Handwritten Devanagari Character Recognition System Using Neural Network

Pahuni Saharan

Research Scholar, Dept.of Computer Science & Engineering, Sat Priya Group of Institutions, Rohtak, Haryana, India

Rupali Malhotra

HOD, Dept.of Computer Science, Sat Priya Group of Institutions, Rohtak, Haryana, India.

Abstract – Handwriting Recognition has been one of the most challenging research areas in field of image processing and pattern recognition in the recent years. Handwriting recognition principally entails optical character recognition. Recognizing Handwritten Hindi Characters is complex task as characters are written in different curves and style. This Paper describes the implementation of one such Devanagari Character Recognition System comprising stages as Preprocessing, Feature Extraction and Classification. This model achieves the accuracy rate of recognition which ranges from70 % to 80 %.

Index Terms – Devanagari Character Recognition, Off-line Handwriting Recognition, Segmentation, Feature Extraction, HDCR.

1. INTRODUCTION

Handwritten character recognition assumes considerable importance because of its applicability to pen-based interfaces and recognition of handwriting on scanned documents. There are two paradigms of handwritten character recognition, known as Offline and Online Recognition System. Off-line handwriting recognition involves the automatic conversion of text in an image into letter codes which are usable within computer and text-processing applications. Online recognition involves the use of pen based input devices to capture the sequence of co-ordinate points as the characters are written. This kind of data is known as digital ink and can be regarded as a digital representation of handwriting.

This paper presents a system created for recognizing Devanagari handwritten characters using Neural Network Toolbox of Matlab. Our main focus is on Hindi Characters.

2. RELATED WORK

Jayadevan R. et.al [1] did a survey of the comparative study of recognition of printed as well as handwritten word recognition by different classification techniques like Artificial Neural Network, Hidden Markov Model, and Support Vector Machine.

Ashutosh Aggarwal et al [2] propose a method for Isolated Handwritten Devanagari Character Recognition. Binarization, noise removal and skeletonization are used as preprocessing steps. Thresholding is used as binarization, median filter used for salt and pepper noise removal. Horizontal and vertical segmentations are used to select individual character. Gradient methods are used as Feature Extraction. The obtained feature is passed to SVM for classification. It gives the recognition efficiency of 94%

Ved Prakash Agnihotri et al [3] Presented zone based features for the recognition of Devanagari script. Enhance the input image using binarization, noise removal, edge detection and dilation. For getting the zonal feature divide the enhanced image in to 54 equal zones. So, 54 features are obtained from each character. The extracted features are classified using genetic algorithm.

3. FEATURES OF DEVANAGARI SCRIPT

Devanagari is the script for Hindi, Sanskrit, Marathi and Nepali languages. Devanagari script is a logical composition of its constituent symbols in two dimensions. "Devanagari" is a compound word with two roots: deva means "deity", and nagari means "city".

Devanagari has 11 vowels and 33 simple consonants. Vowels can be written as independent letters, or by using a variety of diacritical marks called hallant. Devanagari is set of vowel modifiers called matra placed to the left, right, above, or at the bottom of a character or conjunct. Pure consonant (also called half-letters) which when combined with other consonants yield conjuncts. A horizontal line called shirorekha (a header line) runs through the entire span of work [4].

TABLE 1. Vowels and Corresponding Modifiers.

Vowels:	अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ओ	औ
Modifiers:		ा	ि	ी	ु	ू	ृ	े	ै	ो	ौ

TABLE 2. Consonents.

क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट
ठ	ड	ढ	ण	त	थ	द	ध	न	प	फ
ब	भ	म	य	र	ल	व	श	ष	स	ह

TABLE 3. Half Form of Consonents with Vertical Bar.

क्	ख्	ग	घ्		द्		ज्	झ्	ञ्	
		ण्	त्	थ्		ध्	न्	प्	फ	फ्
ब्	भ्	म्	य्		ल्	व्	श्	ष्	स्	

TABLE 4. Some Special Combination of Half-Consonant and Consonant.

क क्क	क ल्ल	घ न्घ	ङ ङ्ख	अ त्स्र	त न्ल	प त न	प ल्ल
व व्व	भ न्भ	म ल्ल	ल ल्ल	श न्श	य द्ध	श त्श	स न्स

TABLE 5. Special Symbols.

क़	ख़	ग़	ज़	फ़	ड़	ढ़	॰	ॱ	:	।	ऽ	ॽ

4. COMPLEXITY OF DEVANAGARI SCRIPT

In Devanagari Scripts, characters are joined by a head line called as ShiroRekha which increases the segmentation difficulty.

Script consists of many characters and stroker classes. It has diverse modifiers such as Chandra bindu, adhaaksharas, virama and diacritics.

Any Word in Devanagari Script can be divided into three zones:

- Upper
- Middle
- Lower

Adhaksharas comes in Middle Zone

Upper and Lower modifiers comes in their respective zones.

Fig. 1.Various zone portioning in Devanagari Script.

5. PROPOSED MODELLING

In the Proposed HDCR, we have taken a handwritten Image from Optical Scanner and converted to Digital Images.

Production of Digitized Image is done after removing disturbances from Image using methods like RGB to Gray, Threshold, Noise Removal and Binarization. Line, Word and Character segmentation is performed to get single character samples for recognition of Devanagari characters through training and testing.

Proposed System comprises of two stages:

- Training
- Recognition.

