IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834, p-ISSN: 2278-8735
PP 05-09
www.iosrjournals.org

Devnagri Character Recognition A Review

Barawkar M. B.[1]
[1](*E&TC Dept. ,RCPIT Shirpur ,NM Univ.Jalgaon, India)*

Abstract: *This paper is a review of Devnagri character recognition. Generally character recognition system is divided into two types as online and offline character recognition system. Process of character recognition involves the common steps as preprocessing, segmentation and feature extraction. Feature extraction is used to classify an unknown handwritten character into one of the known classes. Finally character gets recognized by pattern matching. In case of character recognition system especially Devnagri character recognition is more complicated as it having multiple loops, conjuncts, upper and lower modifiers and the number of disconnected and multistroke characters.*

Keywords – *Classification, Devnagri Character, Feature Extraction, Recognition and Segmentation.*

1. INTRODUCTION

Devnagri is one of the official languages in India and which is used by majority of peoples. It belongs to group of language along with Marathi, Hindi, Bengali, Gujrathi , Konkani and other north Indian languages. [1, 2] Hindi is world's third most commonly used language after Chinese and English. [3] There are approximately 500 million people all over the world who speak and write in Hindi. In Devnagri, all letters are equal, i.e. there is no capital or small letters. Devnagri script consists of thirteen vowel symbols called matras and thirty six consonants called vyanjans as shown in fig. 1 and fig. 2 respectively. Devnagri is written from left to right. The characters are normally aligned below the line of writing. Devnagri script is written in a non-linear fashion and it is written from left to right side. The width of the character is also not constant.

Devnagri Character recognition is very important field as conventional computers will not easily detect each type of characters. Recognition of handwritten or printed character or number is important task in document analysis.

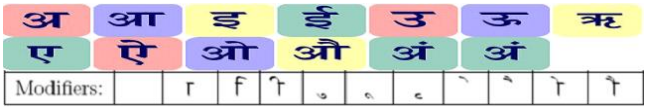

Fig. 1. Vowels and Corresponding Modifiers

Fig. 2: Consonant

1. Complexity of Devnagri Scripts

- All the individual characters are joined by a head line called "ShiroRekha" in case of Devnagri Script. This makes it difficult to isolate individual characters from the word.
- There are various isolated dots, which are vowel modifiers, namely, "Anuswar", "Visarga" and "Chandra Bindu", which add up to the confusion.
- Minor variations in similar characters. For letter which is similar in structure, variations in the hand-written input may give the wrong identification. The letters ma, bha, tha and ya, for example are often confused with each other. Moreover, if the top portion of bha is not detailed, it may be confused with ma or tha. Similarly bha and ya may be confused.
- It contains large number of character and stroke classes.

2. SYSTEM ARCHITECTURE

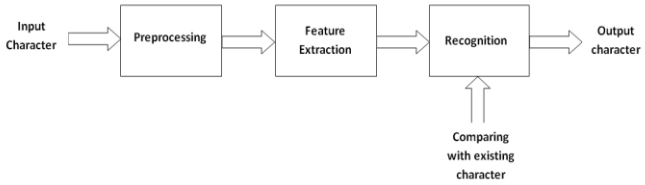

Fig. 3. Block Diagram of HDCR

2.1 Different Stages of Character Recognition

2.1.1 Pre-processing

Before the separated characters are given to the character recognition system they need to be brought in a format that is standard & acceptable to the recognition system. Pre-processing is the name given to a family of procedures as binarization, segmentation and skeletonization. Various Pre-processing Methods are explained below:

- Binirization: Character image get binarized (thresholding) by conversion of a gray-scale image into a binary image i.e. 64×64 pixel for each image.
- Segmentation: Segmentation is the process of separating individual character or pseudo character images from a word image. (Kumar s. 2009). Character segmentation is a key requirement that determines the utility of conventional character recognition systems. It includes line, word and character segmentation. After scanning the document, the document image is subjected to pre-processing for background noise elimination and skew correction to generate the bit map image of the text. Binarized image is then segmented into lines, words and characters.
- Skeletonization: Skeletonization is also called thinning. Skeletonization is nothing but the process of reducing the width of a line, word or character from many pixels wide to just single pixel width. This process can remove irregularities in letters and in turn, makes the recognition algorithm simpler because they only have to operate on a character stroke, which is only one pixel wide. It also reduces the memory space required for storing the information about the input characters and no doubt, this process reduces the processing time too.

