International Journal of Advanced Technology and Engineering Exploration
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
Volume-2 Issue-6 May-2015

Handwritten Hindi Character Recognition System Using Edge detection & Neural Network

Tanuja K*, Usha Kumari V and Sushma T M
Acharya Institute of Technology, Bangalore, India

Abstract

Handwritten recognition has been one of the active and challenging research areas in the field of Image processing. Most of the current work in these areas is limited to English and a few oriental languages. Handwritten characters are written in various curved & cursive ways with different sizes, orientations, thickness and dimensions which is difficult task to recognize the handwritten characters by machine. In this paper a novel technique for Handwritten Hindi Character Recognition System is implemented using Canny Edge Detection technique and artificial neural network. The steps used for Handwritten Hindi character recognition system are – 1.Scanning 2.Preprocessing 3.Segmentation 4.Canny operator 5.Distance transformation 6. Feature Extraction 7. Feed Back Propagation of Artificial Neural Network for Recognition. Experimental result shows that this approach provides better results as compared to other techniques performance of the system in 95%.

Keywords

Canny edge detection, Distance Transformation, Feature Extraction, Neural Network.

1. Introduction

High accuracy Handwritten recognition system is not achieved till date. Hindi handwritten character is inspired to improve the communication with machine by using Hindi language. To provide better results preprocessing step plays an important role in handwritten character recognition system that helps for feature extraction. In this paper we have used edge detection, distance transformation as a part of pre-processing that helps in feature extraction and neural network for machine learning. Edges are important features in an image since they represent significant local intensity changes. They provide important clues to identify the handwritten character. In this paper, Canny Edge detection is used in pre-processing stage for feature extraction. Canny Edge Detector is an optimal edge detection method that detects edges with noise suppressed at same time.

Distance Transformation helps in high level feature extraction. It extracts the center of the handwritten image which helps in obtaining the feature pixels.

Machine learning is a branch of artificial intelligence inspired by psychology and biology that deals with learning from a set of data and can be applied to solve wide spectrum of problems. A supervised machine learning model is given as instances of data specific to a problem domain and an answer that solves the problem for each instance. When learning is complete, the model is able not only to provide answers to the data it has learned on, but also to yet unseen data with high precision. Neural networks are learning models used in machine learning. In this paper neural network is implemented for classification which helps to recognize handwritten character images provided as input. We have used feedback propagation algorithm which works by what is known as supervised training. It first submits an input for a forward pass through the network. The network output is compared to the desired output, which is specified by a "supervisor" and the error for all the neurons in the output layer is calculated. The fundamental idea behind back propagation is that the error is propagated backward to earlier layers so that a gradient descent algorithm can be applied. This assumes that the network can "run backwards", meaning that for any two connected neurons, the "backward weights" must be the same as the "forward weights". Neural network can be very fast at delivering results and may detect connections between seen instances of data that human cannot see.

*Author for correspondence

2. Related Work

Handwritten Hindi character recognition system present by Gunjan et.al [1] used thinning algorithm for feature extraction and back propagation neural network algorithm for recognition which yielded accuracy of 93%. Indian sign language recognition presented by Adhitya et.al [2] used Distance Transformation for feature extraction and neural network for recognition which yielded accuracy of 91%.Content Based Retrieval system using feed forward back propagation neural network by Arvind et.al [3] used edge histogram for edge feature extraction and neural network for recognition which yielded accuracy of 88%. Offline Handwritten Character Recognition with Devnagari Script by Shruti Agarwal et al [4] used template matching algorithm with accuracy of 92.66%. Fuzzy model based recognition of handwritten Hindi numerals and characters by Hanmandlu et al [5] with 92.67% for handwritten Hindi numeral & 90.65% for handwritten Hindi character. Bajaj et al [6] proposed multi-classifier connectionist architecture for increasing the recognition reliability and obtained 89.6% accuracy for handwritten Devnagari numerals.

3. Proposed Handwritten Hindi Character Recognition System

The proposed Handwritten Hindi Character Recognition System consists of 7 stages –

1. Scanning
2. Preprocessing
3. Canny Edge Detection
4. Segmentation
5. Distance transformation
6. Feature Extraction
7. Feed Back Propagation of Artificial Neural Network

1. Scanning: Samples of handwritten Hindi character of different styles are scanned using optical scanner or camera. Scanned images are converted into bitmap image.

2. **Preprocessing:** The pre-processing phase include converting RGB to Gray scale, noise removal, skew detection, slant correction, Binarization ,Morphological Operations, Normalization like processes to make character image easy to extract relevant features and efficient recognition.

