



Devanagari Character Recognition: A Comprehensive Survey

Teja C. Kundaikar¹, J.A Laxminarayana
Computer Engineering¹, Head of Computer Engineering²
Goa College of Engineering, Farmagudi, Goa
tkundaikar10@gmail.com¹, jal@gec.ac.in

Abstract—Devnagari Script is used by over 500 million people around the world. The recognition of Devanagari script is complicated when compared to the recognition of other language scripts. Though different strategies have been proposed by the researchers in the past to identify handwritten characters, the recognition of Devanagari script still offers motivating challenges to research community. This paper analyzes many approaches and challenges concerning Devanagari character recognition.

Keywords—neural network; Devanagari recognition.

I. INTRODUCTION

Devanagari should be given more special consideration for analysis and document retrieval due to its popularity. Although there exists a collection of research works to recognize Devanagari Characters, it is observed that recognition accuracy is not at an expected level. This work mainly focus on challenges occurs in recognition of devanagri characters. The number of strokes and holes, sliding characters, differing writing styles and curves in the devanagri characters are some of the challenges that researches face during recognition.

II. RELATED WORK

In the last half century, the English character recognition was studied and the results were of such type that it can produce technology driven applications. But the same approach cannot be used in case of Indian languages due to the nature of complication in terms of structure and computation. There are many research work done towards the off-line handwritten Devanagari character recognition. The first research work report on handwritten Devanagari characters was published in 1977. Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, D. K. Basu [6] presented Recognition of non-compound handwritten Devnagari characters using a combination of MLP and minimum edit distance. In this work, the overall global recognition accuracy of system using combined MLP is 76.67% when zero percent confusion was considered. 93.27% accuracy was obtained when we considered top 5 choices of the recognition result and with zero percent confusion.

Vedgupt Saraf, D.S. Rao proposed [8] 98.78% recognition accuracy on individual character. Gazal Munjal and Ms. Neha Sahu did study of various techniques to recognize Devanagari characters [6] and concluded that the errors in recognizing handwritten Devanagari characters are mainly due to the incorrect character segmentation of touching or broken characters.

Anilkumar N. Holambe and Dr.Ravinder C.Thool presented [1] comparative study of different classifiers for Devanagari handwritten characters, which consists of experimental assessment of the efficiency of various

classifiers in terms of accuracy in recognition. They have used one feature set and 21 different classifiers for their experiment. It gives an idea of the recognition results of different classifiers and provides new benchmark for future research. Further, they have reported the comparative study of Devanagari handwritten character recognition by using classifiers. They reported highest accuracy of 96% using k-nearest neighbor classification.

Ashwin S Ramteke and Milind E Rane [2] proposed a process of segmentation that includes separating a word, a line, individual character or pseudo character images from a given script image. The large variation in handwriting style and of the script makes the task of segmentation quite difficult. Before proceeding to the process of segmentation preprocessing is need to be done. In the preprocessing step, smoothing of image using median filter, the binarization of image and the scaling are included. The process of segmentation consists of analyzing the digitalized image provided by a scanning device, so as to localize the limits of each character and to isolate them from each other. In the handwritten Devanagari script the space between the words and the characters usually varies. This produces some difficulties in the segmentation process of handwritten Devanagari script. For the segmentation of the handwritten Devanagari script into words, vertical projection profile i.e. the histogram of input image is used in which the zero valley peaks show the space between the words and characters. For the character segmentation of Devanagari, proposed method uses the vertical profiles which separate the base character using clear paths between them. The segmentation accuracy for this method depends upon the proper writing i.e. non-overlapping or characters, proper space between words and characters, proper connection of characters through shirorekha. The segmentation for word gives 98% of accuracy and for characters 97% of accuracy.

