Sanskrit Character Recognition with Improved feed forward BPNN and LSB Matching Patterns

R. Dineshkumar[1], Dr. J. Suganthi[2]

[1]Research Scholar, CSE, KTVR
Knowledge Park for Engineering and Technology, Coimbatore, Tamilnadu, India
[2]Vice Principal and Head, CSE,
Hindusthan College of Engineering and Technology, Coimbatore, Tamilnadu, India

Abstract: Sanskrit documents recognition provides many innovative ideas and challenges to researchers. Handwritten documents of every person varies in styles and consonants, therefore it is very difficult to recognize the handwritten characters. This paper deals with offline Sanskrit handwritten character recognition with feed forward back propagation neural network. This paper provides easy and fast use of feature extraction to obtain good performance and high accuracy. Here we presented a new system which significantly improves current handwriting recognition systems and a method for detection of LSB matching in grayscale images. It also provides differences between neighboring pixels with data embedding for high performance.

I. Introduction

India is a democratic country offering many languages, in that eighteen are official languages. Brahmin scripts are used to develop many languages in India. Sanskrit is an ancient Indo Aryan language with a rich literary tradition. The traditional Sanskrit script is the well-known script in India. It is an ancient language with written materials and no longer spoken. Most of the poetry, scientific, and technical texts are made with a rich tradition of Sanskrit literature. In Hindu religious, Sanskrit is used as a traditional language in the forms of hymns and mantras.

In olden days people recorded their knowledge in various fields in paper documents. The handwritten text written in paper documents ruined over a period of time. It is very difficult to maintain them in the same form. Indian writing system is complex because of combination of characters and consonants. This paper proposes a new approach for converting handwritten Sanskrit script using Unicode.

A method for detection of LSB matching in grayscale images is introduced. It also provides differences between neighboring pixels with data embedding for high performance.

II. Related Work

[1]This paper provides an approach by using, multilayer perceptron (MLP) network with one hidden layer to recognize handwritten English characters. Fourier Descriptor (FD) along with Boundary tracing is used to fragment the features from the character sets. Character identification is done by examining its shape and features with other characters. To obtain higher efficiency in back propagation network in the recognition of handwritten English characters, a study is made to find the number of hidden nodes. Among 1000 samples of handwritten English characters, 500 are used to train the system and the remaining 500 are to test the efficiency of the FD integrated with back propagation network. However, as a result the character recognition accuracy of FD is 94% with minimum training time.

[2] This paper describes a continuous density HMM to recognize a word image. The feature vector is created by scanning the word or image from left to right using a sliding window. Several image frames constitutes to form a single word image, and are called a string. One HMM is raised for each and every

word image. The class conditional probability is calculated for each image or HMM and is used to separate the unknown word image. However, the class that shows highest probability is selected.

[3] In this paper, the handwritten Chinese character recognition system based on the off-line neural network is presented. Recognition rates of 52.44% are obtained from seventeen character sets. From this system, we can achieve liberty to shift, delicate rotations and marginal scaling of the input characters. The Chinese characters are recognized by fusing neural computation and structural representation. Fault tolerance and noise reduction are achieved by neural networks, and structural representation is acquired for relevant construction of Chinese characters. To favor newly founded handwriting styles. However it is not required to remove organized knowledge.

[4] In this paper, work has been performed to recognize Handwritten Tamil Character using a multilayer perceptron with one hidden layer. Fourier Descriptor (FD) technique is used to fragment the feature from the handwritten characters. To obtain higher efficiency in back propagation network for the recognition of handwritten Tamil characters, a study is made to find the number of hidden nodes. Several different handwritten character sets of different age groups are fed as input to educate the system. As a result, recognition accuracy of 97% is achieved by integrating FD with back propagation network, for handwritten Tamil characters.

[5] In this paper, estimation between Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks are done in the quest of Hindi Character Recognition (HCR). The error back propagation algorithm used to educate MLP networks. MLP and RBF networks were included to form an automated HCR system. A trail was undergone with two hundred and forty five samples of five epic writers. It became clearly evident from the output that the error back propagation algorithm showed it supremacy in recognition accuracy and memory usage but it is affected with long term training when compared with RBF networks.

[6] In this paper, a multilayer perceptron (MLP) network with one hidden layer is used for handwritten character recognition, is presented. Border Transition Technique (BTT) and Fourier descriptor (FD) are the techniques used to fragment the features from the character sets. The neural network used to educate back propagation algorithm takes the input values from result of BTT and FD. For handwritten numbers from 0 to 9, the integration of above to techniques provides better recognition accuracy which is up to 96%.

