I.J. Information Technology and Computer Science, **2012, 8, 37-42**
Published Online July 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2012.08.04

Offline Handwritten Devanagari Script Recognition

Ved Prakash Agnihotri
Department of Computer Science and Engineering, LPU Phagwara, India
vedagni86@gmail.com

***Abstract*—** Handwritten Devanagari script recognition system using neural network is presented in this paper. Diagonal based feature extraction is used for extracting features of the handwritten Devanagari script. After that these feature of each character image is converted into chromosome bit string of length 378. More than 1000 sample is used for training and testing purpose in this proposed work. It is attempted to use the power of genetic algorithm to recognize the character. In step-I preprocessing on the character image, then image suitable for feature extraction as here is used. Diagonal based feature extraction method to extract 54 features to each character. In the next step character recognize image in which extracted feature in converted into Chromosome bit string of size 378. In recognition step using fitness function in which find the Chromosome difference between unknown character and Chromosome which are store in data base.

***Index Terms*—** Handwritten Character Recognition, Image Processing, Feature Extraction, Chromosome Bit String.

I. Introduction

Handwriting recognition has been one of the fascinating and challenging research areas in field of image processing and pattern recognition in the recent years [1][2]. It contributes immensely to the advancement of an automation process and can improve the interface between human beings and machine in numerous applications. Several research works have been focusing on new techniques and methods that would reduce the processing time while providing higher recognition accuracy [3].

In general, handwriting recognition classified into two types as off-line and on-line handwriting recognition methods. In off-line recognition, the writing is usually captured optically by a scanner and complete writing is available as an image. 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 methods have been shown to be superior to their off-line counterparts in recognizing handwritten characters due to the temporal information available with the former [4][5]. Several applications including mail sorting, bank processing, document reading 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 [6][7].

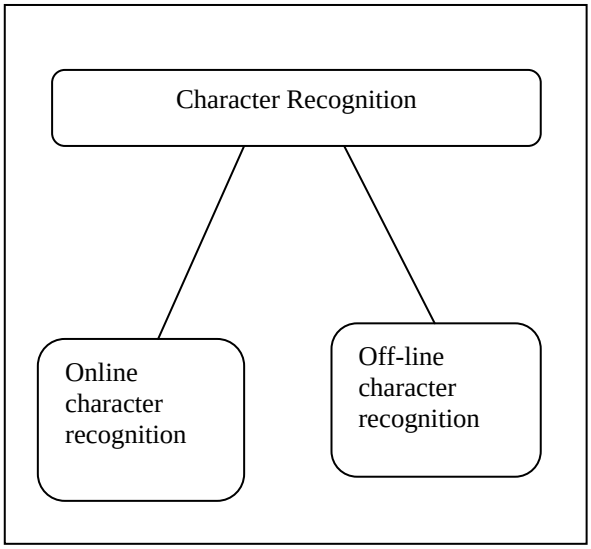

Fig1. Type of Character recognition

The first important step in any handwritten recognition system is pre-processing followed by segmentation and feature extraction. Pre-processing includes the steps that are required to shape the input image into a form suitable for segmentation [8]. In the segmentation, the input image is segmented into individual characters and then, each character is resized into m x n pixels toward training neural network. The selection of appropriate feature extraction method is probably the single most important factor in achieving high performance. Several methods of feature extraction for character recognition have been reported in the literature [9]. The widely used feature extraction methods are Template matching, Deformable templates, Unitary Image transforms, Graph description, Projection Histograms, Contour profiles, Zoning, Geometric

Copyright © 2012 MECS

moments invariants, Zernike Moments, Spline curve approximation, Fourier descriptors, Gradients feature and Gabor features.

An artificial neural network as the backend is used for performing classification and recognition tasks. In the off-line recognition system, the neural networks have emerged as the fast and reliable tools for classification towards achieving high recognition accuracy [10]. Classification techniques have been applied to handwritten character recognition since 1990s. These methods include statistical methods based on Bayes decision rule, Artificial Neural Networks (ANNs), Kernel Methods including, Support Vector Machines (SVM) and multiple classifier combination [11][12].

Handwritten character recognition is not a simple task. Character recognition is complex task, even after writing people are not able to understand but he/she written. So to reach 100% accuracy is very difficult job. Many researchers have done lots of work in this field but, 100% accuracy in not achieved.

