

Survey on various Improved Feature Extraction and Classification Techniques on Handwritten Marathi Character Recognition

Parshuram M. Kamble
Department of Computer Science,
Solapur University,
Solapur, MH, India-413255

Ravindra S. Hegadi
Department of Computer Science,
Solapur University,
Solapur, MH, India-413255

ABSTRACT

Optical Character recognition of Indian scrip is challenging task, due to they have various type of shapes, curves and lines. We found variation in dataset of same writer in same time of same characters. In survey we found the researcher used various morphological, spatial domain techniques in pre-processing to remove noise, slant and skew of handwritten Marathi character. In recently Marathi character feature are extracted using some statistical, geometrical and structural methods. They are applied to classify k-nn, SVM, ANN, Decision Tree classification Methods for handwritten Marathi character classification.

Keywords

Handwritten Marathi character, Classification, Segmentation, Feature extraction.

1. INTRODUCTION

Marathi handwritten character recognition has got lot of application in various areas like in automated form filling, Bank sorting cheque's electronically and postal office [5, 15, 13]. Recognition of handwritten Marathi characters by computer machine is difficult task as the computer can easily recognize compare to typed characters, which can be. In the last half century English printed and handwritten Character Recognition has been rapidly studied and progressed to a level, which has sufficient to develop technological application[5]. But in the term of structure and computation is not same in case for other Indian languages Marathi character recognition is becoming more and more important in the modern world.

It helps human ease their jobs and solve more complex problems over the few past years, the numbers of companies involved in research on handwritten recognition are increasing continually. Marathi being official language of Maharashtra state, also spoken by other state and country, should be given special attention to recognition and study of historical handwritten and printed Marathi literature has been started. [6, 7, 11 14]. Marathi languages belongs to Devnagari script. Devnagari had 13 vowels and 36 Consonants with 14 modifiers[7, 8] as shown in Figure 1. All of this is called basic Characters. In modifiers have verity of different directional marks, lines and curves[1, 15]. Vowels are independent letters by using different combination of modifiers. The combination of modifiers and consonants known as conjuncts[9, 14]. The combination of more than two consonants or vowels characters are known as compound The combine group of characters are

known as have continuous horizontal line in the upper part known as shirorekha (headline). characters[6].

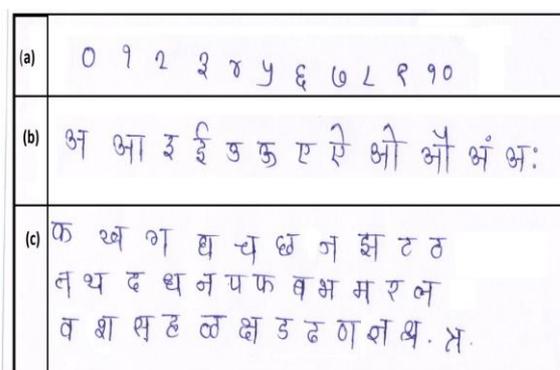


Fig 1: shows handwritten (a) Numerals (b) Vowels and (c) Consonants

The shirorekha is useful to joins next or previous characters of the same word. According to structure of the Devnagari script the text word we partitioned into three zones. The portion of above the shirorekha is upper zone, in this zone may be some components of modifiers found. The middle zone is below the upper zone, in this zone basic characters components is located. Below this zone is Lower Zone in this portion modifiers components or half part of compound characters are found in lower zone. Also in handwritten Marathi language there are vowels, consonants, vowel modifiers, compound characters, and numerals, there are number of similar shaped characters and also writer writes same character in different shape. All these variations make handwritten character recognition a challenging problem to researchers.

2. PRE-PROCESSING

In scanned handwritten document noise like disconnected line segments, bumps and gaps in line due to writing instrument or optical scanning device. Available noise reduction techniques can be filtering, morphological operations. This may be reducing the noise Various spatial and frequency domain filters can be used for smoothing, sharpening, thresholding, removing slightly textured or colored background and contrast adjustment. Morphological operations can be used to remove the noise on the document images due to low quality of paper and ink, as well as erratic hand movement It may be reduce the number of objects and unwanted background information. Normalization methods can be removes the variations of the

writing and obtains standardized data. It can be used to adjust the character size to a certain standard size. Segmentation is an important stage; it can be used in separation of words, lines or a character.

