Wavelet Coefficients and Gradient Direction for Offline Recognition of Isolated Malayalam Characters

Retheesh V V
MTech Computer Science and Engineering
National Institute of Technology Karnataka
Surathkal, India
retheeshvv@yahoo.com

Aneesh G Nath
Department of Computer Science and Engineering
TKM College of Engineering
Kollam, India
aneeshgnath@yahoo.co.in

***Abstract*—**This work attempt to use the wavelet transform coefficients combined with image gradient direction as feature vector for the recognition of isolated handwritten Malayalam characters. It has been established that the number of zero crossings in wavelet transform distinctly characterizes an image. This property has been exploited in this work for the recognition of handwritten characters. The images of 71 characters in Malayalam are considered for the recognition purpose. The segmented image of the symbols are thinned and smoothed for further processing. The feature vector proposed in this work is the combination of number of zero crossings in two level Daubechies (Db4) wavelet transform and gradient direction of the image mapped to twelve regions with each region having 30 degree span. A two level Db4 wavelet transform is applied on each processed symbol and the number of zero crossings in each of 20 sub images are counted and recorded. Gradient direction is combined with this to form the feature vector. Multilayer Perceptron classifier is used for classification. We have obtained an accuracy of 98.8%.

Keywords*-*Handwritten character recognition, Malayalam character recognition, Gradient direction , Wavelet coefficients, MLP.

*****

I. INTRODUCTION

Handwritten character recognition (HCR) is a difficult and complex task and poses challenges in the field of pattern recognition and machine learning. The handwritten characters may not be very clear as they may not have sharp lines, curves may not be smooth as that of printed characters. The non uniformity of character size and the lack of proper orientation in writing adds difficulty in automating the recognition. The 100% recognition of handwritten character may be tedious as most of the people find it difficult to read their own handwritings even after a few days of the writing.

Malayalam is a language spoken in India, especially in the state of Kerala. Malayalam is one of the twenty two scheduled languages in India and is the official Language of Kerala state[1] . It has official language status in Lakshadweep and Mahe also. Malayalam belongs to the Dravidian family of languages, and is spoken by approximately thirty eight million people mainly in Kerala and adjoining areas.

The Malayalam language was first written in Vatteluttu, an ancient script of Tamil. However, modern Malayalam script evolved from Grantha, a script originally used to write Sanskrit. Both Vatteluttu and Grantha evolved independently from Brahmi[2]. The modern Malayalam alphabet consists of thirteen vowels, thirty six consonants, and a few symbols. The vowels and consonants are shown in Fig. 1. The combined letters also exist in the language, which makes the language more tougher for automatic recognition.

Among the various phases in character recognition, feature extraction is the crucial and important one. The quality of the feature set affect the accuracy of the recognition. In this paper we propose a combination of directional information of image gradient with zero crossings in wavelet transform as the feature vector. The Multilayer Perceptron classifier is used since neural network based classifiers are more effective in case of handwritten character recognition [3].

Fig. 1. Malayalam Alphabets(Vowels and Consonants).

The remaining sections of this paper are organized as follows. In section II, an overview of existing study in Malayalam HCR is depicted. The feature extraction scheme used in this work is described in section III. The classification is mentioned in section IV followed by result of the experiment in section V. The section VI concludes this paper.

II. EXISTING METHODS FOR MALAYALAM HCR

Lajish [4] in his work reported that it was the first work in handwritten Malayalam character recognition, in which fuzzy-zoned normalized vector distance features are classified using class modular neural network considering 44 Malayalam characters. The reported accuracy was 83.35%. Raju and K Revathy[5] proposed the use of wavelet packet transformation for the recognition of isolated handwritten Malayalam character. Their work stem from the idea that the count of zero crossings of wavelet transform coefficients clearly characterizes an image. R John [6] proposed a work using 1D wavelet transform of projection profiles as features. In their work, Bindu S Moni and G Raju [7] has utilized twelve directional codes depending on the gradient direction as the feature vector which is processed by Modified Quadratic Discriminant function (MQDF) classifier for character recognition.

Abdul Rahiman M and Rajasree M S[8] has proposed a method based on intensity variations for recognizing online handwritten Malayalam characters. Jomy John et. al[9] have proposed a system which uses both curvature feature and gradient feature in the reduced dimension as the feature vector. Another method proposed in [10] uses Haar wavelet transform for feature extraction and support vector machine for classification. Binu P C and Babu Anto P[11] deals with the recognition of handwritten Malayalam characters using discrete features. The features are extracted from skeletonised images. The skeleton pruning is done by contour portioning with discrete curve evolution with a recognition accuracy of 90.18 %t for 33 alphabet classes. All the works above have used a subset of Malayalam alphabets and to our understanding the study was not extended to combined characters.

