

Feature Extraction and recognition of Modifiers in Handwritten Marathi (Devanagari) Text

P.E.Ajmire¹, R V Dharaskar², V M Thakare³

¹Dept. Of Computer Science,
G S Sci., Arts & Commerce College, Khamgaon(M.S), India

²Former Director, Disha Education Society,
Raipur.(C.G), India

³ Professor and Head,P.G.T.D. Of Comp. Sci.,
S G B Amravati University, Amravati.

Abstract

This research work deals with the feature extraction and recognition of Modifiers in Marathi Text (Devanagari). Many researchers are working on the handwritten Marathi character recognition. As compare to English and other languages Marathi is little bit complex language. Marathi language contains 12 common vowels and 36 consonants. In addition to this there are compound characters. In this research work, the handwritten text is split in three regions namely region above Shirorekha upper region i.e. upper modifier, main character and lower region i.e. lower modifier below the character. Shape features are extracted for upper and lower modifiers. Support vector machine is used as classifier and the average recognition rate is compatible.

Keywords: Modifier, Shape Feature Extraction.

1. INTRODUCTION

The modifier plays an important role in any language. In Marathi (Devanagari) the vowels are modifiers. In addition to this Marathi contains compound characters. This compound character may be the combination of two or more consonants and it may contain modifier. Due to these reasons the task of recognition of handwritten Marathi text becomes more complex. In this research work, the image of handwritten Marathi text is split in three regions. The region above *Shirorekha* is treated as upper modifier. The middle region, the main character and the third region, the region below the main character i.e. lower modifier. The middle region may contain another modifier/vowel. Further, it is segmented and classified accordingly.

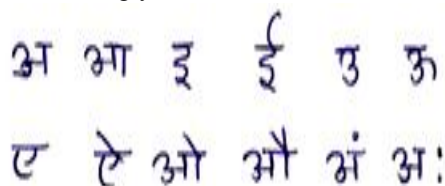


Figure 1 : Marathi Vowels



Figure 2: Marathi Consonants

2. RELATED WORK

The application of pattern recognition to extract proper features of modifiers can lead to through study of shape descriptor. The methodological contents of this paper are to describe a novel approach for pattern recognition method for feature extraction from an image of modifier. This novel method proposed for automated modifier recognition of handwritten Marathi text. Many shape description and similarity measure techniques have been developed in recent past. The approaches of shape-based feature extraction and representation are classified according to their processing approaches [1]. Modified Shape Descriptor (MSD) feature extraction technique is used as descriptive feature to discriminate Objects in an image database [2]. Shape features extracted using gradient operators are also used for image retrieval [3]. Shape features and its analysis method analyze the objects in a scene [4]. Similarly, shape features used for a similarity evaluation method for character patterns with missing shape parts [5]. Handwritten Marathi language uses various feature extraction methods such as moments for vowel recognition [6], chain code histogram and shadow features [7] and [8], and the average recognition rate is compatible. These vowels are used as upper, middle and lower modifiers. Thus the recognition task becomes more complex. In fact the main problem in handwritten Marathi character recognition 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 [9].

3. DATABASE PREPARATION

Standard database of handwritten Marathi (Devanagari) text is not available till today [10]. So, we have developed our own database. The database is prepared by scanning the images of Handwritten Marathi Characters with modifier, which are collected on a special designed sheet, from individuals of various age, gender and professions. Source document are scanned at 300 dpi using HP flatbed scanner and stored in JPG format.



Figure 3 Sample handwritten Lower Modifier



Figure 4 Sample handwritten Upper Modifier with character



Figure 5 Sample handwritten Middle Modifier



Figure 6 Sample handwritten Middle Modifier with Character



Figure 7 Sample handwritten lower Modifier



Figure 8 Sample handwritten Lower Modifier with Character

4. PROPOSED METHOD

In handwritten character recognition system database design and pre-processing are important part of the system. So we make the three parts of the characters and split the modifiers in 3 regions for recognition. We can call this as a micro partition of the character to improve the recognition rate. In this proposed method each regional modifiers are recognized separately. We have performed the morphological operation skeletonization of an image of modifier because in handwritten character the width at the curvature is less as compare to the remaining part of the character/modifier.



Fig.9 Sample Modifiers after Skeletonization

We proposed the following algorithm for the recognition of modifiers in Marathi text.

1. Data Acquisition
2. Pre-processing
3. Split the image in 3 regions.
4. Resize each region.
5. Extract the features for modifiers separately.
6. Classification.

5. FEATURE EXTRACTION

Feature extraction is backbone of any recognition system. At the same time, feature selection and number of features are also important. For this recognition system we use minimum shape and statistical features. The selected features are as follows.

1. Area
2. Solidity
3. Perimeter
4. Form Factor
5. Eccentricity
6. Centroid

For the modifier 'ः', we have 12 values of the above 6 features, so we calculate mean of its corresponding values. The features are extracted before morphing and after morphing.

6. EXPERIMENTAL RESULT

Experimentation is carried out on 15 handwritten Marathi modifiers prepared as database written by writers of various age groups and qualification and of both genders. For classification SVM is used. The result of classification without morphing is given in table 1 and table 2.

Table 1. Accuracy for upper modifier

Char	mu1	mu2	mu 3	mu 4	mu 5	mu 6	mu7
Accuracy %age	100	66.66	50	100	100	100	50
Accuracy %age after Morphing	100	33.33	66.7	100	100	100	66.6

Table 2. Accuracy for middle & lower modifier and average.

