Impact Factor Value 3.441 e-ISSN: 2456-3463
International Journal of Innovations in Engineering and Science, Vol. 2, No.12, 2017
www.ijies.net

Handwritten Marathi Compound Character Recognition using Structural and Statistical Features

Mrs.Snehal S. Golait[1], Dr.L.G. Malik[2], Prof.A.Thomas [3]

[1]*Research Scholar ,Department of Computer Science and Engineering, G.H.Raisoni College of Engineering,Nagpur*
[2]*Former Professor, Department of Computer Science and Engineering, G.H.Raisoni College of Engineering,Nagpur*
[3]*Head of Department, Department of Computer Science and Engineering, G. H. Raisoni College of Engineering, Nagpur*

Abstract **–**Feature Extraction plays a vital role in any character recognition system. It involves measuring those features of the input pattern are relevant to classification. The characteristics of the feature extraction techniques have to be independent of the scalable font characteristics such as type, size, style, tilt, rotation and should be able to describe the complex, distorted, broken characters effectively. A feature vector should be simple, reliable, complete, and compact to recognize any input character with high accuracy similar to human perception. This paper provides a feature extraction for Handwritten Marathi Compound Character using structural and statistical methods. Aside from the similarity of Character, Complex shape researcher founds difficulty to find features of characters. The recognition is carried out using structural and statistical feature extraction and multistage classification scheme. The Proposed system used Edge map as a structural feature extraction method. DFT and DWT used as statistical features extraction method. We apply all techniques on segmented characters able to get more than 90% accuracy in recognition process.

Keywords-DFT,DWT,EdgeMap,Feature Extraction, Extended color map

I. INTRODUCTION

Handwritten character recognition is an important field of Optical Character Recognition (OCR). The objective of OCR is an automatic reading of optically sensed document text materials to translate human readable characters to machine understandable codes.

OCR is popular for its various applications like postal automation, banks , library automation and defense organizations. Handwritten character recognition aims at converting handwritten characters in images into text that can be stored, edited or can be converted into speech. This field of research finds applications in various areas that aim in automation so as to reduce the human efforts like postal automation bank automation form filling etc. Handwritten character recognition for Indian scripts is quite a challenging task due to several reasons. One of the Indian Script is Devnagari Script. Devnagari is third most widely used script, used for Indian languages such as Hindi, Sanskrit, Nepali and Marathi, and is used by more than 500 million people. Unconstrained Devnagari writing is more complex than English language due to the possible variations in the shape, number and direction of the constituent strokes. Devnagari script has 50 characters which can be written as individual symbols in a word. Devnagari Character recognition is complicated process due to presence of multiple conjuncts, loops, lower and upper modifiers and the number of disconnected and multistroke characters, in a word where all characters are connected through Shirorekha. OCR is further complicated by compound characters that makes character separation and identification is very difficult. Samples of Marathi Compound Character as shown in figure 1.

Figure 1: Samples of Marathi Compound Character

OCR work started on printed Devanagari script in early1970s. Veena Bansal and R.M.Sinha worked on printed Devanagari text. First system for hand-written numeral recognitionof Devanagari script was proposed by R. Bajaj , P. M. Patil andT.R. Sontakke also presented an algorithm for handwrittenDevanagari numeral recognition which used concept of scaling, rotation and translation invariant. U. Pal proposed a system for off-line handwritten character recognition of Devanagari using directional information as features. A technique foraccuracy improvement of Devanagari character recognitionsystem was proposed by U. Pal using two features based ondirectional and curvature information in the characters andapplied to the classifiers support vector machines and modified quadratic discriminant function. A comparative study ofvarious features and classifiers used for handwritten Devanagari character recognition was proposede by U. Pal.Work on handwritten Bangla compound characters is carriedout by U. Pal.Recently, some piece of work are found onhandwritten Marathi compound characters. Vamvakas worked on hybrid features which include upper and lower character profile projection features, zoning based features, left and right character profile projection features and distance based features in 2010 [10]. Chacko worked on Wavelet features and chain code features in 2011[11]. Wang & Sajjahar worked on Polar transformed images and Zone based feature extraction in 2011 [13]. Yang and Al-Khateeb worked on Structural features as well as Statistical features in 2011.[14,15]. Rajput & Horakeri worked on Boundary-based descriptors namely crack codes and Fourier descriptors in 2011[16]. Nemouchi worked on Structural features like strokes, concavities, end points, intersections of line segments, loops, stroke relations and statistical features like zoning, invariants moments, Fourier descriptors, Freeman chain code features in 2012[17]. Vidya worked on Cross feature, fuzzy depth ,distance and Zernike moment in (2013) [18].