3. Canny Edge Detection: The purpose of edge detection in general is to significantly reduce the amount of data in an image, while preserving the structural properties to be used for further image processing. In Handwritten recognition system, edge detection plays an important role which help in extracting the feature of each handwritten character. Since Canny edge detection is considered as optimal edge detection the proposed handwritten Hindi character recognition system uses Canny Edge Detection algorithm.

The algorithm runs in 5 separate steps:

1. Smoothing: Blurring of the image to remove noise.
2. Finding gradients: The edges should be marked where the gradients of the image has large magnitudes.
3. Non-maximum suppression: Only local maxima should be marked as edges.
4. Double thresholding: Potential edges are determined by thresholding.
5. Edge tracking by hysteresis: Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge.

4. Segmentation phase: The preprocessed image is the input to this phase which involves character segmentation classifiable module object generally isolated characters or modifiers. Generally practiced segmentations are line segmentation, word segmentation, character segmentation and horizontal segmentation to separate upper and lower modifiers particularly in context to most Indian scripts. In this paper we have used character segmentation after segmentation English representation is given for segmented Hindi handwritten character image.

5. Distance Transformation: Distance transforms plays a central role in the comparison of binary images, particularly for images resulting from feature detection techniques such as edge detection. Distance transformation converts the handwritten binary image into skeletonization that retains the geometry of image.

Distance transformation follows below 3 steps:

- The original (binary) image is converted into feature and non-feature elements.
- The feature elements belong to the boundary of the object.

- The distance map is generated where each element gives the distance to the nearest feature element.
- The ridges (local extremes) are detected as skeletal points.

6. Feature Extraction: After skeletonization of an image using Distance Transformation feature of skeletonized image is extracted to form feature vector that is used for training the neural network.

7. Feed Back Propagation [8] of Artificial Neural Network: Here we are using 3-layer Neural Network [7], feature vectors are used to train the neural network. Supervised training is used both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. Errors are then propagated back through the system which is nothing but feedback propagation, causing the system to adjust the weights which control the network. This process occurs over and over as the weights are continually changed. The set of data which enables the training is called the "training set." During the training of a network the same set of data is processed many times as the connection weights are ever refined.