Mayuri Tathod Ronak Shah S.R Nalamwar, Rajnish Yadav and Minal Shah [4] presented a system to recognize Hand written Devnagari scripts. The system considers handwritten text as an input and recognizes characters using neural network approach in which a character matrix is created of every character in the training phase and the corresponding character is matched during

the recognition phase. The system uses Kohonen neural network algorithm in the recognition phase to recognize the character. The system gives a step by step processing and working of neural networks in different phases. After the recognition of character is done it is replaced by the standard font to integrate information from different handwritings.

III. EXISTING APPROACHES

A. Devanagari Vowels:

- a. The printed form of Devnagari numerals, vowels and consonants are shown in figure 1.

क	ख	ग	घ	ङ	
च	छ	ज	झ	ञ	ऊ
ट	ठ	ड	ढ	ण	अ॥
त	थ	द	ध	न	म
य	र	ल	व	श	भ
ष	स	ह	ळ		ब
अ	आ	इ	ई	उ	फ
ए	ऐ	ओ	औ	अ॥	प
०	०	०	०	०	०
०	०	०	०	०	०
०	०	०	०	०	०

Figure 1: Unicode of consonants, vowel, upper and lower modifier.

B. Preprocessing:

- a. **Binarization:** Binarization is an important preprocessing step which converts gray image into a binary image. It finds the global threshold that minimizes the interclass variance of the resulting black and white pixels.
- b. **Skew Detection:** skew the orientations of the document image to an angle determine using PCA [7].

C. Segmentation:

- a. **Structural:** the structured approach of horizontal and vertical projection of histogram [13]. There are Three steps of segmentation algorithm line, word and character segmentation.

D. Neural Network:

- a. **Kohonen neural networks:** The Kohonen neural network differs considerably from the feed forward back propagation neural network. The Kohonen neural network differs both in how it is trained and how it recalls a pattern. The Kohonen neural network does not use any sort of activation function. Further, the Kohonen neural network does not use any sort of a bias weight. Output from the Kohonen neural network does not consist of the output of several neurons. When a pattern is presented to a Kohonen network one of the output neurons is selected as a "winner". This "winning" neuron is the output from the Kohonen network. Often these "winning" neurons represent groups in the data that is presented to the Kohonen network.

E. Classification:

- a. The approaches such as K- Nearest Neighbour approach, Neural network, SVM classifier are used to classify the Devanagari character features. SVM used for character recognition provide

excellent precision and low recall [9].

IV. CHALLENGES IN DEVANAGRI CHARACTER RECOGNITION

The writer performs various irregularities at writing time that increases the difficulties of recognizing devnagari characters. This irregularities is due to lack of interest, improper handwriting, hasty writing and so on. Some of the irregularities discussed in this paper.

A. Incomplete representation of a vowel symbol:

Handwritten Devanagari characters there is some upper modifier not written properly which is easily interpreted by knower of script but it is difficult to be interpreted by machine as shown in figure

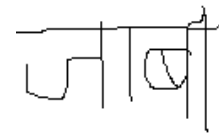


Figure 2: Incomplete upper modifier.

B. Small size of vowel symbol and touching characters:

Some individual write the vowel symbol very small in size where it is completely different angle from the original character so it is difficult to be identified by machine. There are two separate handwritten characters touching each other which make difficult to recognize the character as shown in figure 3.

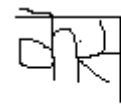


Figure 3: Touching characters and small size of upper modifier.

C. Improper attachment of lower vowel symbol:

In some cases the lower vowel is not written in proper place so it is difficult to recognize the character as shown in figure 4.

With some fine tuning to the neural network model, it is possible to recognize both regular and irregular types of Devanagari characters.

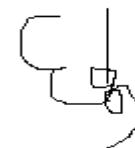


Figure 4: improper attachment of lower vowel symbol.

D. Improper HeaderLine:

Character connected to header line which makes one word. On segmenting [13] each character is identified as separate word although it is one word as shown in figure 5.



Figure 5: Improper header line

E. Overwritten character:

Some times in writing a stroke used to express a character may be faded or broken. To correct a character, a writer over-writes the given character or a part of a character as shown in figure 6, which widens the stroke width and changes the structure of the character that drops the recognition rate.



