributed Algorithm," IEEE Transactions on Communications, vo. 29, no.11, pp. 1694-1701, 1981.

[19] D.J Baker and A. Ephremides, "The Design and Simulation of a Mobile Radio Network with Distributed Control," IEEE Journal on Selected Areas in Communications, vol. 2, no.1, pp. 226-237, January 1984.

[20] A. Ephremides, J.E. Wieselthier and D.J. Baker, "A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling," Proceedings of IEEE vol. 75, no.1, pp.56-73, January 1987.

[21] K.Xu and M. Gerla, "A Heterogeneous Routing Protocol Based on A New Stable Clustering Scheme," Proceeding of IEEE Military Communications Conference, Anaheim, CA, pp. 123-131, October 2002.

[22] R. Nagpal and D. Coore, "An Algorithm for Group Formation in an Amorphous Computers," Proceedings of 10th International Conference on Parallel and Distributed Systems (PDCS'98), Las Vegas, NV, October 1998.

[23] M. Younis, K. Akkaya and A. Kunjithapatham, "Optimization of Task Allocation in A Cluster–Based Sensor Network", Proceedings of the 8th IEEE Symposium on Computers and Communications (ISCC'2003), Antalya, Turkey, June 2003.

[24] A. Manjeshwar and D. Agrawal, "APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks," Proceedings of International Parallel and Distributed Processing Symposium (IPDPS'02), Fort Lauderdale, Florida, pp.195-202, April 2002.

[25] M. Demirbas, A. Arora and V. Mittal, "FLOC: A Fast Local Clustering Service for Wireless Sensor Networks," Proceedings of Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks (DIWANS'04), Palazzo dei Congressi, Florence, Italy, June 2004.

[26] S. Yi, J. Heo, Y. Cho, and J. Hong, "PEACH: Power-Efficient and Adaptive Clustering Hierarchy Protocol for Wireless Sensor Networks," Elsevier Journal on Computer Communications, vol. 30, pp. 2842-2852, October 2007.

AUTHORS PROFILE

Ashok Kumar is working as Associate Professor in the Department of Electronics and Communication Engineering, National Institute of Technology Hamirpur Himachal Pradesh- INDIA. Presently he is pursuing PhD from National Institute of Technology. His current research areas of interest include wireless communication and wireless sensor networks. He has published more than 20 research papers in International/National journals and conferences in these areas. He is life mem member of ISTE.

Dr. Narottam Chand received his PhD degree from IIT Roorkee in Computer Science and Engineering. Previously he received M Tech and B Tech degrees in Computer Science and Engineering from IIT Delhi and NIT Hamirpur respectively. Presently he is working as Associate Professor, Department of Computer Science and Engineering, NIT Hamirpur. He has served as Head, Department of Computer Science & Engineering, during Feb 2008 to Jan 2011 and Head, Institute Computer Centre, during Feb 2008 to July 2009. He has coordinated different key assignments at NIT Hamirpur like Campus Wide Networking, Institute Web Site, Institute Office Automation. His current research areas of interest include mobile computing, mobile ad hoc networks and wireless sensor networks. He has published more than 120 research papers in International/National journals & conferences and guiding six PhDs in these areas. He is member of IEEE, ISTE, CSI, International Association of Engineers and Internet Society.

Dr. Vinod Kapoor was born at Mandi town in Himachal Pradesh, India. He received his BE Degree in Electronics & Communication Engineering from National Institute of Technology (formerly Regional Engineering College), Durgapur, West Bengal, in the year 1987 and Masters Degree in Electronics & Control from Birla Institute of Technology & Science, Pilani (Rajasthan) in the year 1996. He did his Ph. D. from Kurukshetra University, Kurukshetra in the field of Optical Fiber Communication in Sept. 2006. He also obtained his MBA degree with specialization in Human Resources Management in the year 2002.He is presently Professor in the Department of Electronics & Communication Engineering NIT Hamirpur, India. He has coordinated different key assignments at NIT Hamirpur like Training & Placement Officer, Chief Warden (Hostels), organizing short term courses. His research interest includes optical fiber communication and optoelectronics/nano devices. He has published more than 30 research papers in International/National journal & conferences and guiding four Ph.Ds in these areas. He is member of ISTE, IETE and IE.

An hybrid method for the Arabic queries disambiguation to improve the relevance calculation in the IRS

Adil ENAANAI, Aziz SDIGUI DOUKKALI, El habib BENLAHMER
Centre d'études doctorales ST2I, ENSIAS
Rabat, 10000, Morocco

Abstract—In the information systems, the query's expansion brings more benefices in the relevant documents extraction. However, the current expansion types are focused on the retrieve of the maximum of documents (reduce the silence). In Arabic, the queries are derived in many morphosemantical variants. Hence the diversity of the semantic interpretations that often creates a problem of ambiguity. Our objective is to prepare the Arabic request before its introduction to the document retrieval system. This type of preparation is based on pretreatment which makes morphological changes to the query by separating affixes of the words. Then, present all of morphosemantical derivatives as a first step to the lexical audit agent, and check the consistency between the words by the context parser. Finally we present a new method of semantic similarity based on the equivalence probability calculation between two words.

Keywords-Relevance; Information research; Similarity; Semantic gene; Arab treatment.

I. INTRODUCTION

In the information research systems, relevance is function of the similarity degree between the query and the document. However, many functions of similarity are recently proposed. Most of these functions are based on the principle of vector distance where the meaning of words is not supported. However two words whose distance is zero, meaning they are similar. Unlike, there are words of high distance that mark the same meaning (synonyms), or words of distance equal to zero which mean several things.

Functions based on the distance vector are unable to provide the exact value of semantic similarity of terms. Withal, there are many stemming algorithms which contributing to the calculation of relevance by comparing the roots of words. That has yielded good results. But still insufficient, because there are some Arabic words whose roots are written in the same way while their meaning is different.

The difficulty of having a function of semantic similarity lies in the fact that the comparison of meaning between two words is possible only after an inclusion of a valid morphosemantical analysis. Hence, prepare the query before being introduced to the data retrieval system.

II. STATE OF THE ART

Most of the works on the Arabic syntactic analysis have led to the achievement of some laboratory prototypes. Indeed, to our knowledge to date, there is no parser marketed or distributed for scientific research. In the remainder of the State of the art, we present a few of the Arabic language analysis systems.

A. The AraParse system for syntactic analysis of the Arabic unvowled

AraParse is a system to analyze Arab texts in their vowelized, unvowelized or partially vowelized form [1]. The objective is to achieve a core of morpho-syntactic analysis system that can be reused in other applications such as information research and automatic translation.

AraParse is based on linguistic resources in wide coverage. It uses a lexicon of lemmas generated from the DIINAR.1 dictionary [3]. This glossary contains 19 6818 unvowelized bases distributed to 39 000 nominal bases, 79 818 verbal bases and 78 000 derivable bases from 20 000 verbs of DIINAR.1 ([2], [1]).

To recognize unknown sequences or unknown words, this system uses an approximate matching technique implemented with the AGFL formalism and using the priority operator between the alternatives of a rule and regular expressions [1].

Ouersighni [1] proposed the use of AraParse detect and diagnose the faults of accord. He used the accord rules proposed by Belguith [4] in the DECORA system.

B. The system DECORA for detection and correction of the Arabic accord errors

In scientific research focused on the analysis of Arabic language, Belguith [4] proposed a method for detecting and correcting errors in accord. This method has been implemented in the system DECORA. It is based on syntagmatic analysis for the error detection and correction multi-criteria analysis. Extended sentence is defined as a group of one or more original sentences linked to accord between them.

Extended syntagmatic analysis operates in two stages ([5], [6]). The first step is to cut the initial phrase in initial phrases by locating the boundaries between them. This Division is guided by a set of rules using the syntagmatic borders as a means of identification of the original phrases. The second stage, allows building the extended phrases. The constitution of these phrases is guided by indicators of surface and is based on a set of rules to locate the accord links between the original phrases. These rules allow for example to relate the possessive pronouns in the phrases to which they relate, to integrate the verbs in the phrase that contains the subject considered to relate the original phrases that represent anaphoric proposals to the phrase containing the syntactic unit to which it relates.

C. Spoken Arabic Levantin Analyzer

Chiang [7] is interested by the analysis of the Arabic Levantin (AL) (a group of Arab dialects spoken in Syria, Palestine, Western Jordan and the Lebanon). He proposed an approach to translate the AL in Standard modern Arabic (SMA). Then link the sentence in AL to the corresponding analysis in SMA.

Note that the automatic translation is particularly difficult when there is no resource available as the parallel texts or the transfer lexicons. Thus, Chiang is primarily based on a corpus annotated from modern standard Arabic (MSA Treebank) [8] as well as a corpus annotated Arabic Levantine and more specifically that of the Jordanian dialect (i.e., TBPC Treebank [10]).

He built a lexicon contains the AL/ASM pairs of the forms of words. Also he built a synchronous grammar ASM-dialect. He assumes that each tree in the grammar of modern standard Arabic extracted from the MSA Treebank is also a tree of Levantine Arabic given syntactic similarity between the DSO and the AL.

D. A morphological and syntactic Arab text Analyzer

Debili Zouari [9] proposed the automatic construction of a dictionary containing all the inflected forms. This construction is made by a conjugator and a derivator.

The principle of morphological analysis is to make:

- The division of the text to graphical words.

- Research of enclitic and proclitic of the word.

- Verification (for each possible division) of the compatibility (proclitic / enclitic; enclitic / root; root / proclitic).

When consulting the dictionary, Zouari and Debili use the rewrite rules to find the "normal" form of the word.

The parsing process follows the phase of morphological analysis and related on the construction of the frequency binary and ternary matrices of precedence. These matrices are constructed from the annotated start texts "by hand" (this is the learning phase). They are then used to analyze new texts.

E. IRLA analyzer

The IRLA Analyzer is a queries interrogation system in the Arabic natural language [11]. It takes in input an Arabic sentence and translates it as query to run by an operating system. This parser allows to treat a subset of natural language (i.e., essentially imperative sentences), it produces a parenthesed form expressing the semantic of the query [12].

The parser can treat some simple linguistic problems (synonymy, negation, coordination). It is based on the detection of conceptual and linguistic surface indicators at the analysis.

F. Elliptical sentences Analyzer

In its research work on the Arabic analysis, Haddar [13] conducted a parser for the detection and resolution of the elliptical sentences in the Arabic texts. This parser is based on a method of syntactic analysis for verification of the syntactic structures of the proposals. This method uses a formal grammar rules generating verbal proposals written in Arabic. Access to these rules is coordinated with increased transitions (ATN) networks. The parser is coupled with another parser treating with semantic ellipses.

III. CONTRIBUTION

The calculation of relevance in our approach is focused on semantic similarity function which gives a result as a percentage of equivalence between two Arabic words. Knowing that they are written in various derived forms, it had to begin by morphological analysis which returns the origin of the derivative in question. Therefore, the possibility of separate affixes of the word is subsequently obtained by the original non-vocalized of the word which may refer to several meanings. The probable meaning to be just, is that which is on conflict with the user profile. To filter the true meaning, we have developed an automatic profiling system that brings together user queries and implements format indexed in a database. Our approach has given a good result on the morphosemantical ambiguity. In the remainder of this article we will present the various stages of analysis that we introduced in the relevance calculation [17] [18].

G. The morphological analysis

After a sending of query through the meta-search engine, we get a list of results. This list is sorted according to the relevance algorithm used in each data source (search engine).

We begin firstly by re-indexing of the documents founded by a semantic analysis module. Indeed, this module receives three parameters: the document, the query and the user profile. The document and the query are an affected by an in-page modification, which dissects the words to remove affixes [13]. If there are several cases (ex: بطريق = طريق > = بطريق = 'Penguin' or ب طريق = 'by road'), we tests the consistency of each derived, relatively with other words in the query, document and profile. Therefore, we accept just the possibility which is in the current context [5]. The separation of the terms of the request and the document offers more precision to the similarity of the triplet: document, request and profile, Hence, the need to have a flexible, fast and easy method [17].

Stemming used methods are unable to resolve the problem of semantics, because, to return the root of a word means that we can derive several forms to build a set of words that are not necessarily similar on the semantic meaning [6]. For example,

the root of the word 'طريق' is 'طرق'. This root may take several derived forms, same as 'طريقة', 'مطرقة' which do not mean the same thing. Hence, we have to found the word origin by keeping the semantic aspect of all. The method that we have introduced is to dissect the word to draw the origin after applying a light stemming. The origin is later transformed into a singular to test its existence in the dictionary of the Arabic words (ARRAMOOZ ALWASEET dictionary). If the word exists, then we retrieve its definition and we type to construct a semantic entity (SE). These SE are used to test the consistency of the word in a text. For this, presents the following heuristic algorithm:

```
M=Set of arabic words
D=Set of the derived forms
P=Set of the prefixs
S=Set of the suffixs
Mot=The word to derive
Prefix ϵ P
Suffix ϵ S
Fonction derive(Mot, pref, suff, def)
Mot ← Singular(Mot)
IF (Mot in M) AND NOT (Mot in D)
  Add Mot to D
End If
If (Mot is begined by 'ال' ) AND (def=null) AND   NOT ((Mot is ended
by Suffix) OR (suff<>null))
  m ← Mot – 'ال'
  def ← 'ال'
  Derive(m,pref,suff,def)
End If
If (Mot is begined by Prefix) AND (pref=null) AND (def=null)
  M ← Mot-Préfix
  Pref ← Préfixe
  Derive(m,pref,suff,def)
End If
If (Mot is ended by Suffix) ET (suff=null) ET (def=null)
  m ← Mot – Suffix
  suff ← Suffix
  Derive (m,pref,suff, def)
End If
End Function
```

Example: Word = 'بطريقنا' (En: our path)

At the first entry in the *derive* function, the word 'بطريقنا' is not found in in the Set of words M. we test if the Word 'Mot' is begun by a prefix. If yes, we remove the prefix to have the new word 'طريقنا' (En: our path) which will be introduced as a parameter to the recursive function '*derive*'. At its entry, the word 'طريقنا' is undefined (نكرة) and is not in the set of words, and not starting with a prefix. Therefore, we pass to the second test on the suffix. The word ends with the suffix 'نا' (En: our), it also removed to have the newest derived word 'طريق' (En: path) which will also be introduced in the third hierarchical level of the recursive function. At its entry we test again if the word 'طريق' exists in the set of the Arabic words. Now, the word is founded and added in the set D of the derived words.

On returning to the first hierarchical level of recursion, the word 'بطريقنا' must be passed to another test of suffix. Then, we remove the suffix 'نا' to have the word: 'بطريق' (En: Penguin). This last word passes as a parameter to the function 'derive' which tests its existence in the set 'M'. We find also the word 'بطريق' in 'M' and add to the set of derived words 'D'. The following figure illustrates the changes [14].

Figure 1 : The morphological analysis

In this way, the statistical parser and the profiling algorithm receive well presented data. Then, we will see how to use the morphological analysis to give valid semantic presentations [11].

H. Semantic gene building

The semantic gene is an object containing the information needed (from a database) to the disambiguation of the Arabic words. The construction of gene starts at the level of the morphological analysis in the determination of the origin of the word. The following diagram illustrates the format of the semantic gene [17].
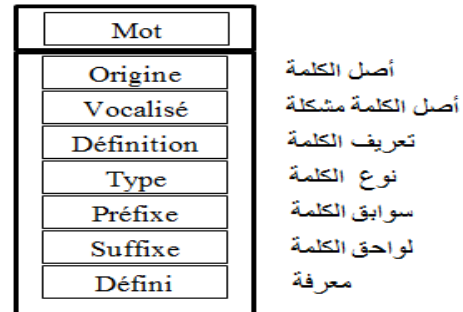

Figure 2 : Structure of the semantic gene

I. Two words queries Analysis

The expression of need is made by a multi-word query. The information system research returns a set of documents that contain all of the semantically valid sentences. These sentences include those that contain the desired word. This Word can mean several things, Hence the problem of semantic ambiguity.

To reduce the effect of this type of ambiguity, we designed a semantic filtering system that recognizes the type of the word based on the rules of constitution of the Arabic sentences. Given the difficulty of semantic analysis of the Arabic sentences, we consider the case of significant sentences of two lemmas. The following table shows the different cases of a sentence of two words semantically consistent [7].

TABLE I. TYPES OF WORDS

Object		Property		Fact	
Type	Acronyms	*Type*	Acronyms	*Type*	Acronyms
صيغة مبالغة	O_{Mob}	اسم تفضيل	P_{Taf}	مصدر	F
منسوب	O_{Man}	اسم مفعول	P_{Maf}		
جامد	O_{Jam}	اسم فاعل	P_{Fa}		
اسم مفعول	O_{Maf}	صفة مشبهة	P_{SM}		
اسم فاعل	O_{Fa}	صفة	P_{S}		
		منسوب	P_{M}		

Note: Just the types listed in the table above are considered.

TABLE II. TABLE ABSTRACT OF THE SYNTHETIC MUSSELS

Prefix of the word1	Word1	R	Word2	Example	Pattern Mot1 R Mot2	Relation between Word1 and word2
(ك، ل، و، ب)	D	∅	D	المدينة القديمة	O P	M2 is P for M1
				الإعادة البطيئة	F P	M2 is P for M1
(ك، ل، و، ب)	I	∅	D	مدينة العرفان	O O	M2 is a specification for M1
				طرح التساؤلات	F F	M1 is an action Applied on M2
				تسيير الشركة	F O	M1 is an action Applied on M2
				أرقى العائلات	P O	M1 is P for M2
				أضعف الاحتمالات	P F	M1 is P for M2
(و)	D	∅	I	المسألة صعبة	O P	M2 is P for M1
				التسارع بطيء	F P	M2 is P for M1
(ك، ل، و، ب)	I	∅	I	سنة سعيدة	O P	M2 is P for M1
				تفكير سليم	F P	M2 is P for M1
				تزويج قاصر	F O	M1 is an action Applied on M2
				أقوى رجل	P O	M1 is P for M2
				أصعب اختبار	P F	M1 is P for M2
(ك، و، ب)	D	ل	D	التعليم ل الشعب	F ل O	M1 is a F applied for M2
				الماء ل القرى	O ل O	M2 is a specification for

Prefix						M1
(ك، ل، و، ب)	D	ب	D	الأمر ب المعروف	F ب O	M2 is a specification for M1
…	…	…	…	…	…	…

We note that there are prohibited cases (as: " O و P"). Therefore, we have designed a set of mussels forming all possible cases of the sentences of two lemmas. This set of mussels is an array of objects where each element describes a phrase (pattern) model. The process of correction is applied firstly on the list of the semantic entities (alimented query) of the user to remove the inconsistent morphosemantical variants [12]. Then, we send the remaining lists for contextual correction system. The latter uses the contextual corpus to refilter the list. The result is one or more lists of consistent semantic entities at the contextual level as at the semantic level. Finally, the research system is receives a suite of semantic genes containing all information that can help the extraction, selection and filtering of relevant documents [8].

There are words that can be objects or properties. Our approach supports the gene by assigning the type of the word [17] [18].

Example:

Let's say that we have the sentence ph = "الحاكم العادل" (En: Just governor).

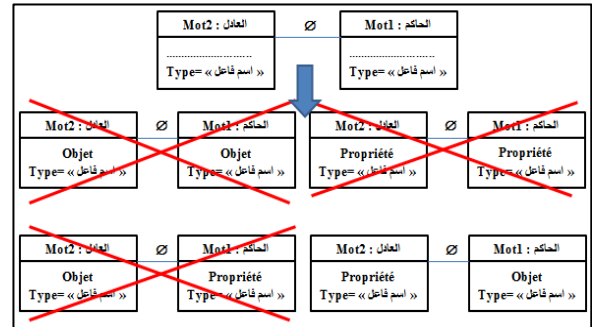Word1= « الحاكم » ; Word2= « العادل » ; R= « ∅ »

Figure 3 : Two words query analysis

The table of mussels shows that the only case which exists in is: Word1 = "object"; Word2 = "property". Therefore, the sentence "العادل الحاكم" is semantically different to "الحاكم العادل", because its components are not similar. In this way, our system will be able to considerate polysemy [9].

This work is an aspect that has largely been addressed to the Latin language (English, French, ...) and even in some work for the Arabic language. Indeed, research based on the user profile to reduce noise and silence in the information research has yielded satisfactory results especially with the modeling of the user profile and the research domain with the notion of ontology. However, the ambiguity in the terms of query cannot guess the domain to choose from. Hence, we must prepare the query to reduce morphosemantical ambiguity, then guess the context from the user profile and

create genes to clarify the semantic field and the context intended by the user. The following diagram illustrates the various steps of our approach [16] [17].
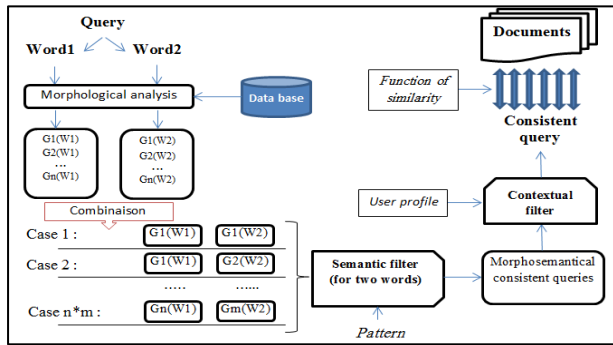

Figure 4 : General diagram of approach steps

IV.   EVALUATION

To test the effectiveness of our method, we have developed a test meta-search engine. The latter uses data sources "Bing", "Yahoo", "Yandex". Then we compared our results with Google results. We have obtained the following table after throwing 100 queries [17].

Table 4: Approach evaluation

	Google	Our system
Average of number of relevant links sorted in the midst of the ten first positions	**6.47**	**8.14**

V.   CONCLUSION ET PERSPECTIVES

In this article, we introduced the concept of the semantic gene that contributes to the Elimination of ambiguity in the information research systems. We also explained how to create the semantic genes from the morphological, contextual and semantic analysis and how to differentiate between homonyms. The automatic profiling is also an interesting factor to approach to the needs of users.

Our target is to automatically create semantic graphs whose semantic genes nodes are very rich in side informational data. Where each node has a context, a definition, a type of Word, a morphological form, a list of successors and a list of predecessors. Finally we wish to develop a meta-research engine which can return optimal results.

REFERENCES

[1]   R. Ouersighni, La conception et la réalisation d'un système d'analyse morpho-syntaxique robuste pour l'arabe : utilisation pour la détection et le diagnostic des fautes d'accord, Thèse de doctorat, université lumière–Lyon2, 2002.