Amulti-feature,multi-classifier scheme for handwritten DevanagariMarathi characters is proposed by S. Shelke and S. Apte [1].G.G. Rajput and S.M. Mali [19] have shown use of Fourier descriptor and normalized chain code to recognize isolated Marathi handwritten numerals and have achieved 98.15% of recognition accuracy. Vijay Rahul pawar and Arun Gaikwad [20] have worked to recognize isolated Marathi character and various structural and statistical features are extracted like end points, middle bar, loop, end bar, aspect ratio etc. and have achieved 93% accuracy. Shushma. Shelke and S. Apte proposed [1] a multi-feature, multi-classifier scheme for handwritten DevanagariMarathi characters.

The Proposed system used Edge map as a structural feature extraction method. DFT and DWT as a statistical feature extraction method.

The remaining sections of this paper are organized as follows. Section 2 elaborates proposed system for feature extraction. Section 3 shows the classification technique for classifying the features of character. Section 4 shows the conclusion made on the basis of recognition accuracy.

II. FEATURE EXTRACTION

The proposed system to recognize handwritten Marathi compound character with structural and statistical features is shown in Figure 2.

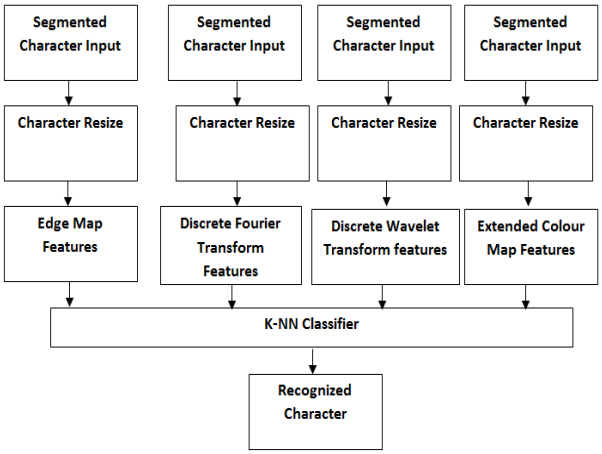


Figure 2: Character Recognition Using Multifeature extraction technique

The main aim of feature extraction is to make improvement in the accuracy and speed of the classifies for the pattern recognition. The extraction of the features of the characters is done in such a way that the complete portion of binary image covered and there is a distinct property associated with the each position. Feature extraction method categories into three types.

1. Structural 2. Statistical 3. Hybrid

1. Structural Features

Characters can be represented by structural features with high tolerance to distortions and style variations. This type of representation may encode some knowledge about the structure of the object or may provide some knowledge as to what sort of components make up that object. Structural features are based on topological and

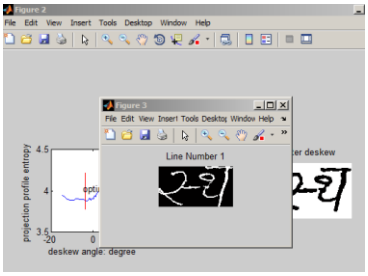
geometrical properties of the character, such as aspect ratio, end points, loops, dots, branch points, junctions, strokes and their directions, inflection between two points, horizontal curves at the bottom or top, etc.

Before extracting features of the Marathi compound character need to see some characteristics of compound characters. For this we need to see some global features of Marathi character such as characters having middle bar, end bar and character with no bar as well as local features such as end points.In the proposed work used edge map and extended color map as a structural feature extraction technique.