Figure 6: overwritten character

V. SOLUTION TO OVERCOME CHALLENGES

The methods discussed in section II are able to successfully recognize handwritten Devanagari characters using a neural network model. It is observed that, said neural network model cannot handle some types of handwritten irregularities. However, with the larger corpora of handwritten texts along Vedgupt saraf, D.S. Rao "Devnagari Script Character Recognition Using Genetic Algorithm for Get Better Efficiency" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2 Issue-4 April 2013.20

VI. CONCLUSION

The proper Devanagari handwritten characters are successfully recognized neural network and classifiers. However, it is observed that, it cannot handle this type of handwritten irregularities unless neural network is been trained by all the type of possible handwritten character, then there are possibility to recognize irregular character.

These irregularities can be overcome by possible solution for each of irregularities as discussed, but it cannot be overcome 100%. The additional processing required to overcome the irregularities to achieve accuracy, increases the processing time. The proposed modifications will increase the efficiency of recognizing the Devanagari characters.

VII. REFERENCES

[1]. Anilkumar N. Holambe, Dr. Ravinder. C. Thool" Comparative Study Of Different Classifiers For Devanagari Handwritten Character ", International Journal of Engineering Science and Technology Vol. 2

(7), 2010, 2681-2689.1

[2]. Ashwin S Ramteke, Milind E Rane"Offline Handwritten Devanagari Script Segmenation" IJ of Science & Technology Research Vol I Issue 4 May 2012.3

[3]. Gazal Munjal, Ms. Neha Sahu Computer Science Department ITM University, Gurgaon, Haryana, India. Astd. Professor. Computer Science Department ITM University Gurgaon Hararyana, Indai. Study of Techniques used for Devanagari Handwritten Character Recognition. International of Research in Engineering and Science (IJRES) ISSN (Online) 2320-9364. ISSN (Print) 2320-9356 June 2013.6

[4]. Mayuri Tathod Ronak Shah S.R Nalamwar, Rajnish Yadav, Minal Shah Department of Computer AISSMS COE, Pune India. "Devnagari Script Recognition Using Kohonen Neural Network" IJIDCS Vol:3 No:1. 2013.8

[5]. Pulkit Goyal IIT Allahabad Amethi Campus Teekarmaphi, Amethi Dist Sultanpur(U.P)-227413, Sapan Diwakar IIIT Allahabad, Amethi Campus Teekarmaphi, Amethi Dist. Sultanpur (U.P)-227413, Anupam Agarwal IIT Allahabad Deoghat Jhalwa Allahabad (U.P) – 211012. "Devanagari Character Recognition towards natural Human Computer Interaction".13

[6]. Sandhya Arora Assistant Professor Department of CSE & IT Meghnad Saha Institute of Technology Kolkata-107 Indai Debotosh Bhattacharjee Department of Computer Science and Engineering Jadavpur University Kolkata-107 India D.K Basu Department of Computer Science and Engineering, M. Kundu Department of Computer Science and Engineering Jadapur University Kolkata-107 India "Recognition of Non-Compound Handwritten Devnagari Characters using a Combination of MLP and Minimum Edit Distance" International Journal of Computer Science and Security (IJCSS) Volume Issue-1.16

[7]. Shivsubramani K, Loganathan R, Srinivasan C J, Ajay V and Soman KP "Enhanced skew Detection Method using Principal Component Analysis".23

[8]. Vedgupt saraf, D.S. Rao "Devnagari Script Character Recognition Using Genetic Algorithm for Get Better Efficiency" International Journal of Soft Computing and Engineering (IJSCE) ISSN : 2231-2307, Volume-2 Issue-4 April 2013.20

[9]. R. Dineshkumar, J.Suganthi "A Research Survey on Sanskrit Offline Handwritten Character Recognition" KTVR Knowledge Park for Engineering and Technology, Hindusthan College of Engineering and Technology Tamilnadu.