International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 6, Issue 1, January 2017 **ISSN 2319 - 4847**

Recognition of Off-Line Handwritten Devanagari Characters Using 64dimensionalFeature Extraction

Prof. L.M.Deshpande[1], Ms. Vimal R. Kshirsagar[2]

[1] (Department of Electronics and Telecommunication Engineering, College of Engineering Osmanabad, Maharashtra, India)

[2] (Department of Electronics and Telecommunication Engineering, College of Engineering Osmanabad, Maharashtra, India)

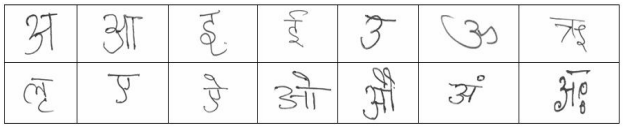
ABSTRACT

Recognition of handwritten characters is a challenging task because of the variability involved in the writing styles of different individuals. In this paper we are concern with features from 64 dimensional feature extraction techniques for the Devanagari character. These features are used for further classification. Histograms of direction chain code of the contour points of the characters are used as feature for recognition [4].From this we get 196 features for future classification. We suppose to be used Multi Layer Perceptron (MLP) based Classifier.

Keywords: 64 dimensional features, weighted majority voting technique (Multi Layer Perceptron)

1. INTRODUCTION

Recognition of handwritten characters has been a popular research area for many years because of its various application potentials. Devanagari is third most widely used script, used for several major languages such as Hindi, Sanskrit, Marathi and Nepali, and is used by more than 500 million people[1].Unconstrained Devanagari writing is more complex than English cursive due to the possible variations in the order, number, direction and shape of the constituent strokes. Devanagari script has 50 characters which can be written as individual symbols in a word (some shown in figure 1). Devanagari Character recognition is complicated by presence of multiple loops, conjuncts, upper and lower modifiers and the number of disconnected and multistroke characters [3],



(a)

(b)

Figure 1 Samples of handwritten Devanagari basic characters (a) Vowels, (b) Consonants

Although first research report on handwritten Devanagari characters was published in 1977 [6] but not much research work is done after that. At present researchers have started to work on handwritten Devanagari characters and few research reports are published recently. Hanmandlu and Murthy [7, 11] proposed a Fuzzy model based recognition of handwritten Hindi numerals and characters and they obtained 92.67% accuracy for Handwritten Devanagari numerals and 90.65% accuracy for Handwritten Devanagari characters.

In this paper, we propose a scheme for unconstrained off-line handwritten Devanagari character recognition based on the feature obtained from chain code histogram. Here the bounding box of a character is segmented into blocks and chain code histogram is computed in each blocks. For future work this chain code features are then fed to the classifier for recognition. Rest of the paper is organized as follows. In Section 2 we discuss about Devanagari language data collection and the preprocessing of the data used for the proposed scheme. Feature extraction procedure is presented in Section 3. In Section 4 discussion about details of the histogram of different writers. Future classifier is discussed in section 5. Conclusion on the paper is given in Section 6 and for our paper we refer he different references in section 7.

2. PROPOSED SCHEME

In our proposed method first we were collected the data from different writers. These data samples are same as shown in figure 1.This collected data is further proposed for normalization.

2.1 DATA COLLECTION

Collected handwritten data from different writers in a paper document are usually captured by optical scanning and stored in a file of picture elements, called pixels. Because of the writing styles of different individuals, characters can have different shapes. As a result recognition of unconstrained handwritten characters becomes a difficult task. So we have to resize the image convert it in to bitmap file by preprocessing as given below.

2.2 PREPOSSESSING

A flatbed scanner was used for scanning. This image is jpg color image we have to convert it into grayscale image and then to two-tone (0 and 1) images (Here '1' represents object point and '0' represents background point) it is threshold binary image. For removing noises from the images, we have used a method discussed in [5].
We have to calculate all top, bottoms, left and right boundaries to resize the no of pixels which are exact multiple of 4 for simplification purpose.

