



Structural Features for Character Recognition System-A Review

P.E.Ajmire
Dept. Of Comp.Sci.
G. S. Sci., Arts & Comm. College, Khamgaon.
peajmire@rediffmail.com

V M Thakare
Dept. Of Comp.Sci.
P.G.T.D, S.G.B.A.U. Amravati
vilthakare@yahoo.co.in

Rajeev Dharaskar
Dept. Of Comp. Sci.,
G H R C O E, Nagpur.
rvdharaskar@yahoo.com

S E Warkhede
Dept. Of Comp. Sci.,
Science College, College, Khamgaon Malkapur.
shankarwarkede@yahoo.co.in

Abstract-This paper presents a review of structural features for character recognition. Structural features are the features that are physically a part of the structure of the character, such as straight lines, arcs, circles, intersections etc. Selection of a feature extraction method is probably the single most important factor in achieving high recognition performance in character recognition systems. Different feature extraction methods are designed for different representations of the characters, such as solid binary characters, character contours, skeletons (thinned characters) or gray-level subimages of each individual character. The feature extraction methods are discussed in terms of invariance properties, reconstruction ability, expected distortions and variability of the characters. When a few promising feature extraction methods have been identified, they need to be evaluated experimentally to find the best method for the given application. In this paper, we discuss the selection of appropriate standard structural features for character recognition

Keywords: Statistical features, Feature Extraction, character recognition.

I. INTRODUCTION

There are many scripts and languages in the world. There are nearly 180 scripts and 6000 languages currently spoken, but not all of them have a writing system and some of them use modified forms of existing writing systems. Character recognition is a classic problem in the field of image processing. Lot of research has been done on recognition of various characters and numerals of different scripts and languages. Many researchers from different disciplines are working in this field. Most of the work done in the field of character recognition is confined to Roman, English, Arabic [1], Japanese [2], Thai [3] and Devnagari [4], Chinese[5] languages. Now a day some efforts have been reported in literature for Devnagari[6], Bangala[7], Tamil[8], Malayalam[9], Gurumukhi[10]. Although there are many scripts and languages in India but not much research work is done for handwritten Marathi characters[11,12].

In fact the main problem in OCR system is the large variation in shapes within a class of character. This variation depends from font styles, document noise, photometric effect, document skew and poor image quality. The large variation in shapes makes it difficult to determine the number of features that are convenient prior to model building. The performance of a character recognition system depends heavily on what features are being used. Though many kinds of features have been developed and their test performances on standard database have been reported, there is still room to improve the recognition rate by developing an improved feature. Selection of a feature extraction method is probably the single most important factor in achieving high recognition performance in character recognition systems. They discussed feature extraction methods in terms of invariance properties, reconstructability and expected distortions and variability of the characters.

II. STRUCTURAL PATTERN RECOGNITION

There are two fundamental approaches for implementing pattern recognition systems, statistical and structural, employ different techniques for description and classification. Statistical approaches to pattern recognition use decision theoretic concept to discriminate among objects belonging to different groups based upon their quantitative features. Structural approaches to pattern recognition use syntactic grammars to discriminate among objects belonging to different groups based upon the arrangement of their morphological (i.e., shape based or structural) features. Structural pattern recognition sometimes referred to as syntactic pattern recognition due to its origins in formal language theory, relies on syntactic grammars to discriminate among data from different groups based upon the morphological interrelationships (or interconnections) present within the data. Structural pattern recognition systems have proven to be effective for data which contain an inherent, identifiable organization such as image data (which is organized by location within a visual rendering) and time series data (which is organized by time). The usefulness of structural pattern recognition systems, however, is limited as a consequence of fundamental complications associated with the implementation of the description and classification tasks.

III. STRUCTURAL FEATURES

The advantage of the structural approach is that it provides a good symbolic description of the image; however, this feature is more useful for synthesis than analysis tasks. Structural features describe a pattern in terms of its topology and geometry by giving its global and local properties. The following figure shows the two approaches.

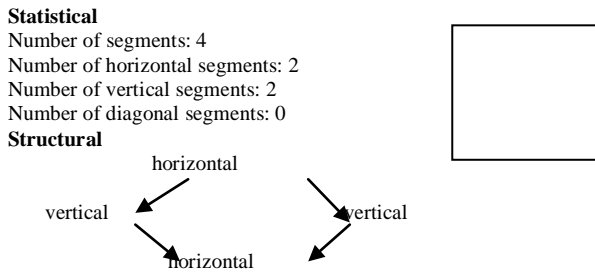


Figure 1. The statistical and structural approaches to pattern recognition applied to a common identification problem.

Some of the main structural features include features like number and intersections between the character and straight lines, holes and concave arcs, number and position of end points and junctions [13]. Ahmad T. Al-Taani and Saeed Al-Hajj[14] proposed method can easily be applied to any application that requires Arabic handwritten character recognition, regardless of its computing power. A variety of simple shapes (e.g. circles, triangles, squares, super ellipses and so on) are optimally fitted to each region within a segmented photograph. The system automatically chooses the shape that best represents the region; the choice is made via a supervised classifier so the “best shape” depends on the subjectivity of a user.[15]. Due to the large variety of the available feature extraction methods, many researchers have turned towards the use of several feature extractors with more complex structures of classification, such as multistage classification schemes or “parallel” combination of multiple classifiers.[16]. Some researchers also used compound feature extraction approach such as extracting various supportive features like affine invariant moments, image thinning, structuring the image in box format based on structural analysis for recognition of offline handwritten Marathi vowels.