Char	mm1	mm2	ml1	ml2	ml3	Avg Accuracy
%age of O/p Classify Correctly	100	100	100	100	100	88.88
%age of O/p after Morphing	100	100	66.7	100	75	84.02

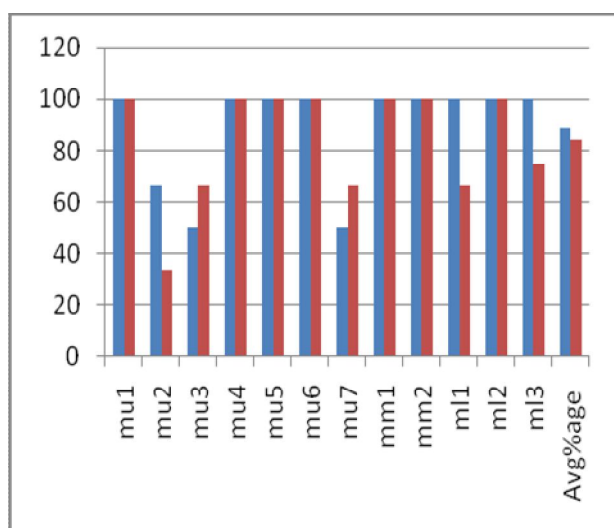


Figure 10. Graphical representation of percentage of accuracy

Average recognition of modifiers in handwritten Marathi text is 88.88 with only 6 features.

7. CONCLUSION AND FUTURE SCOPE

This paper proposes an algorithm of recognition of modifiers in handwritten Marathi text with only 6 features. The recognition rate of only mu2 and mu7 is more after morphing, but the average recognition rate is less than the modifiers without morphing. The recognition rate is compatible. This can be increase by selecting more number and appropriate features for modifier.

REFERENCES

- [1]. Mingqiang Yang, Kidiyo Kpalma, Jo seph Ronsin, "A Survey of Shape Feature Extraction Techniques", Pattern Recognition, IN-TECH, pp.43-90, 2008.
- [2]. A. Srinagesh, K.Aravinda, G.P.Saradhi Varma, A.Govardhan, M. SreeLatha, "A Modified Shape Feature Extraction Technique For Image Retrieval", International Journal of Emerging Science and Engineering (IJESE), Vol.1, Issue-8, June 2013.
- [3]. Dr.H.B.Kekre, P Mukherjee and S Wadhwa, "Image Retrieval with Shape Features Extracted using Gradient Operators and Slope Magnitude Technique with BTC", International Journal of Computer Applications, Vol.6, No.8, September 2010.
- [4]. Sven Loncaric, "A Survey of Shape Analysis Techniques", Pattern Recognition, Vol. 31, No. 8, pp. 983-1001, Elsevier Science 1998.
- [5]. Akihito Kitadai, Masaki Nakagawa, Hajime Baba, Akihiro Watanabe, "Similarity Evaluation and Shape Feature Extraction for Character Pattern Retrieval to Support Reading Historical Documents", Document Analysis Systems (DAS), 10th IAPR. International Workshop on Pattern Recognition-IEEE 2012.
- [6]. P.E.Ajmire and S E Warkhede "Handwritten Marathi Character (Vowel) Recognition" Advances in information mining, Vol. 2, 2010.
- [7]. Arora S., Bhattacharjee D., Nasipuri M., Basu D., and Kundu M., "Recognition of non-compound handwritten Devanagari characters using a combination of MLP and minimum edit distance", International journal of computer science and security, Vol 04, No. 01, pp. 107-120, 2010.
- [8]. Arora S., Bhattacharjee D., Nasipuri M., Basu D., and Kundu M., "Combining multiple feature extraction techniques for handwritten Devanagari character recognition", IEEE, Third International conference on Industrial and information systems, pp. 1-6, 2008.
- [9]. P E Ajmire, R V Dharaskar, V M Thakare and S E Warkhede "Structural Features for Character Recognition System-A Review", International Journal of Advanced Research in Computer Science, Volume 3, No. 3, pp.766-768, May-June 2012.
- [10].Sk Md Obaidullah, Anamika Mondal, Nibaran Das, and Kaushik Roy, "Script Identification from

Printed Indian Document Images and Performance Evaluation Using Different Classifiers”, Applied Computational Intelligence and Soft Computing 2014.

AUTHOR



Prafulla E Ajmire received the M.Sc.(Physics) in 1987, PGDCS in 1988. and M.S. degrees in Software System from Birla Institute of Technology, Pilani 1994, respectively. In 988, he joined the noble profession i.e. teaching. Now,

he is working as head of department of Computer Science at G S Science, Arts & Commerce College, Khamgaon. (M.S.), India.



Dr. Rajiv V Dharaskar, Former Director, Disha Education Society (DIMAT - Disha Technical Campus), Raipur, C.G Former Director, MPGI Group of Institutions Integrated Campus, Nanded Former Professor and Head, PG Department of Computer Science

and Engineering, G H Raison College of Engineering, a TEQIP II benefited Autonomous Institute, Nagpur.(India)



Dr. Vilas M Thakare, Professor and Head in Computer Science, Faculty of Engineering & Technology, Post Graduate Department of Computer Science, Sant Gadgebaba Amravati University, Amravati.

Ph.D.(Computer Science), M.E. (Advance Electronics), M.Sc.(Applied Electronics), Diploma in Computer Management.