

A Comparative Study of Handwritten Marathi Character Recognition

P.E.Ajmire
Deptt. Of Comp.Sci
G. S. Sci., Arts & Comm.
College, Khamgaon

RV Dharaskar
Deptt. Of Comp. Sci
G H R COE, Nagpur.

V M Thakare
Deptt.Of Comp. Sci.,
PGTD, S G B
AmravatiUniversity, Amravati

ABSTRACT

The different pattern recognition models have been proposed in recent years and the different research groups are working on for the recognition result. Handwritten character recognition for any Indian writing system is rendered complex because of the presence of composite characters. Hence the selection of a feature extraction method is probably the most important factor in achieving high recognition performance for Marathi character recognition. The goal of this paper is to present comparative study of various character recognition techniques used for feature extraction and recognition of handwritten Marathi character.

General Term

Pattern Recognition.

Keyword

Feature Extraction, Handwritten character recognition, Pattern matching.

1.INTRODUCTION

Although there are many scripts and languages in India but not much research work is done for handwritten Marathi characters. Marathi handwritten character recognition is the challenging task in the pattern recognition field. Different statistical methods for Marathi handwritten character recognition have been proposed in recent years. Marathi is an Indo-Aryan language spoken by about 71 million people mainly in the Indian state of Maharashtra and neighbouring states. Marathi is also spoken in Israel and Mauritius. Marathi is thought to be a descendent of Maharashtri, one of the Prakrit languages which developed from Sanskrit. Marathi first appeared in writing during the 11th century in the form of inscriptions on stones and copper plates. Marathi was written in Modi script — a cursive script minimizes the lifting of pen from paper while writing. Most writings of the Maratha Empire are in Modi script. However, Persian-based scripts were also used for court documentation. With the advent of large-scale printing, Modi script fell into disuse, as it proved very difficult for type-setting. Currently, due to the availability of Modi fonts and the enthusiasm of the younger speakers, the script is far from disappearing. Now, Marathi is written in the Devnagari script, a set consist of 16 vowels and 36 consonants making a total of 52 letters. It is written from left to right. Devnagari used to write Marathi is slightly different than that of Hindi or other languages. Earlier, another script called 'Modi' was in use till the time of the Peshwas (18th century). This script was introduced by HemadPanta, a minister in the court of the Yadava kings of Devgiri (13th century). This script looked more like today's dravidian scripts and offered the advantage of greater writing speed because the letters could be joined together. Today only the Devnagari script is used which is easier to read but does not have the advantage of faster writing. The script currently used in Marathi is called 'Balbodh' which is a modified version of Devnagari script. From this script,

Marathi language has 12 common vowels and for this the database is created for these vowels. Vowels are the soul of the speech and sounds of the language. In Marathi vowels are usually written in abbreviated symbols or forms. From the 13th century until the mid-20th century, it was written with the Modi alphabet. Since 1950 it has been written with the Devnagari alphabet. Today we have 13 vowels and 36 consonants and there are shown in following figure.

अ	आ	इ	ई	उ	ऊ	ऋ
ए	ऐ	ओ	औ	अं	अः	

Vowels

क	ख	ग	घ	ङ	
च	छ	ज	झ	ञ	
ट	ठ	ड	ढ	ण	
त	थ	द	ध	न	
प	फ	ब	भ	म	
य	र	ल	व	श	
ष	स	ह	ळ	क्ष	ज्ञ

Consonants

Recognition of handwritten characters has been popular research areas since 1870. There are many script and Languages in the world. The researchers have done work on some of them like English, Chinese, Latin, Arabic [1], Japanese [2], Thai [3] and Devnagari [4]. There are many pieces of work have been done towards the recognition of printed Devnagari Character[4] and at present OCR systems are commercially available for some of the printed Indian scripts, many research has been carried out for Bangala[5]. The research of online character recognition started in the 1960s and has been receiving intensive interest from the 1980s. The comprehensive survey of Tappert et al. reviewed the status of research and applications before 1990 and early works of online Japanese character recognition have been reviewed.[6]. A recent comprehensive survey of handwriting recognition, by Plamondon and Srihari, mainly concerns western handwriting. This paper contributes a survey to online Chinese character recognition (OLCCR) since this recognition problem is very

different from western handwriting recognition and it poses special challenge. Among studies on Indian scripts, notable work has been done on recognition of printed Devnagari characters by VeenaBansal and R.M.K Sinha[7]. The researchers suggested contextual post-processing for Devnagari character recognition and text understanding. For handwritten Bengali numeral recognition presented Water Reservoir Analogy Principle for structural features of Bangla numerals[8]. A prototype Devnagari numeral recognition system [9] was implemented using C on Sun's Spark System. For experimentation, samples were collected from a number of different individuals. Four samples of each numeral from eight different persons were used in the training of the neural net. For every feature set, ten SMR modules were trained independently for each numeral. Three to six style classes were obtained depending on the type of numeral. For descriptive features twelve types of segments were identified. Twenty additional samples were used for training the feed-forward superstructure of the Kohonen modules [10] and the integration network. The