2.2 Feature Extractions and Classification

Prior to the task of classification, an important aspect of image analysis is feature extraction. Certain characteristics are used to describe a character; these characters contain a large volume of data in the form of pixel intensity values. The character image can be expressed in the form of a vector of smallest length which is called a feature vector. To recognize many variations of the same character, features that are invariant to certain transformation on the character need to be used. The segmentation of individual characters can itself provides estimates of size and location, but the feature extraction method may often provide more accurate estimates. According to features of each character it gets divided into different classes.

2.3 Recognition

For recognition of character it is compare with stored database if features get match then that character get recognized. Devnagri character recognition involves extraction of features by some defined characteristics to classify an unknown character into one of the known classes. [1]

3. LITERATURE REVIEW

This is a scheme developed for unconstrained off-line handwritten Devnagari character recognition using regular expressions, minimum edit distance method, based on features obtained from chain code. Here the character is converted into chained code encoded string and regular expressions are matched with it. Rejected samples are sent to minimum edit distance classifier. Regular expression matching results in accuracy of 70-75%.To improve the accuracy of the system remaining 25-30% of samples which cannot be identified with regular expressions are tested with minimum edit distance method. The overall recognition accuracy of the proposed scheme for characters is 82% for 50 writers. [4]

Sandhya Arora develop a method for off line handwritten Devnagri recognition system, she used four feature extraction techniques namely, intersection, shadow feature, chain code histogram and straight line fitting features. Shadow features are computed globally for character image while intersection features, chain code

histogram features and line fitting features are computed by dividing the character image into different segments. Weighted majority voting technique is used for combining the classification decision obtained from four Multi Layer Perceptron (MLP) based classifier. Total 4900 samples are given for testing then they got result of 92.80%. [5]

Off-line Devnagri handwritten character recognition is selected for the survey by Dongre and Mankar. They capture the image then preprocess the image by assigning pixels values: 0 or 1 for binary images, 0– 255 for gray-scale images, and 3 channels of 0–255 color values for color images. Then reduce noise, Skew Detection done by making skewed lines horizontal by calculating skew angle and making proper correction in the raw Image. Size normalization by making matrix 32x32 or 64x64 so that all characters have same data size. Thinning done by making boundary detection. Devnagari words can further be splitted to individual character for classification and recognition by removing Shirorekhain. Different feature extraction and classification methods are discussed in detail. [6]

This paper shows a bounded box method proposed for segmentation of documents lines and words and characters. The method is based on the pixel histogram obtained. In line segmentation the global horizontal projection method is used to compute sum of all white pixels on every row and construct corresponding histogram. In word Segmentation the global horizontal projection method is used to compute sum of all white pixels on every column and construct corresponding histogram. In Character Segmentation a slight modification in algorithm of word segmentation is used. It is observed that line segmentation is done with nearly 100% accuracy. Word segmentation is accurate as long as the document contains characters only but character level segmentation needs more effort as it is complicated for Devnagri script. [7]

Quadratic classifier based scheme is proposed for off-line Devnagri handwritten character recognition. The features used for recognition purpose are mainly based on directional information obtained from the gradient. In this paper they use 392 dimensional feature vectors and recognition of characters in quadratic classifier is performed. The overall recognition accuracy of the proposed scheme using 392 dimensional features was 94.24% when zero percent rejection was considered. [8] Structural properties like shirorekha, spine in character is find by first stage approach using differential distance based technique as histogram based method does not work for finding shirorekha, vertical bar (Spine) in handwritten Devnagri characters. Some intersection features of characters which are fed to a feed forward neural network in second stage. 50000 samples are tested they we got 89.12% results. [9] Hidden Markov Model is a natural choice for a handwriting recognition tool. It avoids explicit word segmentation and exploits the holistic approach by combining local features with contextual knowledge. The feature vectors used for recognition are histogram of chain-code directions in the image-strips, scanned from left to right by a sliding window. One HMM is constructed for each word. To classify an unknown word image, its class conditional probability for each HMM is computed. The class that gives highest such probability is finally selected. [10]