[2]   J. Dichy, Morphosyntactic Specifiers to be associated to arabic lexical entries-Methodological and theorical aspects, actes de la conférence ACIDCA'2000, Volume Corpora and Natural Language Processing, Monastir-Tunisie, pp. 55–60, 22–24 Mars 2000.

[3]   J. Dichy, On lemmatization in Arabic : A formal definition of the Arabic entries of multilingual lexical databases, In proceedings of the Arabic Language Processing workshop, Association for computational linguistics (ACL) 39th annual meeting and 10th conference of the european Chapter, Toulouse, Juillet 2001.

[4]   L. Belguith Hadrich, Traitement des erreurs d'accord de l'arabe basé sur une analyse syntagmatique étendue pour la vérification et une analyse multicritère pour la correction, Thèse de doctorat en Informatique, Faculté des Sciences de Tunis, Février 1999.

[5]   L. Belguith Hadrich, A. Ben Hamadou et C. Aloulou, Using the TOPSIS multicriteria method to direct an agreement error correction process : An application to Arabic, Recent Advanced Natural Language Processing II, vol.189, pp. 105–114, John Benjamins, Amsterdam/ Philadelphia, 2002.

[6]   L. Belguith Hadrich, A. Ben Hamadou, Traitement des erreurs d'accord : une analyse syntagmatique pour la vérification et une analyse multicritère pour la correction, Revue d'Intelligence Artificielle (RSTI – RIA), Editions Hermès-Lavoisier, vol. 18, N5 et 6, pp. 679–707, Décembre 2004.

[7]   Chiang, , M.Diab, N. Habash, O. Rambow, S. Shareef, Arabic Dialect Parsing, In Proceedings of the European chapter of the Association of Computational Linguistics EACL, , pp. 369–376, Trento, 2006.

[8]   M. Maamouri, A. Bies, T. Buckwalter, The Penn Arabic Treebank : Building a large-scale annotated Arabic corpus, In NEMLAR Conference on Arabic Language Resources and Tools, pp. 102–109, Cairo, Egypt, 2004.

[9]   Debili et L. Zouari, Analyse morphologique de l'arabe écrit voyellé ou non fondée sur la construction automatique d'un dictionnaire arabe, Cognitiva, Paris, 1985.

[10]   M. Maamouri, A. Bies, T. Buckwalter, M. Diab, N. Habash, O. Rambow, D.Tabessi, Developing and using a pilot dialectal Arabic Treebank, In Proceedings of the Fifth International Conference on Language Resources and Evaluation, LREC'06, Genoa, Italy, 2006.

[11]   N. Fourati et A. Ben Hamadou, The Linguistic Portability of the Natural language Interfaces, 9th CLIN (Computational Linguistics in the Netherlands) Meeting, University of Leuven, 11 December, 1998.

[12]   N. Fourati, A. Ben Hamadou et F. Gargouri, An object-orientated pivot model for the design of human-machine interfaces in natural language, 6th International Conference and Exhibition on Multi-lingual Computing Cambridge, 17–18 April 1998.

[13]   K. Haddar, Caractérisation formelle de l'ellipse et processus de recouvrement de la langue arabe, Thèse de doctorat, Université de Tunis II – Faculté des Sciences de Tunis, juillet 2000.

[14]   El YOUNOUSSI Y (2011), La racinisation de la langue arabe par lesautomates à états finis (AEF), 4th International Conference on Arabic Language Processing.

[15]   Alrahabi M. (2004). Filtrage sémantique de textes en arabe en vue d'un prototype de résumé automatique, JEP-TALN 2004, Traitement Automatique de l'Arabe Fes

[16]   Collet K (2003) Méthode du TALN, traitement automatise du langage naturel, notion de l'indexation automatique, Cours, URFIST Bretagne Loire-Atlantique.

[17]   A. ENAANAI and A.DOUKKALI (2012), An hybrid approach to calculate relevance in the meta-search engines, IJSAT, volume 2, N°3. March 2012.

[18]   Benlahmar H. (2006). A New Solution for Data Extraction: GENE/LONE Method, IJCSNS International Journal of Computer Science and Network Security, Vol 6, N° 7.

AUTHORS PROFILE

Adil ENAANAI was born in 1981 in Fes in Morocco took his Bachelor of mathematical and experimental sciences in 2001; he has continued his university studies at the Faculty of science ben m'sik for having an applied license in computer science. Then he went to the capital to prepare a master of computer engineering research degree. After that, he began his doctoral studies at the CEDoc ENSIAS where the subject is "the meta search in Arabic language".

Also, he obtained the diploma of specialized technician in industrial automatism, and the diploma of the normal superior school in teaching pedagogy. He worked as Professor of computer science for six years which the specialty was the web development and research information on the internet.

Dynamics of Mandelbrot Set with Transcendental Function

Shafali Agarwal
Dept. of Computer Applications
JSSATE, Noida, India

Gunjan Srivastava
Dept. of Computer Applications
JSSATE, Noida, India

Dr. Ashish Negi
Dept. of Computer Science
G. B. Pant Engg. College
Pauri Garwal, Uttarakhand, India

Abstract- **These days Mandelbrot set with transcendental function is an interesting area for mathematicians. New equations have been created for Mandelbrot set using trigonometric, logarithmic and exponential functions. Earlier, Ishikawa iteration has been applied to these equations and generate new fractals named as Relative Superior Mandelbrot Set with transcendental function.**
In this paper, the Mann iteration is being applied on Mandelbrot set with sine function i.e. $sin(z^n)+c$ and new fractals with the concept of Superior Transcendental Mandelbrot Set will be shown.
Our goal is to focus on the less number of iterations which are required to obtain fixed point of function $sin(z^n)+c$.

Keywords-Superior Transcendental Mandelbrot Set; Mann iteration; fixed point; Escape Criteria.

I. INTRODUCTION

Complex dynamics have a new significant development with the explosion of popular interest in the beautiful fractal objects that form the subject matter of the theory [5,6]. The computer-generated images of Julia and Mandelbrot sets bombard mathematicians to investigate the nature of both the Fatou and Julia sets of a given complex function [4].

Though a moment's reflection confirms the origin is a critical point for the cosine and for the sine, the transcendental Julia sets fill major regions of the complex plane [3].

Based on this concept, authors found the region *M* of non-escape corresponds to a principal central bulb set with a fractal series of black hearts, including a series lining the *x*-axis, and that any *c* value in *M* corresponds precisely to Julia set kernels of the corresponding quadratic type [7].

Ereneko [1] studied that for every transcendental functions, the set of escaping points is always non-empty. The set of parameter values of *c* for which the Julia set of *Q* is connected forms the well-known Mandelbrot set.

Mandelbrot set serves as a lexicon for the Julia Set. The location of the parameter *c* with in the Mandelbrot set furnishes information on properties of the corresponding Julia set. There are similarities between magnified positions of the Mandelbrot set and the corresponding filled Julia set holds only near certain *c* values such as the central junctions of the antenna. These are *c* values, for which 0 is eventually periodic; such *c* values are called Misiurcwicz points [10]. In

2004, Rani jointly with Kumar applied Mann iteration to functions and introduces superior iteration in non linear sciences and gave new escape criterions for complex polynomials. Thus authors computed superior Julia set [8] and superior Mandelbrot set [9] for complex polynomials.

Fixed point can be obtained by repeated function iteration or Picard iteration. There are some other iteration processes like Ishikawa iteration and Mann iteration, which is required to obtain weak or strong convergence to a fixed point in case of non-expansive maps, pseudocontractive maps etc [6]. The purpose of this paper is to generate superior Julia set and Mandelbrot set using transcendental sine function. We are generating fractals for $sin(z^n)+c$ using Mann iteration and calculating the fixed points for the same.

II. DEFINITIONS & PRELIMINARIES

The generation of fractal for $sin(z^n)+c$ is much similar to standard quadratic equation of Mandelbrot set but it consists of repeated iterations upto *n* times with respect to sine function. Following are some basic definitions require for subsequent analysis.

A. Definition

(Superior iterates) Let *X* be a non-empty set of real numbers and $f : X \to X$. For x_0 belongs to *X*, construct a sequence $\{x_n\}$ in the following manner [11,12]:

$x_1 = \beta_1 f(x_0) + (1 - \beta_1)x_0$
$x_2 = \beta_2 f(x_1) + (1 - \beta_2)x_1$
.
.
.
$x_n = \beta_n f(x_{n-1}) + (1 - \beta_n)x_{n-1}$
where $0 < \beta_n \leq 1$ and $\{\beta_n\}$ is convergent away from 0.
The sequence $\{x_n\}$ constructed this way is called a superior sequence of iterates, denoted by $SO(f, x_0, \beta_n)$ At $\beta_n = 1$, $SO(f, x_0, \beta_n)$ reduces to $O(f, x_0)$.

B. Definition

(Superior Orbit) The sequences x_n constructed above is called Mann sequence of iteration or superior sequences of iterates. We denote it by $SO(x_0, s, t)$ [11,12].

This procedure was essentially given by Mann, was the first to study it for β_n in 1955.

Since the results obtained in fractal modelling via Mann iterates are the super set of their corresponding fractal models in the Picard orbit. Researchers have since developed superior fractal models for $\beta_n = \beta$, $n = 1, 2,...$, for various values of β.

C. Definition

(Superior Julia sets) The set of complex points SK whose orbits are bounded under superior iteration of a function Q is called the filled superior Julia set. A superior Julia set SJ of Q is the boundary of the filled superior Julia set SK [8].

D. Definition

(Superior Mandelbrot sets) A Superior Mandelbrot set SM for a function of the form $Q_c(z) = z_n + c$, $n = 1, 2,...$, is defined as the collection of $c \in C$ for which the superior orbit of the point 0 is bounded,

SM=$\{c \in C: \{Q_c^k (0): k=0, 1,………\}$ is bounded in SO$\}$ [9].

E. Definition

Suppose x_0 is a fixed point for F. Then x_0 is an attracting fixed point if $|F'(x_0)|<1$. The point x_0 is a repelling fixed point if $|F'(x_0)|>1$. Finally if $|F'(x_0)|=1$, the fixed point is neutral [10].

III. GENERATING THE FRACTAL

The Mandelbrot set is the collection of C points for which the orbit is bounded. The set of those points are known as prisoner set and remaining points comes under escape set. The escape criteria for the function $sin (z^n) +c$ is given as follows:

A. Escape Criteria for Quadratic Function:

For $n = 2$, the escape criteria is depends on a constant value or ($z>=1/\omega$).

B. Escape Criteria for Cubic Polynomial:

The escape criteria for the cubic polynomials by using Mann iteration for n=3, the escape criteria is ($z>= ((1+b)/2 \omega)^{1/2}$).

C. Escape Criteria for General Polynomial:

The escape criteria for the general polynomial equations using Mann iteration procedure for n is $z \geq \left(\dfrac{1}{\omega}\right)^{\frac{1}{n-1}}$.

Note that the initial value z_0 should be infinity, since infinity is the critical point of z for $sin (z^n)+c$. However instead of starting with $z_0 = \infty$, it is simpler to start with $z_1 = c$, which yields the same result. A critical point of $z \rightarrow f(z)+c$ is a point where $f'(z)=0$. The point z in Mandelbrot set for sine function has an orbit that satisfies imag(z) > 50, then the orbit of z escapes [2].

IV. GEOMETRY OF SUPERIOR TRANSCENDENTAL MANDELBROT SETS AND SUPERIOR TRANSCENDENTAL JULIA SETS:

The fractal generated by this iteration process possesses symmetry about x axis in case of all polynomials.

A. Description of Superior Transcendental Mandelbrot Set:

The fractal is symmetrical about x axis for all values of s. Initially for a quadratic polynomial the value of s =1 has been taken and got the fractal with two bulbs. As the value of s has been changed from 1 to 0.3, both bulbs merged together and resultant fractal is not very much sharp. There are very few tiny bulbs attached to primary bulbs for s=0.5. Subsequently we change the value of s to 0.7; the primary bulb is showing too much tiny bulbs with attached decorations.

For a cubic polynomial, fractal shows its beautiful images. Starts with the value of s=0.1, and move towards 0.3, we got an image of lord Vishnu according to Hindu mythology. With the s= 0.5, there is an image of sparkling earthen lamp *(diya)* with its own image has been shown. There is a symmetrical image about x axis for s=0.7.

For a biquadrate polynomial, the fractal is having three primary bulbs and large number of secondary bulbs attached to it. For s=0.7, shape of primary bulbs are approximately same as with the value of s=1 but the number and shapes of secondary bulbs reduced. In case of s=0.3 and 0.5, the shapes of primary bulbs are same but the axis of secondary bulb varies based on above mentioned values of s.

B. Description of Superior Transcendental Julia Set:

Transcendental function sin(z) with superior Julia set, which follows the law of having 2n wings has analyzed, where n is the power of z. The images for all polynomials possesses symmetry about both x and y axis. Here extremely beautiful images of superior Julia set have been generated for different power of polynomial.

An image of superior Julia set has a resemblance with hippocampus for a quadratic polynomial**.** In case of cubic polynomial, a star fish shape has formed with large central body having rotational and reflection symmetry along with axes symmetry. Finally a biquadrate polynomial is having same structure as cubic polynomial with eight wings.

C. Generation of Superior Transcendental   Mandelbrot Sets
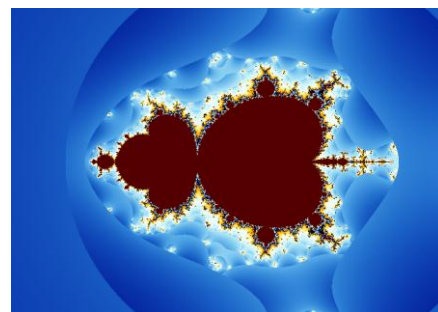
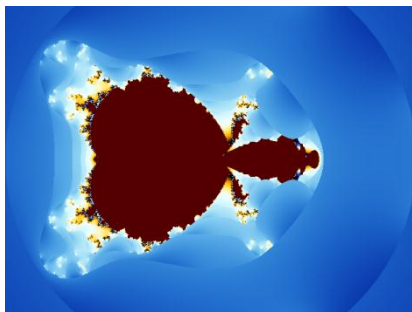For Quadratic function:

Figure 1.　　　F(z) = sin(z)^2+c & s=1
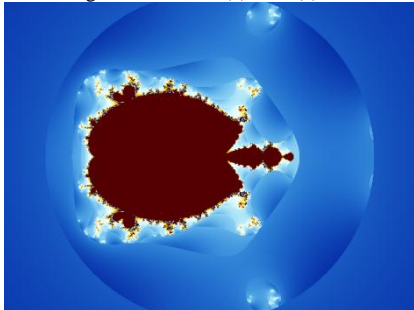
Figure 2.   F(z) = sin(z)^2 +c & s=0.3

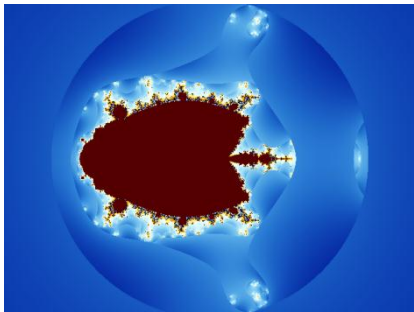Figure 3.   F(z) = sin(z)^2 +c & s=0.5

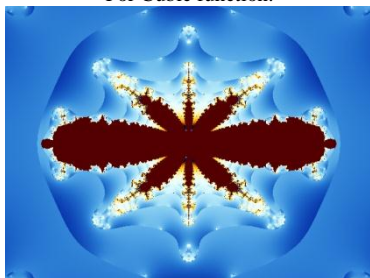Figure 4.   F(z) =sin(z)^2+c & s=0.7

For Cubic function:

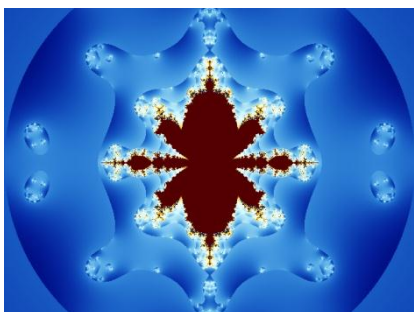Figure 5.   F(z) = sin(z)^3 +c & s=0.1

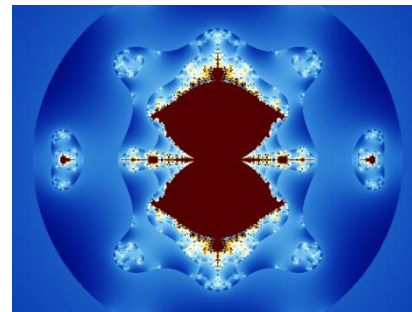Figure 6.   F(z) = sin(z)^3 +c & s=0.3

Figure 7.   F(z) = sin(z)^3 +c & s=0.5
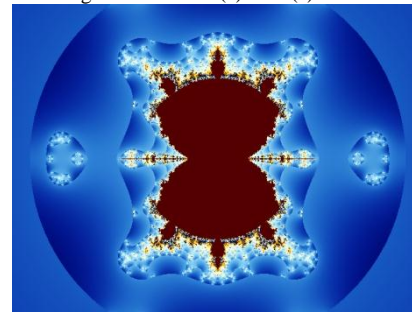

Figure 8.   F(z) = sin(z)^3 +c & s=0.7

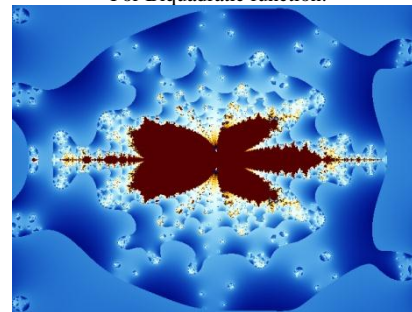For Biquadratic function:

Figure 9.   F(z) = sin(z)^4 +c & s=0.1

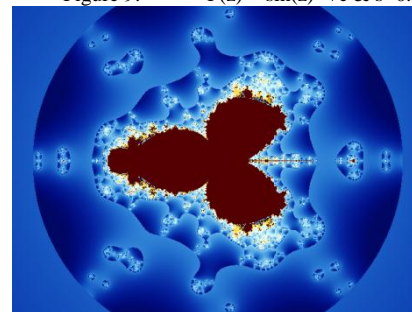
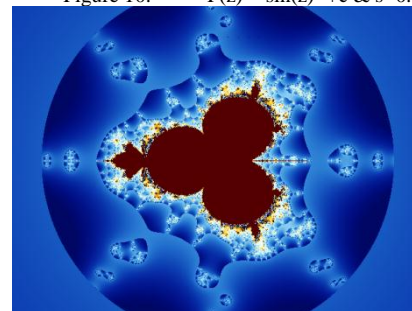Figure 10.   F(z) = sin(z)^4 +c & s=0.5

Figure 11.   F(z) = sin(z)^4 +c & s=0.7

D. Generation of Superior Transcendental Julia sets:

For Quadratic Function:

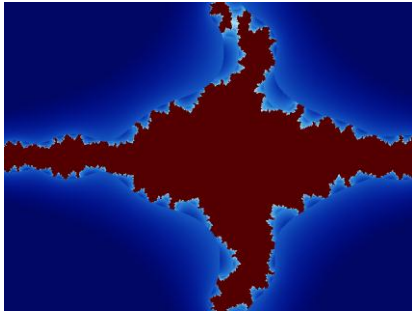

Figure 12. $F(z) = sin(z)^2 +c$,s=0.7 and C=0.01902, 0.17618

For Cubic Function:
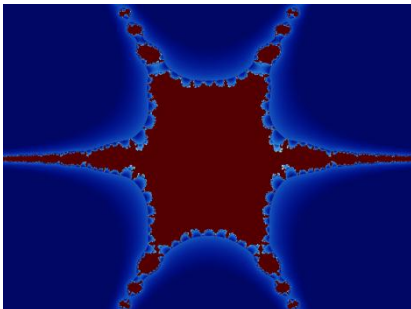
Figure 13.   For $F(z) = sin(z)^3 +c$, s=0.7 and C=0.77083,-0.83i

For Biquadratic function:
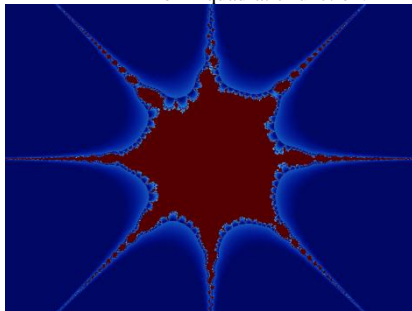
Figure 14.   For $F(z) = sin(z)^4 +c$, s=0.7 and C=-0.025, 0.3125

E. Fixed Points

Fixed points of Quadratic function:

TABLE I.   Orbit of F(z) at s=0.7,

(C=-0.7708333333-0.083333333333i)

Number of Iteration i	F(z)	Number of Iteration i	F(z)
1	0.5396	7	0.5080
2	0.4949	8	0.5080
3	0.5130	9	0.5080
4	0.5061	10	0.5080
5	0.5087	11	0.5080
6	0.5077	12	0.5080
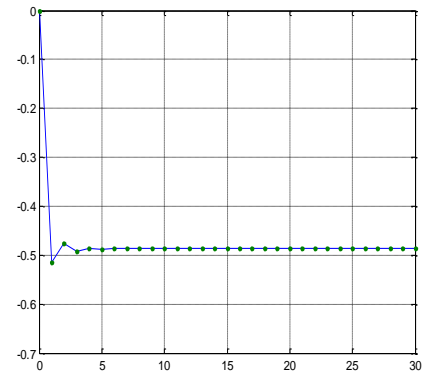
Figure 15.   Orbit of F(z) at s=0.7 for (C=-0.7708333333-0.0833333333333i)

Number of Iteration i	F(z)	Number of Iteration i	F(z)
1	0.0133	7	0.0176
2	0.0169	8	0.0176
3	0.0176	9	0.0176
4	0.0176	10	0.0176
5	0.0176	11	0.0176
6	0.0176	12	0.0176

TABLE II.   ORBIT OF F(Z) AT s=0.7 FOR (C=0.01902+ 0.17618)
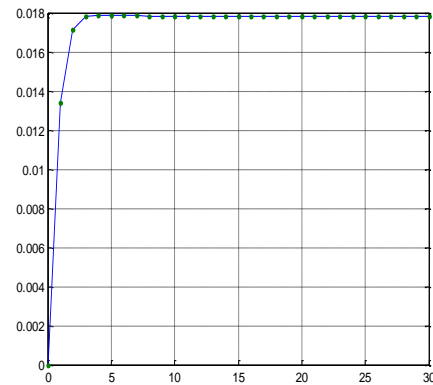
Figure 16.   Orbit of F(z) at s=0.7 for (C=0.01902, 0.17618)

Fixed points of Biquadratic function:

TABLE III.   Orbit of F(z) at s=0.7 for

(C=-0.025+0.3125)

Number of Iteration i	F(z)	Number of Iteration i	F(z)
1	0.0175	7	0.0156
2	0.0213	8	0.0155
3	0.0197	9	0.0154
4	0.0176	10	0.0154
5	0.0164	11	0.0154
6	0.0158	12	0.0154

Figure 17.   Orbit of F(z) at s=0.7 for (c=-0.025+0.3125)

V.   CONCLUSION

In this paper, we have analyzed sine function in Mandelbrot equation with Mann iteration. Superior Julia set possess 2n wings with central black region. Our study shown the striking properties and escape criteria for transcendental function and generated the corresponding fractals using Mann iterates in which most of the images are having symmetry along x axis and y axis. The images revealed their own i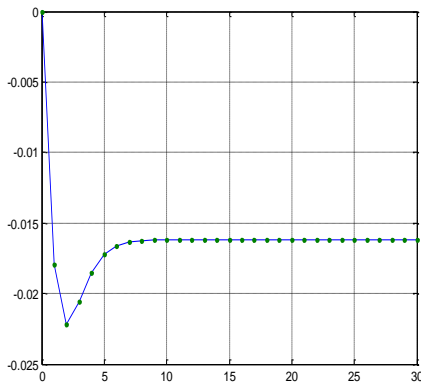dentity. As on a particular value of constant s, an image resembled to Lord Vishnu icon according to Hindu Mythology see Fig. [6]. Another image shows earthen lamp (*diya*) with its own reflection along with real axis see Fig. [7]. We obtained fixed point for quadratic function after 7 iterations, for cubic function after 3 iterations and for biquadratic function after 9 iterations.

