Offline Handwritten Character Recognition with Devnagari Script

Shruti Agarwal[1], Dr. Naveen Hemarjani[2]

[1](Computer Sciene and Engineering Department, Suresh GyanVihar University Jagatpura, Jaipur, India)
[2] (Computer Science and Engineering Department, Suresh GyanVihar University Jagatpura, Jaipur, India)

Abstract : *Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input from sources such as photographs, touch-screens, paper documents and other devices. Written text image may be sensed "off line" from a piece of paper by optical scanning (optical character recognition). Devnagari script has 14 vowels and 33 consonants. Vowels occur either in isolation or in combination with consonants. Apart from vowels and consonants characters called basic characters, compound characters are there in Devnagari script, which are formed by joining two or more basic characters. Coupled to this in Devnagari script there is a practice of having twelve forms of modifiers with each for 33 consonants, giving rise to modified shapes which, depends on whether the modifier is placed to the left, right, top or bottom of the character. The net result is that there are several thousand different shapes or patterns, which makes Devnagari OCR more difficult to develop. Here focus is on the recognition of offline handwritten Hindi characters that can be used in common applications like commercial forms , bill processing systems ,bank cheques, , government records, , Signature Verification ,Postcode Recognition, , passport readers, offline document recognition generated by the expanding technological society .In this project , by the use of templet matching algorithm devnagari script characters are OCR from document images.*

Keywords- Handwriting recognition, Devnagri script, OCR, Template Matching.

I. INTRODUCTION

Hindi handwritten character recognition is the one of the major problem in today's world. Typed Hindi characters are very difficultly recognized by computer machine therefore Hindi handwritten characters are not recognized efficiently and accurately by computer machine. Many researches have been done to recognize these characters and many algorithms have been proposed to recognize characters. Many types of software are in the market for optical Hindi character recognition. For recognizing characters, many processes have to be performed. No single process or single machine can perform that recognition. Artificial neural networks can be used for recognition of characters due to the simplicity of their design and their universality. Hindi character recognition is becoming more and more important in the modern world. It helps human ease their jobs and solve more complex problems. The problem of recognition of hand-printed characters is still an active area of research. With the increasing necessity for office automation, it is imperative to provide practical and effective solutions. All sorts of structural, topological and statistical information has been observed about the characters does not lend a helping hand in the recognition process due to different writing styles and moods of persons at the time of writing. Limited variations in shapes of character are considered.

II. Literature Survey

Although first research report on handwritten Devnagari characters was published in 1977 [1] but not much research work is done after that. At present researchers have started to work on handwritten Devnagari characters and few research reports are published recently. In this paper, implementation is done on matlab which allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran.

Hanmandlu and Murthy [2][3] proposed a Fuzzy model based recognition of handwritten Hindi numerals and characters and they obtained 92.67% accuracy for Handwritten Devnagari numerals and 90.65% accuracy for Handwritten Devnagari characters. Bajaj et al [4] employed three different kinds of features namely, density features, moment features and descriptive component features for classification of Devnagari Numerals. They proposed multi-classifier connectionist architecture for increasing the recognition reliability and they obtained 89.6% accuracy for handwritten Devnagari numerals. Kumar and Singh [5] proposed a Zernike moment feature based approach for Devnagari handwritten character recognition. They used an artificial neural network for classification.