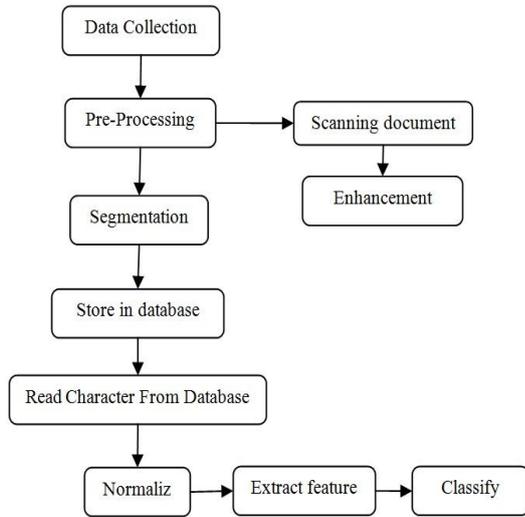


Fig 2: Flow of Handwritten character Recognition

3. FEATURE EXTRACTION

In the handwritten Marathi character recognition feature extraction and selection are robust information from the row data which is effective to accuracy of character recognition. The feature extracted information enhance the class pattern variability during the character classification [1, 6, 11, 12, 13, 14]. In survey we found geometrical, statistical and structural feature extraction Techniques.

The representation of digital character image by statistical distribution of points which are unique identification in the presence of character variation. In survey we found the major statistical feature extraction method Crossing, distance, projection and Zoning [13]. The digital character image map is divided into set of zones in different directions and features of each zone are calculated in crossing feature extraction method and in the zoning method map of character image is divide into number of non-overlapping and overlapping zones.

Then the densities of each zone or different zones are analyzed. In the projection feature of digital character image can be represented by the gray values of pixel are projected onto lines in various directions[11, 5, 13]. This representation creates one-dimensional signal from a two-dimensional image, these can be used as descriptors of digital character image.

3.1 Character Vector

The digital character vector is denoted as X where $X = (f1, f2, \dots, fz)$ where f denotes features and z is the number of zones in which each character is divided[13, 5]. The number of input feature zone is determined by length of the feature vector z . Figure 3 shows the zone of sample handwritten Marathi character and its feature.

3.2 Moment Based

Moment based feature extraction are used to extract in character feature under character invariant condition like translation, rotation, scaling. It is robust performance during presence and absence of noise in character image. moment feature are calculated in different variation on image shape.

3.2.1 Geometric Moment

In a handwritten digital character image with $f(x,y)$ of size $M \times N$ image moment M_{ij} are calculated in equation (1):

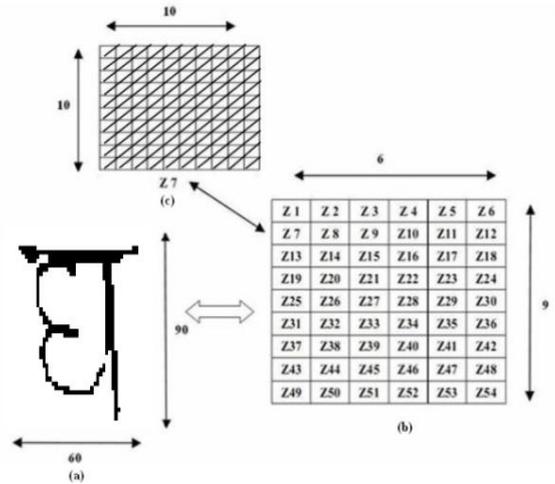


Fig 3 : Procedure for extracting feature from the sample Marathi character.

$$M_{ij} = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} x^i y^j f(x, y) \quad (1)$$

All M_{ij} with $i+j \leq n$, a positive integer are the geometric moments of order $(i+j)$.