The surrounding region of superior Mandelbrot set shown to be invariant cantor set in the form of curve or hair that tends to infinity under iteration in all figures.

REFERENCES

[1]   A. Ereneko, "Iteration of entire function" Dynamical System & Ergodic Theory, Banach Center publ. 23, polish Sc. Publ., Warsaw 1989, 339-345.

[2]   B. B. Mandelbrot, "The Fractal Geometry of Nature", W. H. Freeman, NewYork, 1983.

[3]   Folkert Tangerman, R. L. Devaney, "Dynamics of Entire Function near the Essential Singularity", Ergodic Theory & Dynamical System 64 (19860 489-503.

[4]   H.O. Peitgen and P.H. Richter, The beauty of fractals, Springer Verlag, 1986.

[5]   Kin-Keung Poon, "Fatuo-Julia Theory on Transcendental Semigroups", Bull Austrad. Math Soc. Vol. 58 (1998) [403-410].

[6]   Kin-Keung Poon, "Dynamics on Transcendental Semigroups", (1991).

[7]   M. Abbas, Sagar Hussain Khan, B.E. Rhoades, "Simpler is also better in approximating fixed points", applied Mathematics & Computation 205 (2008), 428-431.

[8]   Mamta Rani & Vinod Kumar, "Superior Julia Set", J. Korea Soc. Math. Educ. Ser. D; Res. Math. Educ., 8(4), 2004, 261-277.

[9]   Mamta Rani & Vinod Kumar, "Superior Mandelbrot Set", J. Korea Soc. Math. Educ. Ser. D; Res. Math. Educ., 8(4), 2004, 279-291.

[10] R.L Devaney, "A First Course in Chaotic Dynmical System: Theory & Experiment", Addision-Wesley, 1992. MR1202237.

[11] R. Rana, Y. S. Chauhan & A. Negi, "Generation of New Fractals for Sin Function", Int. J. Comp. Tech. Appl., vol2(6), 1747-1754 ISSN 2229-6093.

[12] R. Rana, Y. S. Chauhan & A. Negi, "Generation of New Julia sets and Mandelbrot sets for Tangent Function", Computer Engineering & Intelligent Systems, ISSN 2222-2863(online) vol2, No.8, 2011.

Creating a Complete Model of an Intrusion Detection System effective on the LAN

Yousef FARHAOUI

LabSiv, Equipe ESCAM
Faculty of sciences Ibn Zohr University B.P 8106, City
Dakhla, Agadir, Morocco.

Ahmed ASIMI

LabSiv, Equipe ESCAM
Faculty of sciences Ibn Zohr University B.P 8106, City
Dakhla, Agadir, Morocco.

Abstract— the Intrusion Detection Systems (IDS) are now an **essential component in the structure of network security. The logs of connections and network activity, with a large amount of information, can be used to detect intrusions.
Despite the development of new technologies of information and communication following the advent of the Internet and networks, computer security has become a major challenge, and works in this research are becoming more numerous. Various tools and mechanisms are developed to ensure a level of security to meet the demands of modern life. Among the systems is intrusion detection for identifying abnormal behavior or suspicious activities to undermine the legitimate operation of the system. The objective of this paper is the design and implementation of a comprehensive architecture of IDS in a network.**

Keywords—*Intrusion   Detection   System;   Characteristic; Architecture; modele.*

I.   INTRODUCTION

Today, information systems and computer networks are central in modern society. The more data stored and processed, it is more important to secure computer systems. An intrusion is defined as a series of actions that attempt to compromise the integrity, confidentiality or availability of a resource [1]. The intrusion detection Systems (IDS) can be hardware and software that automate the process of observation and analysis of events.

For an IDS to be effective it must run continuously adapt to behavioral changes and large amounts of data, be configurable, do not use too much memory resources of the machine and after system failures, be reusable without new learning [2].

Since their introduction, Cyber-attacks have been a real threat. With their wide variety and specialty, they can have catastrophic consequences. To prevent attacks or reduce their severity, many solutions exist, but no one can be considered satisfactory and complete. The intrusion detection systems are one of the -the most effective solution. Their role is to recognize intrusions or intrusion attempts by users or abnormal behavior by the recognition of an attack from the stream network data. Different methods and approaches have been adopted for the design of intrusion detection systems. An IDS is a tool that complements a wide range of users used to have some level of security. We present here the different architectures of IDS. We will also discuss measures that help to define the degree of effectiveness of IDS and finally the very recent work of standardization and homogenization of IDS.

II.   METHOD FOR DETECTING INTRUSION

Currently, there are two main approaches for intrusion detection. Anomaly detection and misuse detection (Misuse Detection). In the first approach, the normal behavior of network users are known and it is therefore possible to construct profiles representing these behaviors with several features such as network activity, etc.. Once these profiles defined, the intrusions are identified as deviations from normal behavior [1][3][4].

The approach detection of abuse (Misuse Detection) is based on the direct identification of attacks. A signature is a type of attack already known. Intrusion detection is done by comparison of network attacks with signatures [1][3][4].

The advantage of methods based on anomaly detection is the ability to find the unknown intrusions. Once all these methods produce a high rate of false alarms since all deviations from the general behaviors are not necessarily intrusions. For example, a new normal behavior can be seen as a deviation and treated as an intrusion.

On the other hand, in the case of misuse detection, each instance in the data set is labeled "normal" or "intrusion". A learning algorithm is applied to the data labialised, so that each intrusion is characterized as a model (intrusion signatures). We identify a new instance as an intrusion if it looks like an intrusion model. Models (signature) can be created by domain experts.

This is the case in systems based intrusion signatures (signature-database intrusion detection) [5]. These systems are effective   to   search   for   already   known   intrusions. However, these systems are not able to find new intrusions or intrusions for which signatures do not exist.

III.   IDS AND CLASSIFICATION

The IDS-based classification techniques are intended to classify network traffic into two classes: "general" and "intrusion". Classification requires learning.

The accuracy of this learning provides lower false positive rate (normal cases classified as intrusions) and false negative rate (intrusions classified as normal).

Measures used to compare and measure the effectiveness of IDS. The IDS are very important elements in a security strategy; why the choice of the IDS is very critical and must be based on its characteristics. Measures to better choose their

IDS. Donations [6] [7] [8] we can evaluate the IDS based on several criteria such as:

•The rate of false positive and false negative;

• Response by the IDS in an enveironnement overloaded;

• The ability to update the signature database or modify certain signatures;

• …

IV.   OUR INTRUSION DETECTION SYSTEMS MODEL

The study of intrusion detection systems has allowed us to realize the importance of the role of these to its own security policy. Different types of IDS (HIDS, NIDS), each characterized by a certain architecture and method of analysis. The characteristics of the IDS must meet certain requirements; the choice of adopting a certain type relative to another should be based primarily on the needs and constraints of security software and hardware. We can determine the type of IDS according to [7]:

• The location of the IDS (NIDS, HIDS);

• Frequency of use (continuous or periodic);

• The detection method (behavioral or scenario);

• The response of the IDS (passive or active).

In this paper we propose a new architecture for intrusion detection, to mix the two approaches: anomaly approach and misuse detection.

The choice of this approach is essentially based on the fact that the IDS are composed of different modules to be distributed on a set of network station to perform different tasks. The various components of the IDS must be in continuous interaction.

Our model consists of a primary IDS, its role is to organize tasks and manage the various second IDS, which have for role to capture events and the transmissions of the conclusions. The HIDS should be based on user profiles describing their normal behavior. This solution is very interesting since the only information required is the behavior of users in the network. This source of information can be kept updated only in learning phases. However, the disadvantage of this solution is the rate of false positives due to abnormal or unusual behavior of users, who are not necessarily harmful. The NIDS using the scenario approach (misuse detection) uses essentially a database of signatures of known attacks. This source of information allows us to significantly reduce the false positive rate. However, the disadvantage of this solution is the source of information that must be regularly updated. An attack not listed has no chance of being detected by the NIDS.

At the end to take advantage of both approaches (behavioral and scenario) that seem complementary, we chose the design of a hybrid IDS.

A.   The solution description

The core of our IDS generates variations of attack signatures and user profiles in a pseudo-random. This

methodology allows us to upgrade the analyzer to discover possible new attacks or variations of attacks.

B.   Overall architecture of IDS model

Our IDS is composed of (figure 1):

a. *NIDS generate detection based on signatures. These detectors will be used to analyze network traffic.*

b. *HIDS based* on the profiles of normal behavior of users. HIDS generate detectors able to recognize unusual behavior of users.

c. *Administrator* can configure the various parameters of IDS, see the different alerts, and run the learning command.

The components of our solution should be deployed on the output: the NIDS will be installed on the machine that is the proxy network in order to analyze network packets. The HIDS will be deployed on all the machines that consist of the local network.

The use of databases is very important in our model, we opted for the use of three databases:

a. Profiles database contains all information relating to user profiles. The data contained in this database are generated by the HIDS during the learning phase.

b. Database of signatures is the basis of NIDS. It includes all the known attacks by using a certain format. There is no standard for the coding of signatures. The attributes used to represent an attack must be based on the information contained in the packages [6].

c. Database alerts to list all alerts generated by the detectors of the two components of the IDS (HIDS and NIDS). This database will be accessed by the administrator to meet the traces of attacks or the anomalous behavior.

C.   The HIDS architecture

The first step in deploying HIDS is learning phase [8], during which we save the traces of the normal behavior of users by creating a profile for each.

Our HIDS will consist of a supervisor and a set of HIDS slaves to be deployed on all machines the network components.

a. *HIDS supervisor's role:*

• Extract user profiles database;

• Generate the sensors and send them to HIDS slaves;

• Analyze the relationship of slaves and directories HIDS and alerts in a database;

• Sends commands to start the learning phases, analysis, start and stop HIDS slaves.

b. *HIDS slave for:*

• Generate user profiles during the learning phase;

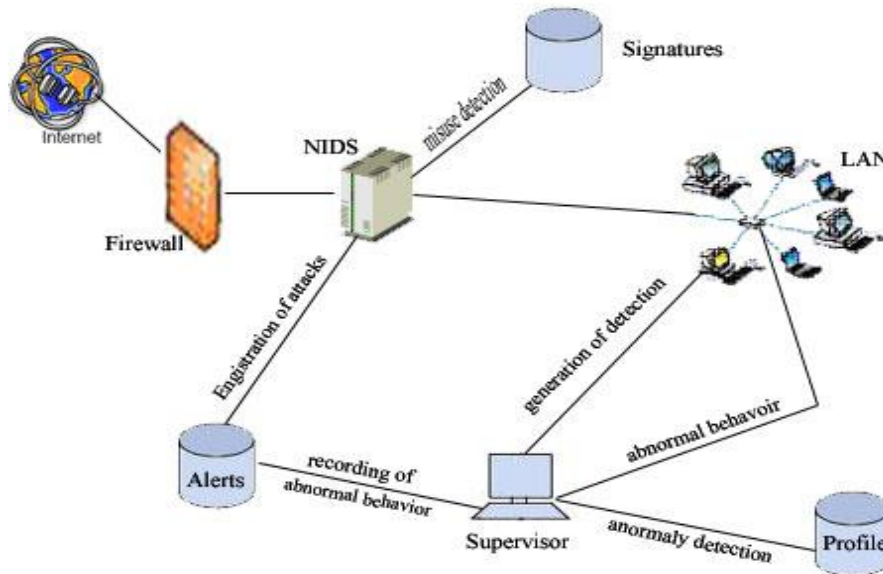• Use of event sensors to extract the current behavior of the user.

Figure 1. Overall scheme of the solution

D. NIDS architecture:

Using the analysis with the scenario approach; The analysis function of our NIDS contains two generation process sensors and their installation for the analysis of packet flows. The stages of execution are:

• Capture packets;

• Extraction and formatting attributes;

> Structuring the data;
> Summarize the data;
> Provide attributes.

• Analysis of attributes;

• Send of reports.

V. THE LIMITATIONS OF IDS

The limits apply to misuse detection techniques like those of anomaly detection. Attacks on the TCP flags. IDS are vulnerable to certain attacks on the TCP flags (TCP flags), such as:

- Sending a SYN false;
- Integration of data with bad sequence number;
- FIN / RST spoofing with wrong sequence number;
- Synchronization after connection;
- Desynchronization before connection [SYN (bad checksum + bad sequence number) and SYN];
- FIN / RST spoofing with bad checksum;
- Data spoofing with bad checksum;
- FIN / RST spoofing with short TTL;

Integration of data with a short TTL, etc.. Placement of the IDS. Without going into details, at the placement of the IDS (Design and implementation), it is interesting to make intrusion detection in the demilitarized zone (attacks against government systems), in network. Private (intrusions into or inside) and behind the firewall (detection of signs from all incoming and outgoing traffic). Each of these positions has its advantages and disadvantages.

The important thing is to identify resources to be protected (major business risks) and what is most likely to be attacked. It should then carefully implement the IDS (settings, etc..) Depending on the investment chosen. Pollution / overload. The IDS can be overloaded or contaminated, significant traffic (the most difficult and cumbersome to analyze possible). A significant amount of Mild attacks can also be sent in order to overload the IDS alerts. Possible consequence of this overload may be the saturation of resources (disk, CPU, memory), packet loss, denial of service or partial.

- Consumption of resources: in addition to the size of log files, intrusion detection is extremely resource-intensive;
- Packet Loss (performance limitation): the transmission rates are sometimes as far exceed the write speed of the fastest hard drives on the market, or even the processing speed of the processors. It is not uncommon for packets are not received by the IDS, and some of them are still received by the destination machine.
- Vulnerability to DoS: An attacker may attempt to cause a denial of service system-level intrusion detection, or at worst operating system of the machine supporting the IDS. Once disabled the IDS, the attacker can try all he likes. For example, the Stick attack is an attempt denial of service attack against the IDS overloading the IDS work at the point of disabling it or at least make it less effective.

Detection time: The detection time is a crucial element for IDS: Intrusion Detection is it done in real time or does it require a delay? What time (a few days ...)? Experience shows that it usually takes some time to identify or reconstruct an attack (analysis time, reaction ...).

Specific limits to the detection of abuse: The main challenges of this technique are as follows.

Definition and maintenance of signatures: All attacks are not detected, depending on the features of the system, the definition of signature, the update of the database, the system load, etc..:

- Limits "human" signatures outdated or poorly designed. Detection of abuse has good design imperatives for attack signatures and a continuous updating of the list of signatures.
- Context of use: sometimes the technology is based on signatures that are not based on the context of use. The result is twofold: many false positifs19 and significant degradation of system performance.
- Even if the method signatures of the body seem to be quite reliable, there are ways to circumvent them.
- Vulnerability to changes: due to its lack of flexibility, detection attack signatures are very vulnerable to mutations. First, in order to define a signature, you must have already faced the attack considered.

On the other hand, some of these signatures are based on characteristics "volatile" a tool, such as wearing a Trojan opens by default or the value of ISN chosen by some hacking tools. But these programs are often either highly configurable, open source is so readily modified. The characteristics used to define the signature are fragile and highly sensitive to changes signatures.

- Lack of definition, new attacks are the IDS without being detected. False positives. Normally, the advantage of detection of abuse should be a low rate of false positives (false alarms) as the criteria of signatures can be precisely defined. Nevertheless, according to sources of information, we read that there is little to a lot of false positives resulting from this technique, particularly regarding:
- The sensitivity / specificity of the IDS: by nature, IDS alerts will go up enormously if they are not configured properly. Full attention must be paid to the establishment of signature rules. The compromise made between the amount of recovery alerts and finesse of the latter is crucial. We must take care to include in the configuration file the file. "Rule" necessary, according to rules established by the firewall. For example, if a service is totally forbidden, it is almost unnecessary to include the signatures associated.

Specific limits on anomaly detection. This technique also involves many complex problems to be solved here is the most commonly mentioned.

Learning / configuration of the IDS: Learning the "normal" behavior is not easy. Automate the reasoning used to think that behavior is "deviant" in relation to that known is a difficult task. By cons, this technique is applied by default by the system or network administrators: If something seems unusual (peak bandwidths, services that fall, file systems that fill faster than usual), the practice is that further research be undertaken. In addition, any abnormality does not necessarily correspond to an attack; it may be a change in user behavior or a change in network configuration. Typically, convergence to a behavioral model "normal" is quite long.

When setting up the IDS, the challenge for the effective detection lies in the choice of metrics, models and in defining the different profiles. For all these reasons, the IDS running through anomaly detection are known to be very long and tedious to configure. Even after an effective configuration, nothing prevents an attacker from knowing guarded "reeducate" such a system by changing its model of progressive convergence towards an abnormal behavior for the analyst, but to actually "normal" to a statistical point of view. False positives. Anomaly detection can detect unknown attacks, but it is not as effective as misuse detection for known attacks. Particular, a high rate of false positives can be met if the setting of the IDS was not carried out carefully

VI. CONCLUSION

In general, the effectiveness of intrusion detection system depends on its "Configurability" (Ability to define and add new specifications attack), robustness (fault tolerance) and the small amount of false positives (false alarms) and false negatives (undetected attacks) it generates. The foregoing paragraphs are aimed both to illustrate the sophistication of today's attacks, to show the complexity of intrusion detection and explain the limitations of current IDS. A struggle between intrusion techniques and IDS is committed, the IDS resulting in a more technical nature of the attacks on IP, and the current attacks requiring IDS to be more complete and powerful. To conclude this article, IDS provide a definite plus for networks in which they are placed. However, their limitations do not guarantee 100% security, unobtainable. You must then tender ... The future of these tools will help fill these gaps by avoiding "false positives" (for IDS) and refining the access restrictions (for IPS)

REFERENCES

[1] Abhinav Srivastava, Shamik Sural, and Arun K. Majumdar. Weighted intratransactional rule mining for database intrusion detection. In PAKDD, pages 611-620, 2006.

[2] J. S. Balasubramaniyan, J. O. Garcia-Fernandez, D. Isaco_, Eugene H. Spa_ord, and Diego Zamboni. An architecture for intrusion detection using autonomous agents. In ACSAC, pages 13_24, 1998.

[3] Wenke Lee, Salvatore J. Stolfo, and Kui W. Mok. Mining audit data to build intrusion detection models. In Knowledge Discovery and Data Mining, pages 66_72, 1998.

[4] Karen Scarfone and Peter Mell. Guide to intrusion detection and prevention systems (idps). Technical Report SP800-94, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, U.S. Department of Commerce, February 2007.

[5] Paul Dokas, Levent Ertoz, Vipin Kumar, Aleksandar Lazarevic, Jaideep Srivastava, and Pang-Ning Tan. Data mining for network intrusion detection. University of Minnesota, Minneapolis, MN 55455, USA, 2002.

[6] Y. Farhaoui, A. Asimi, "Performance method of assessment of the intrusion detection and prevention systems," *IJ E ST , Vol. 3 No. 7 July 2011*

[7] Y. Farhaoui, A. Asimi, «Performance Assessment of the intrusion Detection and Prevention Systems: According to their features: the method of analysis, reliability, reactivity, facility, adaptability and performance», The 6th IEEE international conference Sciences of Electronics Technologies Information and Telecommunication (SETIT 2011), Sousse, Tunisia, 2011.

[8] Y. Farhaoui, A. Asimi, "Performance Assessment of tools of the intrusion Detection and Prevention Systems," IJCSIS , Vol. 10 No. 1 January 2012

Rotation-Invariant Pattern Recognition Approach Using Extracted Descriptive Symmetrical Patterns

Rehab F. Abdel-Kader[1], Rabab M. Ramadan[2] , Fayez W. Zaki[3] , and Emad El-Sayed[1]

1 Faculty of Engineering, Port-Said University, Port-Said, Egypt
[2] College of Computers & Information Technology, University of Tabuk ,Tabuk, KSA
[3] Faculty of Engineering, Mansoura Universit, Mansoura, Egypt

Abstract— **In this paper a novel rotation-invariant neural-based pattern recognition system is proposed. The system incorporates a new image preprocessing technique to extract rotation-invariant descriptive patterns from the shapes. The proposed system applies a three phase algorithm on the shape image to extract the rotation-invariant pattern. First, the orientation angle of the shape is calculated using a newly developed shape orientation technique. The technique is effective, computationally inexpensive and can be applied to shapes with several non-equally separated axes of symmetry. A simple method to calculate the average angle of the shape's axes of symmetry is defined. In this technique, only the first moment of inertia is considered to reduce the computational cost. In the second phase, the image is rotated using a simple rotation technique to adapt its orientation angle to any specific reference angle. Finally in the third phase, the image preprocessor creates a symmetrical pattern about the axis with the calculated orientation angle and the perpendicular axis on it. Performing this operation in both the neural network training and application phases, ensures that the test rotated patterns will enter the network in the same position as in the training. Three different approaches were used to create the symmetrical patterns from the shapes. Experimental results indicate that the proposed approach is very effective and provide a recognition rate up to 99.5%.**

Keywords- Rotation-Invariant; Pattern Recognition; Edge Detection; Shape Orientation; Image processing.