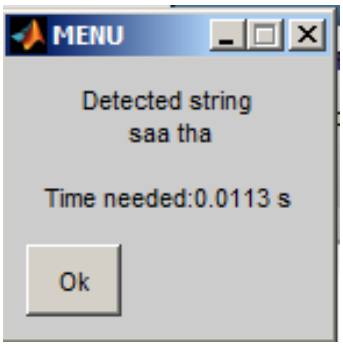
Edge Map

For calculating features with edge map, first we have to find the edge of input handwritten Marathi compound character. In the proposed work we used minutiae detection algorithm for segmentation of compound character. After character segmentation, we find the edge with the canny edge detection algorithm.

After detecting edge, used edge map[12] for feature extraction of character. Results of feature extraction using an edge map is shown in figure 3. Gets the pixels and check the neighboring pixels, and if an edge occurs on the pixels, then the probability of edge increases. This map of the probability of the edges is called as edge map.



(a)



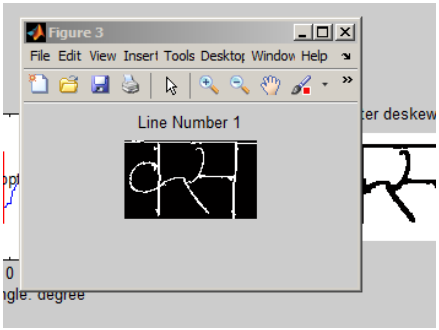
(b)

Figure 3: a) Marathi Compound Character

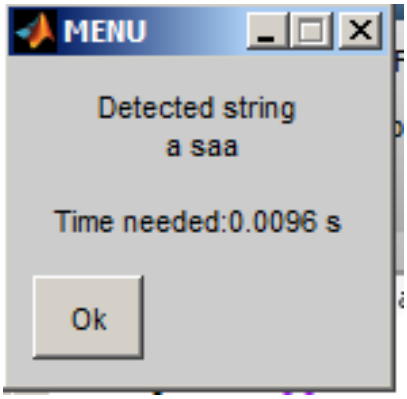
b) Recognized character with edge map

Extended Color Map

Extended Color Map is a graphical representation of distribution of colors within an image. The data contained in a map are obtained by counting the occurrence of each possible color of the respective color model within the image. The result of feature extraction using an Extended color map is shown in figure 4. A method to represent the color feature of an image by counting how many values of each color occur in the image and forming a representational map. A representation of the distribution of colors in an image, derived by counting the number of pixels on each of the given set of color ranges in a typically two-dimensional (2D) or three-dimensional (3D) color space.



(a)



(b)

Figure 4:a) Marathi Compound Character
b) Experimental results of the xtended Color Map

2. Statistical Features

Statistical methods based on planning on how data is collected and selected. It helps to make a hypothesis about the type of data. It is based on the probability theory and hypothesis. The statistical distribution of pixels of an image takes care of the variations in writing

style. In the proposed work used discrete Fourier transform and discrete wavelet transforms as a statistical feature extraction method.

Discrete Fourier Transforms

The first statistical method used for feature extraction is Discrete Fourier Transform (DFT). Fourier transforms are basic method to convert an image from space domain to frequency domain and provide a better alternative to spatial domain filtering. Fourier transforms facilitates to isolate and process particular frequencies among a range of frequencies. It also allows separating low-pass and high-pass frequencies. Fourier method transforms a signal from one domain to another. The Fourier descriptor methods have been successfully applied in character recognition, shape analysis and shape classification. Fourier transforms of discrete images is called Discrete Fourier Transform or DFT. Many algorithms have been devised to compute the DFT. Fast Fourier Transform is one of the most effective transformation algorithms to obtain the DFT . The result of feature extraction using DFT is shown in figure 5.

