IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. X (Mar-Apr. 2014), PP 74-79
www.iosrjournals.org

Handwritten Devanagari Character Recognition using Neural Network

Ms. Seema A. Dongare[1], Prof. Dhananjay B. Kshirsagar[2], Ms. Snehal V. Waghchaure[3]

[1](Computer Department, SRESCOE Kopargaon, India)
[2](Computer Department, SRESCOE Kopargaon, India)
[3](Computer Department, SRESCOE Kopargaon, India)

Abstract: *In this digital era, most important thing is to deal with digital documents, organizations using handwritten documents for storing their information can use handwritten character recognition to convert this information into digital. Handwritten Devanagari characters are more difficult for recognition due to presence of header line, conjunct characters and similarity in shapes of multiple characters. This paper deals with development of grid based method which is combination of image centroid zone and zone centroid zone of individual character or numerical image. In feature extraction using grid or zone based approach individual character or numerical image is divided into n equal sized grids or zones then average distance of all pixels with respect to image centroid or grid centroid is computed. In combination of image centroid and zone centroid approach it computes average distance of all pixels present in each grid with respect to image centroid as well as zone centroid which gives feature vector of size 2xn features. This feature vector is presented to feed forward neural network for recognition. Complete process of Devanagari character recognition works in stages as document preprocessing, segmentation, feature extraction using grid based approach followed by recognition using feed forward neural network.*

Keywords: *Feed forward neural network, handwritten character recognition, image centroid zone, zone centroid zone.*

I. INTRODUCTION

Optical character recognition converts scanned images of printed or handwritten text into digital text. Basically there are two classes of optical character recognition as off line character recognition and on line character recognition. In off line character recognition, writing is captured optically by scanner while in on line character recognition coordinates of successive points are as function of time as well strokes made by user are also considered. Handwritten character recognition is a branch of optical character recognition that converts handwritten input from paper documents into digital text. Handwriting recognition is also can be classified as off line and on line handwriting recognition methods.

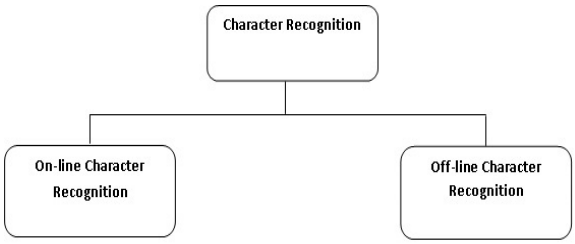

Fig. 1 Types of HCR

Handwritten Devanagari characters are quite complex for recognition due to presence of header line, conjunct characters and similarity in shapes of multiple characters. The main purpose of this paper is to introduce a method for recognition of handwritten Devanagari characters using segmentation and neural networks. The whole process of recognition works in stages as preprocessing on document image, segmentation of document into lines, line into words and word into characters, finally recognition using feed forward neural network. Important steps in any HCR are preprocessing, segmentation, feature extraction and recognition using neural network. [2].

II. RELATED WORK

K. Y. Rajput and S. Mishra have proposed a system for recognizing handwritten Indian Devanagari script. In feature extraction character matrix as an array of black and white pixels of size 30X30 is prepared. Afterwards, the Feed Forward neural network with back propagation is used in learning and recognition process. [3].

S. Arora, D. Bhattacharjee proposed two stage classification approaches for handwritten Devanagari characters. The first stage is using structural properties like detection of shirorekha, spine in character and second stage exploits some intersection features of characters which are presented to a feed forward neural network. Each handwritten character can be adequately represented within 16 segments (each of size 25 X 25 pixels) and hence 32 features for each character can be used as input to neural network [4].

V. Agnihotri proposed Handwritten Devanagari script recognition using neural network. Diagonal based feature extraction is used for extracting features of the handwritten Devanagari script. These feature set is converted into chromosome bit string of length 378.Individual character image of size 90x60 pixels is divided into 54 equal sized zones. Each zone has 19 diagonal lines and the foreground pixels present along each diagonal line is summed to get a single sub feature, thus 19 sub features are obtained from each zone. These 19 sub features values are averaged to form a single feature value and placed in the corresponding zone. Finally, 54 features are extracted for each character [5].

D. Singh, S. Singh and Dr. M. Dutta proposed twelve directional feature inputs depending upon the gradients. This technique can recognize all types of handwritten characters even special characters in any language [6].

N. Sharma, U. Pal, F. Kimura, and S. Pal have proposed a quadratic classifier based scheme for the recognition of offline Devanagari handwritten characters. Features used in the classifier are obtained from the directional chain code information of the contour points of the characters. This technique has achieved 98.86% and 80.36% recognition accuracy on Devanagari numerals and characters, respectively [7].