I. INTRODUCTION

Pattern recognition is one of the most intriguing and active research topics in the field of image processing. Its importance is due to its pervasive presence and influence in a large number of computer vision applications. In spite of the easiness of the recognition of different patterns by the human eyes, it remains challenging to implement automated pattern recognition technique that can be efficiently applied to various shapes. Research effort in this field tend to build robust techniques of pattern recognition which are insensitive to different kinds of transformations, including rotation, translation, and scale. A wide variety of techniques have been proposed to deal with specific or general instances of this problem [1]. Some techniques are based on integral transforms as invariant feature extractors. These transforms are based largely on Fourier analysis, and are probably the most common tools used in invariant pattern recognition [2-7]. Other techniques use moment functions as invariant features by taking quotients and powers of moments such as regular

moments, Zernike moments, the generalized moments and Flusser and Suk's moment for affine transformation invariance [8-13]. Another group of techniques incorporates different designs of neural networks in the invariance problem which rely on the presence of symmetries among the network connections. This includes weight sharing neural network, first-order networks and higher-order neural networks [14-16]. Additional techniques that do not fit easily into any of the above categories include image normalization as in the preprocessing normalization system of Yuceer and Oflazer [17] and the symmetric filter systems [18-20].

In this paper, the proposed system is based on an image preprocessing technique which uses a newly developed efficient shape orientation method to normalize the image rotation by rotating it to a specific rotation angle. Then, the proposed image preprocessor generates a rotation-invariant descriptive pattern from the shape to be used in the training and application phases of the neural network.

II. SHAPE ORIENTATION

Shape orientation has emerged as an important task widely used in the area of image processing. It is very useful in a number of computer vision applications such as robot manipulation and image recognition and image registration. It is usually used as an initial step in many image processing operations. Due to the importance of the shape orientation, several techniques [21–28] are proposed to solve this problem based on different concepts like geometric moments, complex moments and principal component analysis. Some of them are area based which takes into account all points that belong to the shape. Others use only the boundary points. Although it is usually easy to determine shape orientation by the human eyes as shown in Fig. 1, it remains challenging to implement a technique which can be efficiently applied to all shapes because of the variety of shapes and their applications. Each concept used has its strengths and weaknesses relative to the processed image and the application domain. Some techniques have problems in processing symmetric shapes that have several axes of symmetry. For example the most standard technique among the area based techniques determines the shape orientation by finding the axis of the least second moment of the shape. The output axis which minimizes the integral of the squared distances from the shape points [29-31] cannot find the shape orientation of M-fold (M>2) rotationally symmetric shapes. The same problem appears in the boundary

based shape orientation techniques based on either the convex hull points only of the considered shape [32, 33], or the complete boundary points like [34].
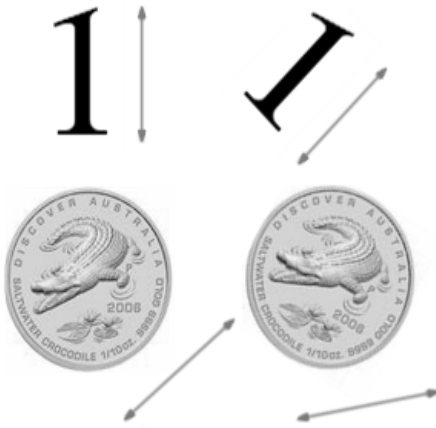

Figure 1.   The shape orientation by human eye

To overcome this problem some techniques use complex or higher order moments were reported [27, 28], however they are usually computationally expensive. Alternative techniques tend to find the orientation of shapes that have equally separated multiple axes depending on the fact that in digitized images those shapes are not perfectly symmetric. This may not be practically useful due to the fact that after rotating the image, the error of the digitization rounding process may change the traced orientation of the original image. For these reasons, in the technique reported here it is concentrated on practical orientable shapes that have a unique determinable orientation. The proposed technique aims to reduce the computational cost by avoiding the use of higher order moments in cases of multiple principal axes shapes.

It is necessary for multiple principal axes shapes, to be orientable, to have non-equal angles' differences between their axes. This feature is used to define a simple method to get average of the axes angles.

III.   THE SHAPE ORIENTATION ALGORITHM

The algorithm comprises four main steps:

1)   Apply Canny edge detection [35] to the gray-level image, and obtain a binary edge map.

2)   Find the center point of the pattern object (x_c, y_c) by calculating the average of the vertical and horizontal coordinates of the points of the edge map.

3)   Find axes of the shape which are the minimum peaks of the absolute summation of perpendicular distances from edge map points to the line passes through center :

a)   The equation of the straight line that passes through the claculated center point (x_c, y_c) with angle θ is given as:

$$ax+by+c=0. \tag{1}$$

Where

$$\begin{cases} a = tan\theta \\ b = -1 \qquad\qquad\quad if\ \theta\ != 90 \\ c = y_c - x_c * tan\theta \\ \\ a = 1 \\ b = 0 \qquad\qquad\qquad if\ \theta\ = 90 \\ c = -x_c \end{cases} \tag{2}$$

b)   The perpendicular distance d_i from any point on the edge map (x_i, y_i) to the predefined straight line is expressed as

$$d_i = \frac{|ax_i + by_i + c|}{\sqrt{a^2 + b^2}} \tag{3}$$

c)   So for each angle θ_j in the range $(0 \le \theta_j < \pi)$ stepped by any suitable step the summation of the perpendicular distances from the edge map points to the predefined line is given as

$$D_j = \sum_{i=1}^{n} \frac{|a_j x_i + b_j y_i + c_j|}{\sqrt{a_j^2 + b_j^2}} \tag{4}$$

Where n is number of edge points and the points are indexed by i= 1, … n and k is the number of the stepped angles straight lines and the lines are indexed by j = 1, … k .

d)   Using resulted curve one can find the shape's axes and their angles. Where the axis is the line with angle θ_j which has one of the minimum peaks of summation D_j from (4).

4)   If the shape has more than one axes the averaging method could be used to find the shape orientation.

The drawn lines in Fig. 2 show the detected axes by applying the proposed technique on "3" as pattern in different positions as an example of single axis shapes.
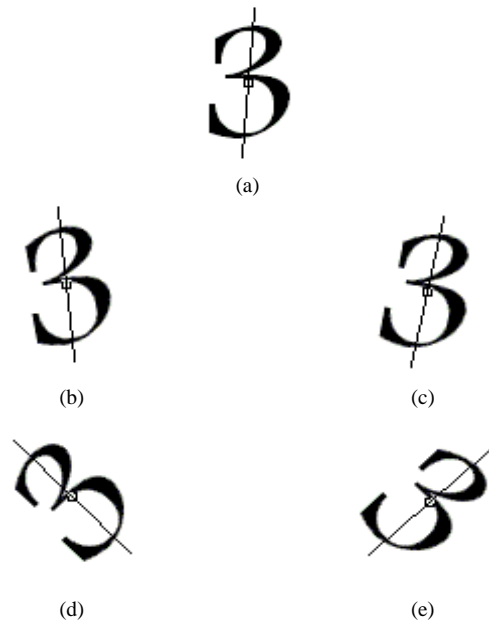

Figure 2.   Samples of dominant direction angle detection

(a) Original pattern without rotation (b) After 10˚ rotation (c) After -10˚ rotation (d) After 45˚ rotation (e) After -45˚ rotation

Fig. 3 shows the summation of the perpendicular distances from the edge map D_j (on the vertical axis) to the line passes through center point with angle θ_j (on the horizontal axis).

As shown in Fig. 3 the orientation angle is the corresponding angle of the single minimum peak value of the curve. It is clear that the phase shift between the original image curve and the rotated image curve is equal to the rotation angle. In Fig. 3(a) the original detected angle is 86˚ and the detected angles after ±10˚ rotation are 76 and 96 with phase shift ±10˚ and also the same in Fig. 3(b). But the rotation angle and the phase shift are ±45˚.

The following points should be taken into account:

- The orientation angle from any image is equal to the output angle after 180˚ rotation because there is no difference between the orientation directions in both cases as shown in Fig. 4(a). It is clear in Fig. 5(a) that the pattern before and after 180˚ rotation has the same overlapped curves which results the same minimum values and output angle.
  The same concept is clear in Fig. 4(b) which shows the two positions of the pattern after ±90˚ rotation and in their overlapped curves in Fig. 5(b) because the angle difference between both of them is 180˚.
  From the above section it is clear that the output angle will repeat itself every 180˚.

Figure 4. The repeated dominant direction every 180˚ (a) The pattern before and after 180˚ rotation (b) The pattern after ±90˚ rotation

(a)

(b)

Figure 5. The overlapped curves of the pattern due to 180˚ angle difference
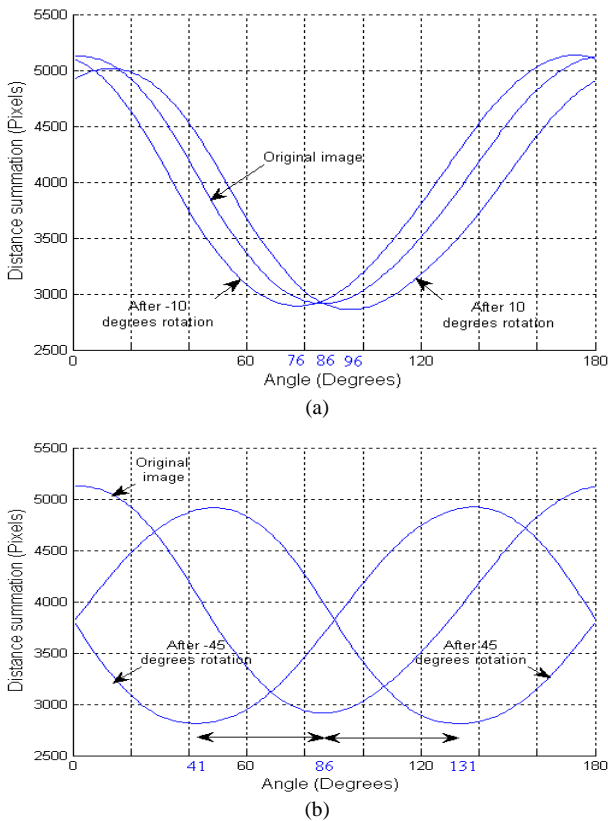
(a) Before and after 180˚ rotation (b) After ±90˚ rotation

(a)

(b)

Figure 3. Relation between distance summation D_j and the angles θ_j (a) Curve of ±10˚ rotated pattern (b) Curve of ±45˚ rotated pattern

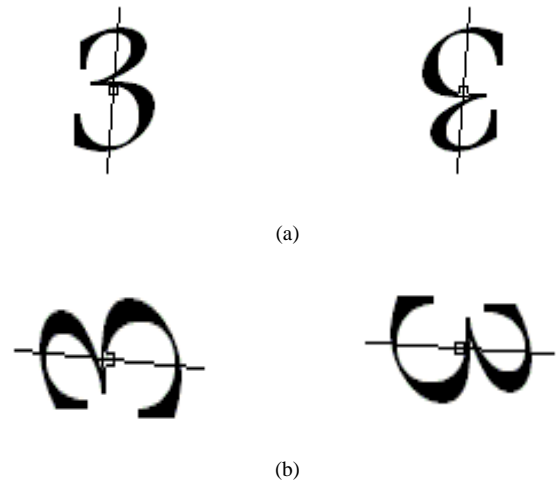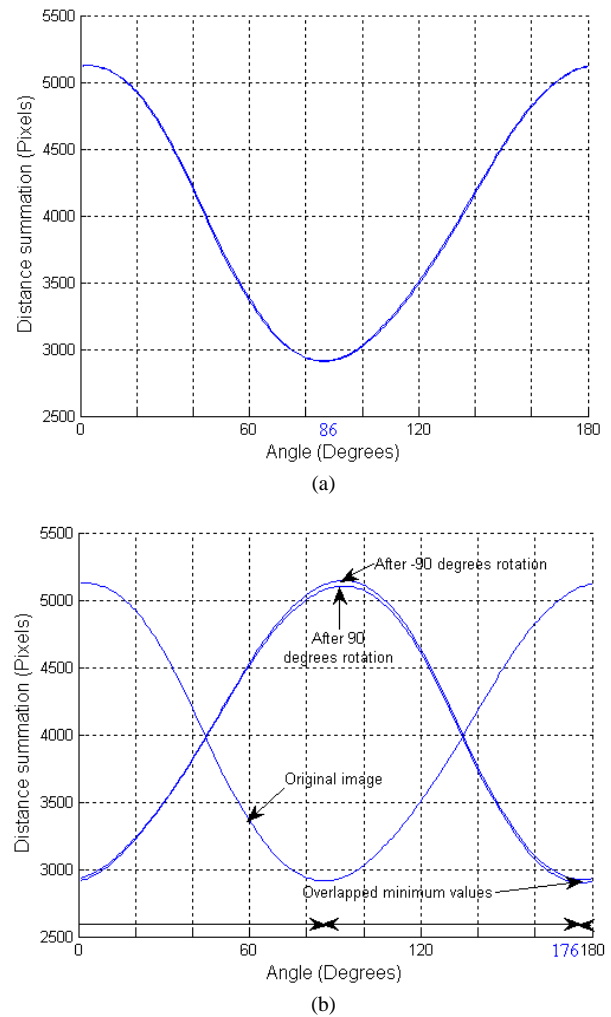- In some cases, the curve may have some fake minimum peaks as a result of any successive decreases and increases in value by error.

  To overcome this problem it is required to get only the true minimum peaks as follows:

  a) Calculate the numerical differential of the curve in order to find the global minimum candidates which are correspond to the points on the differential curve that switch from negative to positive.

  b) To avoid fake candidates we find the depth of each candidate peak as the number of pixels from the candidate peak point to the nearest maximum peak. Then the depths of the found peaks are compared neglecting the candidates with depth less than a factor of the average depth.

IV. THE AVERAGING METHOD

In all previous curves of "3" pattern there is only one minimum peak so it is sufficient to depend on finding the global minimum value of the curve because "3" has only one axis, in some other cases the shape may have multiple axes, so the pattern curve may have two or more minimum peaks as shown in Fig. 6(b) which shows the curve of "X" pattern. This happens because the pattern has two symmetric directions to the center point.

In similar cases, previously reported techniques use higher orders moments which are computationally expensive. In the proposed approach a simple method is adopted to find the average of the axes' angles. The difficulty in using the normal average method is that each axis angle has two probable values θ and $\theta+180$. This is clear from the overlapped curves in Fig. 5. This may cause an error in the calculated average of the rotated shape due to the fact that our detected angles are always between 0° and 180° which means that the detected rotated angle returns to 0° after passing 180°. This may cause a wrong average value if one or more of the angles change to the other direction after rotation, while the other angles do not.

For example the "X" pattern in Fig. 6 has two candidate angles 66° and 114° whose average is 90° so after 100° rotation it is expected that the average will be 190° or 10° (e.g. in the range given above) but the pattern will be as shown in Fig. 7, and the new two candidate angles given as 166° and 34° whose average is 100° which is not as expected in the above argument. This happens because our expected average needs to use the angle 214° instead of 34° but our technique always detects the angles between 0° and 180°.

For any number m of angles there are m different ranges that contain all angles. Each range starts with one of the angles and ends with the former one. It is clear that for the non-equally separated angles there is always a range that is narrower than the others. In our method this feature is utilized to solve the drawback of the normal averaging method. Because the differences between the angles remain the same after rotation, the narrowest range will remain the same. So to find the average we find the narrowest range that contains all angles and calculate the average of this range.

In our previous example before rotation the detected angles are 66° and 114° with two ranges 48 and 132, so by using the narrowest range the average will be 90°. After rotation the detected angles will be 34° and 166° with two ranges 132 and 48, so the angles of the narrowest range will be 166° and 214° with average 190° or 10° as shown in Fig. 7.

(a)

(b)
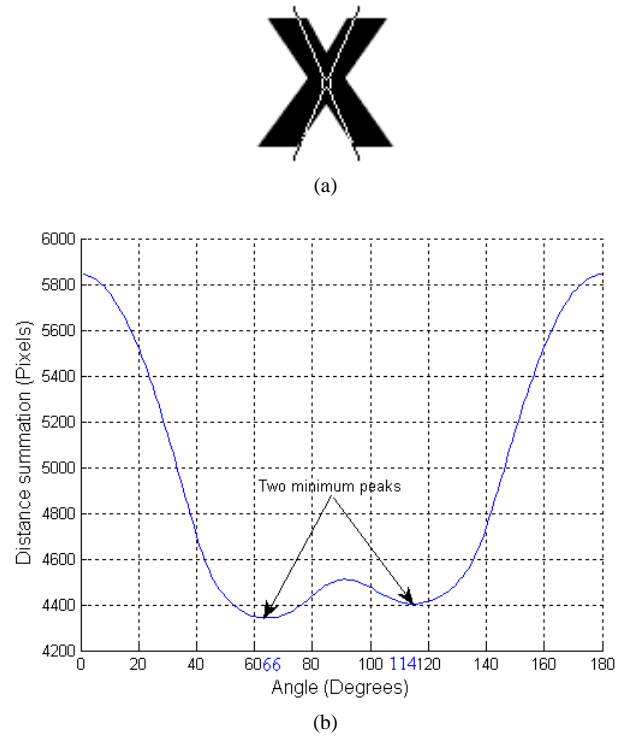
Figure 6.   "X" as a sample of multiple symmetric directions (a) The two symmetric directions (b) The curve with two minimum peaks
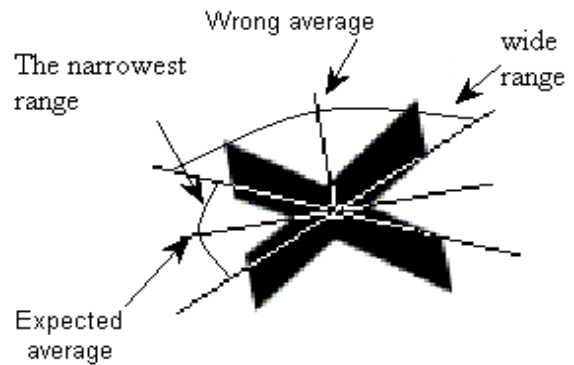
Figure 7.   The average calculation method

V. SHAPE ORIENTATION SIMULATION RESULTS

To test the shape orientation technique introduced here, more than 150 different patterns such as characters, numbers, coins and different shapes with single or multiple axes were used. The patterns in the database were rotated in all positions stepped by 5. Moreover, the patterns have different sizes and

different formats and use 256 gray levels for some patterns and binary images for the others. Fig. 8 shows some of the resulted orientation shapes from the proposed technique. From these results, it can be seen that the proposed technique provide accurate results at low computational cost.

Figure 8.   The detected orientations of different shapes

VI.   PATTERN EXTRACTION

The main goal of the extraction process is to create a rotation-invariant pattern from the shape which could be used in neural network training and testing to eliminate any shape rotation effect. The extracted pattern from the shape should be descriptive as much as possible, so that different shapes should have different extracted patterns to allow one to use those patterns to differentiate between the shapes.

The pattern extraction process begins by applying the shape orientation technique introduced here on the image to calculate the orientation angle of the shape. Then, the system creates a symmetrical pattern about the axis with the calculated orientation angle and the perpendicular axis on it.

The axis of symmetry of any shape should have two identical opposite sides. As the created pattern has two orthogonal axes of symmetry, it should have four identical opposite sides about the center of the shape. And so for any point in the shape to be considered as a symmetrical point, there should exist two another points lying at the same perpendicular distance from this point to the orthogonal axes but on the other sides of the axes.

We used three different approaches to create the symmetrical patterns from the shapes. The first approach is to keep only the symmetrical points about both of the orthogonal axes and eliminate the other points. By using this approach, the extracted pattern is the result of subtracting the nonsymmetrical points from the shape. Fig. 9(c) shows some samples of the extracted patterns using this approach.

Unlike the first approach, the second approach is to complement the nonsymmetrical points by adding their corresponding reflected points about the axes to the shape. In this way the nonsymmetrical points are converted into symmetrical points instead of eliminating them. Fig. 9(d) shows some samples of the extracted patterns using the second approach.

Although, the extracted patterns in the previous two approaches were based on the input shapes; there is still possibility to find two different shapes having the same extracted pattern. In order to reduce this possibility and to make the extracted pattern more descriptive, a hybrid method (third approach) using the extracted descriptive information from the previous two approaches. That is, the extracted pattern by the third approach is the result of subtracting the first method symmetrical pattern from the second method complemented pattern. In this way the possibility of finding two different shapes having the same extracted pattern is reduced because it is difficult for the different shapes to have the same subtracted and complemented patterns in the same time. It is clear from Fig. 9(e) that the extracted pattern using the third approach is the result of subtracting the pattern in Fig. 9(c) from the pattern in Fig. 9(d).

Due to the symmetry of the previous extracted patterns about their axes, the extracted patterns of any two opposite positions of any shape are the same. The importance of this feature is that we can't cancel the effect of shape rotation using only the shape orientation to rotate the shape to any reference angle because the orientation angle of any shape will be the same after 180˚ shape rotation because there is no difference between the orientation directions in both cases as shown in Fig. 4. Therefore, one can't ensure that the resulted rotated shape would have the required reference angle as it could be in its opposite direction. But extracting a symmetrical pattern after rotating the shape to a reference angle will have the same pattern in both cases as shown Fig. 10 which could be used in both of the training and test phases of the neural network. In this way one may ensure that the same pattern would be applied to the network in both phases whatever the rotation angle of the shape is.