Pattern Recognition Models

Approach	Representation	Recognition Function	Typical Criterion
Template matching	Samples, pixels, curves	Correlation, distance measure	Classification error
Statistical	Features	Discriminant function	Classification error
Syntactic or structural	Primitives	Rules, grammar	Acceptance error
Neural networks	Samples, pixels, features	Network function	Mean square error

capability of the Kohonen net in identifying style categories becomes clear from the results in table 1. The Kohonen module for numeral five trained with density features was used for categorising samples of the numeral five written by different individuals. It is expected that a person writes a numeral in one or two distinct ways. Results of the categorization experiment vindicated this fact. Numerals written by an individual have been categorized into at most two style categories. Although results of only one module have been presented here, similar capabilities are observed for other modules as well. It is also found that these style modules have rejection capability (for other numerals) of around seventy percent on average.

The following table shows different pattern recognition models.

2. ANALYSIS OF PRESENT RECOGNITION OF VARIOUS CHARACTERS

2.1 English Character

Long back R.W.S. Tregidgo [11] proposed a system for High Performance English Character Recognition for off-line Handwritten British Post Codes. These are 5-7 alphanumeric codes e.g NAA.2673 after using Enhanced Loci Algorithm on 4800 alphanumeric characters, the overall recognition rate for the numerals alone was 82.4% and for alphabetic characters alone was 75%. The overall recognition rate for alphanumeric characters was 63.1%.

2.2 Address Recognition

Chin Keong Lee and Graham Leedham [12] proposed a system for An Intelligent System for Conflict Resolution in Handwritten Address Recognition. The proposed expert system resolves the conflicts and reduces the error rates by fusing a holistic pattern recognition method with expert knowledge based on a posterior information. The system was evaluated

using 1,071 handwritten Singapore addresses. Experimental results show that the expert system achieved a significant reduction in error rates. Performance was improved from 71.2% correctly sorted, 4.8% reject (cannot sort) and 24.0% error (wrongly sorted) rates using OCR only to 63.7% correctly sorted, 35.7% reject (cannot sort) and 0.6% error (wrongly sorted) rates using the proposed approach. S. S. Loren Schwiebertand Jatinder Singh Bedi [13] presented a paper on "WAVELET PACKET TRANSFORM AND NEURO-FUZZY APPROACH TO HANDWRITTEN CHARACTER RECOGNITION" and it is found that the proposed method is more efficient for handwritten character recognition as well as personal identification compared to energy sorted wavelet transform of character images, since characters contain very few edges in the images. Simulation of characters is done for 3 multi resolution levels using symmlet and results show that this method is more efficient than the methods using only fuzzy logic.

2.3 Arabic character

In American Journal of Applied Sciences 4 (11) of 2007, Ismael Ahmad Jannoud [14] claimed that the "Automatic Arabic Hand Written Text Recognition System" system has been tested using different input Arabic documents. The outputs are directly displayed on a word file which has been opened and each recognized character appended to this file. This word file will have all the text in the Arabic handwritten input document. The proposed system was efficient and the performance was about 90% differs from types of the characters. The biggest recognition rate was for the isolated characters (99%). But the middle characters has worst recognition rate (91%).

2.4 Thai character

Dr. Jarensri L. Mitranont, and UrairatLimkonglap [15] proposed "Using Contour Analysis to Improve Feature Extraction in Thai Handwritten Character Recognition Systems" and they stated that the highest Character Recognition Rate (CRR) reached 95.35%. Additionally, the result of the overall evaluation showed that the THW-CR system generated reliable results for improving the accuracy of the feature extraction rate and the character recognition rate compared with the previous research by 3.62% and 8.33%, respectively.

2.5 Bangla words

T. K. Bhowmik , A. Roy and U. Roy [5] stated that when the feature vector size is 20, that is our L is 8 we have experienced the best result with 88% accuracy, feature vectors of size 14, 24 and 34 also give reasonably good result in their paper entitled "CHARACTER SEGMENTATION FOR HANDWRITTEN BANGLA WORDS USING ARTIFICIAL NEURAL NETWORK".

2.6 Devnagari Script

VeenaBansal, during her Doctorate work on Devnagari Text Recognition, she got the recognition rate of 85% for printed text used from dictionary and 70% when independent classification were used. Similarly VeenaBansal and R. M. K. Sinha proposed algorithm extensively uses structural properties of the script. Statistical information about the height and width of character boxes, which are vertically separate from their neighbors, is used to hypothesize character boxes to be touching character boxes. The recognition rate of 85% has been achieved on the segmented touching characters.