5.1 Training

This involves training the neural network for character recognition. Back Propagation algorithm is used to calculate the Loss function gradient. It is considered to be a supervised learning method. It comprises of propagation and weight update.

Scanning Image:

The image used for training comprises of samples of handwritten Devanagari consonants which is a collection of 4 samples of each of the 33 consonants making *a total of 132.*

Preprocessing:

Pre-processing is the name given to a family of procedures as binarization, segmentation and skeletonization. Preprocessed Image is produced as an end result of Detection of edges in the binarized image using sobel technique, dilation of the image and filling the holes present in it.

Segmentation:

Matlab's function region props, bwlabel,rectangle and imcrop.are used to segment binarized image to Line,word and characters.

Feature Extraction:

It is the most important step of the entire process .It categorizes the different characters based on the common and differing key features such height, width, density, stems, loops etc. The character image can be expressed in the form of a vector of smallest length which is called a feature vector.

Network Training:

An input vector is created by placing first 98 values out of132 of the segmented characters i.e. out of 4 similar samples for the same character, 3 values are collected and an input vector is created. The last value of each character sample is saved in another file for testing. Target vector is a 33x98 identity matrix.

The following parameters are being used for creating the

Network for training:

No. of neurons in Input Layer: 30

No. of Hidden Layer: 2

No. of neurons in each Hidden Layer: 33

Network training parameter epochs: 4000

No. of epochs: 224

Network training parameter goal: 0.1

Transfer Function Used for Layer 1: "Logsig"

Transfer Function Used for Layer 2: "Logsig"

Adaption Learning Function: "Traingdx"

Performance Function: "SSE"

5.2 Recognition

After following the previous steps, we have trained the network and now we will test the network by providing different samples. This is done to analyze the system performance.

Image Acquisition

Digitized Image in any format such as JPEG, GIF, and BMT is fed into the recognition system as Input.

Preprocessing and Feature Extraction

It is done in same way as described in Training section.

Segmentation

In some cases, especially in the case of cursive writing, segmentation is imperative if character level identification is envisaged [5]. When segmentation is performed in a separate step prior to recognition, it is called explicit or external segmentation [6, 7].Segmentation is the process of converting input image in to individual characters which include line segmentation, word segmentation and character segmentation. This is done by inverting the image and replacing the rows with all white pixels into black [1].

Classification

Matlab's function sim is used in this stage. According to features of each character it gets divided into different classes.

A multilayer feed forward neural network consists of a layer of input units, one or more layers of hidden units, and one layer of output units. The output from each layer is the weighted linear summation of all input vectors along with the bias term, passed through some activation function. The network weight adjustment is done by back-propagating the error of the network.
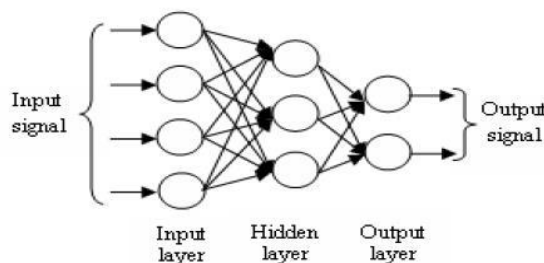

Fig 2: Multilayer Feedforward Neural Network

6. EXPERIMENTAL RESULTS

Neural Network has been trained using known dataset .After training, the recognition system was tested using several unknown dataset and results obtained are shown in section below:

TABLE 6. Accuracy Measure

Characters Included	%ACCURACY
All	69.2
Excluding श ण त	72.3

Devanagari Script's complexity and similarity of characters with each other makes it a bit difficult to analyze the performance

TABLE 7. Similar Characters

Characters Included	Similar Character
य	च थ
त	ण ल

7. CONCLUSION AND FUTURE WORK

A small set of characters are trained and then recognized. Same set of characters were analyzed again and again so as to look for any improvement in results with subsequent recognition and training. Characters for which performance was improved are kept out of testing set in subsequent analysis. It can be concluded that with accuracy can be further improved with more number of sets.

Future Scope

Characters with matras can be included and their recognition can also be implemented. Another research area can be for noisy or hazy characters and tidy written characters.

8. REFERENCES

[1] Bikash Shaw, Swapan Kr. Parui and Malayappan Shridhar, ―Offline Handwritten Devanagari Word Recognition: A Segmentation Based Approach‖, 19th International Conference on Pattern Recognition (ICPR'08), December, 8-11, 2008, Tampa, Florida, USA.

[2] Ashutosh Aggarwal, Rajneesh Rani and Renu Dhir,"Handwritten Devanagari Character Recognition Using Gradient Features", Pattern Recognition (ICPR), pp: 621-624, 2012

[3] Ved Prakash Agnihotri"Offline Handwritten Devanagari Script Recognition", I. J Information Technology and Computer Science, pp: 37-42, 2012

[4] Dongre Vikas J and Vijay H Mankar, "Devanagari Document segmentation using histogram approach", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.1, No.3, pp. 46-53, August 2011.

[5] R. Plamondon and S. N. Srihari, "Online and offline handwriting recognition - a comprehensive survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 63–84, January 2000.

[6] R. M. Bozinovic and S. N. Shrihari, "Off-line cursive script word recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1, pp. 68–83, January 1989.

[7] M. Maier, "Separating characters in scripted documents," in International Conference on Pattern Recognition, (Paris, France), pp. 1056–1058, October 1986.