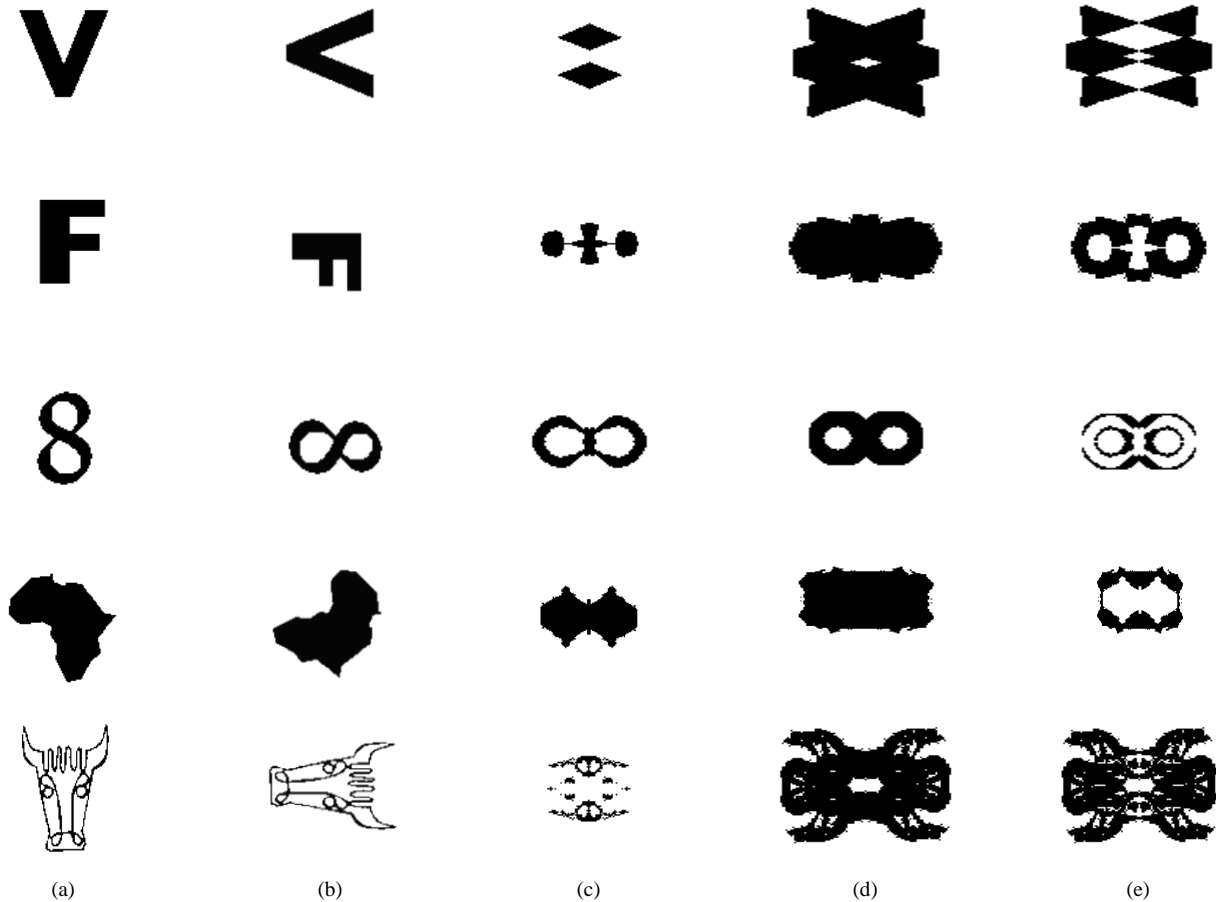|				
(a)	(b)	(c)	(d)	(e)

Figure 9.   some samples of the extracted patterns using the three different approaches (a) The input shape (b) The shape after being rotated to have 0 orientation angle (c) The extracted pattern using the first approach (d) The extracted pattern using the second approach (e) The extracted pattern using the third approach
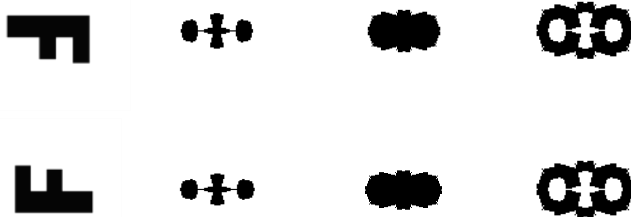

Figure 10. The three extracted patterns of the shape of 'F' in the 0° orientation angle position and in 180° orientation angle position

VII.   ROTAION INVARIANT NEURAL PATTERN RECOGNITION SYSTEM

Fig. 11 shows the different stages of the rotation invariant neural-based pattern recognition system. First, the proposed shape orientation technique is applied on the training image to get its orientation angle θ. Second, the image is rotated to adapt its orientation angle to any fixed reference angle. In our implementation, the system rotates the image by $(180 - \theta)°$ to ensure that the new orientation angle will be 0° or 180°. Third, the system extracts the symmetrical pattern of the shape using one of the three proposed approaches.

Fourth, a feature extraction technique is applied on the pattern to get the input vector of the shape to be applied to the back propagation neural network. In the simulation, the images are down-sampled to a suitable size and then PCA feature selection technique is applied to reduce input vector dimensions while preserving most of the image details.

VIII.   THE SYSTEM EXPERIMENTAL RESULTS

The performance of the system is evaluated using 60 different shapes. An image database containing rotated shapes with different rotation angles from 0˚ to 355˚ stepped by 5˚ was utilized in the experiments. Two different testing models were applied on the system, in the first model only the pattern extracted from the original position of the shape without rotation was used in the training phase.

Table 1 presents the average recognition rates of the system using the different proposed symmetrical patterns with only one pattern per shape used in the training.
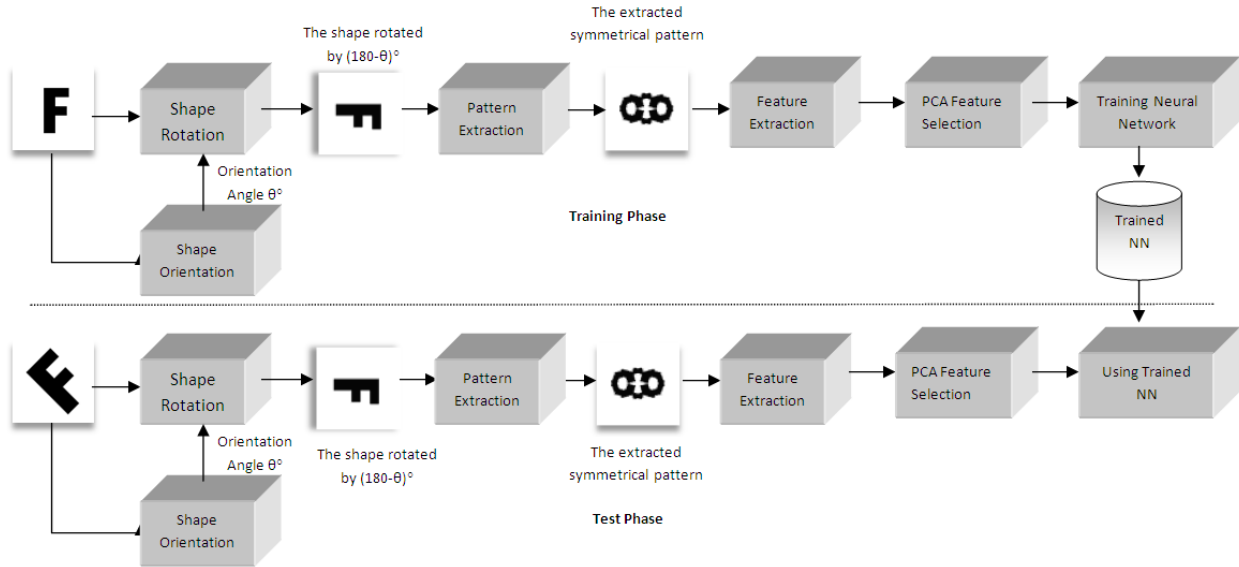
Figure 11. Block diagram of the rotation invariant neural-based pattern recognition system.

TABLE I. PERFORMANCE OF THE PROPSED SYSTEM USING DIFFERENT PATTERN EXTRACTION APPROACHES WITH THE FIRST TESTING MODEL

Pattern Extraction Approach	Recognition Rate
First approach (using only symmetrical points of the shape)	91%
Second approach (complementing nonsymmetrical points of the shape)	89%
Third approach (the hybrid method)	94%

In the second model the neural network was trained using four patterns for every shape extracted at different rotation angles. Table 2 shows the results of the second testing model with high improvement in the performance.

TABLE II. PERFORMANCE OF THE PROPSED SYSTEM USING DIFFERENT PATTERN EXTRACTION APPROACHES WITH THE SECOND TESTING MODEL

Pattern Extraction Approach	Recognition Rate
First approach (using only symmetrical points of the shape)	95%
Second approach (complementing nonsymmetrical points of the shape)	92%
Third approach (the hybrid method)	99.5%

As expected, the third approach has the best recognition rate because it uses the extracted information of both of the other two approaches which makes its extracted patterns much more descriptive and reduces the possibility of having the same pattern from different shapes.

IX. CONCLUSIONS

In this paper, a rotation-invariant neural pattern recognition system based on an image preprocessing technique to extract a rotation-invariant descriptive pattern from the shapes is presented. The system utilizes a newly proposed algorithm for finding the orientation of shapes that have several non-equally separated axes of symmetry. An effective averaging method is presented to calculate the average angle of the shape's axes of symmetry without any need to use higher orders techniques to reduce the computational cost. The orientation angle is used to normalize the shape image and eliminate any rotation effect before the feature extraction phase. Furthermore, three different approaches to extract the symmetrical patterns are proposed.

REFERENCES

[1] Wood, J., Invariant pattern recognition: A review, Pattern Recognition 29 (1996) 1–17.

[2] Rehab F. Abdel-Kader, Rabab M. Ramadan, Rawya Y. Rizk, Rotation Invariant Face Recognition Based on Hybrid LPT/DCT Features, International Journal of Electrical and Computer Engineering 7 (2008) 488.

[3]. V. Fayn, V. Sorokin and V. Vaynshteyn, Continuousgroup, pattern recognition, Engng Cybern. 6(1969) 97-106.

[4] T. M. Caelli and Zhi-Qiang Liu, On the minimum number of templates required for shift, rotation and size invariant pattern recognition, Pattern Recognition 21(1988) 205-216.

[5] R. N. Bracewell, The Fourier Transform and Its Applications, Electrical and Electronic Engineering Series, McGraw-Hill Book Company (1978).

[6] E. Brigham, The Fast Fourier Transform and Its Applications. Prentice-Hall (1988).

[7] M. Clausen and U. Baum, Fast Fourier Transforms, Wissenschaftsverlag, Mannheim (1993).

[8] Y. Li, Reforming the theory of invariant moments for pattern recognition, Pattern Recognition, 25(1992) 723-730.

[9] S. Perantonis and P. Lisboa, Translation, rotation and scale invariant pattern recognition by high-order neural networks and moment classifiers, IEEE Trans. Neural Networks, 3(1992) 241-251.

[10] A. Khotanzad and J.-H. Lu, Classification of invariant image representations using a neural network, IEEE Trans. Acoustics, Speech and Signal Process, 38(1990) 1028- 1038.

[11] A. Khotanzad and Y. H. Hong, Invariant image recognition by Zernike moments, IEEE Trans. Pattern Analysis Mach. Intell, 12(1990) 489-49.

[12] A. Shvedov, A. Schmidt and V. Yakubovich, Invariant systems of features in pattern recognition, Automation Remote Control, 40(1979) 131- 142.

[13] J. Flusser and T. Suk, Pattern recognition by affine moment invariants, Pattern Recognition, 26(1993) 167-174.

[14] L. Fausett, Fundamentals of Neural Networks, Prentice-Hall, Englewood Cliffs, 1994.

[15] J. Hertz, A. Krogh and R. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley (1991).

[16] M. Minsky and S. Papert, Perceptrons. MIT Press, Cambridge, (1969).

[17] C. Yuccer and K. Oflazer, A rotation, scaling and translation invariant pattern classification system, Pattern Recognition, 26(1993) 687-710.

[18] R. Lcnz, Group invariant pattern recognition, Pattern Recognition 23(1990) 199-217.

[19] R. Lenz, Group Theoretical Methods in Image Processing, volume 413 of Lecture Notes in Computer Science, Springer-Verlag (1990).

[20] J. Mas and E. Ramos, Symmetry processing in neural network models, J. Physics A 22(1989) 3379-3391.

[21] J. Cortadellas, J. Amat, F. de la Torre, Robust normalization of silhouettes for recognition application, Pattern Recognition Lett. 25 (2004) 591–601.

[22] V.H.S. Ha, J.M.F. Moura, Afine-permutation invariance of 2-D shape, IEEE Trans. Image Process. 14 (11) (2005) 1687–1700.

[23] J.-C. Lin, Universal principal axes: an easy-to-construct tool useful in defining shape orientations for almost every kind of shape, Pattern Recognition 26 (1993) 485–493.

[24] J.-C. Lin, The Family of Universal Axes, Pattern Recognition 29 (1996) 477–485.

[25] D. Shen, H.H.S. Ip, Generalized affine invariant normalization, IEEE Trans. Pattern Anal. Mach. Intell. 19 (5) (1997) 431–440.

[26] D. Shen, H.H.S. Ip, K.K.T. Cheung, E.K. Teoh, Symmetry detection by generalized complex (GC) moments: a close-form solution, IEEE Trans. Pattern Anal. Mach. Intell. 21 (5) (1999) 466–476.

[27] W.H. Tsai, S.L. Chou, Detection of generalized principal axes in rotationally symmetric shapes, Pattern Recognition 24 (1991) 95–104.

[28] J. Žuni´c, L. Kopanja, J.E. Fieldsend, Notes on shape orientation where the standard method does not work, Pattern Recognition 39 (5) (2006) 856–865.

[29] B.K.P. Horn, Robot Vision, MIT Press, Cambridge, MA, 1986.

[30] R. Jain, R. Kasturi, B.G. Schunck, Machine Vision, McGraw-Hill, New York, 1995.

[31] R. Klette, A. Rosenfeld, Digital Geometry, Morgan Kaufmann, San Francisco, 2004.

[32] H. Freeman, R. Shapira, Determining the minimum-area encasing rectangle for an arbitrary closed curve, Comm. of the ACM 18 (1975) 409–413.

[33] R.R. Martin, P.C. Stephenson, Putting objects into boxes, Computer Aided Design 20 (1988) 506–514.

[34] J. Žuni´c, M. Stojmenovic, Notes on shape orientation where the standard method does not work, Pattern Recognition 41 (5) (2008) 1785–1798.

[35] J. Canny, A Computational Approach To Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8 (6)(1986) 679–698.

An Evaluation of IFC-CityGML Unidirectional Conversion

Mohamed El-Mekawy a
Department of Computer and
Systems Sciences,
Stockholm University, Sweden

Anders Östman b
Faculty of Engineering and
Sustainable Development,
University of Gävle, Sweden

Ihab Hijazi c
An-Najah National University,
Nablus,
Palestinian

Abstract—**Interoperability between building information models (BIM) and geographic information models has a strong potential to bring benefit to different demands in construction analysis, urban planning, homeland security and other applications. Therefore, different research and commercial efforts have been initiated to integrate the most prominent semantic models in BIM and geospatial applications. These semantic models are the Industry Foundation Classes (IFC) and City Geography Markup Language (CityGML) respectively. However, these efforts mainly: a) use a unidirectional approach (mostly from IFC to CityGML) for converting data, or b) Extending CityGML by conceptual requirements for converting CityGML to IFC models. The purpose of this paper is to investigate the potential of unidirectional conversion between IFC and CityGML. The different IFC concepts and its corresponding concepts in CityGML is studied and evaluated. The investigation goes beyond building objects, also including other concepts that are represented implicitly in building schemas such as building objects relations, hierarchies of building objects, appearance and other building characteristics. Due to the large semantic differences between IFC and CityGML standards, the schema mapping is based on a manual pragmatic approach without automatic procedures. The mappings are classified into three categories, namely 'full matching', 'partial matching' and 'no matching'. The result of the study shows that only a few concepts are classified as 'direct matching', a few as well as 'no matching' while most of the building concepts are classified as 'partial matching'. It is concluded that unidirectional approaches cannot translate all the needed concepts from both IFC and CityGML standards. Instead, we propose a meta-based unified building model, based on both standards, which shows a high potential for overcoming the shortages of the unidirectional conversion approaches.**

Keywords-IFC; CityGML; UBM; Evaluation; Unidirectional.

I. INTRODUCTION

Sharing and exchanging spatial information in various disciplines has been a major driving force behind the development of spatial technology and applications in the last decade (Isikdag and Zlatanova, 2009b). 3D city modeling applications is one of the most important products of this development. However, from a technical perspective, building 3D city applications requires common communication standards between design models (focusing on the applications of building industry and fulfilling the requirements of architecture, engineering and construction (AEC) industry) and real-world models (as geospatial information systems that model the built environment around us with their spatial objects) (Pu and Zlatanova, 2006; van Oosterom et al., 2006).

Within the building industry, on the one hand, building information modeling (BIM) has been developed as a research area looking mainly at problems related to information integration and interoperability at the building industry domain (Barrett and Grobler, 2000). Industry Foundation Classes (IFC) has then been developed as a reference model standard for the building industry (IAI, 2011). The IFC standard does not just represent and model building components. It further supports different advanced processes and analyses based on spatial relations among these components. These processes can be scheduled in time for different activities. Different objects are represented by database entities that are characterized by properties such as name, geometry, materials, and so on (Khemlani, 2004). In the geospatial information field, on the other hand, CityGML has been developed as a geospatial model standard that represents geometric as well as entities and non-spatial relationships among entities. With its implementation as an application schema for Geography Markup Language 3 (GML3), CityGML is considered more appropriate for outdoor modeling in different levels of detail (Kolbe and Gröger, 2004). Considering their ability for modeling spatial objects with respect to entities and geometric and non-spatial characteristics, IFC and CityGML are seen today as the two most prominent semantic models for representation of design and real world city objects (Isikdag and Zlatanova, 2009b).

A great deal of research is focused on ways for exchanging of information and bringing IFC and CityGML together towards a unified 3D city model (Kolbe and Bacharach, 2006; Isikdag and Zlatanova, 2009b; Van Berlo and De Laat, 2011). However, these research efforts are mainly achieving integration by two different approaches, namely

1. Transforming IFC building models into CityGML (Nagel, 2007) or generating buildings in CityGML using IFC semantics and components (Isikdag & Zlatanova, 2009b).
2. Extending CityGML by conceptual requirements for converting CityGML to IFC models (Nagel et al., 2009) or using Application Domain Extensions (ADE) (Van Berlo, 2009; Van Berlo and De Laat, 2011) which provide a way to represent the information that is not possible to be presented using the current CityGML classes.

Additional to these efforts, commercial software products and conversion tools from IFC to CityGML such as IfcExplorer (IFCExplorer, 2010) and FME (Safe Software, 2010) have been developed mainly following the first approach where IFC concepts are transformed and forced to be represented by CityGML classes.

The first approach to achieve integration is based on unidirectional transformation of IFC building models into CityGML. It is argued to remain the only valid and more relevant to the goal of integrating and analyzing BIM information in geo-context (Isikdag & Zlatanova, 2009b). This method is also found to be dominant in the IFC-CityGML integration literature.

The second approach to achieve integration, as described above, focus on extending the CityGML model. There are two main methods for extending the CityGML model inside an application. The first method is through introducing generic city objects and attributes, while the second is through application domain extensions (ADEs). The first extension method has some clear limitations. For instance, since additional data types cannot be introduced, there are limited possibilities of introducing generic attributes. In addition, XML parsers have difficulties to validate the layout and occurrences of generic objects and attributes. Although, the second extension method (through ADEs) overcomes the limitation of the first method, other general disadvantages of the extension approach is highlighted, such as: i) moving the details of IFC by extending the CityGML schema results in huge CityGML file size that limit the usability of different 3D city applications, ii) IFC include private (as well as irrelevant) information for a city model. Integrating them within public city models is a problematic issue and iii) the two standards are developed by two different communities that have different goals and requirements. As a consequence, extending one standard to replace the other is not feasible, if both communities and target groups still are to be addressed.

As a third approach, El-Mekawy et al. (2011) proposed a unified building model (UBM) for the integration of IFC and CityGML. The UBM is based on reference ontology and meta-concepts and it aims to act as a mediator in the transformation process. Although the development of the UBM is not fully implemented yet, it shows promising results to this area where an increasing number of indoor and outdoor 3D city models needs a tighter integration.

In order to evaluate the potential of these approaches to data integration, the limitations of each method must be analyzed. Since the second method have general limitations in its feasibility and since the third method yet is not well developed, they are excluded from the analysis. As a consequence, this paper aims to investigate the potential of unidirectional conversion between IFC and CityGML. Different IFC concepts and their corresponding concepts in CityGML are studies and evaluated. The investigation goes further beyond building objects, also including other concepts that are represented implicitly in building schema such as relations between building objects, building hierarchies, appearance and other building characteristics. The purpose is to provide a quantitative analysis and validation of the currently most common unidirectional conversion approaches that are utilized for this type of data integration.

The paper is structured as follows: Section 2 provides an overview of IFC and CityGML, its concepts and modeling paradigms. Section 3 presents the schema matching and mapping approach between the two domains. In section 4 we present the results of the mapping from IFC building to CityGML building model which is organized in two subsections: product extension and shared building elements. In section 5 the research is concluded and direction for further research is provided.

II. BACKGROUND IFC AND CITYGML

A. Industry Foundation Classes (IFC)

IFC is defined as an object oriented specification for exchanging, sharing and re-using information throughout the building industry's life-cycle. It has been initially developed since 1996 by the International Alliance for Interoperability (IAI) and it is still under ongoing process of development by different stockholders (IAI, 2008). The aim of IFC development is to facilitate interoperability in the building industry and sharing information among different participants and stakeholders. It is therefore used to assemble computer readable models that contain data elements and represent parts of buildings with their relevant information. Currently, a wide area of AEC/FM domains (such as architectural design, engineering, building analysis, HVAC, energy simulation, thermal analysis, maintenance and cost estimation) are covered by IFC compliance software vendors (IAI, 2011; Lapierre and Cote, 2008).

It is believed today that IFC is the data model that has widest scope for enabling interoperability within the AEC/FM industry (IAI, 2011). Additionally, with its extensible representation, it is being rapid growing for specific applications in construction, manufacturing and operation tasks within the AEC/FM industry. However, there is no universally accepted building model for IFC yet (Kolbe et al., 2008; Kiziltas et al., 2010). Therefore, primarily based on the work done by the IAI and ISO in form of IFC standard documentation (IAI, 2008), the ISO 16739 standard (ISO 16739) and Benner et al., (2005), an IFC building model is presented in figure 1. Important concepts from those efforts were highlighted and then relationships between them were built. Figure 1 show the proposed IFC building model based on the UML standard notations.
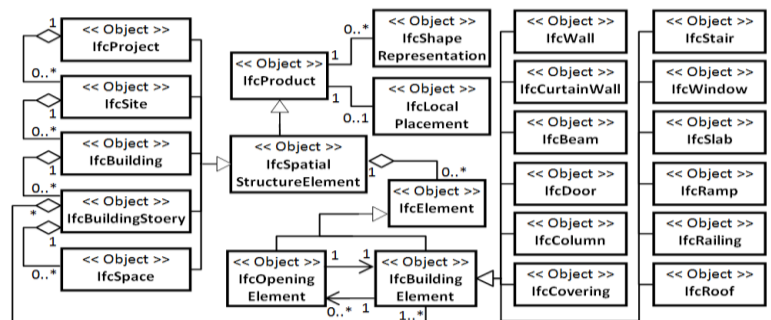

Figure 1. IFC Building Model

Several aspects regarding the IFC schema need to be highlighted. They are discussed as following:

- There are in total about 900 different classes. However most of them have an important role just for defining and linking spatial relationships between objects and their geometrical representation. Berlo (2009) found, in his comprehensive theoretical research, that only 60 to 70 classes out of the 900 can be transformed to GIS. Similarly, most of these classes are used for distributed elements (e.g. heating, ventilation, air conditioning, electrical and equipment elements) and structural calculations (e.g. *IfcStructuralCurveConnection*, *IfcStructuralLinearAction*, *IfcStructuralPointAction*). Consequently, only few are used for the representing the building and architectural elements.

