Marathi Handwritten Character Recognition by using Probabilistic Neural Network Classification

Tulshiram B. Pisal[1] & Parshuram M. Kamble[2]

Abstract: **"Marathi" is the Official language of Maharastra (written in Devanagri script) is most well-liked language after English and Hindi. Marathi handwritten character recognition has got lot of application in various areas like postal office, Bank sorting cheks electronically. Recognition of handwritten Marathi characters by computer machine is difficult task as the computer can easily recognize compare to typed characters, which can be. English Character Recognition (CR) has been widely studied in the last half century and progressed to a level, sufficient to produce technology applications. But same is not the case for Indian languages which are difficult in terms of structure and computation. Digital document processing is acquisition recognition for application to office and library automation, bank, publishing houses communication technology, postal services and many other areas. With never increasing requirement for office automation, it is necessary to provide practical and effective solutions. 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. Devanagri being the national language of India, spoken by more than 500 million people, should be given special attention so that document retrieval and analysis of rich ancient and modern Indian literature can be effectively done.**

Keywords: **Character Recognition, Segmentation, Preprocessing, Off-line Handwriting Recognition, Feature Extraction, Image Classification**

1. INTRODUCTION

Indian Languages Characteristics

In India use multiple languages and scripts with 22 listed languages Marathi, namely, Malayalam, Manipuri (Meithei), Nepali, Oriya, Punjabi, Sanskrit, Santali, Sindhi, Tamil, Telugu and Urdu Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri, Konkani [3].
12 scripts is used to write this Languges.

---

1. Greenfigers College of Computer and Technology, Akluj
pisal.tulashiram123@gmail.com

2. Greenfigers College of Computer and Technology, Akluj
parshu1983@gmail.com

Devnagiri script used to write Marathi, Konkani,Hindi, Nepali, Sanskrit, Bodo, Dogri,Sindhi and Mathili.Urdu Language used script Urdu. Assamese, Manipuri and Bangla languages are written using Bengali script. Gurmukhi script is used to write Punjabi language. Some other languages have their own script. Language scripts not classified into uppercase and lowercase. Most of the Indian languages are derived from Ancient Brahmi and are phonetic in nature and hence writing maps sounds of alphabets to specific shapes. Only urdu language is written right to left and other language is written into left to write. The basic characters comprises of vowels and consonants. Two or more basic characters are combined to form compound characters

Marathi Language Characteristics

Marathi is one of the ancient Indian languages having its own script and it belongs to the languages having Devanagari script. It is the official language of Maharashtra state of India and majority of people speak Marathi in this state. Marathi language characters which are formed by the combination of one of the 12 vowels and 36 consonants resulting in 432 characters. Good handwritten character can recognize by humans correctly. It is very easy task for human and also can do easily by the children. But the same task for machine is not easy. Marathi Character contains complex curves & various shapes. Marathi characters Recognition is hard & complex task. All these consideration make Optical Character recognition (OCR) with Devanagari script very challenging. The eventual goal of designing a character recognition system with an accuracy rate of 100 % is fairly difficult because handwritten characters are similar; they can be written in various styles. In addition, they can be written in various sizes by different writers. Even there is difference in characters written by the same writer at dissimilar time

The paper is organized as follows: Section 2 of the paper contains the pre-processing of numbers and gives a brief description of the Characters selected for recognition and its data set. Feature Extraction method is discussed in Section 3. The Classification method and its algorithm is discussed in Section 4. The experimental details and results obtained are

offered in Section 5. Section 6 contains the conclusion part and further the problem. Section 7 fallows References.

2. OBJECTIVES

1) Handwritten character Recognition is to convert text images into computerized editable text. That can be editing, storing, or can be converting into further process like Marathi script into other language script and speech.

2) Office automation: reduce human efforts like postal, bank, library, publishing house and communication technology.

3) Uniformity of text: different people have different writing style of characters overlapping and interconnection of the neighboring character which can be converted in standard computer Marathi font.

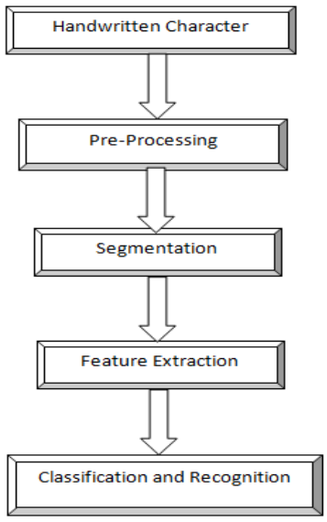
4) It can be easy for rural people they have no idea about use of computer. It can be done in Marathi script through handwritten character which converted into computer editable text.