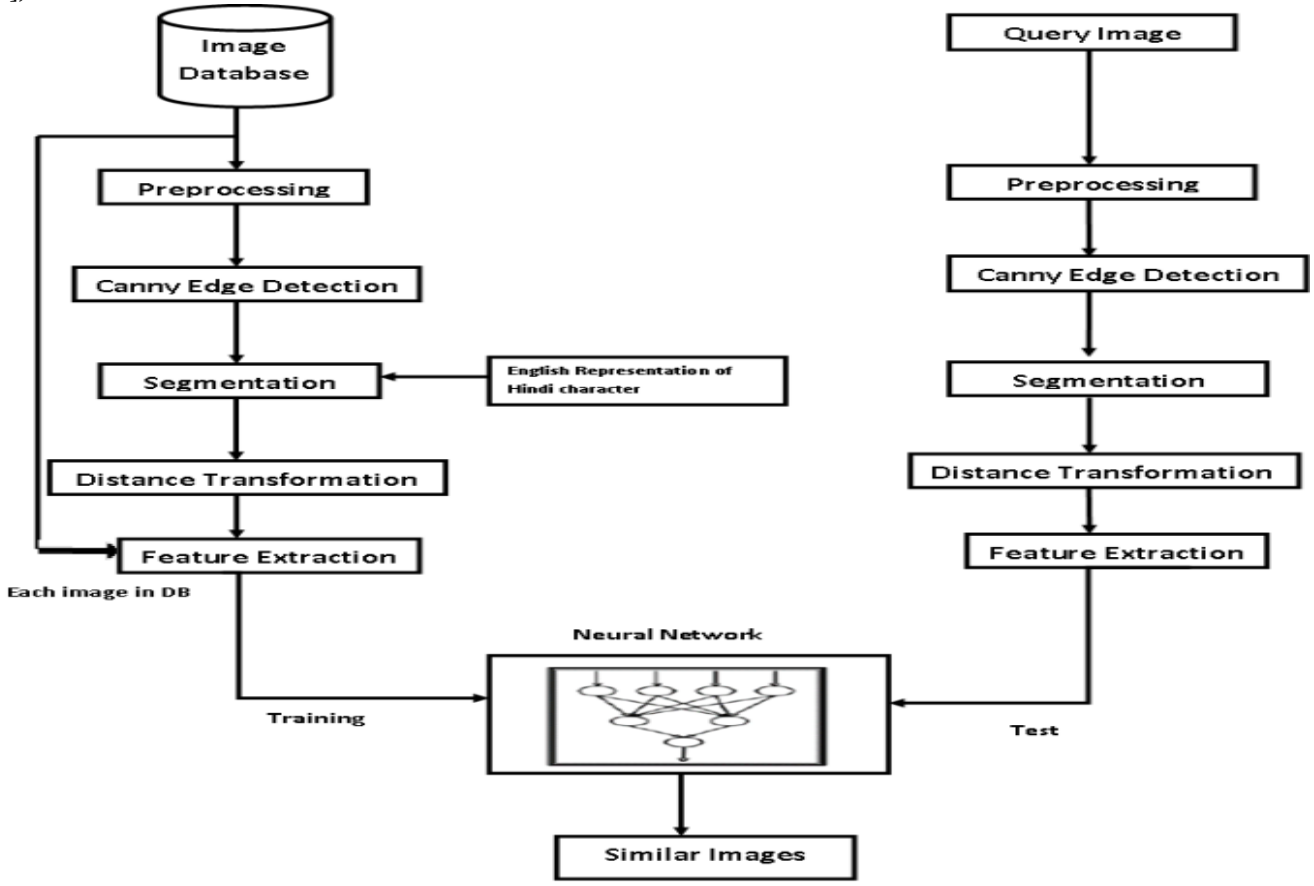

Figure 1: Flow chart of proposed Handwritten Hindi Recognition System

After training the neural network act as classifier here the given input image from the end user after the preprocessing phase, canny edge detection, segmentation, feature vector given to trained neural network matches the image in the database and matched images are displayed on the screen. Figure.1 shows the flow chart of proposed Handwritten Hindi Recognition System.

4. Results

Figure 2 shows the Hindi Handwritten original image given input to the proposed Hindi Handwritten

Recognition system.

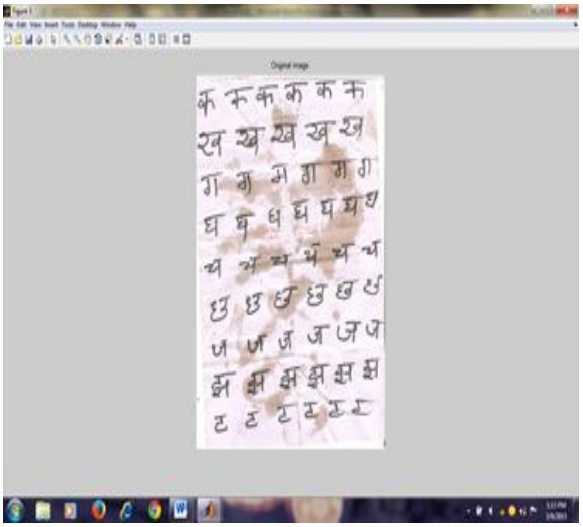

Figure 2: Input image

Figure 3 Shows output after preprocessing and segmentation of each Hindi character with corresponding English representation, canny edge detection of each segmented Hindi character, distance transformation that helps to extract high level features of each character.

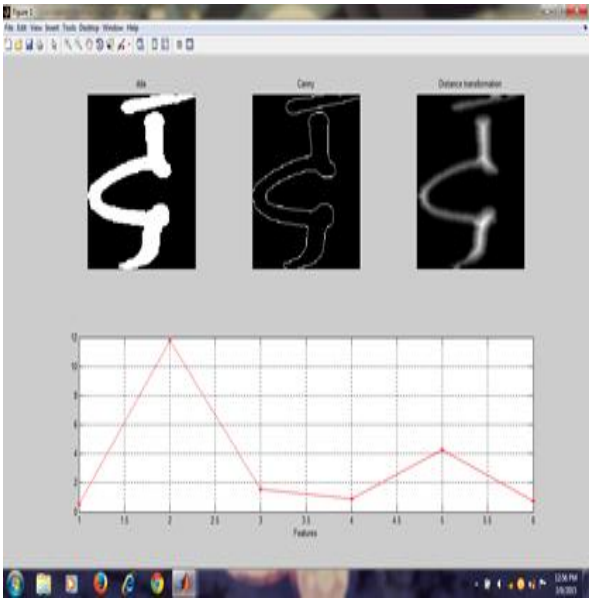

Figure 3: Canny Edge Detection & Distance Transformation with Feature Extraction

Figure.4 shows neural network training phase with epoch of 1500

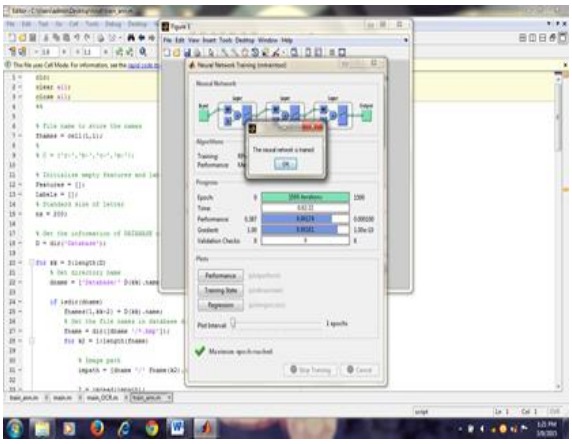

Figure 4: Neural Network getting trained

.