- The structure of IFC is built to support dynamic models. This means more flexibility for users and developers to represent their building data. There are main core elements and containers of elements that store information about building elements in general. All other objects can then be created by using the core elements. This flexibility in IFC structure also explains that there is usually more than one way to connect two different building elements or objects in IFC. *IfcWindow* for example is placed in an opening element (*IfcOpening*) which is connected to a building element such *IfcWall*. The wall surrounds and connected to a space (*IfcSpace*) and the space is connected to the building (*IfcBuilding*). This route in the IFC structure that represents the relationship from an *IfcWindow* towards the *IfcSpace* in which it is located is a user-defined in a specific data file. However, it is not stored in the schema itself. Therefore, the route can be then defined as (*IfcWindow – IfcSpace – IfcBuildingStorey*) or (*IfcWindow – IfcBuildingStorey –IfcSpace*) if for example a space is expanded for more than one storey for the latter route.

- IFC define geometric shape definition of building objects explicitly by solid representation i.e. extrusion or faceted boundary representations. The semantics meanings of entities are directly mapped in IFC, for example the name of IfcWall, IfcBeam or IfcColumn implies its semantic meanings and its explicit geometric representation.

B. City Geography Markup Language (CityGML)

CityGML is an OGC standard, which provides a specification for the representation of 3D urban objects (OGC, 2011). It is the only 3D information model for the exchange of 3D city models. One of the reasons for creating such a model was to enrich 3D city models with thematic and semantic information. The information model of CityGML is an XML-based format implemented as an application schema of Geography Markup language (GML3). The data model has the following features (Becker et al., 2010).

- CityGML seems to provide the best framework for semantic-geometric relations of 3D objects above earth surface (Emgard and Zlatanova, 2008; Groneman and Zlatanova, 2009). It maintains a good taxonomy and aggregations of Digital Terrain Models, sites (including buildings), vegetation, water bodies, transportation facilities, and city furniture. The underlying model differentiates five consecutive levels of detail (LoD), where

objects become more detailed with increasing LoD regarding both geometry and thematic differentiation. In LoDs 2-4 of CityGML the building facade is defined in the form of boundary surfaces, i.e. WallSurface, Roof Surface, Ground Surface or Closing surface. The LoD4 allows the representation of interior building elements, e.g. rooms, furniture, interior wall surfaces.

- In CityGML, the structure of linking building elements is fixed /static. For instance, the window (which is a subclass of the opening class) should be connected to the opening class, the opening is connected to a *WallSurface*, and the *WallSurface* is connected to *Room*.

- CityGML defined geometric representation of building and building objects implicitly. For example a wall is defined by different attributes that define each separate wall surface. They are then combined and their geometric definitions form a complete wall. Another example is beams and columns. They are defined by attributes that define their constructing surfaces and then the surfaces are combined geometrically in the class *IntBuildingInstallation* as interior building installations.

Conversely to the IFC, a CityGML Building Model has been produced in the CityGML standard (Kolbe, 2008). The building model (shown in Figure 3) is an excerpted version from the CityGML standard in which only the used conversion concepts to IFC are represented i.e. *BuildingFurniture*, and *IntBuildingInstallation* are not represented. UML standard notations are used for describing the CityGML building model.
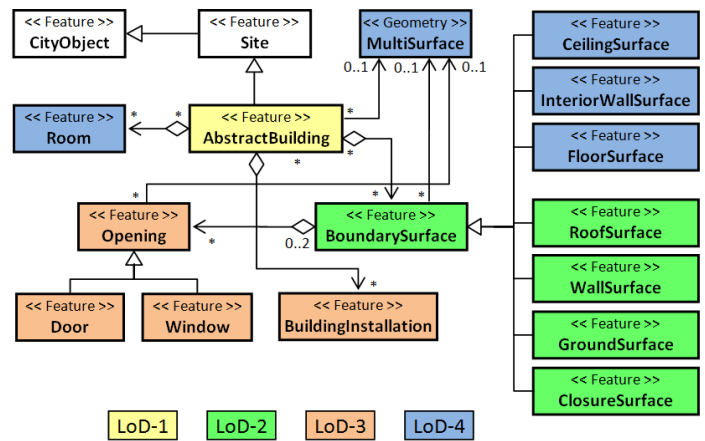

Figure 2. CityGML Building Model

III. SCHEMA MATCHING AND MAPPING OPERATIONS

A schema transformation is characterized by the mapping functions that specify the transformations from one domain to another. If all information in the target schema can be derived from the information in the source schema, mapping functions may be specified that ensures a complete transformation fulfilling the needs of the target schema. In order to investigate the potential of an undirected transformation between data in IFC and CityGML, the properties of the schema matching and schema mapping operations are studied. The matching between IFC and CityGML, is here based on a pragmatic and manual approach. No automatic mapping procedure or matching between ontologies has been made. This is mainly due to the

large semantic differences between both standards. The aim of the schema matching is to specify which entities and attributes correspond to each other. The focus of this paper is on building objects and their relationships as specified by LoD4 in CityGML. In this context, the entities of the IFC Interoperability Layer – specifically shared building element – and core layer – specifically product extension – constitute the starting point for this study of the direct conversion process.

The shared building elements are the major elements, which constitutes the architectural design and the structure of a building. The building elements (e.g. wall, beam, column, slab, roof, stair, ramp, window, door and covering) are the main components of the raw building (or carcass) which is central for the exchange of project data. The Product Extension further specialize the concepts of a (physical) product, i.e. a component likely to have a placement within the project context. The spatial project structure is introduced within the *IfcProductExtension*. This structure defines the site (the land area where the building is constructed), the building or building parts, the building storey and the space and its decomposition structure. Relationships between the spatial structure and the building elements are defined, such as spatial containment and space boundaries. Other relationships like building elements, opening elements, furnishing elements, distribution elements (including heating, ventilation, air conditioning, electrical and equipment elements) and transportation elements are also defined.

The schema matching operation investigates which entities and attributes correspond to each other. Additionally, it investigates the semantic meaning of different building objects and the associated loss of information based on the direct transformation. The schema mapping operations investigate the relationships and hierarchies between different building objects and how they can be mapped from the IFC schema. The mapping is a specification describing how data is to be transformed from one model to another. Specifications of the relationships that cannot be represented or created from the source model will be used as indicators of the quality of the transformation.

- A full (direct) matching is defined as a matching where a single element in the target model (CityGML) can be directly reconstructed from one or several elements in the source model. Unfortunately, such matching is rare.
- A partial matching is defined as matching where a single element in the target model (CityGML) can be partially reconstructed from one or several elements in the source model. Such matching is found in most of the cases.
- A no matching is defined as a matching where a single element in the target model (CityGML) cannot be reconstructed at all from the information provided in any of the elements in the source model.

IV. MAPPING

Following the focus of this paper (discussed in Section 3), we keep the evaluation in this section divided in two parts: the product extension and the shared building elements. This gives more insights on both groups and provides modularity of the discussion as a scientific contribution.

A. Evaluation of Product Extension

The product extension discipline in IFC defines the spatial structure of buildings and the relationships between the spatial structure of a building and the building elements. The objects in the product extension are the site where the building/s/(parts of buildings) are placed, building as a structure, building storey and space. These objects are then also containing the architectural objects within a spatial structure. The product extension objects specialize the concepts of a physical product which are the components that represent shape and placement of building elements within a project context. There are 50 different classes/entities in the product extension data model (Table 1). These entities define 5 different concepts, namely:

- the spatial project structure (that defines site, the building, the building storey, and the space),
- the element (that has the types of building element, opening element, furnishing element, distribution element and transportation element),
- the grid (defines a constraint placement for elements),
- the port (defines connectivity between elements), and
- the annotation (captures additional annotations to a building model or plan such as explanatory text, dimensioning, etc.).
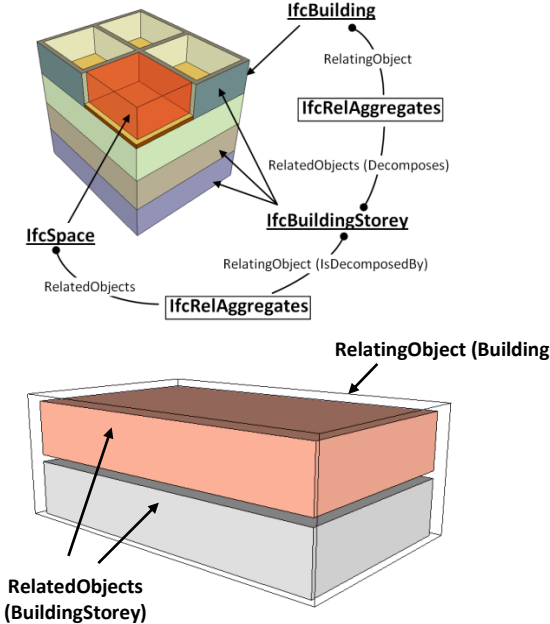
TABLE I. DIFFERENT 50 CLASSES/ENTITIES IN THE PRODUCT EXTENSION DATA MODEL
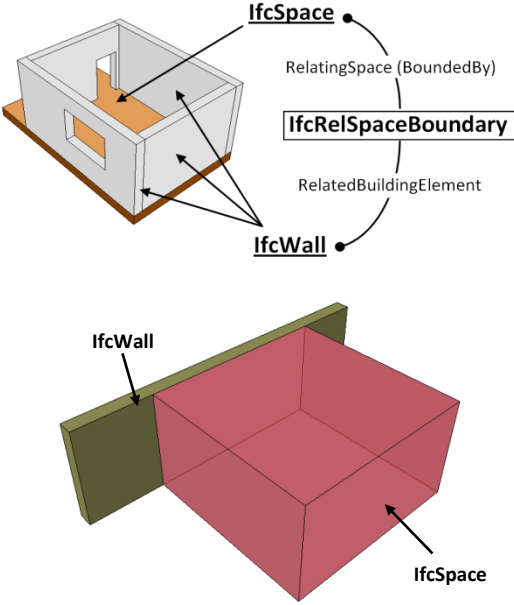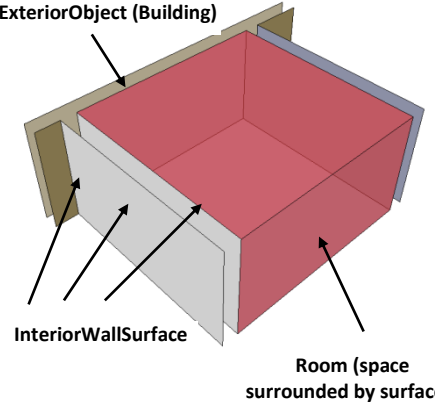
IfcAnnotation	IfcProjectionElement
IfcBuilding	IfcRelAssociatesMaterial
IfcBuildingElement	IfcRelConnectsElements
IfcBuildingElementProxy	IfcRelConnectsPortToElement
IfcBuildingElementProxyType	IfcRelConnectsPorts
IfcBuildingElementType	IfcRelConnectsWithRealizingElements
IfcBuildingStorey	IfcRelContainedInSpatialStructure
IfcCovering	IfcRelCoversBldgElements
IfcCoveringType	IfcRelCoversSpaces
IfcDistributionElement	IfcRelFillsElement
IfcDistributionElementType	IfcRelProjectsElement
IfcElectricalElement	IfcRelReferencedInSpatialStructure
IfcElement	IfcRelServicesBuildings
IfcElementAssembly	IfcRelSpaceBoundary
IfcElementQuantity	IfcRelVoidsElement
IfcElementType	IfcSite
IfcEquipmentElement	IfcSpace
IfcFeatureElement	IfcSpaceType
IfcFeatureElementAddition	IfcSpatialStructureElement
IfcFeatureElementSubtraction	IfcSpatialStructureElementType
IfcFurnishingElement	IfcSystem
IfcFurnishingElementType	IfcTransportElement
IfcGrid	IfcTransportElementType
IfcOpeningElement	IfcVirtualElement
IfcPort	IfcZone

■ Concepts

■ Structural Objects

■ Basic Types of Elements

□ Classes defining spatial relationships between building objects.

Most of the 50 entities are used for defining and linking, spatial relationships between building objects and structure calculations. Therefore, not all of them can be used as spatial object in the GIS context. However, some of their properties can be transformed to CityGML attributes. The applicable classes are discussed in the matching below (Table 2).

TABLE II.   MATCHING IFC AND CITYGML CLASSES BASED ON THE PRODUCT EXTENSION CONCEPTS

IFC	CityGML	Matching
  /* Spatial structure element definitions */ #4=IFCBUILDING (); #8=IFCBUILDINGSTOREY (); #9=IFCBUILDINGSTOREY (); ….. ….. /*Assignment of two storeys to the building */ #10=IFCRELAGGREGATES ('**…**', $, $, $, #4, (#8, #9)); ….. …..	< ?xml version="1.0" ... ... > < bldg:Building xmlns:gml = "..." [...] > < bldg:Room > < bldg:BoundedBy > < bldg:WallSurface > < bldg:Opening > < bldg:Window/ > < /bldg:Opening > < /bldg:WallSurface > < /bldg:BoundedBy > < /bldg:Room > ….. ….. ….. ….. < /bldg:Buil ding >	**Building Structure** • In IFC, a building is smoothly structured by breaking down the building into storeys and into spaces that form a specific storey. However, in CityGML, there is no explicit definition of spaces or stories. • *IfcBuildingStorey* is defined geometrically as an elevation and typical representation of (nearly or in most cases) horizontally aggregated spaces (*IfcRelAggregates*) that are separated vertically. Storeys in CityGML can be represented as an explicit aggregation of all building features on a certain height level. • A building in CityGML consists of rooms, where room is a space surrounded by different boundary surfaces. Storeys are not explicitly defined but they can be represented as an explicit aggregation of all building features according to arbitrary user defined criteria on a certain height level.
The matching between these two different structures is seen as 'no matching'. However, the needed information for defining the target structure can be extracted from entities that represent the hierarchy in the source model. Additionally, in CityGML, the proper way of linking objects in such a structure is strictly defined as (*Window*, *WallSurface*, *Room*, *Building*) the same all the time. Conversely, the spatial structure in IFC is user-defined and to be stored in a specific data file. It can be described as (*IfcWindow*, *IfcOpening*, *IfcWall*, *IfcSpace*, *IfcBuilding*) to highlight different openings and spaces if spaces are the focus of an application or (*IfcWindow*, *IfcStorey*, *IfcBuilding*) if storey's are in the focus. A 'no matching' here is referred to the difficulties of following a single data model of IFC or CityGML for hierarchical structure.		

IfcSpace ●

RelatingSpace (BoundedBy)

IfcRelSpaceBoundary

RelatedBuildingElement

IfcWall ●

IfcWall

IfcSpace

ExteriorObject (Building)

InteriorWallSurface

Room (space surrounded by surfaces)

Building Spaces

- The *IfcSpace* class defines all volumes and areas that are bounded in some way. This includes rooms that are bounded by different building elements. A building in CityGML is represented by rooms representing all type of spaces as semantic objects for modeling the free spaces inside a building.

- Space in IFC is geometrically associated to a building storey and may be divided into partial spaces. The space is bounded by related building elements (*IfcRelSpaceBoundary*) that surround the space. However, in CityGML, rooms are not objects in itself. Instead they are represented by boundary surfaces that topologically close the rooms.

- The *IfcSpace* is geometrically represented with multiple representations as usually a combination of Sweeping, CSG and rarely in a BRep. Rooms in CityGML are, however, usually geometrically represented by *GML:Solid* or *GML:MultiSurface*.

/* Spatial structure element definitions */

#285=IFCSPACE ();
#1000=IFCWALL ();
.....
.....

/*Assignment of wall to the space */

#1036=IFCRELSPACEBOUNDARY ('**...**', #6, $, $, #285, #1000, #1037, .PHYSICAL., .INTERNAL.);
.....
.....

Although the differences between space concepts in both IFC and CityGML are apparent, there is a 'partial matching' between them. In IFC, there is no concept of specific *Room* object as in CityGML. *IfcSpace* class defines all volumes and areas that are bounded. This definition includes rooms that are bounded by different building elements but also other bounded volumes such as corridors etc. As all IFC objects, Sweeping/CSG geometry is used for spaces and their elements which requires conversions to BRep geometric models that are used in CityGML. To create spaces in CityGML, information from *IfcWall*, *IfcRoof* and *IfcSlab* that form the boundaries of rooms are used. Information about both *Ceiling* and *Floor* in CityGML can be acquired from *IfcSlab* where a slab may represent both a ceiling for a storey and a floor for another storey on the top of it (for example, a ceiling for the 2nd storey is a floor for the 3rd storey).

Building Windows and Doors

- *IfcDoor* and *IfcWindow* are referred from the *IfcOpeningElement*. The *IfcOpeningElement* is a model element that is used to describe spatial representation of openings by their geometry and semantics. Each opening then may contain one or more doors or windows i.e. *IfcOpeningElement* in the IFC model can refer to multiple *IfcDoor* and *IfcWindow* elements. The opening in CityGML is, however, defined as class *Opening*. The *Door* and *Window* classes are defined as subclasses of the *Opening* class.

- As it can be noticed in next figures (upper-left), *IfcDoor* and *IfcWindow* are represented by defining two relationships. The first relationship (*IfcWall_IfcRelVoidsElement _IfcOpeningElement*) represents how the opening element is spatially attached to the building element (i.e. *IfcWallStandardCase* in this example). The second relationship however (*IfcOpeningElement_IfcRelFillsElement_IfcWindow/IfcDoor*) represents how the door or window elements semantically fill the opening (*IfcOpeningElement*).

- In CityGML, the opening elements (e.g. door or window) are represented by the *gml:MultiSurface* geometry. Doors and windows are represented as Sweeping, CSG or in a form of finer BRep composing of different surfaces similar to the wall and slab representations. Thus, the window or door is represented in CityGML by combined surfaces (like a wall or slab) that are attached to the interior or exterior surfaces of a building element (e.g. wall) or in between. Therefore, CityGML does not store additional semantic information about the opening elements such as width, height, panel type, panel number, and the opening direction where they can be stored in IFC.

This is the only case that shows direct matching where a CityGML feature (Window or Door) can be reconstructed from a single IFC element (*IfcWindow* or *IfcDoor*). The CityGML opening (window or door) can be reconstructed from the element voiding relationship (*IfcRelVoidsElement*) and the surfaces of a window or door can be then reconstructed from the element filling relationship (*IfcRelFillsElement*).

The two standards, IFC and CityGML, represent both 3D semantic models of buildings. They are both object oriented and represent the building objects (e.g. slabs, walls, spaces) based on their semantic rules. The data models for the IFC and CityGML standards are associated with detailed 3D geometry. The data models provide a clear hierarchy for organizing the representation of building structure.

In IFC, the entire spatial structure is subsumed from *IfcProject* class which is the uppermost container of all building information. Under the *IfcProject*, there are two mandatory levels should be defined, *IfcBuilding* and *IfcBuildingStorey*.

There are other optional classes as levels in the hierarchy that can be subsumed from the project. They are IfcSite, *IfcBuildingSection* and *IfcSpaces*. The site may contain zero or several (0 .. *) buildings.

A building has at least one storey (1 .. *). Each building storey may have zero or more spaces (0 .. *). All building elements are assigned to the building storey in which they are located. If building elements (or spaces) span through several storeys, then they are assigned to the lowest storey among them in which they are based. CityGML in LoD4, however, provides a different hierarchy for the building structure than the IFC. A building in CityGML is composed of rooms which are enclosed by surrounding surfaces. Storeys are not explicitly defined in the structure. However, they can be modeled using the CityGML generic grouping mechanism.

In IFC, the users can define their own structure for representing different spaces of a building. Users can generate their route of connections with attached clarification regarding the reference service systems in buildings in the structure starting from a wall or slab and linking them to a space then storey the and ending by a building. There is, however, no unique way to connect a specific IFC object to another.

The user-defined route of connections is something we do not find in the IFC schema. Instead, they are stored in a specific data file (remember the optional levels such as *IfcSpace*). For example, *IfcFlowSegment* can be used for defining the containment relationship between *IfcWall* and *IfcSpace* based on the information that spaces are enclosed by walls. Similarly, *IfcSpace* can be spatially related to *IfcStorey* and then to *IfcBuilding*.

These kinds of spatial connections are defined in CityGML as more static and explicit. The relationship between *Window*, *WallSurface*, *Room* and *Building* is always the same. However, the wall class as a building element does not exist in CityGML. It is therefore, difficult for software implementations to transform data from IFC to CityGML. Supported by the claims of Van Berlo & De Laat (2011), this can also be a clear reason for why not all the 50 entities with their properties in the IFC Product Extension discipline can be transformed to CityGML model.

B. Evaluation of Shared Building Elements

In the shared building elements schema of IFC, most of the architectural design elements of a building are defined. These elements include wall, beam, column, slab, floor, roof, stair, ramp, window, door and covering. It is worthwhile here to say that the reason of naming these elements as the shared building elements is their use and involvement in other disciplines of the IFC schema and buildings in general.

There are 33 different classes/entities in the shared building elements schema (Table 3). These entities define the subtypes of *IfcBuildingElement* which is defined in the *IfcProductExtension*. The architectural design of a building structure is constituted by those subtypes.