On the basis of the head line, a word image is segmented into pseudo characters. These individual sub images are then recognized by HMM classifier generating a pseudo character string. The training and test databases here consist respectively of 22500 and 17200 images of handwritten words of 100 words classes collected from 436 different writers. [10] A novel segmentation based approach is proposed for recognition of offline handwritten Devnagri words. Stroke based features are used as feature vectors. 8 scalar features are extracted from each vertical and horizontal stroke. These features represent the shape, size and position of a stroke with respect to the pseudo character image. A hidden Markov model classifier is used for recognition. A string edit distance algorithm is used for recognizing the word.[11]A comparative study of Devnagri handwritten character recognition using Projection distance, subspace method, linear discriminant function, support vector machines, modified quadratic discriminant function, mirror image learning, Euclidean distance, nearest neighbour, k-Nearest neighbor, modified projection distance, compound projection distance, and compound modified quadratic discriminant function are used as different classifiers and four sets of feature is presented. MIL classifier provided best results among all the 12 classifiers. [12]

Offline Handwritten Devnagri Character Recognition is done by using different feature as Chain code histogram, four side views, shadow are extracted and fed to Multilayer Perceptrons as a preliminary recognition step. Finally the results of all MLP's are combined using weighted majority scheme. [13] Gradient and curvature based feature extraction method is used and comparison of Nearest Neighbor, K-Nearest Neighbor, Euclidian Distance-based K-NN, Cosine Similarity -based KNN, Condensed Nearest Neighbor, Reduced Nearest neighbor, Farthest like neighbor and Nearest unlike Neighbor is done on gray level images which is handwritten images. [14] Kailash S. Sharma, A. R. Karwankar and Dr. A.S. Bhalchandra develop a system which can recognize an online handwritten Devnagri character. Two layer self organizing map is used. Network is trained by unsupervised learning. [15] The recognition is carried out using multistage feature extraction and

classification scheme. The initial stages of feature extraction are depends on the structural features and then the classification of the characters is done as per their parameters. The final stage of feature extraction employs Radon transform and Euclidean distance transform and applied to two separate feed forward back propagation neural networks. [16] All features of each input character is stored in a feature vector. Thus feature vectors of all characters in the database are constructed. A feature vector for the test sample is also constructed. Minimum edit distance algorithm is used for final recognition. [17] Satish Kumar suggest three tier strategies to recognize the hand-printed characters of Devangri script. In primary and secondary stage classification, the structural properties of the script are exploited to avoid classification error. The results of all the three stages are reported on two classifiers i.e. MLP and SVM [18] Invariant moments is used for the feature extraction .Recognition rate increases if a character is divided in a systematic manner and features of each divided part are used in recognition system. The three methods of division are suggested in the paper for Recognition of Handwritten Devanagri Numerals. The Gaussian Distribution Function has been adopted for classification. [19], [20] Two approaches are mainly used in handwritten character recognition. First is segmentation-based approach and the other is segmentation-free approach (holistic approach). In the first approach, the words are initially segmented into characters or pseudo characters, and then, recognized. As a result, the success of the recognition module depends on the performance of the segmentation technique. The second approach treats the whole word as a single entity and it recognizes without doing explicit segmentation. [2]

4. DISCUSSION AND CONCLUSION

Only during recent years, research toward Indian handwritten character recognition is getting increased attention although the first research report on offline handwritten Devnagri characters was published in 1977(Sethi and Chatterjee 1977) In India large amount of historical documents and books either handwritten or printed in Devnagri script .They should be digitized for better access, sharing, indexing, etc. This will definitely be help for other research communities in India in the areas of social sciences, economics, and linguistics. From the survey, it is observed that the errors in recognizing printed Devnagri characters are mainly due to improper segmentation of touching or broken characters. Because of upper and lower modifiers of Devnagri text, maximum areas of two consecutive lines may also overlap. If proper segmentation of such overlapped portions are done then accuracy will be improve. If standard database is not available then according to writing style of different person accuracy may affect. The recently, some efforts have been reported toward building benchmark databases to enhance the quality of OCR-related research in India.

REFERENCES

[1] S. B. Patil, G. R. Sinha and K.Thakur, Isolated Handwritten Devnagri Character Recognition using Fourier Descriptor and HMM. *International Journal of Pure and Applied Sciences and Technology, volume 8, No.1*, 2012, 69-74.

[2] R.Jayadevan, S. R.Kolhe, P. M. Patil and U. Pal, Offline Recognition of Devanagari Script: A Survey, *IEEE Transactions on Systems, Man, And Cybernetics-Part C: Applications and Reviews, volume 41, No. 6*, 2011,782-796.