3.2.2 Central Moment

This feature used to calculate translation of $M \times N$ sized digital image plane. it is to be mapped on to square which was defined was $x \in [-1, +1]$ and $y \in [-1, +1]$. Invariance with respect to position of the object in the digital character image which used to calculate the central moments of the mapped digital image.

$$\mu_{ij} = \sum_{x=-1}^1 \sum_{y=-1}^1 (X - \bar{X})^i (Y - \bar{Y})^j f(x, y) \quad (2)$$

where $\bar{X} = \frac{m_{10}}{m_{00}}$, and $\bar{Y} = \frac{m_{01}}{m_{00}}$ are the component of the

centroid.

4. CLASSIFICATION

The few techniques for handwritten character recognition are proposed by the researchers. In handwritten character Recognition system tremendously use the data mining and pattern recognition Methodology, which are useful to assign unknown predictor to trained class. A very good survey on classification and feature extraction for Character recognition of Devnagari script can be found in [5, 13]. In handwritten Marathi character classification techniques can be classified as Support Vector Machine (SVM), Neural Networks algorithms and Combination classifier. This approaches are neither necessarily independent Marathi Character recognition nor disjoint from each other.

4.1 Neural Network

In Optical character recognition classification is like human beings can identify document and character by their learning and experience, which is related to heuristic logic. There are various types of neural networks, which are used for digital

characters classification and recognition. In neural network have massively parallel adaptive neural activation function [13], also neural network is a calculating structure. Due to their parallel computing architecture it can performed in a higher rate as compared to other classical techniques. It can adapt to change in the feature of digital image data and train the descriptors value of input vector[12, 14]. Neural network fed the output of previous node to next one in the network and the final result will be depending upon on the complex interaction between all of nodes. The existing training of neural networks approaches are Boltzman, competitive learning, Hebbian, Radial basis function and error correction. They covers continuous and binary input values for unsupervised and supervised learning. The neural network architectures have been classified as feed forward, Artificial, back propagation and feedback network. As per review of various researcher prepare the neural network in their classification systems used multilayer perception of the feed forward network and the SOM of the feedback network[9, 5, 13].

5. SUPPORT VECTOR MACHINE

The objective of any classification and recognition machine system depends upon the learning is achieve the good generalization performance. In character recognition have finite amount of training samples by striking a balance between the goodness of fit on given training dataset and perform the ability of the machine to achieve error-free recognition on other datasets. With this concept as the basis, SVM's have provided to achieve good reputed performance with no prior knowledge of the data. The principle of an SVM is to map the input data onto a higher dimensional feature space nonlinearly related to the input space and determine a separating hyper plane with maximum margin between the two classes in the feature space [5, 13, 1]. A SVM is a maximal margin hyper plane in feature space but by using a kernel function in original space. The optimal separating hyper plane can be determined without any computations in the higher dimensional feature space by using various kernel function in the input space. Many researchers used SVM successfully viz. kale et. al [1, 13, 5].

6. RESULT AND DISCUSSION

We found in survey paper kale et. al [1] worked on database of Handwritten Devnagari (Marathi) Characters samples of 9600 basic character. They applied Zernike moment which represent the character features, SVM and K-NN applied for classification. Recognition of handwritten Marathi numerals using multi-layer feed-forward neural network proposed by R.S. Hegadi and Kamble P. M. In this work they normalized all digital numerals image in 7x5 pixels using cubic interpolation. After these resized numerals were converted in to a one-dimensional vector with 35 values. These 1-D vector values were used as input to train the feed-forward neural network. for these experiment they used on 1000 set of handwritten Marathi numerals.

Table 1. Table shows the comparison result of Researcher.