In this paper, a diagonal feature extraction scheme for the recognizing off-line handwritten characters is used. In the feature extraction process, resized individual character of size 90x60 pixels is further divided into 54 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 features are converted into Chromosome bit string of size 378. Feature is extracted from each zone of size 10x10 is converted into 7 bit string so, there are 54 zone in each character image so, 54x7=378 bit string is used to represent each character image.

The paper is organized as follows, in Section II; the proposed recognition system is presented. The feature extraction procedure adopted in the system is detailed in the section III. Section IV describes the classification and recognition. Section V presents the experimental results and finally, the paper is concluded in section VI.

II. Application of Offline Handwritten Character Recognition System

Some of the important applications of offline handwritten character recognition are listed in the following section:

A. Cheque Reading

Offline handwritten Character recognition is basically used for cheque reading in banks. Cheque reading is very important commercial application of offline handwritten character recognition. Handwritten character recognition plays very important role in banks for signature verification and for recognition of amount filled by user.

B. Postcode Recognition

Handwritten character recognition system can be used for reading the handwritten postal address on letters. Offline handwritten character recognition system used for recognition handwritten digits of postcode. HCR can be read this code and can sort mail automatically.

C. Form processing

HCR can also be used for form processing. Forms are normally used for collecting public information. Replies of public information can be handwritten in the space provided.

D. Signature verification.

HCR can also used for identify the person by signature verification. Signature identification is the specific field of handwritten identification in which the writer is verified by some specific handwritten text.

III. Devanagari Script Recognition System

In this section, the proposed recognition system is described. A typical handwriting recognition system consists of pre-processing, segmentation, feature extraction, classification and recognition, and post processing stages. The schematic diagram of the proposed recognition system is shown in Fig.3

A. Image Acquisition

In Image acquisition, the recognition system acquires a scanned image as an input image. The image should have a specific format such as .jpeg, .bmp etc. This image is acquired through a scanner, digital camera or any other suitable digital input device.

B. Pre-processing

The pre-processing is a series of operation performed on the scanner input image. It essentially enhances the image rendering it suitable for segmentation. The various tasks performed on the image in pre-processing stage shown in Fig. 2. Binarization process converts a gray scale image into a binary image using thresh holding technique. Detection of edges in the binarized image using canny technique, dilation the image are the operations performed in the last two stages to produce the pre-processed image suitable for segmentation [13].

Copyright © 2012 MECS

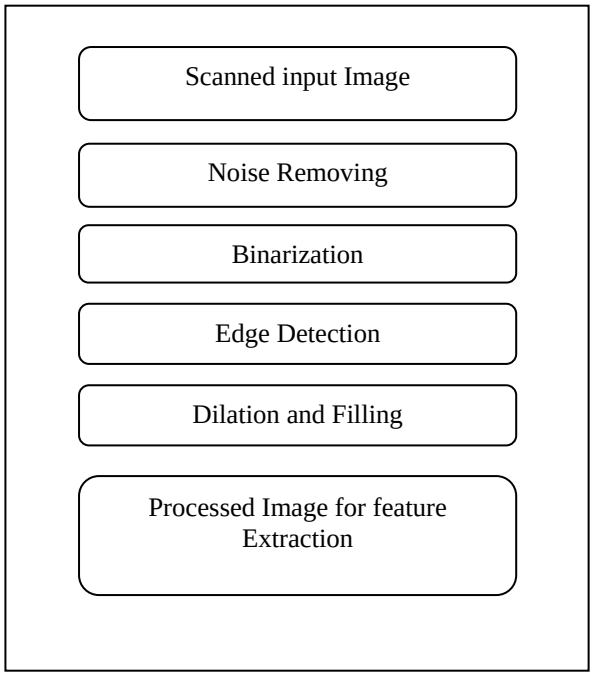

Fig. 2 Pre-processing of handwritten character

C. Segmentation

In the segmentation stage, an image of sequence of characters is decomposed into sub-images of individual character. In the proposed system, the pre-processed input image is segmented into isolated characters by assigning a number to each character using a labeling process. This labeling provides information about number of characters in the image. Each individual character is uniformly resized into 90x60 pixels for classification and recognition stage.