III. DEVNAGARI SCRIPT

Hindi is world's third most commonly used language after Chinese and English, and there are approximately 500 billion people all over the world who speak and write in Hindi. It is the basic script of many languages in India, such as Hindi and Sanskrit. It is indisputable that Devnagari has the most accurate scientific basis. For a long time, it has been script of Indian Aryan languages. It is even now used by Sanskrit, Hindi, Marathi and Nepali languages. Hindi is the world's widely spoken language and since it's script is Devnagari, so it's the most popular script. As Hindi has been declared the national language by constitution of Indian, Devnagari has got the status of national dialect. In the beginning, Hindi was declared as the state language and Devnagari as the state script of other major states such as Himachal, Haryana, Rajasthan, Madhya Pradesh, Bihar, Uttaranchal, etc. Presently, it is found that Devnagari script is the most scientific script. Since every script is developed from Brahmi script so, Devnagari has connection with almost every other script. In Devnagari, all letters are equal, i.e. there is no concept of capital or small letters. Devnagari is half syllabic in nature.
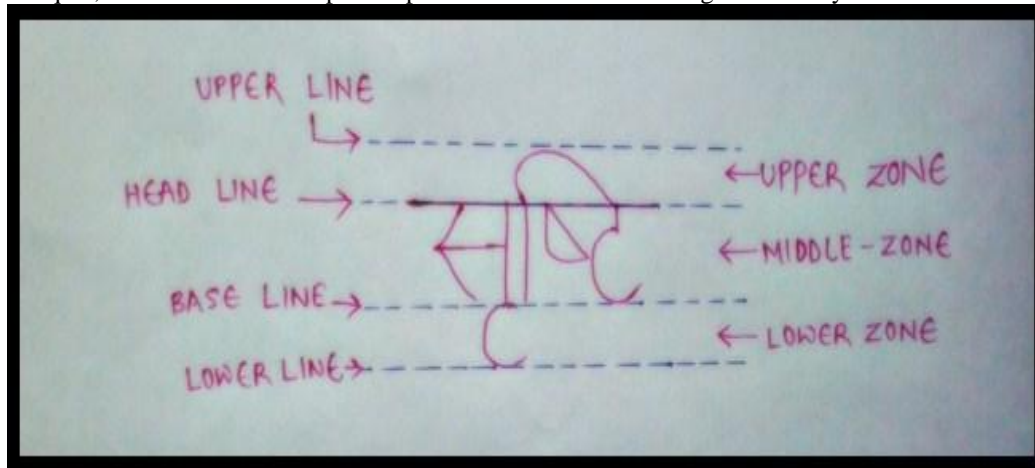
IV. Optical Character Recognition (Ocr)

OCR is the acronym for Optical Character Recognition. This technology allows a machine to automatically recognize characters through an optical mechanism. Many objects are recognized by human beings in this manner. "Optical mechanism" are the Eyes while the brain "sees" the input, according to many factors the ability to appreciate these signals varies in each person. Reviewing the variables, the challenges faced by the technologist developing an OCR system can be understood easily. Documents are in the form of papers which the human can read and understand but it is not possible for the computer to understand these documents directly. In order to convert these documents into computer process able form, OCR systems are developed. OCR is the process of converting scanned images of machine printed or handwritten text, numerals, letters and symbols into a computer process able format such as ASCII. OCR is an area of pattern recognition and processing of handwritten character is motivated largely by desire to improve man and machine communication.

V. Proposed Algorithm

The system performs character recognition by exploring the feature of templates matching for its ability to recognize handwritten Hindi characters.

The following steps are followed:

1- A database of Hindi handwritten character is created in different handwritings from different peoples.
2- Preprocessing of training image.

    a) Binarization of image using function
" bw= im2bw(Ibw,level)".
    b) Edge detection of image using function
"iedge=edge(uint8(BW2))".
    c) Dilation of image using function
" se = strel('square',2) ;
iedge2=imdilate(iedge,se) ".
    d) Region filling of image using function
"ifill = imfill(iedge2,'holes')".
    e) Character detection in image using
" [Ilabel num] = bwlabel(Ifill);

          Iprops = regionprops(Ilabel);
          Ibox = [Iprops.BoundingBox];
          Ibox = reshape(Ibox,[4 num]); "

3-Extraction and Scaling the normalized characters to 50*50 scale using boundary value analysis

"img{cnt} = imcrop(Ibw,Ibox(:,cnt));

bw2 = imgcrop(img{cnt});

charvec = imresize(bw2,[50 50]);"

4- Templates generation using image averaging and saving templates in templates.mat file which is used in matching phase.

5-Binarizing test image and matching it with templates and generating a result.txt file containing recognized characters.

The system performs character recognition by exploring the feature of templates matching for its ability to recognize handwritten Hindi characters.

The scope of the proposed system is limited to the recognition of a single character.

VI. Results

       Handwritten Devanagari Character sets are taken from 'test.bmp' image. Steps are followed to obtain best accuracy of input handwritten Hindi character image given to the system. First of all, training of system is done by using different data set or sample. And then system is tested for few of the given sample, and accuracy is measured. For each character, feature were computed and stored in templates for training the system.