Author	Feature Extraction Method	Classification Techniques	Accuracy
R.S. Hegadi and Kamble P.M. (2014)	Normalized 1-D Vector	ANN	97.00%

Karbhari V. Kale et. al. 2014	Zernike moment	SVM and K-NN approach	98.32%
P.M Kamble and R.S. Hegadi (2015)	Pixel Connected based area, eccentricity, orientation,	Minimum Distance	94.38 %

Kamble P. M. and Hegadi R.S. [12] proposed by Basic handwritten Marathi character recognition using statistical Method. They computed connected pixel based feature such as orientation, eccentricity, area of character and minimum distance classifier was used for classification. in this experiment they used 8000 basic handwritten Character. Table 1 shows the comparatively result and applied methods to calculate features and classification of handwritten Marathi character.

7. CONCLUSION

In this paper we attempted few work to compare the performance of various algorithms. They have difficulty of effective recognition are determined by the nature of the material in the experiment. Still lot of research is needed to perform the challenges in optical character recognition. In order to achieve high accuracy in complex cases in character recognition, we have prepare segmentation, pre-processing and classification. We found in survey the current research models not only the basic character but also text, phrases, word, and entire page document. The number of references are provide to more understanding of the character recognition system approaches listed below.

8. REFERENCES

- [1] K. V. Kale, P. D. Deshmukh, Y Shrinivas, V Chavan and Majharoddin M. Kazi, " Zernike moment Feature Extraction for Handwrittem Devnagari(Marathi) Compound Character Recognition", *International Journal of Advanced Reserch in Artificial Intelligence*, Vol. 3, No.1, 2014.
- [2] Nafiz Arica, Fatos T. and Yarman Vural. "An Overview of Character Recognition Focused on off-line handwriting", C99-06-C-203, *IEEE*, 2000.
- [3] R.M.K. and Sinha, "Rule Based Contextual Post-processing for Devnagari Tex Recognition", *Pattern Recognition*, Vol. 20, No. 5, pp.475-485, 1987.
- [4] S. Kumar and C. Sinha, " A Study of Zernike Moments and its use in Devnagari Handwritten Character Recognition", *Int. Conf. on Cognition and Recognition*, pp-514-520, 2005.
- [5] U. Pal and B. B. Chaudhuri, " Indian Script Character recognition: A survay", *Pattern Recognition*, Vol. 37, PP.1887-1899, 2004.
- [6] B.V. Dhandra and Mallikarjun Hangarge, "Global and Local Features Based Handwritten Text Words and Numerals Script Identification", *Intl. Conf. on Computational Intelligence and Multimedia Applications*, PP 471-475. 2007.
- [7] Latesh Malik and P.S. Deshpande, "Recognition of printed Devnagari characters with regular expression in finite state models", *International workshop on machine intelligence research, GHRCE Nagpur, India*, 2009.
- [8] M. Hanmandlu and Pooja Agrawal, "A Structural Approach for Segmentation of Handwritten Hindi Text", *International Conference on Cognition and Recognition*. pp 589-597.2007.

- [9] Muharrem Mercimek, Kayhan Gulez and Tarik Veli umcu, "Real object recognition using moment invariants", *Sadhana* Vol. 30, Part 6, pp. 765–775, Dec. 2005.
- [10] Harikesh Singh, R. K. Sharma, "Moment in Online Handwritten Character Recognition", *National Conference on Challenges & Opportunities in Information Technology (COIT-2007) Gobindgarh*. March 23, 2007. ,
- [11] R. S. Hegadi and P. M. Kamble, "Recognition of marathi handwritten numerals using multi-layer feed-forward neural network," in *World Congress on Computing and Communication Technologies (WCCCT)*,. IEEE, 2014, pp. 21–24.
- [12] P. M. Kamble and R. S. Hegadi, "Handwritten marathi basic character recognition using statistical method," in *Emerging Research in Computing, Information, Communication and Applications*. Elsevier 2014, Vol. 3, 2014, pp. 28–33.
- [13] Dongre, Vikas J., and Vijay H. Mankar. "A review of research on Devnagari character recognition." *arXiv preprint arXiv:1101.2491* (2011).
- [14] Kamble P M and Hegadi R S, "Handwritten Marathi Character Recognition Using R-HOG Feature", *Elsevier Procedia Computer Science*, Vol.45 pp.266–274, 2015.