Indian Journal of Science and Technology, *Vol 8(1), 65–69, January 2015*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Sanskrit Character Recognition System using Neural Network

R. Dineshkumar[1*] and J. Suganthi[2]

[1]Faculty of Information Technology, Anna University, Chennai, Tamil Nadu, India; me.dineshkumar@gmail.com
[2]Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India.

Abstract

In the fast moving world with the amazingly growing technology, character recognitions play a wide role by providing more scope to perform research in OCR techniques. Sanskrit handwritten recognition has been one of the challenging research areas in the field of pattern recognition. Character recognition is the electronic translation of scanned images of handwritten or printed text into a machine encoded text. The character recognition is a standout amongst the most generally utilized biometric attributes for authentication of persons and document. In this paper proposed an off line handwritten character recognition framework utilizing feed forward neural network. A handwritten Sanskrit character is resized into 20x30 pixels and this character is used for training the neural network. After the training process, the same character is given as input to the neural network with different set of neurons in hidden layer and their recognition accuracy rate for different Sanskrit characters has been calculated and compared. The results of the proposed system yields good recognition accuracy rates comparable to that of other handwritten character recognition systems.

Keywords: Classification, Feed Forward Neural Network, Handwritten Sanskrit Character Recognition, Image Extraction, Pre-Processing

1. Introduction

With emergence of the digital content the need for the development of an OCR engine with high performance has become essential. The idea of OCR is to analyze a document image by page, words and characters. To identify the exact characters, these characters are compared with image patterns. Character recognition can either be done from printed documents or from handwritten documents. Sanskrit handwritten is more complicated than other related works in offline mode, because Sanskrit letters have more consonants and modifiers.

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. Sanskrit is a phonetic language which consists of 48 characters (15 vowels, 33 consonants). A Sanskrit document is written from left to right.

2. Related Work

[1]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.

[2]This paper have proposed a system to recognise both Devanagari and English handwritten numerals. They used a set of global and local features, which were derived

**Author for correspondence*

from the right and left projection profiles of the numeral image. They tested their system on both Devanagari and English numerals independently and found that recognition rate for Devanagari numeral was better. For Devanagari numerals they found recognition rate of 89% and confusion rate of 4.5%. For English set they found recognition rate of 78.4% and confusion rate of 18%.

[3]This paper have used three different types of features namely density, moment and descriptive component features to recognize the Devanagari numerals. They also used three different neural classifiers and finally the outputs of three classifiers were combined using a connectionist scheme. Thus they obtained classification rate of 89.68% by combining all classifier.

[4]This paper have proposed majority voting scheme for multi-resolution recognition of hand-printed numerals. They used the features based on wavelet transforms at different resolution levels and multilayer perceptron for classification purpose. They achieved 97.16% recognition rate on a test set of 5000 Bangla numerals.

[5]This paper have used ANN and HMM for recognition of handwritten Devanagari numerals. In their proposed scheme they obtained 92.83% recognition rate.

[6]This paper have proposed a modified quadratic classifier based scheme towards the recognition of off-line handwritten numerals of six popular Indian scripts namely Devanagari, Bangla, Telugu, Oriya, Kannada and Tamil scripts for their experiment. The features used in the classifier were obtained from the directional information of the numerals. They obtained good accuracy for the scripts respectively.

[7]This paper have proposed general fuzzy hyper line segment neural network to recognize handwritten Devanagari numerals. They proposed a rotation, scale and translation invariant algorithm and reported high recognition rate.

[8]This paper has used a method based on invariant moment and the divisions of numeral image for recognition of handwritten Devanagari numerals. They adopted Gaussian Distribution Function for classification and achieved 92% success rate.

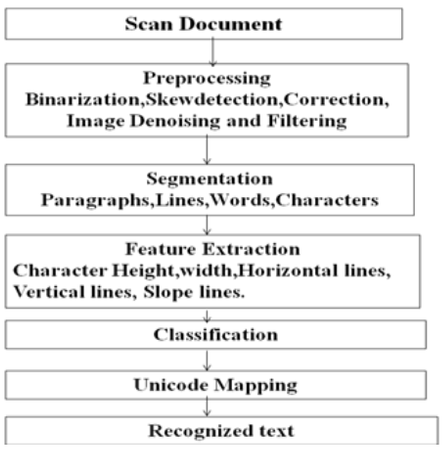

Figure 1. Block diagram.

3. Proposed Method

The steps involved in character recognition, namely pre-processing, segmentation, feature extraction and classification.

3.1 Pre-Processing

Before performing character recognition there are numerous tasks to be completed. First step in pre-processing is to scan the handwritten document and converted into a suitable format. The advantage of pre-processing a handwritten character image is to organize the information so as to make the task of recognition simple. Pre-processing has different types of sub processes to clean the document image and make it appropriate to carry the recognition process correctly. The sub processes which get involved in pre-processing are:

- Binarization.
- Noise reduction.
- Normalization.
- Skew correction, thinning and slant removal.