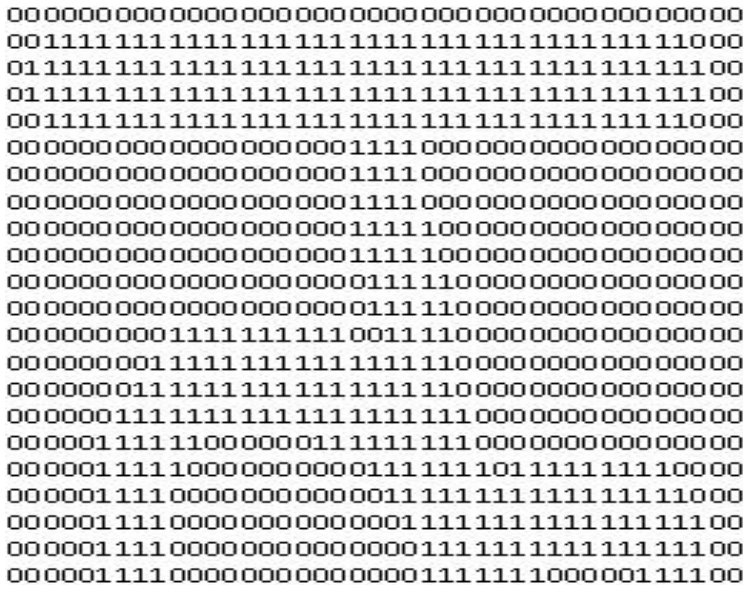


Figure 2 Binary image

3. FEATURE EXTRACTION

Here we use 64 dimensional features [8]for our recognition purpose. The feature extraction techniques are described below.

3.1 64 Dimensional Feature Extraction

The above scaled image which is having pixel size in exact multiple of 4 used for finding out the features. First find the contour points of the image by the canny method. The Canny method finds edges by looking for local maxima of the gradient of I. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method detected true weak edges as shown in fig.3.

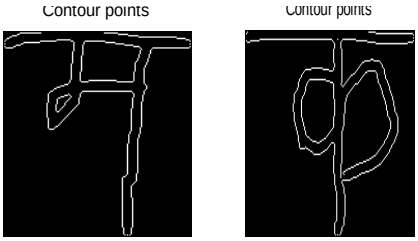


Figure 3 It shows contour points

This above bonding box is divided into 4x4 blocks which results into 16 blocks. Then for each block 4 dimensional features are computed. In each of these blocks, the direction chain code for each contour point is noted and the frequency of the direction codes is computed. Here we use chain code of four directions only [directions 0 (horizontal), 1 (45 degree slanted), 2(vertical) and 3 (135 degree slanted)]. See Fig 4(b) for illustration of four chain code directions.

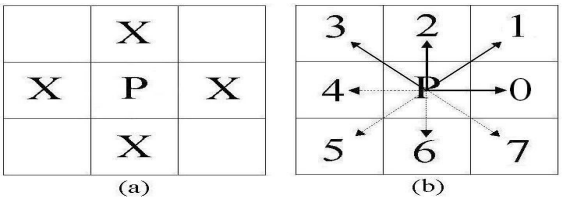
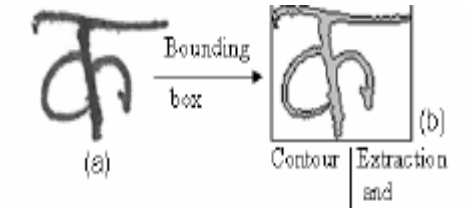
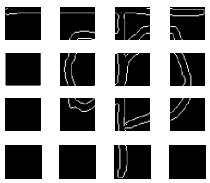


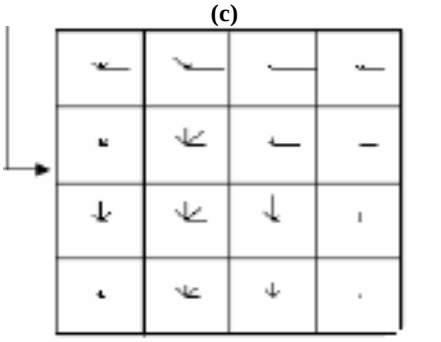
Figure 4 (a) For a point P and its four neighbors are shown by 'X', (b) For a point P the direction codes for its eight neighboring points are shown.