TABLE III.   DIFFERENT 33 CLASSES/ENTITIES IN THE SHARED BUILDING ELEMENTS DATA MODEL

IfcBeam	IfcRampFlight
IfcBeamType	IfcRampFlightType
IfcColumn	IfcRelConnectsPathElements
IfcColumnType	IfcRoof
IfcCurtainWall	IfcSlab
IfcCurtainWallType	IfcSlabType
IfcDoor	IfcStair
IfcDoorLiningProperties	IfcStairFlight
IfcDoorPanelProperties	IfcStairFlightType
IfcDoorStyle	IfcWall
IfcMember	IfcWallStandardCase
IfcMemberType	IfcWallType
IfcPlate	IfcWindow
IfcPlateType	IfcWindowLiningProperties
IfcRailing	IfcWindowPanelProperties
IfcRailingType	IfcWindowStyle
IfcRamp	

    Building Elements

▭ Classes that define spatial characteristics and representations of building objects

Similar to the *IfcProductExtension*, many of these 33 entities are used for defining spatial characteristics and representations of building objects.
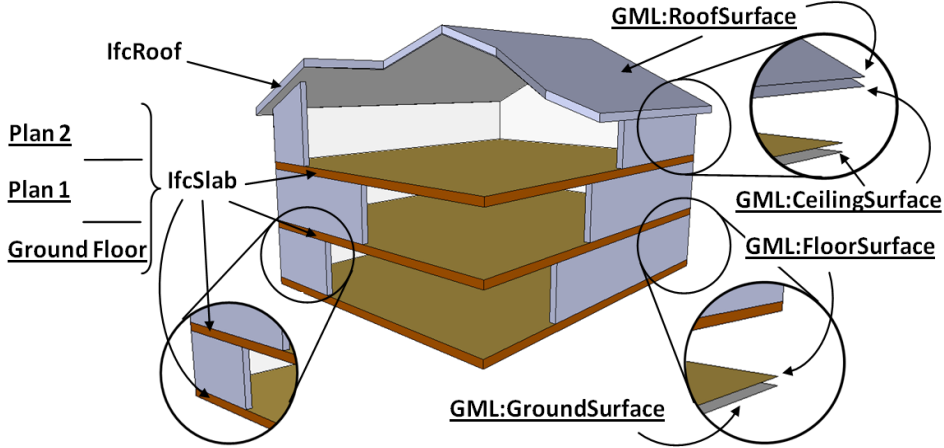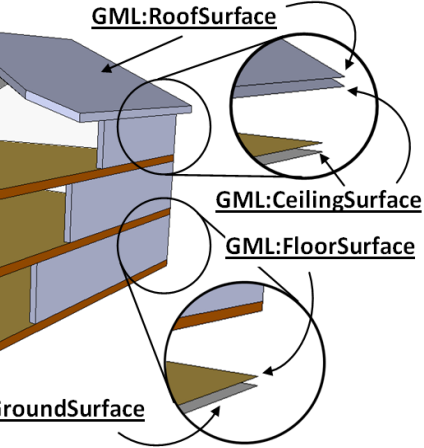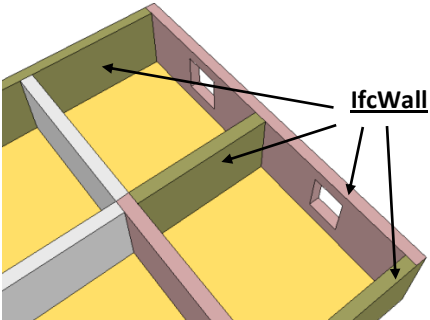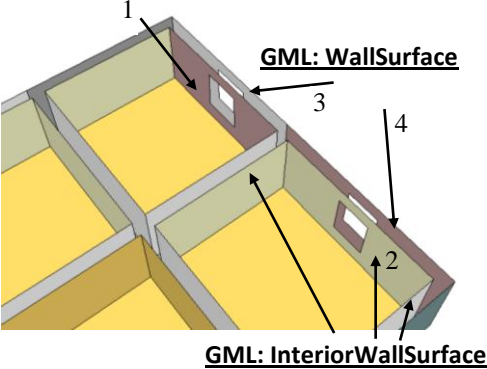
Therefore, not all of them can be used in the GIS context, and hence, the applicable classes are discussed in the mapping below (Table 4).

V.   DISCUSSION AND CONCLUSIONS

The potential of the unidirectional conversion between IFC and CityGML is mainly focused on the investigation of how much information in the BIM (represented by IFC) can be transferred to the GIS context (represented by CityGML). The reason for that is the very rich semantic model of IFC which still has more detailed schemas than the CityGML (Van Berlo & De Laat, 2011; Isikdag & Zlatanova, 2009b) for modeling specific buildings. Additionally, this has been shown in most of the unidirectional efforts transforming IFC building models to CityGML (e.g. Nagel, 2007; IFG, 2007; Isikdag & Zlatanova, 2009b).
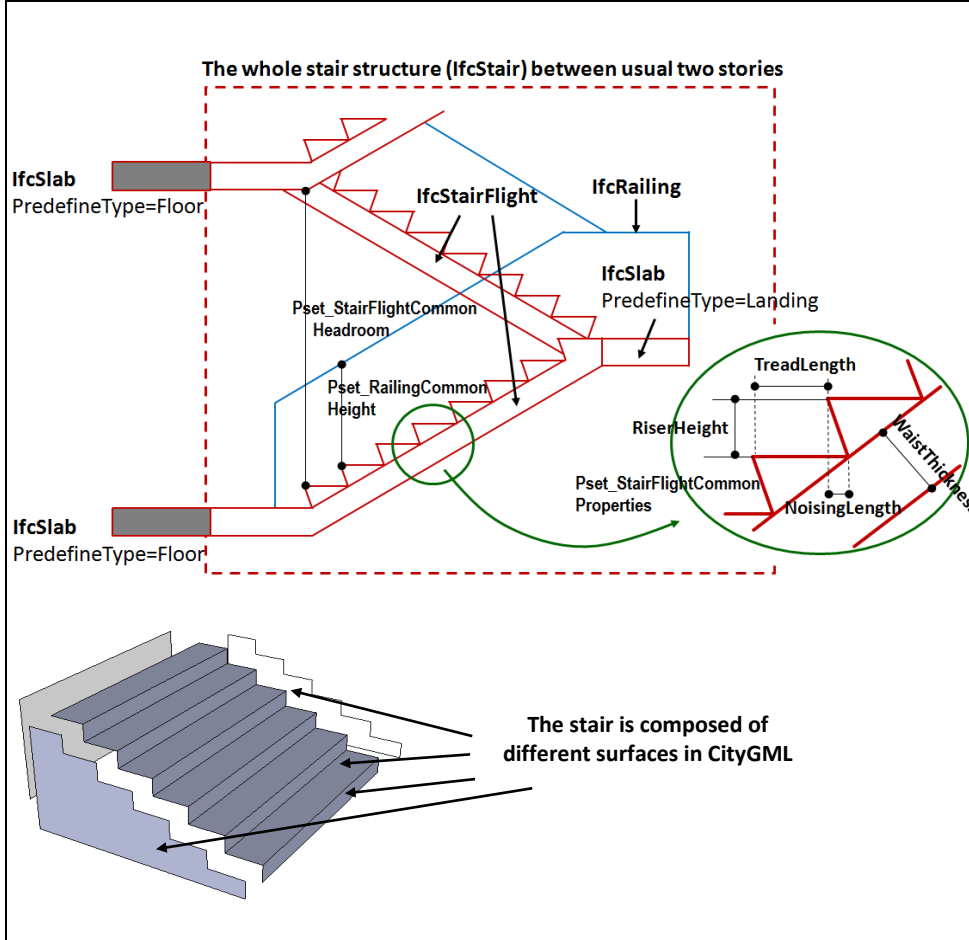
The results of our paper shows that the information provided in IFC that can be transferred to GIS is mainly included in the *IfcProductExtension* and *IfcSharedBuildingElement* domains. Combined, they have 83 classes/entities in their schemas, 50 for the *IfcProductExtension* and 33 *IfcSharedBuildingElement*.

TABLE IV.   MATCHING IFC AND CITYGML CLASSES BASED ON SHARED BUILDING ELEMENTS CONCEPTS

IFC	CityGML	Matching
		Building Slabs and Coverings • *IfcRoof* and *IfcSlab* represent complete building elements as roofs and slabs respectively. *IfcRoof* represents the total roof and acts as a container for the all parts of the roof aggregated by the *IfcRelAggregates* relationship. *IfcSlab* represents a constructed component that encloses a space vertically in most of the cases. For any space in the building, it may provide the lower construction of a space (floor) or the upper one (roof slab). • The flooring and roofing slabs in CityGML are classified in different types; *RoofSurface, CeilingSurface, FloorSurface and GroundSurface* (Figure XXX). However, all of them are represented by their split, but geometrically connected surfaces in which different building spaces are surrounded by these surfaces (e.g. one slab has its lower surface as a *CeilingSlab* for a building space in Plan N, and its upper surface as a *FloorSurface* for a building space in plan N+1). • *IfcRoof* and *IfcSlab* are geometrically solid representations of a roof and slab respectively where such building element exists. However, in CityGML, the surfaces (*Roof, Ceiling, Floor and Ground*) represent the visible parts of the building elements from inside the space.

The matching case in the above mentioned elements is seen as a 'partial matching'. We deal in these examples with 1:m (one-to-many) matching relationships. For instance, *IfcSlab* is the only source information for reconstructing the CityGML Ceiling and Floor surfaces i.e. a slab may represent both a ceiling for a storey and a floor for another storey on the top of it (for example, a ceiling for the 2nd storey is a floor for the 3rd storey). The ceiling and floor surfaces in CityGML are then reconstructed from different slabs in IFC. Looking from the other side, an *IfcSlab* represented the ground floor of a building can be reconstructed by combining information from both CityGML Floor and Ground surfaces representing n:1 (many-to-one) matching relationships.

IFC	CityGML	Matching
		Building Walls • *IfcWall* represents a complete wall as building element. In the CityGML, the wall is split into geometrically connected surfaces in which different building spaces are surrounded by these surfaces. • *IfcWall* is geometrically a solid representation of a wall volume where the building element exists. In CityGML, the surface WallSurface represents the visible part of the wall from inside the space.

Similar to the concepts of slabs and ceilings, the "partial matching" is seen visible for mapping the building walls between IFC and CityGML. For example, two or more CityGML surfaces have to be combined to form a single IFC element. For example, in the next figure (CityGML part), the interior wall surfaces 1, 2 and the exterior surfaces 3, 4 represent only one wall object and to be mapped to the corresponding IfcWall elements. This represents n:1 (many-to-one) example from CityGML to IFC. The 1:n can also be seen from the other way around. Furthermore, the relationship n:m is important to be considered for the case of curtain walls. A curtain wall in CityGML is represented as surfaces spanning for the whole wall. However, in IFC, the building is structured into storeys that require walls to vertically respect storey's boundaries, and hence, each CityGML wall surface is then divided into parts.

The whole stair structure (IfcStair) between usual two stories

The stair is composed of different surfaces in CityGML

Building Installations

- All spatial objects that can be components of a building architecture or structure are defined in IFC as building elements. However, in CityGML, only the main building structural objects are considered as building elements. Other spatial elements including some structural elements (e.g. columns, beams) are considered in CityGML as internal installations and represented by *IntBuildingInstallation* class.

- *IntBuildingInstallation* class in CityGML represents all the objects inside a building with a specialized function or semantic meaning. However, *IntBuildingInstallations*, which is different from *BuildingFurniture*, are permanently attached to the building structure and cannot be moved. Typical examples of these installations are interior stairs, ramps, railings, beams, columns, pipes and radiators.

- The geometry of *IntBuildingInstallation* objects are represented, as can be noticed from the next figure (lower), in different surfaces that are aggregated and defined by *gml:Geometry* type to build the installation object similar to the building elements (e.g. wall, slab). However, internal installations in IFC are represented as building elements having their own representations with geometrical entities. For the *IfcStair*, the geometric representation is defined by *IfcProductDefinitionShape* that allows multiple geometric representations of the stair components. The way of representing the whole stair is user-defined. It can either be an aggregation of all stair components (then the stair geometry is the sum of the representation of all components within the aggregate) or individually (then independent geometric representations for all components are used separately).

- In CityGML, the *IntBuildingInstallation* objects can either be associated to a specific room or to the complete building/building part. For each *BuildingInstallation* object, there are three optional attributes, class, function and usage. The class attribute defines a general classification for the object. The function and usage attributes define prospectively the nominal and real functions of a building installation. The geometry of the *IntBuildingInstallation* objects are represented by *gml:Geometry* type.

The "partial matching" is clearly observed in all building installations. A clear example for that is the n:1 relationship for building components that spans over other components and penetrate different building spaces. The (stairs, columns) and beams, are examples of these components that spans over different storeys and rooms respectively. These building installations are represented in CityGML as different surfaces. Then the same beam can be seen in different rooms as represented in two thematic building installation objects with separated surface geometries. Columns and stairs are similarly observed from different vertically structured spaces (rooms) in two or more different storeys. Therefore they are also represented in different thematic building installation objects with own geometries. This clearly conforms the results of (Nagel et al., 2009) who suggest that in order to aggregates different CityGML surfaces into one IFC building element, we do not only need the semantic information. Instead, geometric-topological (such as parallelism, perpendicularity, distance, and adjacency) relationships are also important to be analyze

In the *IfcProductExtension*, the 50 classes include mostly entities that are used for defining the structure of how building objects are spatially connected or related. They, however, have important information for defining how the building objects are spatially represented. In addition to that, entities like *IfcBuildingStorey* and *IfcSite* have information that is not used in CityGML which causes clear information loss when converting from IFC model to CityGML model. Table 1 shows all the 50 entities. In our study, we have focused on the transferred IFC entities to CityGML. Out of these 50 entities, it is found that only four IFC classes can be mapped directly to CityGML in either partial or full matching. These four classes are: *IfcBuilding*, *IfcFurnishingElement*, *IfcSpace*, *IfcAnnotation*. *IfcBuilding* can be mapped to *_AbstractBuilding*. IfcFurnishingElement can be mapped to the GML BuildingFurniture. The geometry of furniture is usually represented in IFC model as a BRep. The information in IfcFurnishingElement can be used to generate the CityGML BuildingFurniture and to be represented by the *gml:Geometry*. Although there is a need for geometry conversion from Sweeping/CSG to BRep, the information in *IfcSpace* can be transferred for generating the *Room* object in CityGML. *IfcAnnotation* is usually represented in IFC model as a graphical representation within the geometric context of a building or a project. Annotation adds a meaning to objects in forms of (e.g. text, line drawings, hatching) and they are modeled by the IfcRelContainedInSpatialStructure relationship. The information in IFCAnnotation can be transferred to the geometry annotation in CityGML.

All the concepts that are related to the product extension of IFC are found to have matching possibilities with CityGML, except the building structure as discussed in Section 4.1.

The relation between the opening elements and the building element e.g. wall, slab is found to be full matching as windows and doors with their opening can be reconstructed in the target model from the information provided in the source model in both sides. The space (from IFC) and room (from CityGML) concepts and their relationship to their bounding objects cannot be transformed fully without a need for additional information. Therefore, it is considered as partial match. The concepts related to the building structure and its hierarchy, no matching is observed as the structure of both IFC and CityGML building models are built on different concepts.

Therefore, the unidirectional approach seems not to be able or having capabilities for supporting the transformation of concepts in the product extension discipline, and substantial amounts of information will get lost through such transformation approach. A mapping summary of the product extension concepts is provided in Table 5.

TABLE V.   MAPPING SUMMARY OF THE PRODUCT EXTENSION CONCEPTS

Concept	Investigation Result
Building Structure	No match
Building Spaces	Partial match
Building Windows and Doors	Full match

In the *IfcSharedBuildingElement*, the 33 classes define the subtypes of *Ifcbuildingelement* that are defined in the *IfcProductExtension*. These entities, which are the central for the exchange of project data, are represented in the Table 6. For each element, a geometric use definition is introduced in order to define the correct application of available shape representation types for the element. However, not all of them can be used in the GIS context as they are used for defining the spatial characteristics, representations of building objects and structure calculations. Nine entities out of the 33 in *IfcSharedBuildingElement* are found to be directly and partially mapped to CityGML. The nine entities are: *IfcBeam*, *IfcColumn*, *IfcDoor*, *IfcRailing*, *IfcRoof*, *IfcSlab*, *IfcStair*, *IfcWall* and *IfcWindow*. Similar to the entities of *IfcProductExtension*, the non-transformed-to-GIS entities have important information for defining the spatial representation of building objects. In addition to that, there are other entities (e.g. *IfcCurtainWall*, *IfcRamp*) that have attributes and information to be used in CityGML building.

Based on the investigation done on the *IfcSharedBuildingElement* data model, all the transformed entities are found to be partial matching. They can be transformed using additional information from the relationships entities/classes or processing of the spatial data (e.g. splitting, aggregating). As a result, a unidirectional conversion between IFC and CityGML implies data loss and database processing. Table 6 provides a summary for this investigation.

TABLE VI.   MAPPING SUMMARY OF THE SHARED BUILDING ELEMENTS CONCEPTS

Concept	Investigation result
Building Slabs and Coverings	Partial match
Building Walls	Partial match
Building Installations	Partial match

VI. CONCLUSIONS

The IFC and CityGML standards represent indoor and outdoor spatial objects of a building. In order to fulfill the demands in urban planning applications and construction analysis, it is important to integrate these standards. However, existing approaches do not provide complete integration because they mostly offer a unidirectional transformation i.e., from IFC to CityGML.

The need for integrating indoor IFC information of buildings and outdoor CityGML built-environment information is apparent. However, different findings, shown in this paper, need to be considered while combining data from both standards. These findings form the following conclusions:

- Most of the CityGML objects can be mapped from the IFC model by partial matching, which implies that additional information or complex processing are required in the mapping. One example is slabs in IFC which should be mapped to several different surfaces in CityGML. How to extract parts of a certain slab and map it to the correct class in CityGML is not a straight forward task. Probably manual processing is required to map the correct surfaces and the whole process is not easily automated.
- The IFC standard is flexible in the sense that structures are user-defined. As a consequence, it can accept a variety of 3D building models from different communities. The 3D building models of today are however scattered over different data holders, in the public as well as the private

sector, in different systems, different conceptual models, different data formats, different data schemas, different levels of detail and different quality. Having different user-defined structures in all these different organizations for serving different purposes makes the mapping to CityGML a complex process especially if automation is needed because the main reason for this is that the structures in CityGML are static.

- In order to overcome the limitations of the existing unidirectional approaches, a new approach is required in order to achieve interoperability. It is recognized that both IFC and CityGML have been developed for different purposes. In addition, CityGML was not originally designed to fully comply to the semantics of the IFC standard. Therefore, CityGML needs to be overloaded with additional information to be smoothly matched and integrated with IFC. One solution may be the proposed unified building model (UBM), which is based on reference ontology and meta-concepts, see figure 3.
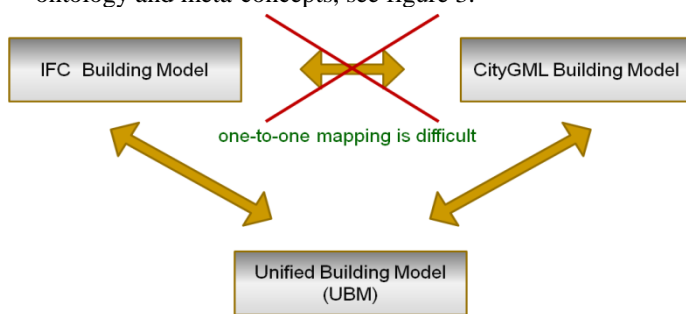

Figure 3.   The UBM as a Meta Model

We believe that such an approach might overcome the limitations identified in this paper. By bringing both IFC and CityGML to the UBM in a common model, both standards can be smoothly integrated without a need for conversion. In addition, the number of unified applications, which requires an integration of indoor and outdoor data, may increase without affecting existing users of either model. Additionally, the UBM might have the potential to be extended to support applications that neither IFC nor CityGML can support.

REFERENCES

[1] Barrett, D. and Grobler, F., 2000. Understanding the different purposes of IFCs and aecXML in achieving Interoperability. http://www.iai-na.com [last accessed Feb-2012].

[2] Becker, T., Nagel, C. and Kolbe, T., 2010. UtilityNetworkADE,-Core Model-Draft version. http://www.citygmlwiki.org/index.php/CityGMLUtility- NetworkADE [last accessed Feb-2012]

[3] Benner, J., Geiger, A. and Leinemann, K. 2005. Flexible Generation of Semantic 3D Building Models. In: Gröger, G. et al. (Eds.): Proceedings of the 1st International Workshop on Next Generation 3D City Models, Bonn. www.eurosdr.net/km_pub/no49/workshops_docs/citymodels_june_05/p aper03_benner.pdf [last accessed Feb-2012]

[4] El-Mekawy, M. Östman, A. and Shahzad, M.K., 2011. Towards Interoperating CityGML And IFC Building Models: A Unified Model Based Approach, 3D Geoinfo Conference, Berlin, Germany, Lecture Notes in Geoinformation and Cartography series (Springer-Verlag, Heidelberg).

[5] Emgard, KL. and Zlatanova, S., 2008. Design of an integrated 3D information model. In: Coors, Rumor, Fendel & Zlatanova S.(eds.),

Urban and regional data management: UDMS annual 2007 (pp. 143-156), Taylor & Francis Group, London, UK.

[6] Groneman, A. and Zlatanova, S., 2009. TOPOSCOPY: a modelling tool for CITYGML. In: Onsrud H, R van de V (Eds.), GSDI Association, pp. 1-13.

[7] IAI, 2011. BuildingSMART. http://www.iai-tech.org/ [last accessed Feb-2012].

[8] IAI, 2008. IFC Model documentation webpage. International Alliance for Operability. http://www.iai-international.org/Model/R2x3_final/index.htm [last accessed Feb-2012].

[9] IfcExplorer 2008. IfcExplorer CityGML Export. http://www.ifcwiki.org/index.php/IfcExplorer_CityGML_Export [last accessed Feb-2012].

[10] Isikdag, S. and Zlatanova, S., 2009b. Towards defining a framework for automatic generation of buildings in CityGML using BIM, in Lee and Zlatanova (eds.), 3D geo-information sciences, LNG&C, Springer Verlag, pp. 79-96.

[11] ISO 16739, 2011. Industry Foundation Classes, Release 2x, Platform Specification. http://www.iso.org/iso/iso_catalogue/catalogue_tc/ catalogue_detail.htm?csnumber=38056 [last accessed Feb-2012].

[12] ISO 19107. 2011. Geographic information -- Spatial schema. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm ?csnumber=26012 [last accessed Feb- 2012].

[13] ISO 19109. 2011. Geographic information -- Rules for application schema. http://www.iso.org/iso/catalogue_detail.htm?csnumber=39891 [last accessed Feb- 2012].

[14] Khemlani, L., 2004. Analysis, Research and Review of AEC Technology. http://www.aecbytes.com/feature/2004/IFCmodel.html [last accessed Feb-2012].