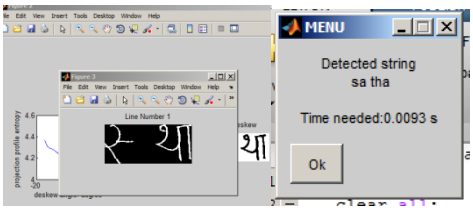


Figure 5 : Experimental result for DFT

Discrete Wavelet Transform(DWT)

The second statistical feature extraction method used in proposed work is Discrete Wavelet Transform (DWT). Wavelet transform decomposes a signal into a set of basic functions called wavelets. We can also say that Wavelet transforms are based on wavelets or small waves. In the proposed system we used Haar wavlets for calculating the features. In Haar transforms, signal is divided into two subsignals. First sub signal is the set of first trend signal and second sub signal is the set of first fluctuation signal. The result of feature extraction using DWT is shown in figure 6.

Database Creation

As from the literature review there is no such standard database available for handwritten Marathi Compound characters. For database creation taken the samples of handwriting of 400 peoples. We scan the image of text document and crop the image of character and store as jpg file. The sample of database is shown in figure 7.

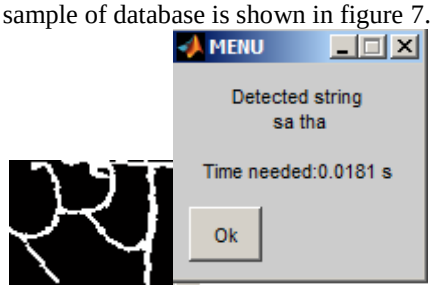


Figure 6: Experimental result for DWT

Database Creation

As from the literature review there is no such standard database available for handwritten Marathi Compound characters. For database creation taken the samples of handwriting of 400 peoples. We scan the image of text document and crop the image of character and store as jpg file. The sample of database is shown in figure 7.

Figure 7: Sample Database

III. CLASSIFICATION

K-NN classifier is one of the popular techniques used for classification due to their learning and generalization abilities. The pseudo code for the Marathi Character recognition using K-NN classifier is shown in figure 8 and the overall performance of compound character is shown in Table 1.

```
max_support = 0;
 max_index = 0;
for db_count=1:size(features,2)
  feat = features(:,db_count);
  support = corr2(feat,char_features');
  if(support > max_support)
     max_support = support;
     max_index = db_count;
  end
end
```

Figure 8: Pseudo Code for K-NN classifier

Table 1: Performance of compound character recognition

Sr No.	Feature extraction Technique	Resize features	Recognition Time in Sec. Time in sec.	Overall Recognition
1	Edge Map	64 × 64	0.036sec.	93.55%
2	Discrete Fourier transform	64 × 64	0.0009sec	92.41%
3	Discrete Wavelet transform	64 × 64	0.024 Sec.	94.92%
4	Extended Color map	64 × 64	0.024 Sec	85.35%

IV. CONCLUSION

Compound character is one of the features of the Marathi script. This paper discussed about various structural and statistical feature extraction techniques. The proposed system used the discrete Fourier transform, discrete wavelet transform, edge map and extended color map as a feature extraction technique for recognizing Marathi compound character. From the experimental results and recognition accuracy conclude that with Discrete wavelet Transform getting highest recognition results.

ACKNOWLEDGMENT

This paper is made possible through the help and valuable support of my guide Dr.Latesh Malik. Finally, I sincerely thanks to my parents and God.