3.1.1 Binarization

The method of transforming a gray-scale image into a black and white image through thresholding is binarization[9]. Thresholding concepts are usually used by researchers to

Figure 2. Cropped image.

extract the foreground image from background image. Histogram based thresholding approach can also be used to convert a gray-scale image into a two tone image. In contrast to this Adaptive Binarization method can be used to identify the local gray value contrast of the Image. This helps in extracting the text information from low quality documents.

3.1.2 Noise Removal

Digital images are usually prone to so many types of noises. Noise can be termed as a document image which is due to poorly photocopied pages. Median Filtering[10], Wiener Filtering method[11] and morphological operations can be performed to remove noise[12]. To replace the intensity of the character image[13] Median filters are used. Gaussian filters can be used to smoothing the image[14].

3.1.3 Normalization

The process of converting a random sized image into a standard size is normalization. The Roi-Extraction[15], Bicubic interpolation[16], linear size normalization[17] and Java Image Class[18] normalization are used to get the single structural element from the image.

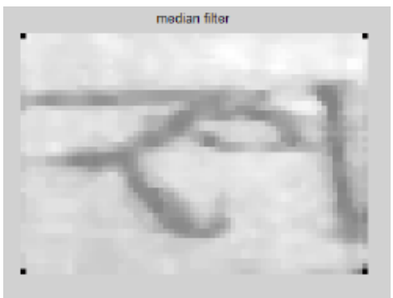

Figure 3. Median filter image.

Figure 4. Threshold image.

3.1.4 Skew Correction, Thinning and Slant Removal

Thinning is a pre-process that results in single pixel width image to recognize the handwritten character easily. Skew is inevitably introduced into the incoming document image during document scanning. For correction of the slant, angle stroke, width and vertical scaling Normalization[18], Fourier Spectrum is used.

3.2 Segmentation

After pre-processing, the image is passed to the segmentation phase. It is one of the most basic phases in character recognition which is used to segment the input image into individual glyphs. The input images is fragmented into individual characters and are resized into mxn pixels as presented in [19] 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[19].

3.3 Feature Extraction

Individual image glyph is considered and extracted for features such as character height, width, horizontal lines, vertical lines, slope lines, circles, arcs etc**.** The feature extraction method is most important in achieving high recognition performance. Its main goal is to obtain the most relevant information from the original data and quickly perform tasks such as image matching and retrieval. In [20] several methods of feature extraction for character recognition have been reported.

3.4 Classification

3.4.1 Artificial Neural Network

The network needs to be trained first with some predefined standard character patterns to perform the recognition task. BPNN algorithm is used for this, which is considered as the unsupervised form of learning method where every neuron competes with each other in the basis of their activation value. The connection weights towards the winner neuron get adjusted during training process. Some random values are assigned initially to all the connection weights, during the training process these values are converged to some fixed values. The training process is similar to an unsupervised training method.

The network training parameters are:

Input nodes: 70
Hidden nodes: 1
Training Algorithm: Feed forward NN
Training function: Mean squared error

3.4.1.1 Average Error (%)

MSE: Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

3.4.1.2 Classifier Accuracy

The graph in Figure 6 shows the accuracy of the classifiers for the Sanskrit characters which have been trained in neural networks.

Table 1. Average error %

Classifier	Error (%)
Feed Forward NN	3.99e–10

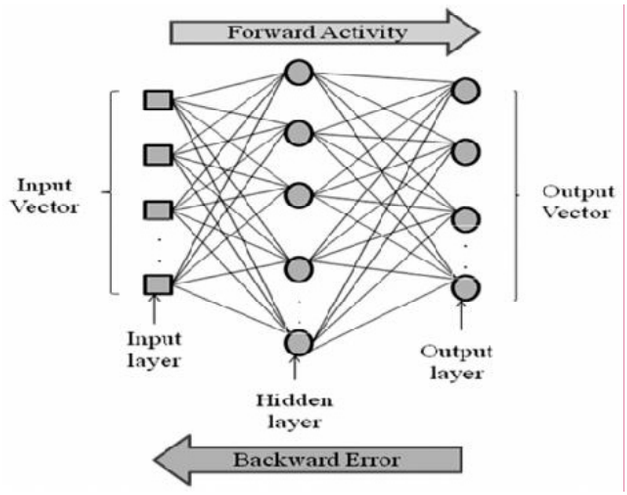

Figure 5. Artifcial Neural Network.

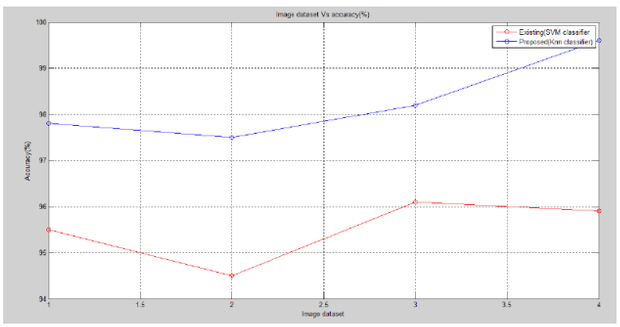

Figure 6. Classification accuracy.