III. DEVANAGARI SCRIPT

3.1 Properties of Devanagari Script

Devanagari script has features different from other languages. Devanagari character set has 13 vowels, 36 consonants and 10 numerals with optional modifier symbols. Characters are organized into three zones as upper, middle and lower zone. Core characters are positioned in middle zone, while optional modifiers in upper and lower zones. Two characters may be connected to each other. In Devanagari script, the concept of uppercase and lowercase characters is absent. Fig. 2 represents Devanagari character set. It represents Devanagari character modifier set. Modifiers are optional symbols arranged in upper and lower zones.

3.2 Issues Regarding Recognition of Devanagari Script

Some reasons that cause recognition of Devanagari characters difficult are as:

1. In Devanagari Script individual characters are connected by header line (Shirorekha) which makes segmentation of individual character is quite difficult.
2. Characters may be connected to form conjuncts for which separation is complex.
3. Presence of modifiers makes segmentation difficult.
4. Some Devanagari characters are similar in shape.

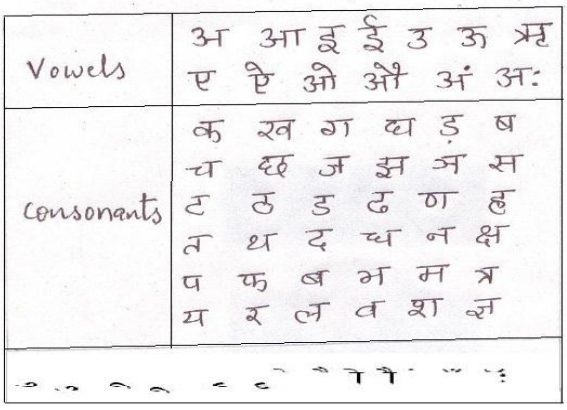

Fig. 2 Devanagari Character Set

IV. PROPOSED SYSTEM

HCR works in stages as preprocessing, segmentation, feature extraction and recognition using neural network. Preprocessing includes series of operations to be carried out on document image to make it ready for segmentation. During segmentation the document image is segmented into individual character or numeric image then feature extraction technique is applied on character image. Finally feature vector is presented to the neural network for recognition.

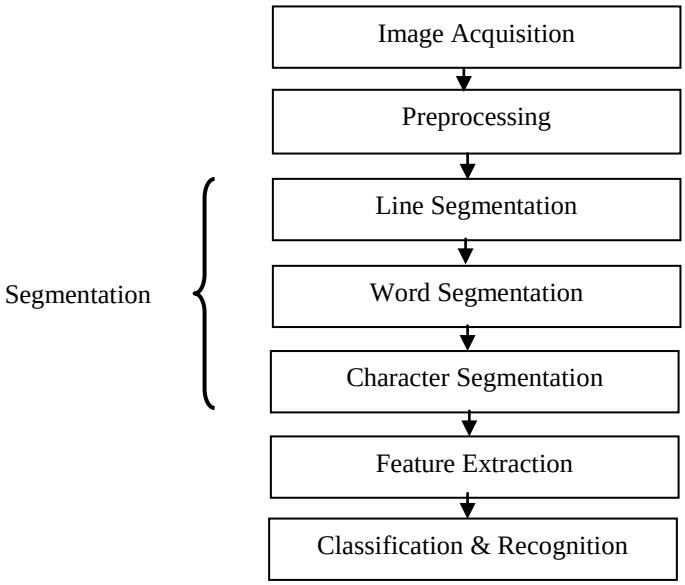

Fig. 3 Block Diagram of system

4.1 Preprocessing

The preprocessing consists of series of operations as grayscale conversion, noise removal, and binarization. After selecting Devanagari document image, color image is converted into gray scale. Unwanted contents are removed from image. Then binarization is applied on gray scale image.

4.2 Segmentation

Once Image preprocessing is done it is necessary to segment document into lines, line into words and word into characters. When individual character has been separated from document we can extract features from it for recognition.

4.3 Feature Extraction

For feature extraction we will use grid or zone based approach which is the combination of image centroid zone and zone centroid zone of individual character or numerical image. In this technique individual character or numeric image is divided into n equal sized grids or zones, then average distance of all pixels with respect to image centroid or grid centroid is computed. In combination of image centroid and zone centroid approach it computes average distance of all pixels present in each grid with respect to image centroid as well as zone centroid which gives feature vector of size 2xn features. Three variances of this approach can be used as:

4.3.1 Image Centroid Zone

Compute the centroid of image (numeral/character). Individual character image (100x100) is divided into 100 equal zones where size of each zone is (10x10) then compute the average distance from image centroid to each pixel present in the zones/grid. Thus we can get 100 feature values for each character.

Algorithm1: Image Centroid Zone (ICZ) based feature extraction.
Input: Preprocessed individual character/numerical image.
Output: Extract features for classification and recognition.