[7] In this paper the problem of detecting spatial domain least significant bit (LSB) matching steganography in grayscale images, which has proved much harder than for its counterpart, LSB replacement. We use the histogram characteristic function (HCF), introduced by Harmsen for the detection of steganography in color images but ineffective on grayscale images. Two novel ways of applying the HCF are introduced. Calibrating the output using a down sampled image and computing the adjacency histogram instead of the usual histogram.

[8] In this paper it gives a detector for color images, which has its basis in the effect that the embedding algorithm has on the occurrences of close pairs of colors. Westfeld comments that the detector can be applied to grayscale images by converting triplets of gray pixels into the red, green, and blue (RGB) components of a single color pixel, but in practice, this detector performs barely better than a random decision.

III. Sanskrit Recognition System

The Sanskrit Recognition system consists of stages as Preprocessing, Segmentation, Feature Extraction, and Classification.

3.1 Preprocessing

The processes which get involved in pre-processing are Denoising, Binarization, Skewing and Thinning.

3.1.1 Denoising

Noise reduction is the preprocessing phase in character recognition and it is used to eliminate inappropriate marks in the image without clearing its edges to make a clean and minimum noise image [10].

3.1.2 Binarization

The conversion of gray scale image to binary image with global threshold technique [9]

3.1.3 Skewing

Skew detection is the process of finding whether the angle of orientation of the character is with ±15 degree. And the process of correcting the skewed image by simple image rotation with respect to the true horizontal axis [11].

3.1.4 Thinning

A binary image is thinned to produce another binary image with the thickness of one pixel by making tidy up the result of edge detectors.

3.2 Segmentation

The input images is fragmented into individual characters and are resized into mxn pixels are called segmentation as presented in [11] The binarized image is checked for interlined space and horizontal lined space. If interlined space is detected then it is fragmented into separate paragraphs and if horizontal lined space is detected then it is fragmented into separate word followed by characters using character decomposition [11].

3.3 Classification

In recognition system, classification is the decision making part which takes input from segmentation and defines the character [9]

3.3.1 BPNN

Log sigmoid activation function is used in the hidden layer, and one of the character is to be found in the output layer, hence it called as competitive layer. The feature vector is denoted as p where p =(x1, x2, x3 ...xd) and x denotes features and d is the number of zones into which each character is divided. The length of the feature vector d is used to find the number of neutrons for input. The number of neurons in the output layer is found by the total number of characters t.

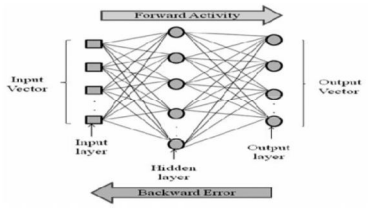

Fig 1.Artifcial Neural Network

The network training parameters are:
Input nodes: 70
Hidden nodes: 1
Output nodes: 70
Training Algorithm: Feed forward back propagation neural network
Training and Adaptive Learning Perform function: Mean squared error
Training Goal Achieved: Yes
Training Epochs: 8605
Training Momentum Constant: 0.9
Training Learning Rate: 0.01
Training for all characters may take approximately 8000 epochs.

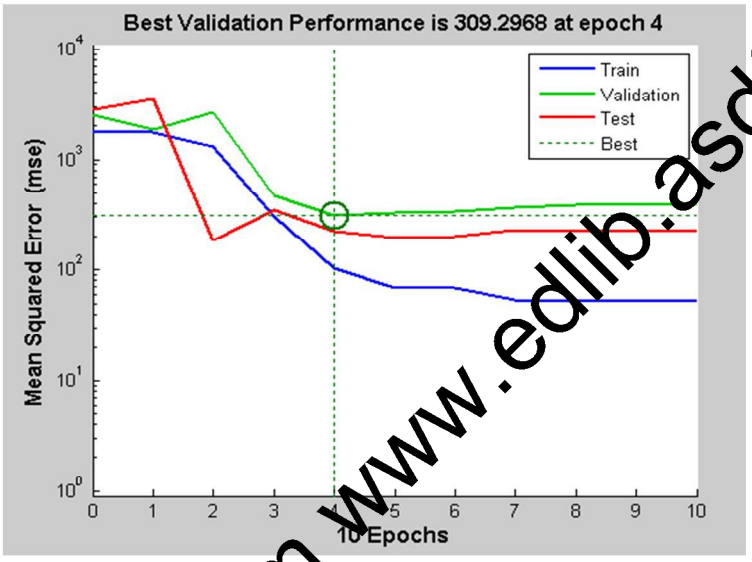

Fig 2. Training Performance

Table 1.Decrease in MSE for 48x8x4 network using learning rate 0.1

EPOCH	MSE
500	32.9
2300	0.264
5300	1.56e-05
6300	7.79e-10
8605	3.99e-10

Table 2. Summary of the results achieved by the proposed methods

Classifier	No.of Alphabets with recognition greater than 90%	Alphabets with recognition greater than 90%
ANN	43	Consonants – 28 Vowels -15 Total – 43 characters