[3] B. Shaw, S.K. Parui and M. Shridhar , A Segmentation Based Approach to Offline Handwritten Devnagari Word Recognition *International Conference on Information Technology*, 2008, 256-257.

[4] Dr. P. S. Deshpande, L. Malik and S. Arora, Fine Classification & Recognition of Hand Written Devnagari Characters with Regular Expressions & Minimum Edit Distance Method, *Journal of Computers, volume 3, No. 5*,2008, 11-17.

[5] S. Arora, D. Bhattacharjee, M. Nasipuri, D. Basu and M. Kundu , Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition, *IEEE Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA*, 2008, 1-6.

[6] V.J. Dongre and Mankar V. H. , A Review of Research on Devnagari Character Recognition, *International Journal of Computer Applications ,volume 12, No.2* ,2010., 8-15.

[7] V.J. Dongre and Mankar V. H. , Devnagari Document Segmentation Using Histogram Approach, *International Journal of Computer Science, Engineering and Information Technology , volume 1, No.3*, 2011,46-53.

[8] U. Pal, N. Sharma, T. Wakabayashi and. F. Kimura , Off-Line Handwritten Character Recognition of Devnagari Script, *Ninth International Conference on Document Analysis and Recognition, volume 2*, 2007, 496-500.

[9] S. Arora, D. Bhattacharjee, M. Nasipuri and L. Malik, A Two Stage Classification Approach for Handwritten Devanagari Characters, *International Conference on Computational Intelligence and Multimedia Applications, volume 2*, 2007, 399-403.

[10] B.Shaw, S.K.Parui and M. Shridhar, Offline Handwritten Devanagari Word Recognition: A Segmentation Based Approach. *19th International conference on Signal Processing & Analysis* , 2008, 1-4.

[11] B. Shaw, S.K. Parui and M. Shridhar, Offline Handwritten Devanagari Word Recognition: A holistic approach based on directional chain code feature and HMM, International Conference on Information Technology, 2008, 203-208.

[12] U. Pal, T.Wakabayashi and F. Kimura, Comparative Study of Devnagari Handwritten Character Recognition using Different Feature and Classifiers. *10th International Conference on Document Analysis and Recognition*, 2009, 1111-1115.

[13] S. Arora, D. Bhattacharjee, M. Nasipuri, D. Basu and M. Kundu and L.Malik , Study of Different Features on Handwritten Devnagari Character, *Second International Conference on Emerging Trends in Engineering and Technology, ICETET-09*, 2009,929-933.

[14] A.N. Holombe, S. N.Holombe and R. C. Thool, Comparative Study of Devanagari Handwritten and printed Character & Numerals Recognition using Nearest-Neighbor Classifiers , *3rd IEEE International Conference on Computer Science and Information Technology ,volume 1*, 2010, 426 – 430.

[15] K.S. Sharma, A.R. Karwankar and A.S. Bhalchandra , Devnagari Character Recognition Using Self Organizing Maps, *IEEE International Conference on Communication Control and Computing Technologies*, 2010, 687 – 691.
[16] S. Shelke and S. Apte, A Novel Multi-feature Multi-Classifier Scheme for Unconstrained Handwritten Devanagari Character Recognition, *12th International Conference on Frontiers in Handwriting Recognition*, 2010, 215-219.
[17] A. Desai, L.Malik and R. Welekar, A New Methodology for Devnagari Character Recognition, volume 1, No.1, 2011,56-60.
[18] S. Kumar, A Three Tier Scheme for Devanagari Hand-printed Character Recognition, *World Congress on Nature & Biologically Inspired Computing, Coimbatore, India*, 2009, 1016-1021.
[19] R. J. Ramteke and S. C. Mehrotra, Recognition of Handwritten Devanagari Numerals, *International Journal of Computer Processing of Oriental Languages, volume 48, No. 8*, 2008, 1-9.
[20] I. K. Sethi and B. Chatterjee, Machine recognition of hand printed Devanagari numerals, *J. Inst. Electron. Telecommunication. Eng., volume 22*, 1977, 532–535.
[21] Ramteke R. J., 2010. Invariant Moments Based Feature Extraction for Handwritten Devanagari Vowels Recognition. International Journal of Computer Applications, volume 1, No. 18, 1-5.