5) To study Marathi character recognition through English scenario.

6) To help Maharashtrien people to recognize Marathi characters in computer application.

3. METHODOLOGY

Model for Handwritten Character Recognition:



1) Handwritten Character:

It contains vowels and corresponding Modifiers, Consonants, Half Form of Consonants with Vertical bars, Special Combination of Half-Consonants and Consonants and some Special Symbols. All the Character has a horizontal line at the upper part known as “Shirorekha” or “Headline”. It can be scanned by using flatbed scanner.

2) Pre-Processing:

2.1) Noise Reduction:

The noise some optical scanning device or the writing instrument, causes disconnected line segments, bumps and gaps in line. Available noise reduction techniques can be filtering, morphological operations. This may be reducing the noise.

2.1.1) Filtering:

Various spatial and frequency domain filters can be used for smoothing, sharpening, thresholding, removing slightly textured or colored background and contrast adjustment.

2.1.2) Morphological Operations:

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

2.2) Thresholding

It may be reduce the number of objects and unwanted background information.

2.3) Normalization:

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.

3) Segmentation:

Segmentation is an important stage; it can be used in separation of words, lines or a character.

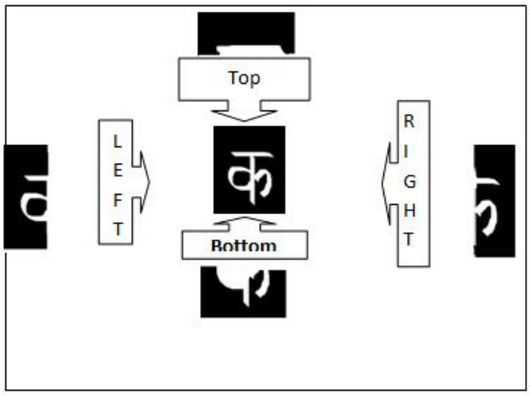
4. FEATURE EXTRACTION

The directional density estimation features, features based on water reservoir principle, maximum profile distance features, and fill-hole density features are the structural features can be used for the Character recognition.

4.1) Directional density estimation

The outer directional density of pixels can be counted row/column wise until it touches the outer border of the character in the four directions viz. left, right, top, and

bottom directions as shown in Fig. 4. It also exhibits the corresponding directional pixels considered in the count as black band area.



Directions for Density estimation and pixels consideration

4.2)Water Reservoir:

The cavity regions of the component where water will be stored considered as reservoirs. *Top (bottom) Reservoir*: The reservoir obtained when water poured from top (bottom) of the component. *Left (right) Reservoir*: The water stored cavity regions of the component, when water is poured from left (right) side of the component will be the left (right) reservoir. Top, bottom, left, and right reservoir of Characters illustrated in Fig. 5.

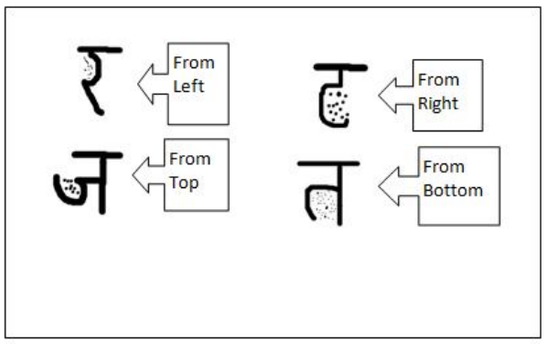


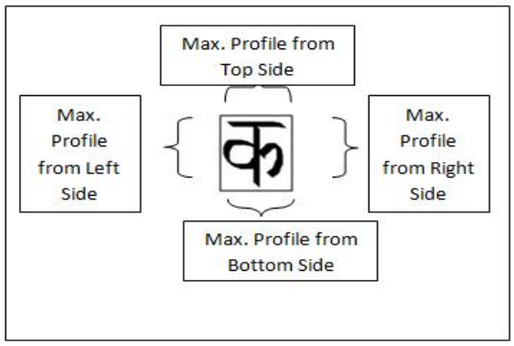
Figure 5: Water Reservoirs in Characters

4.3) Fill-hole Density:

The looping area of the Characters can be filled with ON pixels and feature consider the fill-hole density.

4.4) Maximum profile distances:

After fitting the bounding box on each Character, their profiles can be computed in four directions. While computing the profile, we will be considering only 40% of the middle area in four directions of the bounding box. The maximum profile obtained in four directions, and the profile features are illustrated in Fig. 6.