We assume chain code of direction 0 and 4, 1 and 5, 2 and 6, 3 and 7, are same. Thus in each block we get an array of four integer values representing the frequencies and those frequency values are used as feature. Histogram of the values of these four direction codes in each block of a Devanagari character is shown in Fig.5 (d). Thus, for 4x4 blocks 4x4x4=64 features are computed. To normalize the features we compute maximum value of the histograms from all the blocks. We divide each of the above features by this maximum value to get the feature values between 0 and 1.For example one sample of chain code and its count shown in fig.4 (e)





(c)



(d) 64 Dimensional Features

Chain Code	Count
0	1
1	0
2	11
3	3

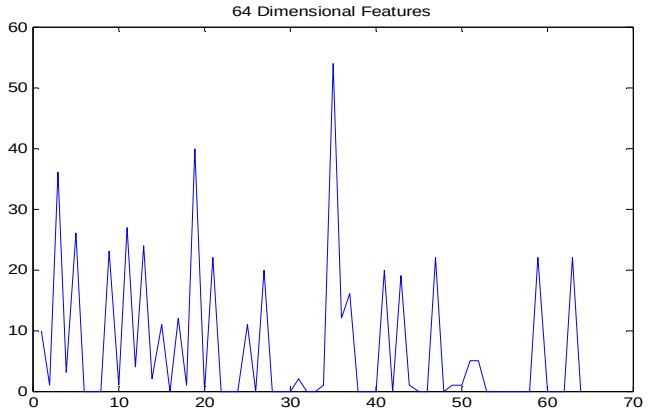
(e)

Figure 5 Pictorial representation of the 64 dimensional feature extraction process for a sample Devanagari character. (a) Two tone image of a Devanagari character, (b) Bounding box of the character. (c) Contour of the character shown in black color and the bounding box is segmented into4 X 4 blocks. (d) Chain code histogram (e) sample chain code and its count.

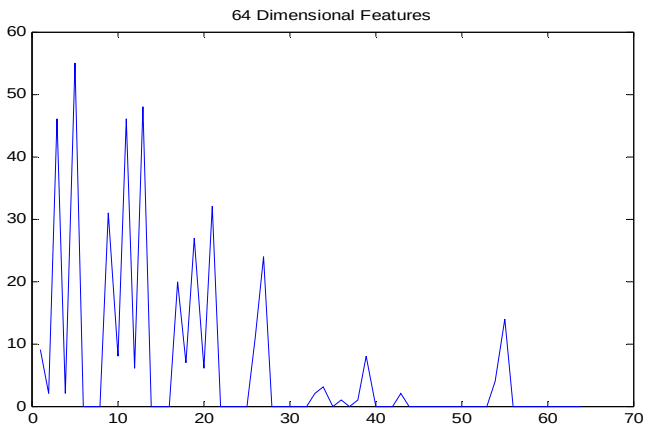
The chain code for the character contour will yield a smooth, unbroken curve as it grows along the perimeter of the character and completely encompasses the character. When there is multiple connectivity in the character, then there can be multiple chain codes to represent the contour of the character. From this algorithm we are computing features from single character,if we have to compute features from compound character then we have first segmented that particular character by using another algorithm.That process is called as segmentation[9].

4. RESULT AND DISCUSSION

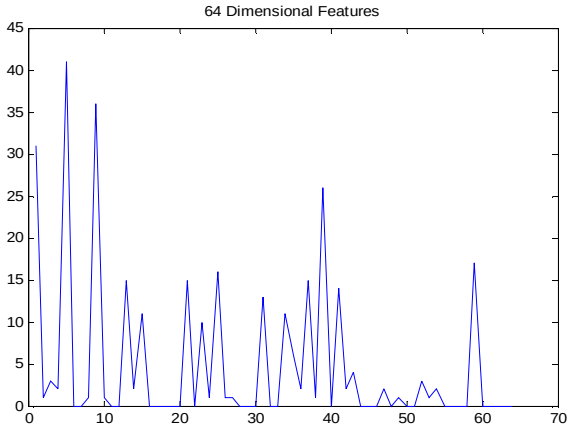
In this section we are going to discuss about the result that we get after performing feature extraction. As we know from above 64 dimensional feature extraction is used here for computing features of handwritten Devanagari character from different writers. Data used for the present work was collected from different individuals. For example we are showing the histogram of one character image written by different writers. These histogram means 64 dimensional features which are represented graphically.For the same character the graph is looking similar that is it having near about same values but not exactly same as shown in fig 5.