The test.bmp is taken for testing
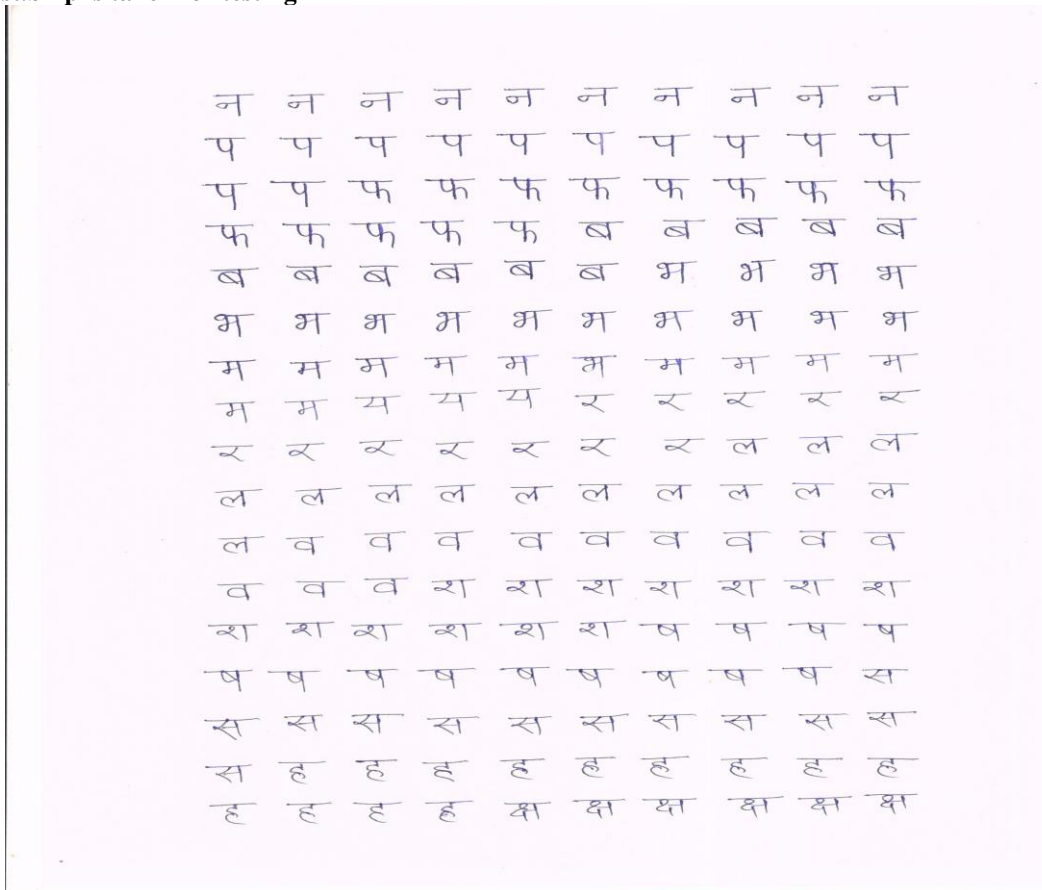
And text file generated as-
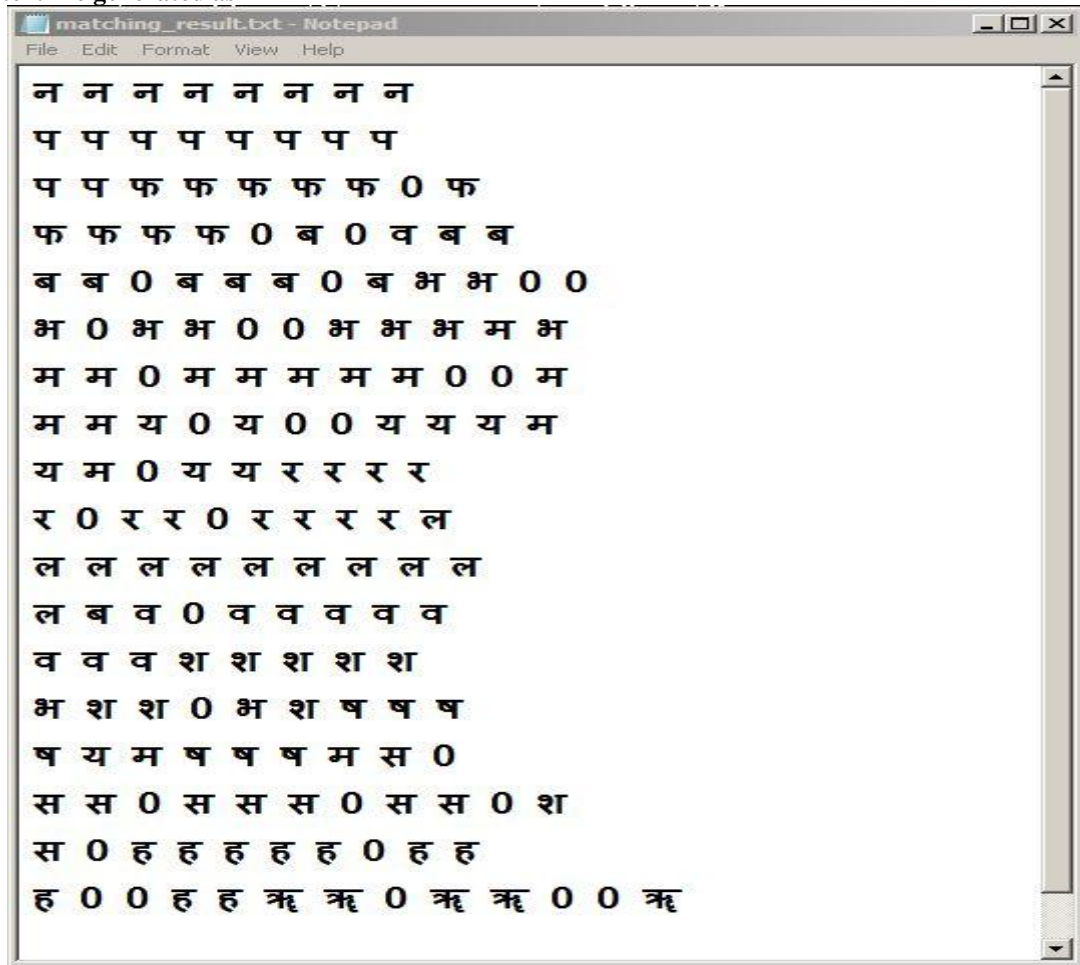
% Accuracy = No of characters found correctly *100
Total no of patterns.

