

# Handwritten Devnagari Character Recognition System

Neha Avhad<sup>1</sup>, Shraddha Darade<sup>2</sup>, Neha Gawali<sup>3</sup>, Apurva Matsagar<sup>4</sup>

<sup>1234</sup>Department of Computer Eng. Dept., KKWIEER, Nashik, India  
Email: <sup>1</sup>neha.avhad394@gmail.com, <sup>2</sup>daradeshreddha167@gmail.com,  
<sup>3</sup>ngawali93@gmail.com, <sup>4</sup>apurvamatsagar178@gmail.com

**Abstract:** Optical Character recognition which is the original method of character recognition many times gives poor recognition rate due to error in character segmentation. Segmentation is a very important task in every OCR system. OCR system separates the scanned handwritten image text documents into lines, words and characters. The accuracy of OCR system mainly depends on segmentation algorithm and noise removal technique being used. Segmentation of handwritten Devnagari text is difficult when compared with printed Devnagari or printed English or any other printed document, because of its structural complexity and increased character set. It contains vowels, consonants and half form consonant. This system addresses the segmentation of handwritten Devnagari text document, the most popular script of Indian sub – continent into lines, words and characters. Mainly artificial neural network technique is used to design to pre-process, segment and recognize Devnagari characters.

**Keywords:** Devnagari Character Recognition, Offline Handwritten Recognition, Segmentation, Feature Extraction

## 1. Introduction

Handwritten character recognition System is the ability of a computer to receive and recognize handwritten input from any sources such as paper documents, photograph and other devices. Handwritten Devnagari Characters are more complex for recognition than corresponding English characters due to many possible variations in number, order, direction and Shape as well as writing style of the constituent strokes. Every individual has his own style of writing. Any individual having a very good knowledge of the script of a language can easily read some words written on a paper, though those are written in very bad manner, on the basis of his or her mental dictionary. Such words cannot be easily read by a machine as there may be various irregularities caused in expressing these words which are not easy to handle by a machine. Due to very strange styles of writing, a lot of difficulties are faced in machine recognition process. Handwriting recognition contributes immensely to the advancement of an automation process and can improve the interface between human beings and machine in numerous applications. The main purpose of this dissertation work is to introduce a method for recognition of offline handwritten Devnagari characters using segmentation and Artificial neural networks. The process of recognition includes segmentation of line into word and word into characters and then recognized through feed-forward neural network using back propagation algorithm. We will take the scanned handwritten document as input. Then apply segmentation algorithm to divide document into the line, word and into the character. Then apply recognition algorithm base of feed foreword neural network [1].

## 2. Current Market Survey:

HCR is an intelligent Optical character recognition capable of handling the complexity of writing, writing environment, materials, etc. Many possible techniques have been developed in the OCR. This technique based on extracting various features of handwritten characters and providing these features as input to an Artificial Neural Network.

Handwriting character recognition classified into off-line and on-line handwriting recognition methods. In off-line recognition method, the writing is usually captured by a scanner and complete writing is available in an image format. But, in the on-line system the two dimensional coordinates of successive points are represented as a function of time and the order of strokes made by the writer are also available. The on-line method is superior than off-line methods in recognizing handwritten characters due to the temporal information available with the former. HCR has several applications like in government office, bank processing, document reading, mail sorting and postal address recognition require off-line handwriting recognition systems. As the result, the off-line handwriting recognition continues to be an active area for research towards exploring the newer techniques that would improve recognition accuracy.

K. Y. Rajput and Sangeeta Mishra have proposed a system for recognizing hand written Indian Devnagari script. This system considers handwritten images as an input, separates the lines, words and then characters step by step and then recognizes the character using artificial neural network approach, in which Creating a Character Matrix and a corresponding Suitable Network Structure is essential. Afterwards, the Feed Forward Algorithm used for entire working of a neural network, followed by the Back Propagation Algorithm which required for Training, Modifying Weights and calculation of errors. Once the characters are recognized they can be replaced by the standard fonts to integrate information from diverse sources [4].

J.Pradeep, E.Srinivasan and S.Himavath has proposed a diagonal feature extraction scheme for the recognizing off-line handwritten characters. In the feature extraction process, resized individual character of size 90x60 pixels is further divided into fifty four equal zones, each of size 10x10 pixels. The features are extracted from the pixels of each zone by moving along their diagonals. This procedure is repeated for all the zones leading to extraction of 54 features for each character. These extracted features are used to train a feed forward back propagation neural network employed for performing classification and recognition tasks. Extensive simulation studies has shown that the recognition system using diagonal features provides good recognition accuracy while requiring less time for training [5].

N. Sharma, U. Pal, F. Kimura, and S. Pal have proposed a system which is based on recognition of offline Devnagari handwritten characters. The features which are used in the classifier are obtained from the directional chain code information of the contour points of the characters. The bounding box of a character is used to segment blocks and the chain code histogram is computed in each of the blocks. Based on the chain code histogram, 64 dimensional features are used for recognition. These chain code features are given to the quadratic classifier for recognition. In this system fivefold cross-validation technique is used for result computation [6].