IV. LSB MATCHING

LSB matching embedding algorithm is as follows. Convert the secret data into a sequence of bits. Consider each and every pixel of the cover image: if the LSB of the next cover pixel does not the next bit of secret data, randomly add or subtract one from the cover pixel value; otherwise, do nothing. When the number pixel in the cover image is larger than length of the secret message then changes are made throughout the image uniformly and it is confirmed by pseudo-random permutation technique. When the pixel is saturated then the maximum range of pixel value will be increment or decrement.

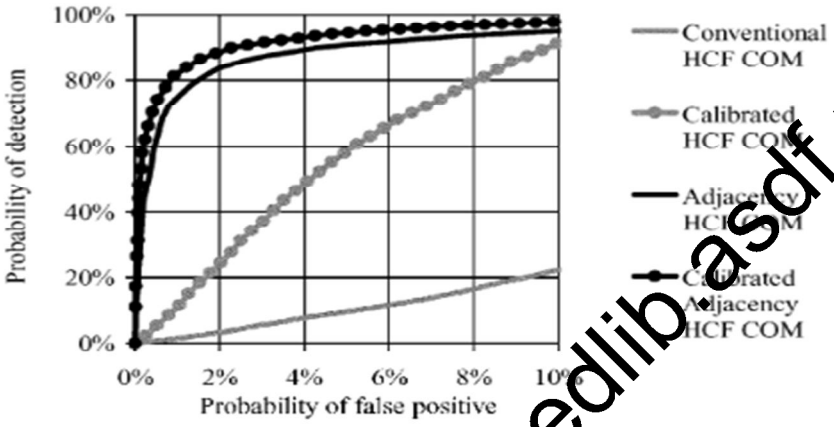

Fig 3.False Positive Rate

5. Conclusion

We have given two methods as character recognition and security which includes Sanskrit character recognition with feed forward back propagation neural network with good performance and high accuracy and also with LSB matching in grayscale images, also provides differences between neighboring pixels with data embedding for high performance with normal false positive rate. It is highly likely that calibration will improve the reliability of detection.

References

1. Anita Pal, Dayashankar Singh" Handwritten English Character Recognition Using Neural Network" International Journal of Computer Science & Communication,Vol. 1, No. 2, July-December 2010, pp. 141-14
2. Bikash Shaw, Swapan Kumar Parui, Malayappan Sridhar, "Offline handwritten Devanagari word recognition: A holistic approach based on directional chain code feature and HMM" IEEE, 2008.
3. D.S. Yeung, "A Neural Network Recognition System for Handwritten Chinese Character Using Structure Approach," Proceeding of the World Congress on Computational Intelligence, **7**, pp. 4353-4358, Orlando, USA, June 1994.
4. Sutha, J. "Neural Network Based Offline Tamil Handwritten Character Recognition System", conference on Computational Intelligence and Multimedia Applications, 2007, IEEE, page(s) 446-450.
5. Verma B.K, "Handwritten Hindi Character Recognition Using Multilayer Perceptron and Radial Basis Function Neural Network", IEEE International Conference on Neural Network, **4**, pp. 2111-2115, 1995.
6. Y.Y. Chung, and M.T. Wong, "Handwritten character recognition by fourier descriptors and neural network" proceedings of IEEE Tencon-speech and Image Technologies for C, Vol 1,Dec 1997.

7. Andrew D. Ker “Steganalysis of LSB Matching in Grayscale Images”, IEEE signal processing letters, Vol. 12, No. 6, JUNE 2005.
8. A. Westfeld,“Detecting low embedding rates,”in Proc. Inf. Hiding Workshop, Springer LNCS, vol. 2578, 2002, pp. 324–339.
9. J. Pradeep, E. Srinivasan, S. Himavathi “Neural Network Based Recognition System Integrating Feature Extraction and Classification for English Handwritten” IJE TRANSACTIONS B: Applications Vol. 25, No. 2, (May 2012) 99-106
10. Shubhra Saxena, P. C. Gupta “A Novel Approach of Handwritten Devanagari Character Recognition through Feed Forward Back Propagation Neural Network “International Journal of Computer Applications (0975 – 8887) Volume 51– No.20, August 2012.
11. Vijay Laxmi Sahu and Babita Kubde,”Offline English Hand Written Character Recognition Using Neural Network”, International Journal of Scientific Research in Computer Science (IJSRCS) Vol. 1, Issue. 2, Sep. 2013.