REFERENCES

1. *Sushama Shelke, Shaila Apte, "A Multistage Handwritten Marathi Compound Character Recognition Scheme using Neural Networks and Wavelet Features " International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 4, No. 1, March 2011.*
2. *Mrs.Snehal Golait,Dr.Latesh Malik, "Handwritten Marathi Compound Character Segmentation using Minutiae Detection Algorithm", International Conference on Recent Trends in Computer Science and Engineering (ICRTCSE 2016), Published by Elsevier B.V.*
3. *Dr.AmitabhWahi, undaramurthy.S,Poovizhi.P,"Recognition of Handwritten Tamil Characters using wavelet", International Journal of Computer Science & Engineering Technology (IJCSET).*
4. *Sushama Shelke, Shaila Apte, "Multistage Handwritten Marathi Compound Character Recognition Using Neural Networks ",Journal of Pattern Recognition Research 2 (2011) 253-268.*
5. *Roli Bansal, Priti Sehgal & Punam Bedi," Effective Morphological Extraction of True Fingerprint Minutiae based on the Hit or Miss Transform", International Journal of Biometrics and Bioinformatics (IJBB), Volume (4) : Issue (2).*
6. *Tanuja K *, Usha Kumari V and Sushma T M,"Handwritten Hindi Character Recognition System Using Edge detection & Neural Network ",International Journal of Advanced Technology and Engineering Exploration ISSN (Print): 2394-5443 ISSN (Online): 2394-7454 Volume-2 Issue-6 May-2015.*
7. *Vijaya Rahul Pawar, Arun Natha Gaikwad," Structural and Statistical Feature Based Multistage Recognition Approach for Handwritten Devanagari Script Recognition ", International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 6 (2014) pp. 528-534*
8. *Neeraj Pratap,Dr. Shwetank Arya, "A Review of Devnagari Character Recognition from Past to Future ", International Journal of Computer Science and Telecommunications, Volume 3, Issue 6, June 2012.*
9. *Mrs.Vinaya. S. Tapkir,Mrs.Sushma.D.Shelke, "OCR For Handwritten Marathi Script ", International Journal of Scientific & Engineering Research Volume 3, Issue 8, August -2012 ISSN 2229-5518.*
10. *Vamvakas, G., B. Gatos, and S.J. Perantonis, Handwritten character recognition through two-stage foreground sub-sampling. Pattern Recognition, 2010. 43(8): p. 2807-2816.*
11. *Chacko, B., et al., Handwritten character recognition using wavelet energy and extreme learning machine. International Journal of Machine Learning and Cybernetics, 2012. 3(2): p. 149-161.*
12. *Mrs. Snehal Golait and Dr.Latesh Malik, "Segmentation of Handwritten Marathi Compound Character using Morphological Operation ",Journal of Emerging Technologies Image Processing and Networking, Volume 8, Issue 1, July 2016,pp 107-113 ,ISSN No: 0973-2993.*
13. *Wang, X. and A. Sajjanhar, "Polar Transformation System for Offline Handwritten Character Recognition", in Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2011, R. Lee, Editor. 2011, Springer Berlin Heidelberg. p. 15-24.*
14. *Yang, Y., X. Lijia, and C. Chen, English Character Recognition Based on Feature Combination. Procedia Engineering, 2011. 24(0): p. 159-164.*
15. *AlKhateeb, J.H., et al., Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking. Pattern Recognition Letters, 2011. 32(8): p. 1081-1088.*

16. *Rajput, G. and R. Horakeri, Handwritten Kannada Vowel Character Recognition Using Crack Codes and Fourier Descriptors, in Multidisciplinary Trends in Artificial Intelligence, C. Sombattheera, et al., Editors. 2011, Springer Berlin Heidelberg. p. 169-180.*

17. *Nemouchi, S., L. Meslati, and N. Farah, Classifiers Combination for Arabic Words Recognition: Application to Handwritten Algerian City Names, in Image and Signal Processing, A. Elmoataz, et al., Editors. 2012, Springer Berlin Heidelberg. p. 562-570.*

18. *Vidya V, Indhu T R, Bhadran V K,R Ravindra kumar, "Malayalam Offline Handwritten Recognition using Probabilistic Simplified Fuzzy ARTMAP", Advances in Intelligent Systems and Computing Volume 182, 2013, pp 273-283.*

19. *G G. Rajput and S. M. Mali. Fourier descriptor based Isolated Marathi Handwritten numeral Recognition. International journal of Computer Application.2010, pp. 5120/724-1017.*

20. *V. R. Pawar, "Performance Evalution of Multistage Offline Marathi Script Recognition System," vol. 88, no. 4, pp. 33–39, 2014.*

21. *Mrs. Snehal Golait , Dr.Latesh Malik , " Review on feature extraction technique for hand written Marathi compound character recognition ", ICETET-13,Dec 2013,Nagpur ,pp-113 -114, 978-1-4799-2560-5, IEEE.*