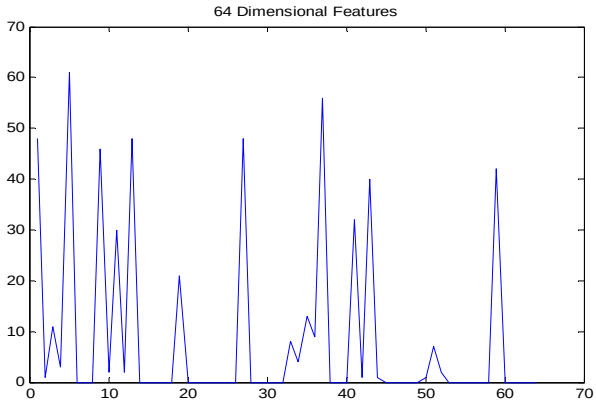
(a)



(b)



(c)



(d)

Figure 6 Histogram of character  from different four writers

These features are used for training as the number of data image is more it will give you better result of classification.So that we are going to use as much as possible data for classification.

5.FUTURE SCOPE

In this section we discuss about classifier that we will going to use. At this stage all characters are non compound, single characters so segmentation is not required. We will going to propose a Multi-Layer Perceptron (MLP) neural network based classification approach for the recognition of Devanagari handwritten characters. In general Multilayer perceptron using a backpropagation algorithm are the standard algorithm for any supervised learning pattern recognition process. The multilayer perceptron consists of three or more layers an input and an output layer with one or more hidden layers.
The features obtained from feature extraction will fed as an input to MLP. This is for our future work.

6. CONCLUSION

From proposed method we get 64 feature for training purpose. These are fed to classifier for the future work. In future again we plan for another one or more feature extraction methods in combination with MLP classifier or more different classifiers.

REFERENCES

[1]. U. Pal, B. B. Chaudhuri, "Indian Script Character recognition: A survey", Pattern Recognition, vol. 37, pp. 1887-1899, 2004.
[2]. Mohamed Cheriet, Nawwaf Kharma, Cheng-Lin Liu, Ching Y. Suen, "Character Recognition Systems:
[3]. L. Koerich, Large Vocabulary off-line handwritten word recognition. PhD thesis, Ecole de Technologic Superieure, Montreal- Canada, August 2002.

[4]. T. Wakabayashi, S. Tsuruoka, F. Kimura and Y. Miyake, "Increasing the Feature size in handwritten Numeral Recognition to improve accuracy, System and Computers in Japan, Vol.26, No.8, pp.35-44, 1995.
[5]. B. B. Chaudhuri and U. Pal, "A complete printed Bangla OCR system", Pattern Recognition, vol. 31, pp. 531-549, 1998.
[6]. I.K. Sethi and B. Chatterjee, "Machine Recognition of constrained Hand printed Devnagari", Pattern Recognition, Vol. 9, pp. 69-75, 1977
[7]. M. Hanmandlu and O.V. Ramana Murthy, "Fuzzy Model Based Recognition of Handwritten Hindi Numerals", Intl.Conf. on Cognition and Recognition, pp. 490-496, 2005.
[8]. N. Sharma, U. Pal*, F. Kimura**, and S. Pal " Recognition of Off-Line Handwritten Devnagari Characters Using QuadraticClassifier", P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 805 – 816, 2006.© Springer-Verlag Berlin Heidelberg
[9]. Naresh Kumar Garg, Lakhwinder Kaur,and M. K. Jindal "Segmentation of Handwritten Hindi Text", International Journal of Computer Applications (0975 – 8887)Volume 1 – No. 4, ©2010.
[10]. M. Hanmandlu, O.V. Ramana Murthy, Vamsi Krishna Madasu, "Fuzzy Model based recognition of Handwritten Hindi characters", IEEE Computer society, Digital Image Computing Techniques and Applications, 2007