4. Result and Discussion

This paper describes all the necessary steps for offline handwriting recognition system.

The experimental results obtained in recognizing the handwritten Sanskrit characters using Artificial Neural network are summarized .The recognition accuracy and performance efficiency obtained for the Neural network method are also discussed. Larger dataset for character and input sample the result for the complete system can be further improved and enhanced.

5. Conclusion

An off-line handwritten character recognition system with Neural Network for recognizing handwritten Sanskrit alphabets has been described in this paper. The proposed system will find useful applications in recognizing the handwritten names, reading documents and conversion of any handwritten document into structural text form. Accuracy level obtained is 98%. The work can be extended to recognize characters or numerals of some other languages also and to recognition of word and text to audio conversion.

6. References

1. Shaw B, Parui SK, Sridhar M. Offline handwritten Devanagari word recognition: a holistic approach based on directional chain code feature and HMM. IEEE. 2008.
2. Lehal GS, Bhatt N. A recognition system for Devanagari and English handwritten numerals. Proc ICMI, Springer; 2000. p. 442–9.
3. Bajaj R, Dey L, Chaudhuri S. Devnagari numeral recognition by combining decision of multiple connectionist classifiers. Sadhana. 2002 Feb; 27(Part 1):59–72.
4. Bhattacharya U, Chaudhuri BB. A majority voting scheme for multiresolution of hand printed numerals. Seventh International Conference on Document Analysis and Recognition (ICDAR); 2003. p. 16.
5. Bhattacharya U, Parui SK, Shaw B, Bhattacharya K. Neural combination of ANN and HMM for handwritten Devanagari numeral recognition. Tenth International Workshop on Frontiers in Handwriting Recognition; 2006.
6. Pal U, Wakabayashi T, Sharma N, Kimura F. Handwritten numeral recognition of six popular Indian scripts. Proceedings of International Conference on Document Analysis and Recognition (ICDAR); 2007. p. 749–53.

7. Patil PM, Sontakke TR. Rotation, scale and translation invariant handwritten Devanagari numeral character recognition using general fuzzy neural network. Pattern Recogn. 2007 Jul; 40(7):2110–7.
8. Ramteke RJ, Mehrotra SC. Recognition of handwritten Devanagari numerals, Int J Comput Process Orient Lang. 2008.
9. Shanthi N, Duraiswami K. Performance comparison of different image size for recognizing unconstrained handwritten Tamil character using SVM. J Comput Sci. 2007; 3(9):760–4.
10. Sigappi AN, Palanivel S, Ramalingam V. Handwritten document retrieval system for Tamil language. Int J Comput Appl Tech. 2011; 31.
11. Asthana S, Haneef F, Bhujade RK. Handwritten multiscript numeral recognition using Artificial Neural Networks. Int J of Soft Computing and Engineering. 2011 Mar; 1(1). ISSN:2231–2307.
12. Rajashekararadhya SV, Vanaja RP. Zone-based hybrid feature extraction algorithm for handwritten numeral recognition of two popular Indian script. World Congress on Nature and Biologically Inspired Computing. 2009. p. 526–30.
13. Sutha J, Rama RN. Neural network based offline Tamil handwritten character recognition system. International Conference on Computational Intelligence and Multimedia; 2007. p. 446–50.
14. Paulpandian T, Ganapathy V. Translation and scale invariant recognition of handwritten Tamil characters using hierarchical neural networks. IEEE Int Symp Circ Syst. 1993; 4:2439–41.
15. Subashini A, Kodikara ND. A novel SIFE-based codebook generation for handwritten Tamil character recognition. 6th IEEE International Conference on Industrial and Information Systems (ICIIS); 2011. p. 261–4.
16. Shanthi N, Duraiswami K. A novel SVM-based handwritten Tamil character recognition system. Pattern Analysis and Applications. 2010; 13(2):173–80.
17. Ramanathan R, Ponmathavan S, Thaneshwaran L, Nair AS, Valliappan N. Tamil font recognition using gabor and support vector machines. International Conference on Advances in Computing, Control and Telecommunication Technologies; 2009. p. 613–5.
18. Sarveswaran K, Ratnaweera. An adaptive technique for handwritten Tamil character recognition. International Conference on Intelligent and Advanced Systems; 2007. p. 151–6.
19. Indra GR, Iyakutti K. An attempt to recognize handwritten Tamil character using Kohonen SOM. Int J of Advanced Networking and Applications. 2009; 01(03):188–92.
20. Jagadeesh KR, Prabhakar R. An improved handwritten Tamil character recognition system using octal graph. Int J Comput Sci. 2008; 4(7):509–16. ISSN 1549–3636.