Algorithm: Method Begins
Step 1: Divide an input image in to n equal sized grids.
Step 2: Compute centroid of image.

Step 3: Compute distance between the image centroid and each pixel present in the grid.
Step 4: Repeat step 3 for the entire pixels present in the zone/grid.
Step 5: Computation of average distance between these points present in image.
Step 6: Repeat this procedure for all grids.
Step7: Obtaining n such features for classification and recognition.
Ends.

4.3.2. Zone Centroid Zone

Similarly in ZCZ we can divide an image into n equal sized grids and calculate centroid of each grid. Then compute the average distance from the grid centroid to each pixel present in grids. There could be some grids that are empty then the value of that particular grid is assumed to be zero. We can repeat this procedure for all grids present in image(numeral/character).

Algorithm2: Zone Centroid Zone (ZCZ) based feature extraction.
Input: Preprocessed individual character/numerical image.
Output: Extract features for Classification and Recognition.

Algorithm: Method Begins
Step 1: Divide an input image in to n equal sized grids.
Step 2: Compute centroid of each grid.
Step 3: Compute the distance between the grid centroid and each pixel present in the grid.
Step 4: Repeat step 3 for the entire pixels present in the zone/grid.
Step 5: Computation of average distance between these points present in image.
Step 6: Repeat this procedure for all grids.
Step7: Obtaining n such features for classification and recognition.
Ends.

4.3.3 Combination of ICZ and ZCZ

This system uses combination of both (ICZ+ZCZ) feature extraction system. For this we will compute the centroid of image (numeral/character) then we will divide an image into n equal size grids. Then compute average distance from image centroid to each pixel present in the zones/grid. Compute the average distance from the zone centroid to each pixel present in grid. We can repeat this procedure for all grids present in image (numeral/character).In this system we will get (2 x n) features.

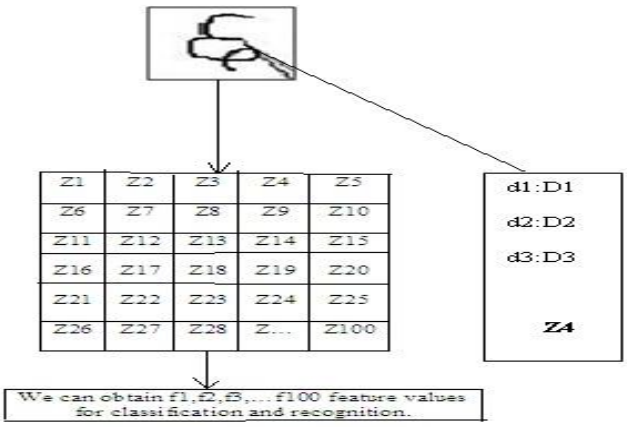

Fig. 4 Feature Extraction from Devanagari Numeral Image "six"

Fig. 4 show an illustration of procedure to extract features from Devnagari numeral "six". Individual character image of size 100 X 100 is divided 100 zones of equal size.d1,d2,d3,...,dn are distances from image centroid similarly D1,D2,D3,..., Dn are distances from the zone centroid, then compute average distances between these points separately. This gives 2 feature values for each grid. Same procedure is repeated sequentially for each of grid. With combination of ICZ and ZCZ we will have two feature we will have two feature values per grid which gives 200 feature vector provided no of zones are as 100[1].

4.3.4 Recognition of Characters using Neural Network

The backend used for performing recognition is neural network. In the off-line recognition system, the neural network is fast and reliable tool in order to achieve high recognition accuracy. This module will implement Artificial Neural Network using error back propagation (EBP) algorithm. In this EBP ANN, we will use n input neurons where n = length of the extracted features from character. Architecture has only two hidden layer for error handling and for the communication.

- Input nodes: n input neurons.(2Xn for Image centroid zone and zone Centroid zone).
- Output nodes: 59 (13vowels, 36 consonants and 10 numerals).
- No of Hidden layers:1.
- Training algorithm: Error Back Propagation.

V. RESULT

We have implemented preprocessing on scanned image of handwritten Devanagari document. In which color image is converted into gray scale image. Noise is removed from image. Finally image is converted into binary image in which each of the pixels is either black or white. Once image has been preprocessed segmentation is applied in which image is segmented into lines, line is segmented into words and word into individual character image. Fig. 5 shows scanned image of Handwritten Devanagari Document.