4.1 Performance Table

The table 1 shows the result of each handwritten Hindi character recognized by proposed handwritten Hindi recognition system.

Table 1: Result Chart

S .No	English representation of Hindi Character Recognition	Accuracy
1.	a	100%
2.	aa	95%
3.	ana	95%
4.	Aou	100%
5.	ba	95%
6.	Bha	100%
7.	Cha	100%
8.	Chha	91%
9.	Da	94%
10.	Dda	100%
11.	Ddha	91%
12.	Dha	100%
13.	E	100%
14.	Ee	91%
15.	Ga	100%
16.	Gha	95%
17.	Ha	95%
18.	Ja	100%
19.	jha	91%
20.	Ka	100%
21.	Kha	91%
22.	La	90%
23.	Ma	90%

24.	Na	94%
25.	O	94%
26.	Pa	94%
27.	Pha	95%
28.	Ra	90%
29.	Raa	95%
30.	Ro	100%
31.	Ru	89%
32.	Sa	89%
33.	Sha	100%
34.	Ta	94%
35.	Tha	94%
36.	Tta	94%
37.	Ttha	90%
38.	U	90%
39.	Uu	100%
40.	Va	100%
41.	Ya	90%
Average Performance		**95.02%**

The caption for a figure appears below the figure; for a table, above. Do not be afraid to use lengthy figure and table captions better that than confusing or incomplete ones.

5. Conclusion & Future work

In this paper the scope of the proposed system is limited to single character recognition. When compared to the performance of various techniques used for Handwritten Hindi character recognition the proposed technique of using the canny edge detection, distance transformation for feature extraction and neural networks with back propagation algorithm for Handwritten Hindi character recognition system has resulted in 95% of accuracy. In future the work can be extended to Handwritten Hindi word recognition.

References

[1] Gunjan Singh and Sushma Lehr , "Recognition of Handwritten Hindi Characters using Back propagation Neural Network " International Journal of Computer Science and Information Technologies, Vol. 3 (4) , pp. 4892-4895, 2012.

[2] Adithya V.Vinod P. R. ,& Usha Gopalakrishnan,"Artificial Neural Network Based Method for Indian Sign Language Recognition" IEEE Conference on Information and Communication Technologies ,ICT 2013, 978-1-4673-5758-6, 2013.

[3] Arvind Nagathan, Manimozhi & Jitendranath Mungara." Content Based Retrieval system using feed forward back propagation neural network" IJCSNS International Journal of Computer Science and Network Security, vol.14, No.6, June 2014.

[4] Shruti Agarwal, Naveen Hemarjan" Offline Handwritten Character Recognition with Devnagari Script ", IOSR-JCE, Volume 12, June 2013.

[5] 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.

[6] 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.

[7] M. Lamari, M. Bhuiyan, and A. Iwata. "Hand alphabet recognition using morphological pca and neural networks". In Proc. IJCNN, pp. 2839-2844, 1999.

[8] F. Li, ands. Gao ," Character Recognition System Based on Back-Propagation Neural Network," International Conference on Machine Vision and Human-Machine Interface (MVHI), pp.393-396, 2010.

Tanuja K is an Asst. Prof., ISE Dept., Acharya Institute of Technology, Bangalore, Karnataka, India. She has done her M.Tech in Computer Science and Engineering from VTU, Karnataka. Her area of interest is Image processing on Character Recognition.
Email:tanujak@acharya.ac.in

Usha Kumari V is an Asst. Prof., ISE Dept., Acharya Institute of Technology, Bangalore, Karnataka, India. She has done her M.Tech in Computer Science and Engineering from VTU, Karnataka. Her area of interest is Image processing on Character Recognition.

Sushma T M is an Asst. Prof., ISE Dept., Acharya Institute of Technology, Bangalore, Karnataka, India. She has done her M.Tech in Computer Science and Engineering from VTU, Karnataka. Her area of interest is Image processing on Character Recognition.