FUZZY LOGIC APPROACH FOR OFF-LINE HANDWRITTEN SINHALA CHARACTER RECOGNITION

P. S. Nilanthi and S. R. Kodituwakku

Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya

Introduction

The segmentation and recognition of handwritten characters is a key problem in the development of OCR/ ICR systems. Many methods have been proposed in the recent years. Most of the recent work examines the applicability of soft computing techniques for recognition. Various segmentation algorithms have been proposed (Batuwita and Bandara, 2005, Bandara *et.al.*, 2002). These include segmentation of disconnect upper case English characters; disconnect Sinhala characters and uppercase connected English characters (Hewavitharana et.al., 2002, Karunanayaka *et.al.*, 2005, Fernando and Kodikara, 2003, Rajapakse *et.al.*, 2004). Artificial neural networks and statistical approaches such as Hidden Markov Models are the techniques used for recognition. Offline Sinhala character recognition is a challenging task because of the special structure of Sinhala characters. There are few attempts on Sinhala character recognition. A Neural network based technique presented in (Rajapakse *et.al.* 2004) consists of three major components: pre-processor, pattern recognizer and post-processor. The pre-processor handles all the manipulations necessary for preparing the input character before forwarding it for recognition. The pattern recognizer analyses the input pattern and the post-processor combines the results of each neural network to compose the character. Algorithms for thresh holding, noise reduction and skew correction of Sinhala Handwritten Words are presented in (Karunanayaka *et.al.* 2005). Those are used to improve the accuracy of segmentation and recognition. The authors reported that these algorithms have 97.2% accuracy. Off-line Sinhala Handwriting recognition using Hidden Markov Models is presented in (Fernando and Kodikara, 2003). It was found that the system is 64% accurate. To obtain such accuracy the letters should be written only in a pre-formatted paper. This is a drawback of the system. This paper describes a fuzzy logic based recognition technique for offline handwritten Sinhala characters. The salient feature of this technique is that it segments characters into meaningful segments so that the fuzzy characteristics can be applied on resultant segments for the purpose of recognition.

Materials and Methods

We investigated the applicability of Fuzzy techniques for Sinhala character recognition. The recognition system consists of the following activities.

Binarizaion

In order to reduce unwanted noise and to obtain the character skeleton, the binarization process was used. The value of a pixel point is set to be 1 or 0 if the intensity value of a certain pixel is above or below a threshold intensity value.

Segmentation and Feature Extraction

Sinhala characters mainly consist of arcs and of varying shapes. These arcs make it difficult to segment the character by travelling along the character skeleton. So the character skeleton area is segmented into regions based on the minimum and maximum coordinate values of each character image. For each character, row pixel density count and column pixel density count are taken and stored in an array. Figures 1 and 2 (Appendix) depict how these counts are taken. After storing the pixel counts relevant fuzzy membership values are calculated using the fuzzy membership function shown in Figure 3 (Appendix).

Linguistic terms VVS, VS, S, M, H, VH, VVH stand for *'very very small'* ,*'very small'*, *'small'*, *'medium'*, *'high'*, *'very high'*, *'very very high'* . Fuzzy membership values are calculated five times for a character written differently and all the values are recorded. Finally a rule base is constructed based on extracted features to recognize characters.

Results and Discussion

A prototype system developed was tested with a sample Sinhala characters written by different persons. The test results indicated that the system successfully identified smoothly written characters with more than 70% accuracy. However, some problems occur when there are characters with similar shape, but they differ by a small part. The system reported an average of 65% accuracy.

Conclusions

Fuzzy logic can be used for Sinhala character recognition with a good segmentation technique. Segmentation of Sinhala characters is difficult compared to the segmentation of English characters. The system based on the proposed segmentation responds well to the training set of characters. It showed a higher recognition rate for characters which does not tend to conflict with another character. The recognition rate of identifying a smoothly written character is higher compared to others. The system shows a higher rate of identification for the characters which are written smoothly. Sometimes it fails to identify characters written in different shapes. On the average, the proposed system achieves 65% accuracy. The accuracy may be improved by grouping characters according to shapes before applying fuzzy rules.

References

Bandara, G.E.M.D.C., Pathirana, S.D. and Ranawana, R. M. (2002). Use of Fuzzy Feature Descriptions to Recognize Handwritten Alphanumeric Characters. In: Proceedings of 1st Conference on Fuzzy Systems and Knowledge Discovery, Singapore, 269-274

Batuwita K.B.M.R. and Bandara G.E.M.D.C. (2005). "An Online Adaptable Fuzzy System for

Offline Handwritten Character Recognition," in the Proceedings of Fuzzy Logic, Soft Computing and Computational Intelligence--11th World Congress of International Fuzzy Systems Association (IFSA 2005). China, Springer Tsinghua, 2:1185-1190.

Fernando H.C. and Kodikara N.D (2003). A Database of handwritten Text Recognition Research in Sinhala Language, Proceedings of the Seventh International Conference on Document Analysis and Recognition, 1262-1264.

Hewavitharana S., Fernando H. C. and Kodikara N.D. (2002). "Off-line Sinhala Handwriting Recognition using Hidden Markov Models", Proc. of Indian Conference on Computer Vision , Graphics & Image Processing (ICVGIP) 2002, Ahmedabad, India, 266-269.

Karunanayaka M.L.M, Marasinghe C.A. and Kodikara N.D. (2005). Thresholding, Noise Reduction and Skew correction of Sinhala Handwritten Words, MVA2005 IAPR Conference on Machine Vision Applications, May 16-18, Tsukuba Science City, Japan.

Rajapakse R. K., Weerasinghe A. R. and Seneviratne E. K. (2004). A neural network based character recognition system for Sinhala script. Department of Statistics and Computer Science, University of Colombo.

Appendix

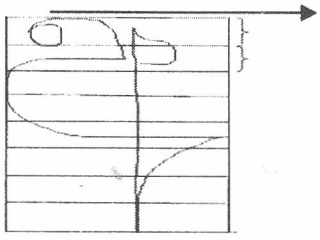

Fig. 1. Horizontal pixel density counts

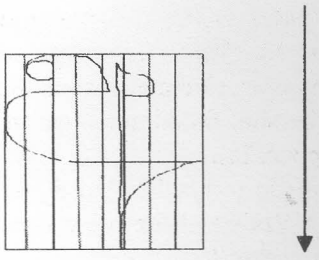

Fig. 2. Vertical pixel density counts

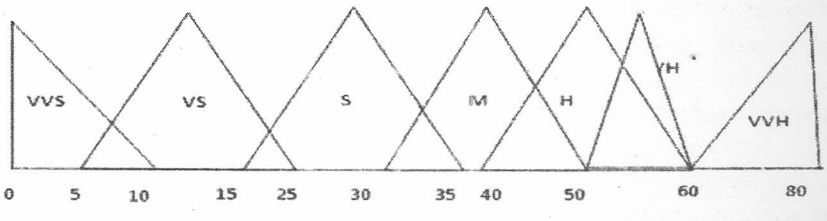

Fig. 3. Fuzzy membership functions for pixel density count