Maximum profile distances from all sides of Bounding Box

5. CLASSIFICATION

Probabilistic Neural Network (PNN) Classification: Probabilistic neural network is a kind of radial basis network suitable for classification problems. The PNN classifier provides a good generalization ability and fast learning capability, which may be using in handwritten character recognition system. This network uses a radial basis transfer function to calculate layers
Output from its net input. The architecture of Probabilistic Neural Net will be made up of four units viz. input units, Pattern units[Class 0…52], Summation units and Output units as shown in the Figure 7.

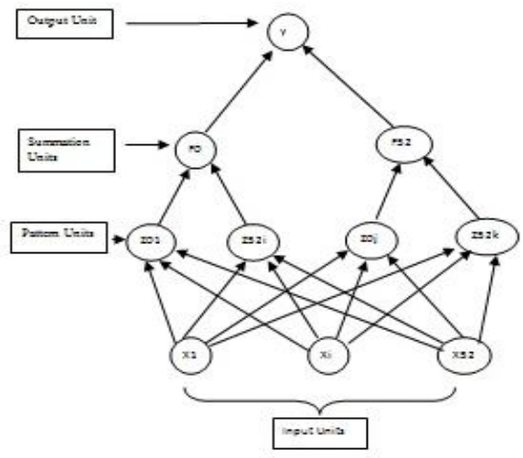


Figure 7: Architecture of Probabilistic Neural Net

Input: Isolated Binary Characters.
Output: Recognition of the Characters.
Step 1: Preprocess the input image to eliminate the noise using filter and invert the image.
Step 2: Fit the minimum rectangle-bounding box for an input image and crop

the Characters.

Step 3: Extract the structural features and stored them in the Library.

Step 4: Train the PNN classifier with feature vector stored in the library.

Step 5: Classify the test images to its appropriate class label using PNN classifier with various radial values.

6. CONCLUSION

In literature survey the Marathi recognition was done by individual vowel and consonant recognition. This reduces the number of characters to be recognized from 432 to just 36 consonants and 12 vowels. Total Characters of 36+12=48 single shapes need to be recognized. I can be found that the feature set most suited for vowels is different from those for consonants. I can be also finding a nonlinear relation in the recognition power of both the features and so can be used together as a single feature set for the recognition. I can be investigating an overall architecture for HCR incorporating all these, attempting to improve the overall accuracy further. Such a structure will help to exploit further domain information in the recognition process.

7. REFERENCES

1) P.B. Khanale and S.D. chitins, "Handwritten Devanagari Character Recognition Using Artificial Neural Network", *Journal of Artificial Intelligence,* vol. 4, no. 1, pp. 55-62, 2011.

2) Sushama Shelke and Shaila apte, "A Multistage Handwritten Marathi Compound Character Recognition Scheme using Neural Networks and Wavelet Features", *Int. J. of Signal Proc., Image Proc. and Pattern Rec.,* Vol. 4, No. 1, pp. 81-94, 2011.

3) Sushama Shelke and Shaila apte, "Multistage Handwritten Marathi Compound Character Recognition Using Neural Networks", *Journal of Pattern Recognition Research,* vol. 2, pp. 253-268, 2011.

4) P.E. Ajmire and R.V. Dharaskar and V M Thakare, "A Comparative Study of Handwritten Marathi Character Recognition", *NCIPET,* pp. 26-28, 2012.

5) Vikas J Dongre Vijay H Mankar, "A Review of Research on Devnagari Character Recognition", *Int. J. of Comp. App.,* Vol. 12, No. 2, pp. 8-15, 2010.

6) C.Namrata Mahender and K.V. Kale, "Structured based Feature Extraction of Handwritten Marathi Word", *Int. J. of Comp. Applications,* Vol. 16, no- 6, pp. 42- 47, 2011.

7) Bilan Zhu, Xiang-Dong Zhou and Cheng-Lin Liu, "A robust model for on- line handwritten Japanese text recognition", *Springer,* no 13, pp.121-131, 2010.

8) Tian-Fu Gao, Cheng-Lin Liu, "High accuracy handwritten Chinese character Recognition using LDA-based compound distances", *Elsevier,* vol. 40, 2008. pp. 3442-3451.

9) Cheng-Lin Liu "Lexicon-Driven Segmentation and Recognition of Handwritten Character Strings for Japanese Address Reading" *IEEE Tran. On Pattern Analysis and Machine Intelligence,* Vol. 24, no. 11, pp-1427-1437 2002.