[15] Kiziltas, S., Leite, F., Akinci, B. and Lipman, R. 2009. "Interoperable Methodologies and Techniques in CAD", CAD and GIS Integration, Auerbach Publications. ISBN 978-1-4200-6805-4

[16] Kolbe, T., 2008. Representing and Exchanging 3D City Models with CityGML, In: Jiyeong Lee and Sisi Zlatanova, Lecture Notes in Geoinformation and Cartography, 3D Geo-Information Sciences, Springer Berlin Heidelberg.

[17] Kolbe, T. and Bacharach, S., 2006. CityGML: An Open Standard for 3D City Models, http://www.directionsmag.com/articles/citygml-an-open-standard-for-3d-city-models/123103 [last accessed Feb- 2012].

[18] Kolbe T, Gröger G., 2004. Unified Representation of 3D City Models. Geoinformation Science Journal, Vol.4, No.1.

[19] Kolbe, T., Nagel, C. and Stadler, A., 2008. CityGML – A Framework for the Representation of 3D City Models from Geometry Acquisition to full Semantic Qualification. In: Proc. of ISPRS Congress 2008 in Bejing, China.

[20] Lapierre, A. and Cote, P., 2008. Using Open Web Services for urban data management: a testbed resulting from an OGC initiative offering standard CAD/GIS/BIM services, in Coors, Rumor, Fendel & Zlatanova (eds.): Urban and Regional Data Management; Taylor and Francis, London, pp. 381-393. Proceedings of the 26th UDMS, October 10-12, Stuttgart.

[21] Nagel, C., Stadler, A. and Kolbe, T., 2009. Conceptual Requirements for the Automatic Reconstruction of Building Information Models from Uninterpreted 3D Models, Academic Track of Geoweb 2009 Conference, Vancouver.

[22] Nagel, C., 2007. Conversion of IFC to CityGML; Meeting of the OGC 3DIM Working Group at OGC TC/PC Meeting, Paris (Frankreich), July 2007.

[23] OGC, 2011. http://www.opengeospatial.org/ [last accessed Feb- 2012]

[24] Pu, S. and Zlatanova, S., 2006. Integration of GIS and CAD at DBMS level, In: Fendel & Rumor (Eds); Proceedings of UDMS'06 Aalborg, Denmark May 15- 17, TU Delft, pp. 961-971.

[25] Safe Software, 2008. FME Desktop Translator/Converter Software http://www.safe.com/products/desktop/formats.php [last accessed Feb-2012].

[26] Van Oosterom, P.J.M., Lemmen, C.H.J., Ingvarsson, T., van der Molen, P., Ploeger, H.D., Quak, W., Stoter, J.E. and Zevenbergen, J. 2006. The

core cadastral domain model. In: Computers, Environment and Urban Systems: an International Journal, 30 (2006)5, pp. 627-660.

[27] Van Berlo, L., 2009. CityGML extension for Building Information Modelling (BIM) and IFC. Free and Open Source Software for Geospatial (FOSS4G), Sydney.

[28] Van Berlo, L. and De Laat, R., 2011. Integration of BIM and GIS: The Development of the CityGML GeoBIM Extension. 3D GeoInfo Conference, Berlin, Germany, Lecture Notes in Geoinformation and Cartography Series (Springer-Verlag, Heidelberg).

Growing Cloud Computing Efficiency

Dr. Mohamed F. AlAjmi, PhD
Vice Dean-Quality and Development
Head of Quality and E-Leaning units
Prince Sultan College of EMS
King Saud University
Riyadh, Saudi Arabia

Dr. Arun Sharma Head, Department
of Computer Science, Krishna
Institute of Engineering and
Technology, Ghaziabad-201206,
INDIA

Shakir Khan* M.Sc (Computer
Science) Researcher at E-Learning
and Distance Learning Deanship,
King Saud University, Riyadh Saudi
Arabia Nationality Indian

Abstract— **Cloud computing is basically altering the expectation for how and when computing, storage and networking assets should be allocated, managed and devoted. End-users are progressively more sensitive in response time of services they ingest. Service Developers wish for the Service Providers to make sure or give the ability for dynamically assigning and managing resources in respond to alter the demand patterns in real-time. Ultimately, Service Providers are under anxiety to build their infrastructure to facilitate real-time end-to-end visibility and energetic resource management with well grained control to decrease total cost of tenure for improving quickness. What is required to rethink of the underlying operating system and management infrastructure to put up the on-going renovation of data centre from the traditional server-centric architecture model to a cloud or network centric model? This paper projects and describes a indication model for a network centric data centre infrastructure management heap that make use of it and validates key ideas that have enabled dynamism, the quality of being scalable, reliability and security in the telecommunication industry to the computing engineering. Finally, the paper will explain a proof of concept classification that was implemented to show how dynamic resource management can be enforced to enable real-time service guarantee for network centric data centre architecture.**

Keywords- Cloud Computing; Data centre; Distributed Computing; Virtualization; data centre architecture.

I. INTRODUCTION

The random demands of the Web 2.0 era ingrouping with the desire to better utilize IT resources aredriving the need for a more dynamic IT infrastructure that canrespond to speedily changing requirements in real-time. This need for real-time enthusiasm is about to basically alter the data centre background and alter the IT infrastructure as we know it. In the cloud computing era, the computer can no longer be assumed in standings of the physical insertion i.e. the server or box, which households the processor, memory, storage and related components that establish the computer. In its place the "computer" in the cloud perfectly includes a group of physical work out resources i.e. work stations, retention, network bandwidth and storage, possibly circulated physically across server and geographical borders which can be planned on demand intoa dynamic consistent entity i.e. a "cloud computer", that can develop or shrink in real-time in order to promise the desired levels of potential sensitivity, performance, scalability, consistency and safety to any application that runs

in it. What is really supporting this alteration today is virtualization technology, more precisely hardware assisted server virtualization. At an ultimate level, virtualization technology allows the abstraction or decoupling of the request payload from the original physical resource. What this typically means is that the physical resources can then be carved up into logical or virtual resources as needed. This is acknowledged as provisioning. By introducing a suitable management infrastructure on top of this virtualization functionality, the provisioning of these logical resources could be made dynamic i.e.the logical resource could be made bigger or smaller in accordance with demand. This is known as dynamic provisioning. To enable a true "cloud" computer, each single computing element or resource should be proficient of being enthusiastically provisioned and succeeded in real-time. Currently, there are numerous holes and areas for development in today's data centre infrastructure before we can attain the above vision of a cloud computer.

A. *Server useful systems and virtualization*

Whereas networks and storage resources appreciates to advances in network facility management and SANs have already been proficient of being virtualized for a while, only now with the broader acceptance of server virtualization, do we have the complete basic foundation for cloud computing i.e. all computing properties can now be virtualized. Subsequently, server virtualization is the catalyst that is now motivating the transformation of the IT infrastructure from the traditional server-centric computing architecture to a network centric cloud computing architecture. When server virtualization is done, we have the capability to generate whole logical (virtual) servers that are free of the fundamental physical infrastructure or their physical position. We can postulate the computing, network and storage resources for all logical server (virtual instrument) and even transfer workloads from one virtual machine to another in real time (live migration).

All of this has aided deeply to convert the cost structure and competence of the data centre. Despite the many assistances that virtualization has allowed, we are still to realize the complete potential of virtualization with respect to the cloud computing. This is because:

1) *Usual server centric operating systems were not planned to manage collective spread resources:*

The Cloud computing example is all about optimally involving a set of scattered computing resources while the server-centric computing example is about devoting resources to a specific application. The server-centric example of computing fundamentally ties the application to the server. The work of the server operating system is to commit and guarantee to obtain ability of all accessible computing resources on the server to the application. If another application is installed on the same server, the operating system will once again manage the entire server resources to confirm that each application remains to be checked as if it has access to all available resources on that server. This model was not designed to allow for the "dial-up" or "dial down" of resource allocated to an application in response to change workload demands or business priorities. This is the reason that load-balancing and clustering was introduced.

2) Current hypervisors do not supply sufficient division between application management and physical supply management:

Today's hypervisors have just interposed themselves one level down below the operating system to enable multiple "virtual" servers to be hosted on one physical server.While this is great for consolidation, once again there is no way for applications to manage how, what and when resources are assigned to themselves without having to concern about the management of physical resources. It is our observation that the current generation of hypervisors which were also born from the era of server-centric computing does not define hardware management from application management much similar the server operating systems themselves.

3) Server virtualization does not yet allow contribution of scattered resources:

Server virtualization currently permits a single physical server to be structured into multiple logical servers. However, there is no way for example to generate a analytical or computer-generated server from resources that may be physically placed in separate servers. It is true that by virtue of the live migration capabilities that server virtualization technology enables, we are intelligent to move application loads from one physical server to another possibly even geographically distant physical server. However, moving is not the similar as sharing. It is our contention that to enable a truly distributed cloud computer, we must be able efficiently to share resources, no problem where they exist in purely based on the potential constraints of applications or services that consume their sources.

B. Storage set of connections & virtualization

Before the production of server virtualization, storage networking and storage virtualization permitted many improvements have been done in the data centre. The key improvement was the introduction of the FibreChannel (FC) protocol and Fibre Channel-based Storage Area Networks

(SAN) which delivered great speed of storage connectivity and dedicated storage solutions to allow such profits as server-less backup, point to point reproduction, HA/DR and presentation optimization outside of the servers that run applications. However, these pay backs have come with improved management complication and costs.

C. System virtualization

The virtual networks now applied inside the physical server to switch between all the virtual servers to provide a substitute to the multiplexed, multi-patched network channels by trucking them nonstop to WAN transport, thus shortening the physical network infrastructure.

D. Function creation and binding

The existing method of exhausting Virtual Machine images that contain the application, OS and loading disk images is once again born of a server-centric computing model and does not provide itself to enable supply across mutual resources. In a cloud computing pattern, applications should preferably be built as a collection of facilities which can be integrated, disintegrated and distributed on the fly. Each of the services could be measured to be individual procedure of a larger workflow that establishes the application. In this way, individual services can be arranged and provisioned to improve the overall performance and potential requirements for the application.

II. PLANNED SUGGESTION STRUCTURAL DESIGN MODEL

If we were to purify the above interpretations from the previous section, we can realize that a couple of key subjects emerging. That is:

A. *The next generation architecture for cloud computing must entirely decouple physical resources management from virtual resource management.*

B. *Supply the proficiency to intervene between applications and resources in real time.*

As we stressed in the earlier section, we are still to attain perfect decoupling of physical resources management from virtual resource management but the outline and improved acceptance of hardware assisted virtualization (HAV) as a significant and essential step towards this objective. Thanks to HAV, a next generation hypervisor will be capable to achieve and truly guarantee the identical level of access to the fundamental physical resources. Moreover, this hypervisor should be proficient of handling both the resources situated locally inside a server as well as any resources in other servers that may be situated somewhere else physically and linked by a network. Once the controlling of physical resources is decoupled from the virtual resource management. The necessity for a mediation layer that mediates the distribution of resources between various applications and the shared distributed physical resources becomes obvious.

Figure 2 Position Architecture Model for Next Generation Cloud Computing Infrastructure
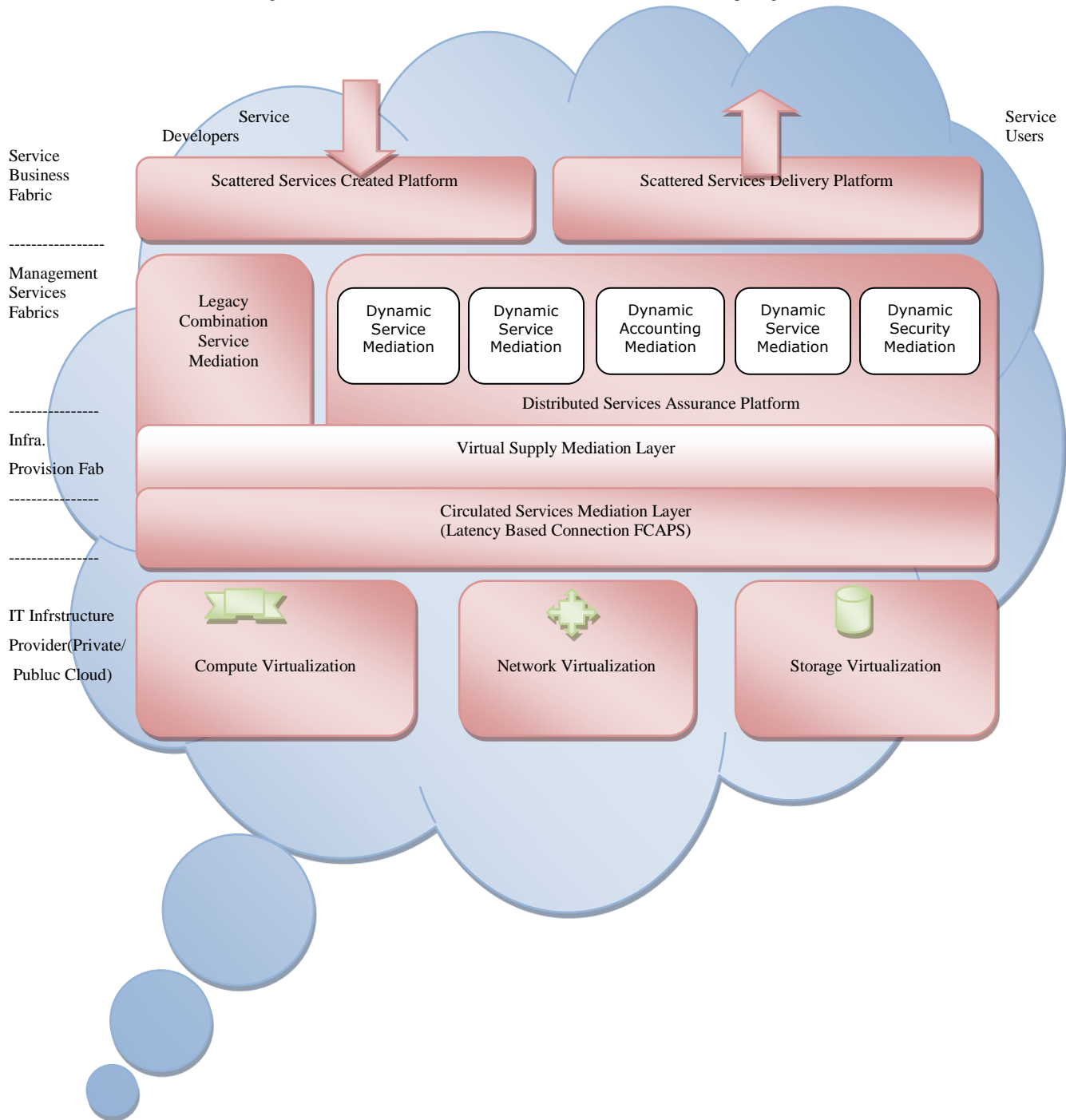
III. INFRASTRUCTURE PROVISION FABRICS

This layer includes two pieces. Together with the two components allow a computing resource "dial-tone" that delivers the basis for provisioning resource fairly to all applications in the cloud.

A. Scattered services mediation

This is a FCAPS based (Fault, Configuration, Accounting, Performance and Security) abstraction layer that enables

autonomous self-management of every individual resource in a network of resources that may be distributed geographically.

B. Virtual supply mediation layer

This gives the ability to create logical virtual servers with a level of service guarantee those assurances resources such as number of CPUs, memory, bandwidth, latency, IOPS (I/Ooperations per second), storage through put and capacity.

C. Circulated services Assurance Platform

This layer will allow for creation of FCAPS-managed virtual servers that pack and host the desired choice of OS to allow the loading and execution of applications. Since the virtual servers implement FCAPS-management, they can give automated mediation services natively to guarantee fault management and reliability (HA/DR), performance optimization, accounting and security. This describes the management dial-tone in our orientation architecture model.

D. Scattered Services Delivery Platform

This is basically a workflow engine that implements the application which we described in the previous section, is preferably composed as business workflow that organizes a number of distributable workflow elements. This describes the services dial tone in our reference architecture model.

E. Scattered Services Creation Platform

This layer gives the tools that developers will utilize to generate applications defined as group of services which can be composed, decomposed and scattered on the fly to virtual servers that are automatically shaped and run by the distributed services assurance platform.

F. Legacy Combination Services Mediation

This is a layer that gives addition and support for existing or legacy application in our reference architecture model.

IV. DEPLOYMENT OF THE SUGGESTION MODEL

Any generic cloud service platform requirements must address the needs of four categories of stake holders:

1) *Infrastructure suppliers,*
2) *Service suppliers.*
3) *Facility Developers*
4) *End Users.*

Below we explain how the reference model described will affect, benefit and are set up by each of the above stake holders.

A. Infrastructure suppliers

These are vendors who give the underlying computing, network and storage resources that can be fixed up into logical cloud computers which will be dynamically forced to deliver extremely scalable and globally interoperable service network infrastructure. The infrastructure will be utilized by both service creators who develop the services and also the end users who use these services.

B. Service suppliers

With the employment of our innovative reference architecture, service providers will be capable to promise both service developers and service users that resources will be obtainable on demand. They will be capable effectively to determine and meter resource utilization end-to-end usage to allow a dial-tone for computing service while management Service Levels to meet the availability, performance and security needs for each service. The service provider will now handle the application's link to computing, network and storage resource with suitable SLAs.

A. Facility developers

They will be able to develop cloud based services using the management services API to configure, monitor and manage service resource allocation, availability, utilization, performance and security of their applications in real-time. Service management and service delivery will now be integrated into application development to allow application developers to be able to specify run time SLAs.

C. End users

Their demand for selection, mobility and interactivity with sensitive user interfaces will continue to rise. The managed resources in our reference architecture will now not only permit the service developers to generate and distribute services using logical servers that end users can dynamically provision in immediate to respond for changing needs, but also provide service providers the ability to charge the end-user by metering correct resource handling for the required SLA.

V. CONCLUSIONS

In this paper, we have explained the needs for implementing a real dynamic cloud computing infrastructure which contains a group of physical computing resources i.e. processors, memory, network bandwidth and storage, potentially dispersed physically through server and geographical limits which can be controlled on demand into a dynamic reasonable entity i.e. "cloud computer", that can develop or reduce in size immediately in order to give surety about the desired levels of latency, sensitivity, performance, scalability, consistency and security to any application that runs in it. We worked out few key areas of shortage of current virtualization and management technologies. Particularly we explained detail importance of sorting out physical resource management from virtual resource management and why current operating systems are not designed and hence it was suitable to provide this ability for the distributed shared resources especially of cloud deployment. We also painted the need for FCAPS-based (Fault, Configuration, Accounting, Performance and Security) service "mediation" to give global administration functionality for all networked physical resources that include a cloud, irrespective of their allocation across a lot of physical servers in different geographical locations. We then projected an indication architecture model for a distributed cloud computing mediation (management) platform which will outline the foundation for making the possibility of next generation cloud computing infrastructure. We proved how this infrastructure will involve as well as advantage key stake holders such as the Infrastructure providers, service providers, service developers and end-users.

Description in this paper is considerably different from most current cloud computing solutions that are nothing more than hosted infrastructure or applications accessed over the Internet. The proposed architecture in this paper will significantly change the current setting by enabling cloud computing service providers to give a next generation infrastructure platform which will recommend service developers and end users exceptional control and enthusiasm in real-time to assure SLAs for service latency, availability, performance and security.

REFERENCES

[1] Adriana Barnoschi, "Backup and Disaster Recovery For Modern Enterprise", 5th International Scientific Conference, Business and Management' 2008, Vilnius, Lithuvania.http://www.vgtu.lt/leidiniai/leidykla/BUS_AND_MANA_2008/infcommunication/ 630-635-G-Art-Barnoschi00.pdf

[2] David Chisnall, "guide to the xen hypervisor", First edition, Prentice Hall, Press, NJ, 2009.

[3] Gartner's 2008 Data centre Conference Instant Polling Results: Virtualization Summary – March 2, 2009.

[4] Graham Chen, Qinzheng Kong, Jaon Etheridge and Paul Foster,"Integrated TMN Service Management", Journal of Network and Systems Management, Springer New York, Volume 7, 1999,p469-493.

[5] Jason A. Kappel, Anthony T. Velte, Toby J. Welte, "Microsoft Virtualization with Hyper-V", McGraw Hill, New York, 2009.

[6] RaoMikkilineni, Vijay Sarathy "Cloud Computing and Lessons from the Past", Proceedings of IEEE WETICE 2009, FirstInternational Workshop on Collaboration & Cloud Computing, June 2009.

[7] Rajkumar Buyyaa, Chee Shin Yeoa, SrikumarVenugopala,James Broberga, and Ivona Brandicc, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems, Volume 25, Issue 6, June 2009, Pages 599-616.

AUTHORS PROFILE

Dr Mohammed Fahad AlAjmi was born in Kingdom of Saudi Arabia. He received his Ph.D in Pharmacy from King Saud University, Riyadh, Saudi Arabia in 2007.He chaired many position in the university and currently working as vice dean for quality and development in Prince Sultan College for EMS affiliated to King Saud University. To date he taught many pharmacy students, more than 30 courses. Students' level varies from primary to undergraduate levels.

Dr. Arun Sharma, alumni of IIT Roorkee and Thapar University, received his M.Tech. (Computer Science and Engineering) from Punjabi University, Patiala, INDIA and Ph.D. (Computer Science) Thapar University, Patiala, INDIA. Currently, he is working as head of the department of Computer Science and Engineering Department in KIET school of engineering and Technology at Ghaziabad, India. His areas of interests include Software Engineering, Soft Computing and Database Systems. He has published a number of papers in international Journals and Conferences including IEEE, ACM, Springer, WILEY and others in India and abroad. Dr. Sharma is an active member of IEEE, ACM and Computer Society of India. He is also a member of Board of Studies (BoS) of Mahamaya Technical University (MTU), Noida. He is also on the panel of subject experts and examination for various Universities like IGNOU, BBA University (Central), Lucknow, GGSIP University, Delhi, Thapar University, Patiala, and others. He is also an active member of Editorial Board and Review Committee of several Journals including Journal of Computer Science (USA), International Journal of Computer Science and Security (Malaysia), Research Journal of Information Technology, USA and others.

Shakir Khan was born on 5th Feb, 1978 at Kallanheri in Saharanpur district UP, India. He is working as a Researcher in College of Electronic Learning in King Saud University, Kingdom of Saudi Arabia. He received his Master of Science in Computer Science from Jamia Hamdard (Hamdard University), New Delhi, India in the year 2005, and PhD computer Science scholar in Manav Bharti University, Solan (HP) India since 2010. He is member of IEEE. He has actively attended many international conferences and published various research papers in National and International conferences as well as journals. His current areas of interests are in Cloud Computing, Software Engineering, Data Mining and E Learning. Apart from that he worked in the field of Software Development in different MNC companies at Noida India.