

## **DEVANAGARI CHARACTER RECOGNITION (DCR)**

**Bipin Pandey<sup>1</sup>, Pallavi Parab<sup>2</sup>, Priya Nikam<sup>3</sup>, Ms. Manali Patekar<sup>4</sup>**

<sup>1,2,3</sup>*Computer Engineering, University of Mumbai, St. John College of Engineering And Technology, Aldel Education Trust, Mumbai, (India)*

<sup>4</sup>*Assistant Professor, Computer Engineering, University of Mumbai, St. John College of Engineering and Technology, Aldel Education Trust, Mumbai, (India)*

### **ABSTRACT**

*In India most people use Devanagari for documentation. The modern society needs input text in computer readable format. we proposed the Optical Character Recognition (OCR) for Devanagari text using ANFIS. ANFIS is Artificial Neuro Fuzzy Interference System. ANN, is an artificial intelligent system able to be trained to perform a particular function by calculating the values of the connections that is weight between these elements. In our project ANN is used to train the system to extract the characters depend upon their features and match with database for recognition. Histogram and Affine moment invariant technique is used. Using ANFIS technique we will develop a system that extract characters, recognize character and covert that input text into computer readable format. Recognition rate of all Devanagari character is near about 98% .*

**Index terms:** *Training of NN, Kohonen network, feature extraction, Affine moment invariant, Segmentation*

### **I. INTRODUCTION**

Character recognition is process to classify the input character according to predefined dataset. Optical character recognition is process by which the scanned handwritten document to computer readable format. Devanagari character recognition is quite complicated task because every Devanagari character has its own different features. Devanagari originated from earliest Brahmi characters through various transformations. The script has a complex composition of its constituent symbols. In Devanagari there are 13 vowels Fig.1(a) and 34 consonants Fig.1(b) with 14 modifiers of vowels and of “rakar,” as shown in Fig. 1(a) symbols. spaced out from the vowels and consonants, there are compound (composite) typeset in most of Indian scripts as well as Devanagari, which are formed by combining two or more basic characters.

In this paper we propose a system which take handwritten Devanagari text in image format , extract features from it and convert into computer readable format. Neural network is technology which is used in pattern recognition process. We implement steps of DCR system like image preprocessing, segmentation, feature extraction and classification using neural network.

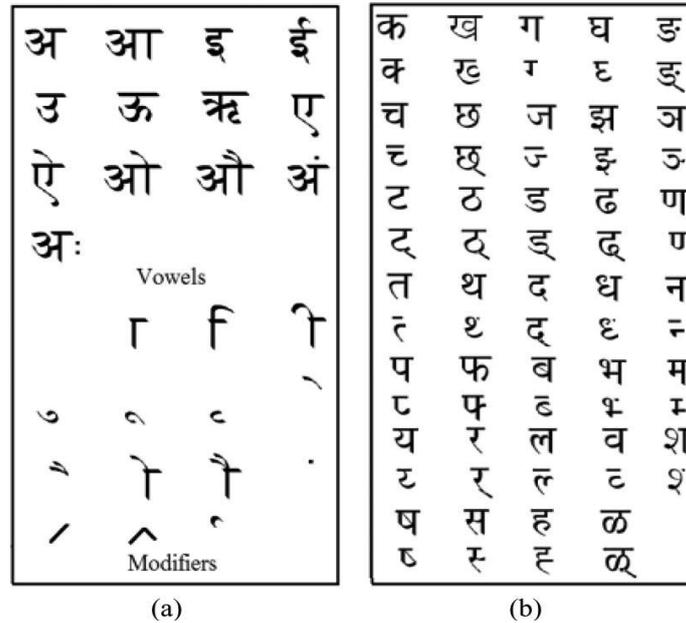


Fig.1.(a)Vowels and modifiers of Devanagari scripts  
(b) Consonants and their corresponding half forms in Devanagari script