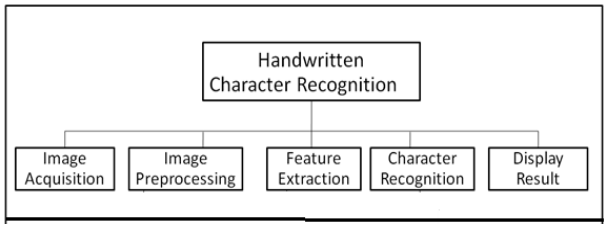

Fig.3 Proposed recognition system.

Preprocessing Algorithm

1. start
2. I=imread('sample1.jpg'); Read image using imread(); method
3. I=im2bw(I,level) convert into binary image
4. Detecting Edge using Sobel Mehtod
5. Imwrite(I,'sample.bmp','bmp');
6. End

IV. Feature Extraction Method

In this stage, the features of characters that are crucial for classifying them at recognition stage are extracted. This is an important stage as its effective functioning improves the recognition rate and reduces the misclassification [14]. First of all try to understand Devanagari script, Devanagari script written left to right along a horizontal line. Its basic set of symbols consists of 34 consonants ('vyanjan') and 11 vowels ('svar'). Characters are joined by a horizontal bar that creates an imaginary line when Devanagari text is suspended and no spaces are used between words. A single or double vertical line called 'Puran Viram' was traditionally used to indicate the end of phrase or sentence. In part, Devanagari owns its complexity to its rich set of conjuncts. The language is partly phonetic in that a word written in Devanagari can only be pronounced in one way, but not all possible pronunciations can be written perfectly. Diagonal feature extraction scheme for recognizing off-line handwritten characters in proposed in this work. Every image of size 90x60 pixels is divided into 54 equal zones, each of size 10x10 pixels (Fig.4(c)). The features are extracted from each zone pixels by moving along diagonals of its respective 10x10 pixels. Each zone has 19 sub-features values are averaged to form a single feature value and placed in the corresponding zone (Fig. 4(b)). This procedure is sequentially repeated for the all the zones. There could be some zones whose diagonals are empty of foreground pixels. The feature values corresponding to these zones are zero. Finally, 54 features are extracted for each character.

Divide Character Image into Zones Algorithm

1. Start
2. Read image I=imread('sample.bmp');
3. Resize image into 90x60 B=resize(I, [90 60]);
4. Find number of row and column [c1 c2]=size(B);
5. Set bs=10; Block size (10x10)
6. Set=nob=(c1/bs)*(c2/bs); Total number of 10x10 block
7. Set k=0; kk=0;
8. For i=1:(c1/bs)
    a. For j=1:(c2/bs)
    b. K=k+1;
    c. Block(:,:,kk+j)=B((bs*(i-1)+1:bs*(i-1)+bs),(bs*(j-1)+1:bs*(j-1)+bs)); imwrite(Block(:,:,kk+j),strcat(int2str(k),'.bmp'),'bmp');
    d. end
    e. kk=kk+(c1/bs);
9. end