Gita Sinha, Mrs. Rajneesh Rani, Prof. Renu Dhir proposed an overview of Feature Extraction techniques for off-line recognition of isolated Devnagari numeral recognition. This system presents Zone based approach which is the combination of image centroid zone and zone centroid zone of numeral or character image. In ICZ algorithm (Image centroid zone) character is divided into n equal zone and then image centroid and the average distance from character centroid to each zones/grid/boxes present in image is calculated. Similarly, in ZCZ algorithm (zone centroid zone) character image is divided into n equal zones and centroid of each zones/boxes/grid and average distance from zone centroid to each pixel present in block/zone/grid is calculated. Support Vector Machine is used for subsequent classifier and recognition purpose [2].

### 3 Devnagari Script:

Devnagari script is different from Roman script in several ways. This script has two-dimensional compositions of symbols: core characters in the middle strip, optional modifiers above or below core characters. Two characters may be joined to each other. While line segments (strokes) are the predominant features for English, most of the characters in Devnagari script are formed by holes, curves and strokes. In Devnagari language script, the concept of uppercase, the lower-case characters, is absent. But the alphabet itself contains more number of symbols than that of English. Marathi is an Indo-Aryan language spoken by about 71 million people, mainly the Marathi people of western and central India [8]. It is the official language of the state of Maharashtra. Marathi is thought to be a descendent of Maharashtra, one of the Prakrit languages which developed from Sanskrit. We know that the Handwriting style varies from person to person. It has a large character set with curves and lines in the shape formation, which may be over lapping (touch) in a word. Touching characters can touch each other at different position because of individual writing styles vary greatly. Following are the various regions of a Devnagari script [11, 15]

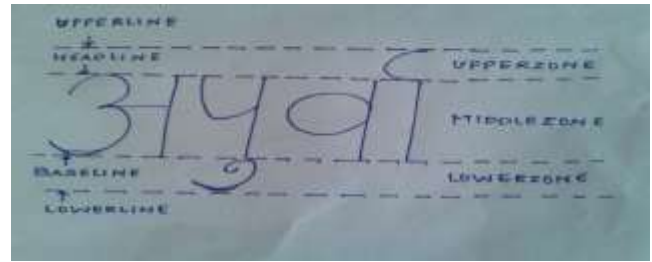


Fig 1: Various zones portioning in Devnagari Script.

Devnagari Script has 13 vowels (scar) and 36 consonants (Vyanjan) [2] and 10 numerals along with modifier symbols. All the individual characters are joined by a header line called —Shiro Rekhal which makes it difficult to isolate individual characters from the words. There are various vowel modifiers which add up to the confusion [3]. Minor variations in similar characters can be there in the handwriting.

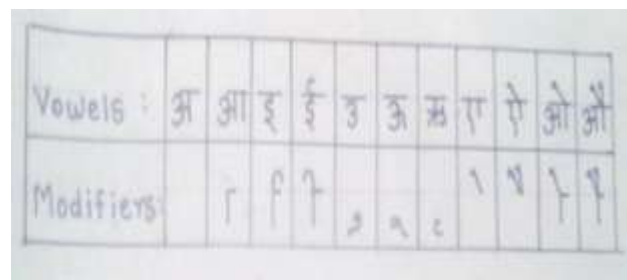


Fig 2: Vowels and corresponding Modifiers.

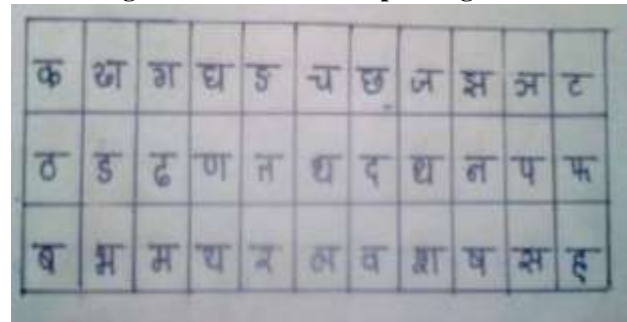
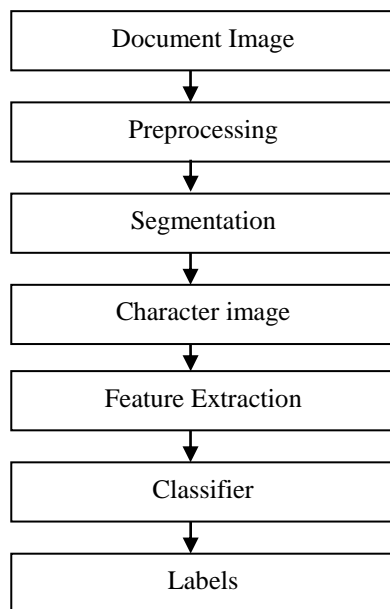


Fig 3: Consonants.