## II. RELATED WORK

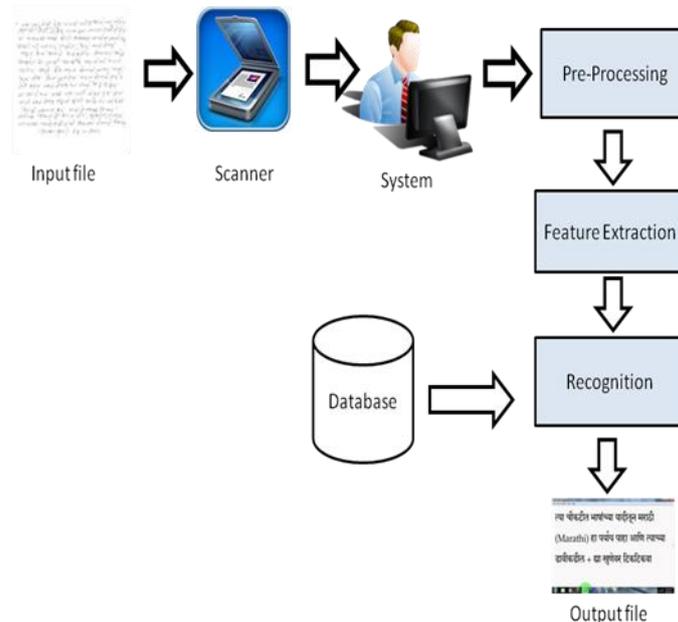
In[3] these paper presents an algorithm for segmentation of touching Devanagari characters into its constituent symbols and characters. Proposed algorithm extensively uses structural properties of the script. The information about the height and width of character , which are vertically separated from their neighbors, is used to hypothesize character boxes to be touching character boxes. The recognition of characters is 85% has been achieved on the segmented touching characters.

The disadvantage of these paper is that not recognized the touching character of devnagari .

In[4] these paper the fuzzy system is used for fast and flexible detection of the text. It specially uses the binarize technique which detects and recognizes the character. There are no restrictions like other algorithms. It gives 85% detection of character correctly.

The main disadvantage of the paper is it does not provide accuracy to segmented character it gives error while segmenting the touching character.

## III. PROPOSED WORK



**Fig: System Architecture**

As shown in fig the input file is given to system. The software pre-process the image which is in the color image is converted into binary image. After image preprocessing the process image is used to extract features of character in that image. The segmented character is recognized using artificial neural network. The database and extracted featured is compared and output is given as character.

### 3.1 Processing steps

Devanagari character recognition has different phase, namely:

1. Image preprocessing
2. Segmentation
3. Training network
4. Database Input
5. Recognition

## IV. IMAGE PRE-PROCESSING

Image processing is a type of signal processing in which the input is an image, such as text image, any document image; the output of image processing is an image which is converted into 2D image. It is an important step of applying number of procedures for smoothing, filtering, enhancing, dilation etc. for making a digital image usable by an otsus algorithm to improve the readability of optical character recognition system.

The various stages involved in preprocessing are:

- Convert image to grayscale image
- Convert grayscale image to binary image
- Apply median filter

- Remove the header line of image

## V. SEGMENTATION

The most important part in the OCR of Devnagari character is segmentation. The image is taken and segmentation method is applied.

Firstly line segmentation is performed by using below algorithm

1. Convert the binary image into matrix form which contain the values as 1 and 0 in which 1 represent background and white as image
2. Now take the complement of binary matrix so that 1 represent image and 0 as background
3. Calculate the row in which it has sum  $\geq 1$  and assign the start variable as that row value in which sum  $\geq 1$
4. Calculate the row in which it has sum = 0 and assign the end variable as that row value in which sum = 0
5. Now by using subimage fuction in java get the sub image from main image with height from start position to end position
6. Repeat the above steps from 3 to 5 till it reaches to height of image.

Secondly the word segmentation is performed with above result by using below algorithm

1. Convert the line image into matrix form which contain the values as 1 and 0 in which 1 represent background and white as image
2. Now take the complement of binary matrix so that 1 represent image and 0 as background
3. Calculate the column in which it has sum  $\geq 1$  and assign the start variable as that column value in which sum  $\geq 1$
4. Calculate the column in which it has sum = 0 and assign the end variable as that column value in which sum = 0
5. Now by using subimage fuction in java get the sub image from line image with width from start position to end position
6. Repeat the above steps from 3 to 5 till it reaches to width of image.