10. End (Dividing into Blocks and write each block into image)

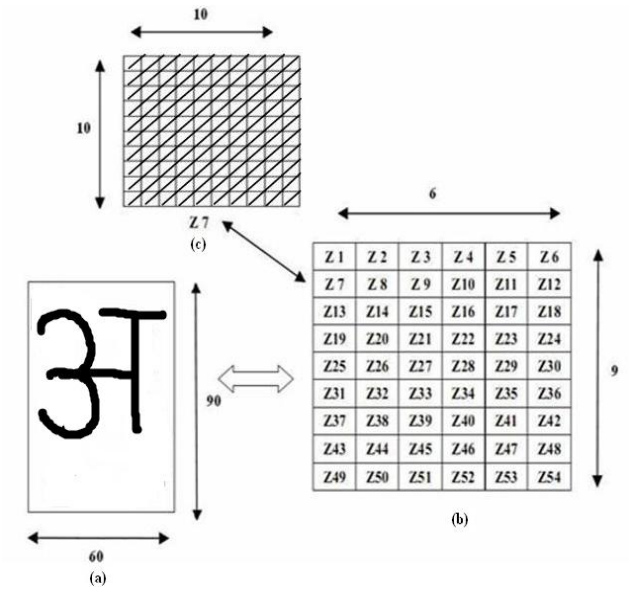

Fig. 4(a) Procedure for extracting feature from characters

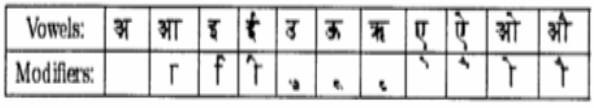

Fig. 4(b) Vowels and corresponding modifiers

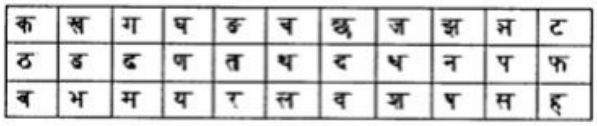

Fig. 4(c) Consonant

V. Classification and Recognition

The classification stage is the decision making part of a recognition system and it use the features extracted in the previous stage. It is very important stage, in this stage input is classified that in which class particular input is belong. Feature which are extracted from pervious stage are used to form chromosome, after that this chromosome bit string is used to recognize the Devanagari character image. This phase consists of two functions. 1) Chromosome function generation. 2) Chromosome fitness function.

Each of function has the following details:-

Chromosome function generation: **-** The Chromosome generation functions generate Devanagari character Chromosome of length 162 bit, by combining all feature extracted from pervious phase. The sample Chromosome for Devanagari Character " क " is "01100001110000000001101101110101001011011001 11001001001100110001010110111010100100010110111101000100010110111000100111000111010010011001000111101011101011001100" where 0s is show that none of feature is shown in character image, and 1s is showing that feature present in the character image.

Chromosome String Generation Algorithm

1. Start
2. Clear global bitString;
3. Set global bitString;
4. For i=1:54
5. I=imread(strcat(int2str(i),'.bmp'));
6. Set A=spdiags(I);
7. Set Avg=0;
8. Find Size of A [r c]
9. For j=1:c
10. For k=1:r
11. Avg=Avg+A(k,j);
12. End
13. End
14. DiagonalAverage(i)=ceil(Avg/19);
15. bitString=horzcat(bitString,dec2bin(DiagonalAverage(i),7));
16. End

Chromosome fitness function: **-** *The* Chromosome bit string from the Chromosome generation function is used to recognize Devanagari character by comparing the fitness value of an unknown character with all the Devanagari character which is store in database during training process. The highest fitness value is the recognition result. The fitness value is calculated as follow:

$$fitness\ value = \sum_{i=1}^{n} S - L \quad (1)$$

Where S is Chromosome bit strings which are store in Database and L is Chromosome bit string of unknown character.

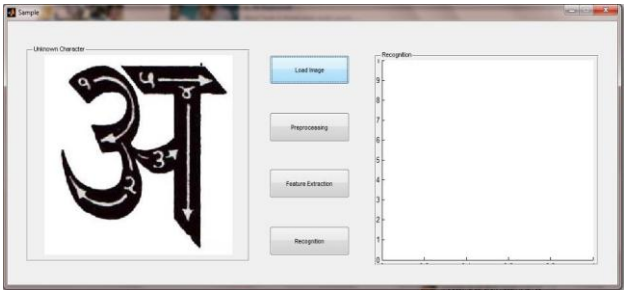

Fig. 5: User Interface

Copyright © 2012 MECS

VI. Design

The HDCR system interface is shown in Figure 13. The interface has 2 panels. The first panel show unknown character which system try to recognize. The next one is result panel. This panel is used to show the picture of a character which is a result of recognition. The HDCR Interface contains 4 Buttons. First Button1 (Load Image) is used to upload scanned image which want to recognize and show on panel number 1. Button 2 (Pre-processing) performs pre-processing function so, it become easy to extract feature from image. Next Button 3 (Feature Extraction) extracts all essential features from the selected image and store in the form of bit String of size 162. At last Button 4(Recognition) recognize unknown character and show result in result panel as image.

VII. Results and Discussion

The recognition system has been implemented using Matlab 7.10.0 (R2010a). The scanned image is taken as dataset /input. The experiment is conducted on more than 1000 characters. The testing characters are separated into data sets, the training data set and testing data set. The training set contains 904 characters and testing set contains 204 characters. The precision of offline Devanagari system is 85.78% match, 13.35% mismatch.