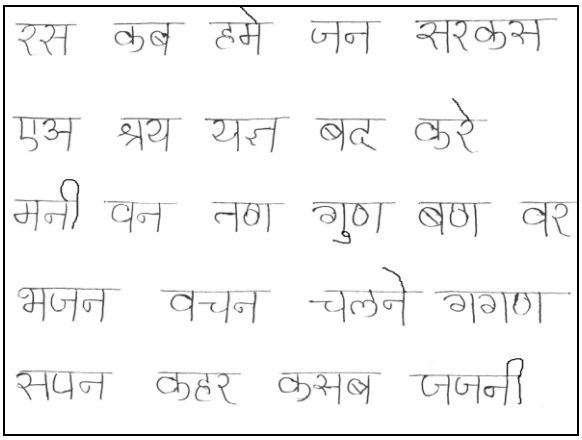

Fig.5 Input Document Image

In line segmentation, we scan each horizontal pixel row starting from the top of document. The lines are separated where we find a row with no black pixels. This row acts as a separation between two lines. After line segmentation, we scan vertical pixel column, words can be separated by looking for the column with no black pixels. Fig. 5 shows result after word segmentation. These words are input for character segmentation.

Fig. 6 Word Segmentation of Document Image

In character segmentation of line 1 is represented in fig. 7.

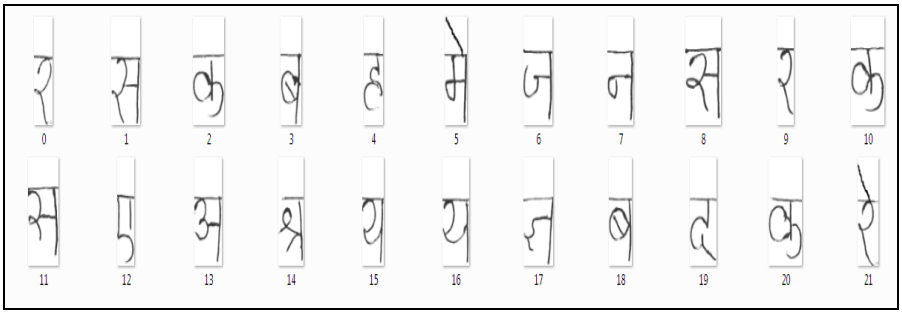

Fig. 7 Character Segmentation of Document Image

As individual character has been separated, character image can be resized to m x n pixels. Individual character image is divided into n equal sized grids. Grid based feature extraction can be applied to generate feature vector. Accuracy of system depends on number of feature values presented to neural network, but larger feature set increases number of grids as well distance computation. Design of neural network is based on number of input, output neurons, number of hidden layers and performance function. Neural network must be trained on large dataset to improve precision but larger dataset places the limit on the speed of recognition. So our aim is to achieve perfect balance between number of feature values, size of dataset, speed of recognition and accuracy of recognition.

VI. CONCLUSION

Development of HCR for Devanagari script OCR is a challenging task. Here, we are designing a method which does the segmentation of handwritten characters and recognition using neural network. The attempt is to improve the performance in terms of time and to get better accuracy. It has been found that recognition of handwritten Devanagari character is quite difficult due to presence of shirorekha, conjunct characters and similarity in shapes for multiple characters. This system needs to be tested on a wider variety of images containing characters in diverse fonts and sizes. This work can be extended to character recognition for other languages.

ACKNOWLEDGEMENT

I am very much thankful to my respected project guide and Head of Dept. Prof. D. B. Kshirsagar, for his ideas and help proved to be valuable and helpful during the creation of this dissertation work. I am also thankful to our P.G. Coordinator Prof. P. N. Kalavadekar, for helping me while selecting and preparing dissertation work. I would like to thank all the faculties who have helped me during my dissertation work. Lastly, I am thankful to my friends who shared their knowledge in this field with me.

REFERENCES

[1] G. Sinha, Mrs. R. Rani, Prof. R. 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.

[2] M. Patil, V. Narawade, Recognition of Handwritten Devanagari Characters through Segmentation and Artificial neural networks , *International Journal of Engineering Research and Technology (IJERT)* , ISSN:2278-0181,Vol. 1 Issue 6, August 2012.

[3] K. Y. Rajput, S. Mishra, Recognition and Editing of Devanagari Handwriting Using Neural Network, *Proceedings of SPIT-IEEE Colloquium and International Conference*, Vol. 1.

[4] S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik and B. Portier, A Two Stage Classification Approach for Handwritten Devanagari Characters.

[5] V. Agnihotri, Offline Handwritten Devanagari Script Recognition, *IJITCS*, 2012, 8, 37-42.

[6] D. Singh, S. Singh and Dr. M. Dutta, Handwritten Character Recognition Using Twelve Directional Feature Input and Neural Network, *International Journal of Computer Applications*, 0975-8887, Vol. 1.

[7] N. Sharma, U. Pal, F. Kimura and S. Pal, Recognition of Off-Line Handwritten Devanagari Characters Using Quadratic Classifier.

[8] V. Bansal, R. Sinha, Segmentation of touching and fused Devanagari characters , *Pattern Recognition* 35,875-893, 2002.