IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. I (Jan – Feb. 2015), PP 60-64
www.iosrjournals.org

Artificial Neural Network For Recognition Of Handwritten Devanagari Character

Ms. Seema A. Dongare[1], Prof. Dhananjay B. Kshirsagar[2], Mr. Nitin J.Khapale[3], Ms. Amruta D. Pawar[4]

[12]*(Computer Department, SRESCOE Kopargaon, India)*
[3]*(E&TC Department, KBP Polytechnic Kopargaon, India)*
[4]*(IT Department, KBP Polytechnic Kopargaon, India)*

Abstract: *Information gathering is one of the most important thing in the digital world, many organizations still use handwritten documents for storing their information. So it is important task to convert this information in to digital world. Handwritten character recognition helps user to convert their handwritten documents in to digital information. Purpose of this dissertation work is to implement zone based feature extraction method which is combination of image centroid zone and zone centroid zone of individual character image. In feature extraction using zone based approach, individual character image is divided into n equal size zones then average distance of all pixels with respect to image centroid or zone centroid is computed. In combination of image centroid and zone centroid approach it computes average distance of all pixels present in each zone 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.*

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. Handwritten character recognition is a field of OCR. Basically there are two classes of HCR 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 represented as function of time as well strokes made by user are also considered. Handwritten character recognition is an intelligent optical character recognition capable of handling the complexity of writing, writing environment, materials, etc.

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.

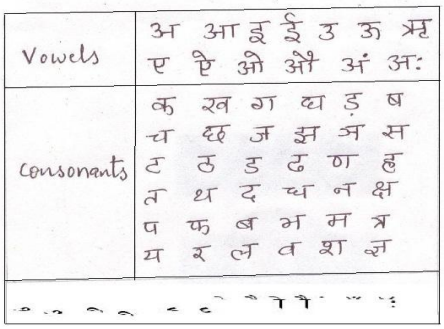

Fig. 1 Devanagari Character Set

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.

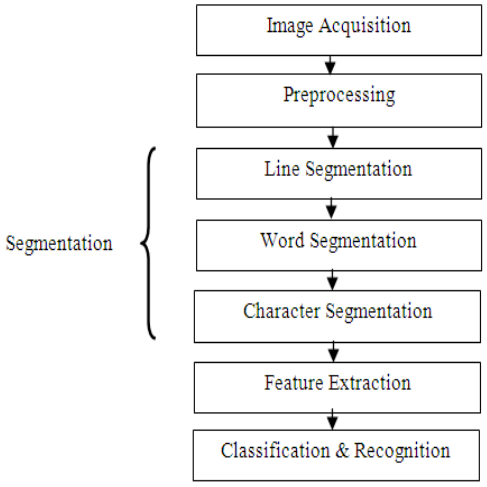

Fig. 2 Block Diagram of system

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 have implemented grid based approach which is the combination of image centroid zone and grid centroid zone of individual character or numerical image. In this technique individual character image is resized to a size of 256X256 and then divided into 16 equal sized grids each of size 64X64. Average distance of all pixels with respect to image centroid and zone centroid is computed. It gives feature vector of size 2X16 features.

Algorithm: Image Centroid Zone (ICZ) and zone centroid zone (ZCZ) feature extraction.

Input: Preprocessed individual character image.

Output: Extracted features for classification and recognition.

Algorithm:

Begins

Step 1: Divide an input image in to n equal sized grids.

Step 2: Compute distance between the image centroid and each pixel present in the grid.

Step 3: Compute the distance between the grid centroid and each pixel present in the grid.

Step 4: Repeat step 2 and 3 for the entire pixels present in the grid.

Step 5: Computation of average distance between these points with respect to image centroid.

Step 6: Computation of average distance between these points with respect to grid centroid.

Step 7: Repeat this procedure for all grids.

Step 8: Obtaining 2Xn such features for classification and recognition.

Ends

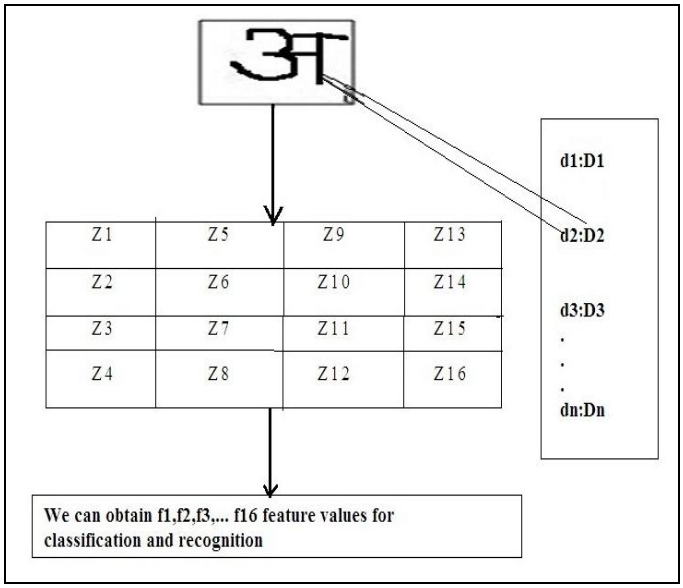

Fig. 3 Feature Extraction from Devanagari Numeral Image "A"

Fig. 3 shows an illustration of procedure to extract features from Devanagari character "A". Let 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 zone. Same procedure is repeated sequentially for each of zone. With combination of ICZ and ZCZ we will have two feature values per zone which gives 32 feature values provided no of zones are as 16 [1].

4.3.4 Recognition of Characters using Neural Network

The back end 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.

Fig. 4 shows structure of feed forward neural network with one input, hidden and one output layer. We have implemented feed forward neural network with following specification.