VIII. Issue Regarding System

Handwritten Devanagari character recognition is discussed here. It has been found that recognition of handwritten Devanagari characters is very difficult task. Follow are main reasons for difficulty in recognition of Devanagari Characters:-

- Some Devanagari character similar in shape
- Different or even the same writer can write differently times depending upon pen or pencil.
- The character can be written at different location on paper or its window.
- Character can be written in different fonts

IX. Future Work and Scope

A simple off-line handwritten Devanagari script recognition system using feature extraction, namely, diagonal feature extraction is used. The main objective in this research paper is to genetic algorithm technique in Devanagari Character recognition. In this research paper I use genetic algorithm to generate Chromosome bit string from the 54 feature which in extracted using diagonal based technique. The success of any recognition system is depends on feature and classifier which is used to classify the unknown input to well define class. So this need more complex method like mutation, etc. technique used to optimize the solution using genetic algorithm. Using genetic algorithm which is less use in Devanagari and any other Indian languages is more productive. But as the concern of handwritten character each person write character in its own way, so well define structure is not applicable in this type of problem. My future work is to recognize character of Devanagari script using Neural Network and recognize Devanagari word is step forward in Handwritten Word recognition.

Acknowledgement

With deep sense of gratitude I express my sincere thanks to my esteemed and worthy supervisor **Mr. Tejinder Thind** in the Department of Computer Science and Engineering for his valuable guidance in carrying out this work under his effective supervision. My greatest thanks are to all who wishes me success especially my brother **Dr. Anil Agnihotri** whose support and care makes me complete this work**.**

References

[1] S. Mori, C.Y. Suen and K. Kamamoto,"Historical review of OCR research and development," Proc. Of IEEE, vol. 80, pp. 1029-1058, July 1992.

[2] S. Impedovo, L. Ottaviano and S. Occhinegro,"Optical character recognition", International Journal Pattern Recognition and Artificial Intelligence, vol. 5(1-2), pp. 1-24, 1991.

[3] V. K. Govindan and A.P. Shivaprasad,"Character Recognition-A review," Pattern Recognition, vol. 23, no. 7, pp. 671-683, 1990.

[4] R. Plamondon and S. N. Srihari,"On-line and Off-line handwritten character recognition: A comprehensive survey," IEEE Transactions o Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 63-84, 2000.

[5] N. Arica and F. Yarman-Vural,"An Overview of Character Recognition Focused on Off-line Handwriting",IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2001, 31(2), pp. 216-233.

[6] U. Bhattacharya, and B. B. Chaudhuri,"Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals,"IEEE Transaction on Pattern analysis and machine intelligence, vol. 31, no. 3, pp. 444-457, 2009.

[7] U. Pal, T. Wakabayashi and F. Kimura,"Handwritten numeral recognition of six popular scripts," Ninth International conference on

Copyright © 2012 MECS

Document Analysis and Recognition ICDAR 07, vol. 2, pp. 749-753,2007.

[8] R. G. Casey and E. Lecolinet,"A Survey of Methods and Strategies in Character Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligance, vol. 18, no. 7, pp. 690-706, july 1996.

[9] Anil K. Jain and Torfinn Taxt,"Feature extraction methods for character recognition-A Survey," Pattern Recognition, vol. 29, no. 4, pp. 641-662, 1996.

[10] R. G. Casey and E. Lecolinet,"A Survey of Methods

[11] C. L. Liu, H. Fujisawa,"Classification and learning for character recognition: Comparsion of methods and remaining problems," International Workshop on Neural Networks and Learning in Document Analysis and Recognition, Seoul, 2005.

[12] F. Bortolozzi, A. S. Brito, Luiz S. Oliveira and M. Morita,"Recent advances in handwritten recognition," Document Analysis, Umapada Pal, Swapan K. Parui, Bidyut B. Chaudhuri, pp. 1-13.

[13] Rafael C. Gonzalez, Richard E. woods and Steven L. Eddins, Digital Image Processing using MATLAB, Pearson Education, Dorling Kindersley, South Asia, 2004.

[14] S. V. Rajashekararadhya, and P. Vanajaranjan," Efiiciency zone based feature extraction algorithm for handwritten numeral recognition of four popular south-Indian scripts," Journal of Theoretical and Applied Information Technology, JATIT, vol. 4, no. 12, pp. 1171-1181,2008.

Ved Prakash Agnihotri: Ved Prakash Agnihotri a science graduate from HPU, Shimla. He has done MCA from Punjabi University, Patiala in 2009. He is currently a research scholar of LPU, Phagwara. His interests are in Machine Learning, Pattern Recognition and Artificial Intelligence areas.