### 4 Proposed Handwritten Devnagari Character Recognition System:

In the proposed system we have taken scanned Handwritten Devnagari document as a input and then converted into the gray scale image and then apply noise removal techniques to remove noise .After that this image is converted into binary form and then apply segmentation on it to segment document into line, words and character .Then we have used neural network techniques for recognition of individual Devnagari character through Training and Testing.



**Fig 4: Block Diagram of HCR**

#### 4.1. Training

Training part train the network to recognized characters correctly .The training part is performed by an image of all 33 Devnagari character to achieve better accuracy.

##### 1. Image Preprocessing

The pre-processing is a series of operations performed on the scanned input image. It essentially enhances the image rendering it suitable for segmentation. Noise is removed from input image. Image is converted to gray scale. Binarization process converts a gray scale image into a binary image using threshold technique to produce the pre-processed image suitable for segmentation.

##### 2. Segmentation

In the segmentation stage, an image of sequence of characters is decomposed into sub-images of individual character. In this system, the pre-processed input image is segmented into isolated characters. Each individual character is uniformly resized into  $n \times n$  pixels for classification and recognition stage.

##### 3. Feature Extraction

For feature extraction we will use Zone based approach which is the combination of image centroid zone and Zone centroid zone of numeral/character image. In image centroid zone character is divided into  $n$  equal zone and then image centroid and the average distance from character centroid to each zones/grid/boxes present in image is calculated. Similarly, in zone centroid zone character image is divided into  $n$  equal zones and centroid of each zones/boxes/grid and average distance from zone centroid to each pixel present in block/zone/grid is calculated.

#### 4.2. Recognition

After the completion of training part the system is tested upon some samples. This requires analysing system performance. Recognition part includes following steps:

##### 1. Image Acquisition

In this step, Scan binarized image is taken as an input to the recognition system. The image can be in any form such

as jpeg, BMT etc. The image is acquired through a scanner.

##### 2. Preprocessing

Preprocessing is same as a training phase Preprocessing.

##### 3. Segmentation

During segmentation phase, to recognized word correctly it is divided into separate character by removing its headline by horizontal as well as vertical scanning. After that image is inverted by replacing the rows with all white pixels into black once. After this the segmentation is same as training phase.

##### 4. Feature Extraction

Feature extraction is done in the same way as in the training part and the recognized characters are stored in text format.

## Conclusion

In Character Recognition System performance is largely depends upon the character recognized in a word .And also Devnagari language having the script's complexity and most of the characters are similar so it create difficulty to analysed character perfectly. Because of similarity in character the performance of the system as well as efficiency and Accuracy decreases. So in Handwritten Devnagari character recognition system ,using back propagation algorithm neural network is trained ,and then testing is performed .To increase the accuracy of the system those character who are difficult to recognized ,these character are further analysed in training part and tested with different set. And again the accuracy is analysed, few character which shows no improvement are kept out of the testing. And again testing was performed on some new images sets written by different people, and by using this approach accuracy of the network is increases in many cases. It can be concluded that as the network is train with the large no of sample sets, the accuracy of character recognition system was increases.

## Acknowledgement

With deep sense of gratitude we would like to thanks all the people who have lit our path with their kind guidance. We are grateful to these intellectuals who did their best to help during our project work. And lastly we thanks to our all friends and the people who are directly or indirectly related to our project

## References

- [1] Mitrakshi B. Patil, Vaibhav Narawade, Recognition of Handwritten Devnagari Characters through Segmentation and Artificial neural networks , International Journal of Engineering Research & Technology (IJERT) , ISSN: 2278-0181, Vol. 1 Issue 6, August 2012.
- [2] Gita Sinha, Mrs. Rajneesh Rani, Prof. Renu Dhir , Recognition Using Zonal Based Feature Extraction Method and SVM Classifier , International Journal of Advanced Research in Computer Science and Software Engineering , ISSN: 2277 128X, Volume 2, Issue 6, June 2012.
- [3] Pulkit Goyal, Sapan Diwakar, Anupam Agrawal, Devnagari Character Recognition towards natural Human-Computer Interaction.

- [4] K. Y. Rajput ,Sangeeta Mishra, Recognition and Editing of Devnagari Handwriting Using Neural Network , Proceedings of SPIT-IEEE Colloquium and International Conference, Vol. 1.
- [5] J.Pradeep, E.Srinivasan and S.Himavathi , Diagonal Based Feature Extraction for Handwritten Alphabets Recognition System Using Neural Network , International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, Feb 2011.
- [6] N. Sharma, U. Pal, F. Kimura and S. Pal, Recognition of Off-Line Handwritten Devnagari Characters Using Quadratic Classifier.
- [7] Andrew Hellman and Jennifer Greene, Head First C#, O'Reilly.
- [8] Daniel Solis Illustrated C# 2012, Apress, 2012.