Thirdly the character segmentation is performed with above result by using below algorithm

1. Convert the word image into matrix form which contain the values as 1 and 0 in which 1 represent background and white as image
2. Now take the complement of binary matrix so that 1 represent image and 0 as background
3. Calculate the row in which it has sum is maximum from rest of row and assign that row value to max variable
4. Now make row of max position to 0 i.e it identify the header line of image.
5. Calculate the column in which it has sum  $\geq 1$  and assign the start variable as that column value in which sum  $\geq 1$
6. Calculate the column in which it has sum = 0 and assign the end variable as that column value in which sum = 0
7. Now by using subimage fuction in java get the sub image from line image with width from start position to end position

8. Repeat the above steps from 3 to 5 till it reaches to width of image.

## VI. TRAINING NETWORK

The kohonen network is used to train the system. In which we gave image of character as input and assign it to their associate character. In these we have taken for all the images which is collected by written from many user so that we can make dataset of that and train the network so we get proper output. In these we create network which correctly recognized the character properly.



Fig: character image

As above image we assign above image with क variable like that we assign all devnagari images with their respective character and then we train our network so that when we gave that image as input it should give corresponding character of the that image.



Fig.3 Some handwritten numerals, characters and sentences in Devanagari script

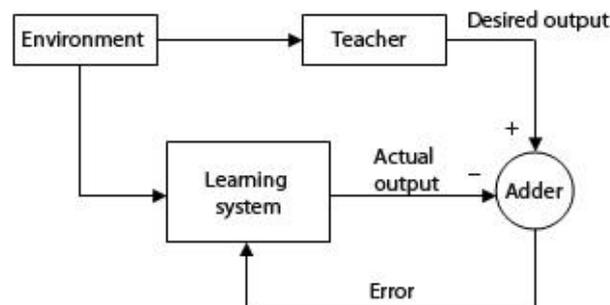
## VII. DATABASE INPUT

In database each devanagari character is stored in image format. Each character is stored in standard format, that is each having same font face, font size, and style. The font face we are using is Kruti dev 014 and font size is 14. In database the images which are stored are trained by using neural network to give equivalent character for that image. And also in database features are stored of each devanagari character. The above fig.3 shows that handwritten devanagari character are stored in database and the sentences which are there are being processed and the feature are compared with database in which handwritten character are being stored.

## VIII. RECOGNITION

Recognition is used to recognize the extracted devanagari character input. This recognition is performed using Artificial neural network. The segmented character is given to neural network as input and it gives recognized devanagari character as output. The target vector of neural network is database in which features of character is stored.

Training and recognition phase of the ANN has been performed using Kohonen network algorithm .



It uses supervised learning method for training the neural network. In these it adjust the input by comparing with output. The system uses these neural network for training the system.

## VIII. CONCLUSION

Character recognition of Devanagari character is difficult task because of its various shapes, features and size. This paper proposes system that recognize Devanagari character using ANN which is simple and efficient method for classification and recognition. The accuracy of this system is 100% for single characters and for complex characters (touching characters) is 96%.

क

When we talk about single character it gives image It gives output as क and for words of devnagari character it gives almost 96% result.

Our proposed system does 99 % segmentation part of devnagari character printed as well as handwritten character and recognized single character as 100 % and touching character as 96%. The overall accuracy of proposed system is 98%.

## REFERENCES

- [1] D.Trier ,A.K.Jain ,T.Taxt , “Feature Extraction Method for Character Recognition-A Survey” ,Pattern Recognition
- [2] B. B. Chaudhuri and U. Pal, “Skew angle detection of digitized Indian script documents,” IEEE Trans. Pattern Anal. Mach. Intell.
- [3] Veena Bansal & R.M.K. Sinha, “Segmentation of Touching Characters in Devanagari
- [4] Mohanad Alata — Mohammad Al-Shabi,” TEXT DETECTION AND CHARACTER RECOGNITION USING FUZZY IMAGE PROCESSING”, Journal of electrical engineering.

# International Conference on Recent Innovations in Engineering and Management

Dhananjay Mahadik Group of Institutions (BIMAT) Kolhapur, Maharashtra

(ICRIEM-16)

23rd March 2016, [www.conferenceworld.in](http://www.conferenceworld.in)

ISBN: 978-81-932074-5-1

- [5] U. Pal and B. B. Chaudhuri, "Indian script character recognition: A Survey", Pattern Recognition.
- [6] D.Trier ,A.K.Jain ,T.Taxt , "Feature Extraction Method for Character Recognition-A Survey" ,Pattern Recognition.
- [7] U. Bhattacharya and B. B. Chaudhuri, "Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals," .
- [8] Veena Bansal and R.M.K. Sinha, "A Complete OCR for Printed Hindi Text in Devnagari Script".