Dr. P. S. Deshpande, Mrs. Latesh Malik, Mrs. SandhyaArora [16] mentioned in "Handwritten Devnagari Character

Recognition Using Connected Segments And Minimum Edit Distance” that they achieved 95% of accuracy for hand written Devnagari Characters.

2.8 Marathi Handwritten Numerals

R. J. Ramteke, P. D. Borkar, S. C. Mehrotra [17] describes an approach based on invariant moments for recognition of isolated Marathi Handwritten Numerals and their divisions. The proposed technique is independent of size, slant, orientation, translation and other variations in handwritten characters. Handwritten Marathi Characters/Numerals are more complex for recognition than corresponding English characters due to many possible variations in order, number, direction and shape of the constituent strokes. The Gaussian Distribution Function has been adopted for classification. The success rate of the method is found to be 87%. Similarly, for the vowel character recognition work the author[18] treats isolated Characters as an image of 40X40 pixel size. Seven invariant central moments of the image and two additional feature sets are evaluated. 10 samples of each number from 20 different writers have been sampled and prepared a database has been made. Seven central invariant moments are evaluated for each character and its parts by dividing it by two different ways. In all, there are 14 features corresponding to each character. The Gaussian Distribution Function has been adopted for classification.

3. 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.Mitrpanont, 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] T.K.Bhowmik, A Roy & U Roy “Character Segmentation for Handwritten Bangla Words using Artificial Neural Network” Proc.1st IAPR TC3 NNLDAR, 2005.
- [6] Cheng-Lin Liu, Member, Stefan Jaeger, and Masaki Nakagawa “Online Recognition of Chinese Characters: The State-of-the-Art” IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 26, NO. 2, FEBRUARY 2004.
- [7] Veena Bansal and R. M. K. Sinha “Segmentation of Touching Characters in Devanagari” Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India.
- [8] U. Pal A. Belaid B. B. Chaudhuri “A System for Bangla Handwritten Numeral Recognition” Indian Statistical Institute, Culcutta.
- [9] Cheng-Lin Liu, Member, Stefan Jaeger, and Masaki Nakagawa “Online Recognition of Chinese Characters: The State-of-the-Art” IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 26, NO. 2, FEBRUARY 2004.
- [10] R. Bajaj , L. Dey and Santanu Chaudhury “Devnagari numeral recognition by combining decision of multiple connectionist classifiers” Sadhana Vol. 27, Part 1, pp.59-72, February 2002.
- [11] R.W.S. Tregidgo and A.C. Downton “High Performance English Character Recognition for off-line Handwritten British Post Codes” Dept. of Electronic System Engineering, University of Essex, UK.
- [12] Chin Keong Lee and Graham Leedham “An Intelligent System for Conflict Resolution in Handwritten Address Recognition” International Journal of Information Technology, Vol. 12 No. 1.
- [13] SREELA SASI LOREN SCHWIEBERT, ” Dept. of Electrical Engineering, University of Idaho, Moscow & JATINDER SINGH BEDI Dept. of Electrical Engineering and Computer Engineering, Wayne State University, Detroit, “WAVELET PACKET TRANSFORM AND NEURO-FUZZY APPROACH TO HANDWRITTEN CHARACTER RECOGNITION”
- [14] Ismael Ahmad Jannoud “Automatic Arabic Hand Written Text Recognition System” American Journal of Applied Sciences 4 (11):p.no. 857-864, Science Publications 2007.
- [15] Jarernsri L. Mitrpanont, and Urairat Limkonglap “Using Contour Analysis to Improve Feature Extraction in Thai Handwritten Character Recognition Systems” Seventh International Conference on Computer and Information Technology, pp. 668-673 IEEE, 2007.
- [16] Dr. P. S. Deshpande, Mrs. Latesh Malik, Mrs. Sandhya Arora “Handwritten Devnagari Character Recognition Using Connected Segments And Minimum Edit Distance” IEEE 2007.
- [17] R. J. Ramteke, P. D. Borkar, S. C. Mehrotra “Recognition of Isolated Marathi Handwritten Numerals: An Invariant Moments Approach”, Proceedings of the International Conference on Cognition and Recognition.
- [18] Ajmire P.E. and Warkhede S.E. “Handwritten Marathi character (vowel) recognition” Advances in Information Mining, ISSN: 0975-3265, Volume 2, Issue 2, 201
- [19] 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.
- [20] Ovid Due Trier, Anil K Jain and Torfinn Taxt “Feature extraction methods for Character Recognition-A Survey” Pattern Recognition, Vol 29, NO.4 pp 641-662. 1996.
- [21] Richard G Casey and Eric Lecolinet “A Survey of methods and Strategies in Character Segmentation” IEEE Transaction on Pattern Analysis and Machine intelligence. Vol 18, NO.7 July 1996