% Accuracy = 139 *100
150
% Accuracy = 92.66%

The given system gives good accuracy with noise free scanned images.

The accuracy is also good as both train and test documents are written by same person so hand writing is same.

VII.   Conclusions

The scope of the proposed system is limited to the recognition of a single character.Offline handwritten Hindi character recognition is a difficult problem, not only because of the great amount of variations in human handwriting, but also, because of the overlapped and joined characters. Recognition approaches heavily depend on the nature of the data to be recognized. Since handwritten Hindi characters could be of various shapes and size, the recognition process needs to be much efficient and accurate to recognize the characters written by different users.There are few reasons that create problem in Hindi handwritten character recognition. Some

characters are similar in shape (for example भ and म). Sometimes characters are overlapped and joined. Large numbers of character and stroke classes are present there. Different, or even the same user can write differently at different times, depending on the pen or pencil, the width of the line, the slight rotation of the paper, the type of paper and the mood and stress level of the person.The character can be written at different location on paper or in window Characters can be written in different fonts.

References

[1]   I.K. Sethi and B. Chatterjee, "Machine Recognition of constrained Hand printed Devnagari", Pattern Recognition, Vol. 9, pp. 69-75, 1977.
[2]   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.
[3]   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.
[4]   Reena Bajaj, Lipika Dey, and S. Chaudhury, "Devnagari numeral recognition by combining decision of multiple connectionist classifiers", Sadhana, Vol.27, part. 1, pp.-59-72, 2002.
[5]   Satish Kumar and Chandan Singh, "A Study of Zernike Moments and its use in Devnagari Handwritten Character Recognition", Intl.Conf. on Cognition and Recognition, pp. 514- 520, 2005.