- Input nodes: 32 input neurons (2 X 16 for combination of ICZ and ZCZ)
- Output nodes : 41 output neurons.
- No of Hidden layer : 1 with 4 nodes.
- Training algorithm : Error Back Propagation.
- Performance function : Mean Square Error.

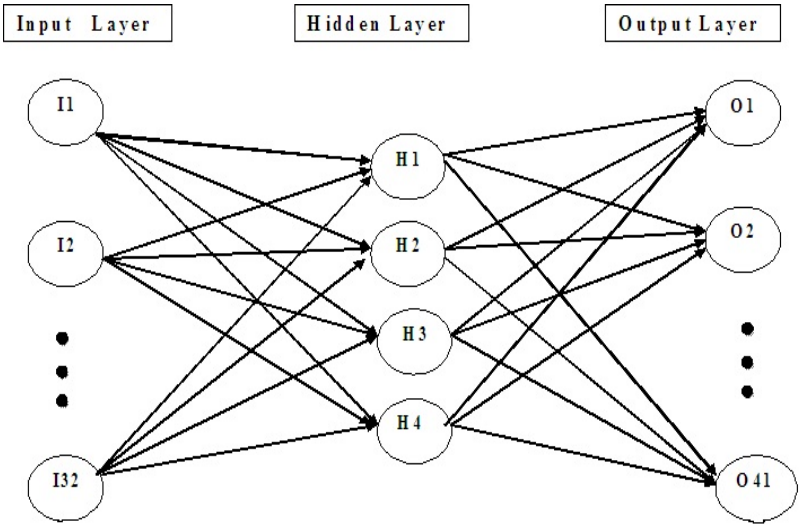

Fig. 4 Neural Network Structure

V. Result And Discussion

We have studied various research papers which reveal that the difficulty of handwritten character recognition has two aspects. The first is attributed to the writer variations in style, size, shape, ink color, ink flow and thickness, digitization imperfections etc. The second is the deficiencies of the particular method used for feature extraction. With consideration of these two aspects neural network training and testing has been performed and observations are recorded.

We have created synthetic dataset consisting of total 2460 patterns out of which we have used 1640 patterns for training and 820 patterns for test. We have divided this dataset into two datasets as Dataset1 and Dataset2 which contains 820 and 1640 patterns of each alphabet respectively. Testing has been performed on separate dataset containing 20 patterns of each alphabet total of 820 patterns. During testing we have recorded accuracy in terms of number of characters recognized accurately with percentage accuracy for each character.

From Table1 we can analyze that with increase in size of dataset used to train neural network, accuracy of recognition increases. Accuracy of the system depends on writing order and style, font size and color, thickness.

Table1. Combined Result Set for Dataset1 and Dataset2

Sr. No.	Character	% Accuracy Training: Dataset1	% Accuracy Training: Dataset2
1	k	70	80
2	ba	55	80
3	[	55	80
4	na	55	80
5	t	55	75

From these two we can observe that when size of training dataset increases recognition accuracy increases. Neural network accuracy depends on size of dataset used to train it. Larger the dataset increases accuracy but decreases speed of recognition. For few characters accuracy remains same or slight increase has been observed, reasons behind this can be due to similarity in shapes of multiple Devanagari characters, or may be simply due to variations in writing styles. From Fig. 5 we can conclude with increase in size of dataset used to train neural network, accuracy of recognition increases.

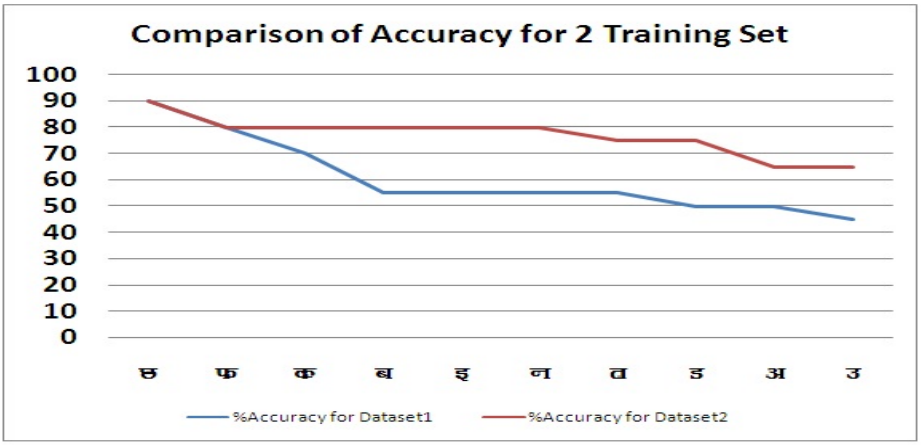

Fig. 5 Comparative Accuracy Analysis for Dataset1 and Dataset2

The dataset consists 2460 patterns of Devanagari alphabets. We have used zone based feature extraction techniques on these samples in which we have extracted 32 feature values from each character image. We have obtained different recognition accuracy on different characters. First we have tested system on 20 test patterns of each alphabets, the highest accuracy obtained for character is 90. Then we have tested it on test document with alphabets here maximum accuracy we have achieved is 77. And finally testing has been done on different documents where we get average accuracy as 75%.

VI. Conclusion

Development of handwritten Devanagari OCR is a challenging task in Pattern recognition area. Here, we proposed a method which does the segmentation of handwritten document into lines, words and characters. And further recognition process will be done with the help of neural network. This system needs to be tested on a wider variety of images containing characters in diverse fonts and sizes. It has been found that recognition of handwritten Devanagari characters is quite difficult task due to presence of header line, conjunct characters and similarity in shapes for multiple characters.

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 e 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.