A new online structural pattern recognition approach has been proposed[17] to recognize the handwritten digits; the primitives are determined by identifying the changes in the slope’s signs around the zero and the infinity values (break points). This technique is independent of the type of drawing (upward or downward). Verma[18] proposed a method for cursive handwriting recognition. He used a contour code feature and a rule-based segmentation in the recognition process

IV. CLASSIFICATION TECHNIQUES

The classifier for a structural pattern recognition system is composed of a set of syntactic grammars, one for each group among which discrimination is being performed, and a parser. The identification generated by the classifier is the group whose associated syntactic grammar successfully parses the primitives extracted from the data. An adjudication scheme is necessary to resolve the situation where there is more than one successful parse. The main difficulty in developing the classifier for a structural pattern recognition system lies in constructing the syntactic grammars. Since the grammars embody the precise criteria which differentiate among the groups, they are by their very nature domain and application specific. Complicating matters is the lack of a general solution for extracting structural features from data, causing the primitives used within the grammars to vary among domains, identification problems, and pattern

recognition systems. Grammar inference techniques can be used to construct automatically a grammar from examples, but these methods can fail in the most general cases such as when the grammar for a group is context free. A fuzzy membership function is used to estimate the likelihood of the primitives and a tree-like classifier based on extracted feature points, primitives and fuzzy membership is applied for classification of numerals.

V. CONCLUDING REMARKS

A structural approach to pattern recognition necessarily requires generating a morphological description of the data under analysis and then performing classification based on that description. A general approach to producing such a morphological description of image data is one that finds an optimal piecewise approximation by fitting a sequence of structures selected from a library of domain independent structure types. The task of techniques generating a morphological description of a data set, therefore, becomes one of identifying a sequence of structures that approximate the data such that the difference between the approximation and the data is minimized.

The paper has concentrated on an appreciation of principles and methods. We have not attempt to compare the effectiveness of algorithms, or to discuss the crucial topic of evaluation. In reality, it would be very difficult to assess techniques separated from the systems for which they were developed. We believe that the hybrid approach may give the better recognition of character or numeral. The standard database and advance techniques will be carried out to shed light on this issue.

VI. REFERENCES

- [1] I.A. Jannoud “Automatic Arabic Hand Written Text Recognition System” American Journal of Applied Sciences 4 (11): 857-864, 2007.
- [2] Tour Wakahara, Y. Kimura & Mutsuo “ Handwritten Japanese Character Recognition Using Adaptive Normalization by Global Affine Transformation.” Proc. 6th ICDAR Vol., Issue , 2001.
- [3] J. L .Mitranont, U. Limkonglap “Using Countour Analysis to improve Feature Extraction in Thai Handwritten Character Recognition Systems” Proc. 7th IEEE ICCIT , 2007.
- [4] U. Pal, N.Sharma, T.Wakabayashi and F. Kimura. “Off-line Handwritten character recognition of Devnagari Script” 9th ICDAR, 2007.
- [5] Bing Feng, Xiaoqing Ding, Ypushou Wu “Chinese Handwriting Recognition using Hidden Markov Models” p no 1051-4651/02 IEEE 2002.
- [6] H. Swethalakshmi1, Anitha Jayaraman1, V. Srinivasa Chakravarthy2, C. Chandra Sekhar “Online Handwritten Character Recognition of Devanagari and Telugu Characters using Support Vector Machines” Indian Institute of Technology Madras, Chennai - 600 036, India.
- [7] T. K. Bhowmik, A Roy & U Roy “Character Segmentation for Handwritten Bangala Words using Artificial Neural Network” Proc.1st IAPR TC3 NNLDAR, 2005.

- [8] R. Indra Gandhi, K Iyakutti, "An attempt to recognize Handwritten Tamil Character using Kohonen SOM" 188-192 Vol. 1(3) IJANA 2009.
- [9] Bindu S Moni, G Raju "Modified Quadratic Classifier and directional features for Malayalam Character recognition", NCCSE 2011.
- [10] Kartar Singh Siddharth, Renu Dhir, Rajneesh Rani, "Handwritten Gurumukhi Character Recognition Using Zoning Density and Background Directional Distribution Features" ,1036-1041 Vol. 2 (3) , IJCSIT 2011.
- [11] P.E.Ajmire and S E Warkhede "Handwritten Marathi Character (Vowel) Recognition" Advances in information mining, Vol 2, 2010.
- [12] P E Ajmire, R V Dharaskar and V M Thakare " A Comparative study of Handwritten Marathi Character Recognition", pp 29-32, NCIPET 2012.
- [13] M. K. Jindal, R. K. Sharma and G. S. Lehal "Structural Features for Recognizing Degraded Printed Gurumukhi Script" 978-0-7695-3099-4/ IEEE 2008.
- [14] Ahmad T. Al-Taani and Saeed Al-Ha "Recognition of On-line Arabic Handwritten Characters Using Structural Features", Journal of Pattern Recognition Research 2010.
- [15] Yi-Zhe Song, P. L. Rosin, P. M. Hall and John Collomosse, "Arty Shapes" Computational Aesthetics in Graphics, Visualization, and Imaging (2008).
- [16] L. Heutte, T. Paquet, J.V. Moreau, Y. Lecourtiera, C. Olivier "A structural statistical feature based vector for handwritten character recognition" Pattern Recognition Letters, Elsevier Science B.V.1998.
- [17] Ahmad T. Al-Taani "An Efficient Feature Extraction Algorithm for the Recognition of Handwritten Arabic Digits", World Academy of Science, Engineering and Technology, 2006.
- [18] Brijesh Verma, "A Contour Code Feature Based Segmentation For Handwriting Recognition". Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003).