III. FEATURE EXTRACTION SCHEME

This work proposes a combination of number of zero crossings in wavelet transform coefficients and gradient direction as feature vector for the recognition of handwritten Malayalam characters. We have considered 71 characters in Malayalam which includes combined characters for this study. The proposed solution is illustrated in Fig. 2. The segmented symbols are thinned and normalized to have same physical dimension for feature extraction phase to operate on. A two level wavelet transform(Db4) is applied to the processed symbols and the number of zero crossings in each of 20 sub images are counted and recorded.

An image gradient is the directional change in the intensity in an image. It measures how intensity is changing. There are two components i) The magnitude of the gradient describes how quickly the intensity is changing and ii) the direction of the gradient tells us the direction in which the intensity is changing most rapidly. The gradient vector is formed by combining the partial derivative of the intensity in the X and Y direction, which can be represented by equation below:

$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \tag{1}$$

where

$$\frac{\partial I(x,y)}{\partial x} = \lim_{\Delta x \to 0} \frac{I(x+\Delta x, y) - I(x,y)}{\Delta x} \tag{2}$$

and

$$\frac{\partial I(x,y)}{\partial y} = \lim_{\Delta y \to 0} \frac{I(x, y+\Delta y) - I(x,y)}{\Delta y} \tag{3}$$

Eqn. (2) and (3) above compute the gradient in X and Y direction respectively. In discrete case we can take differences at one pixel interval and the above equation can be approximated as:

$$G_x = \frac{\partial I(x,y)}{\partial x} \approx \frac{I(x+1, y) - I(x-1, y)}{2} \tag{4}$$

$$G_y = \frac{\partial I(x,y)}{\partial y} \approx \frac{I(x, y+1) - I(x, y+1)}{2} \tag{5}$$

where G_x and G_y are the gradient in X and Y direction respectively. The strength **G** and direction Θ of the gradient can be then computed using the equation (6) and (7) respectively.

$$G = \sqrt{G_x^2 + G_y^2} \tag{6}$$

$$\Theta = \arctan\left(\frac{G_y}{G_x} \right) \tag{7}$$

ISSN: 2321-8169
2514 – 2517

The gradient direction thus obtained is mapped to twelve regions at a displacement of 30^0 for the ease of computation. The mapped direction is then combined with the number of zero crossings obtained in 20 sub images of wavelet transform coefficients. This combined set is the feature vector which is fed to the classifier.

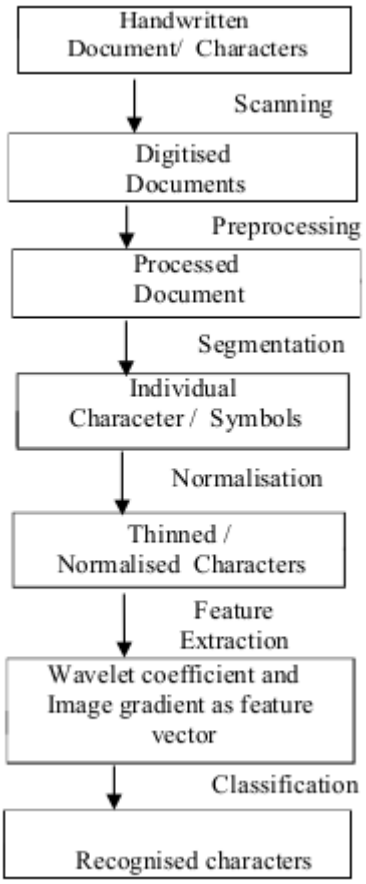

Fig. 2. Block diagram of proposed solution.

Algorithm

i. Apply thinning and normalize the segmented symbols
ii. Apply two level wavelet transform to each symbol using Db4 filter.
iii. Count the number of zero crossings in each 20 sub - images and record it.
iv. Calculate the gradient and gradient direction as in Eqn. (4), (5) and (7).
v. Map the directions into 12 regions at a displacement of 30^0.
vi. Merge the same with the recorded zero crossings in step (iii) to form the feature vector.

The feature vector is then fed to Multilayer Perceptron classifier for training and later for testing.

IV. CLASSIFICATION

A Multilayer Perceptron (MLP) is a feed forward artificial neural network model that maps input sets to appropriate output classes. MLP uses a supervised learning technique called back propagation for training the network.

V. RESULT OF THE EXPERIMENT

It is observed that count of zero crossings in various sub-images in 2D wavelet transform vary considerably from each characters. Hence it is evident that the same can form a rich feature set for recognition. However it is noted that effective rate of recognition when zero crossing alone is considered is in the range of 75 to 80 %. Fig. 3 below shows the plot of number of zero crossings obtained for the most similar Malayalam characters. The Malayalam characters are given some unique code for the purpose of illustration as shown in Fig. 4.

The classification accuracy obtained when gradient direction combined with number of zero crossings used as feature vector is 98.8%

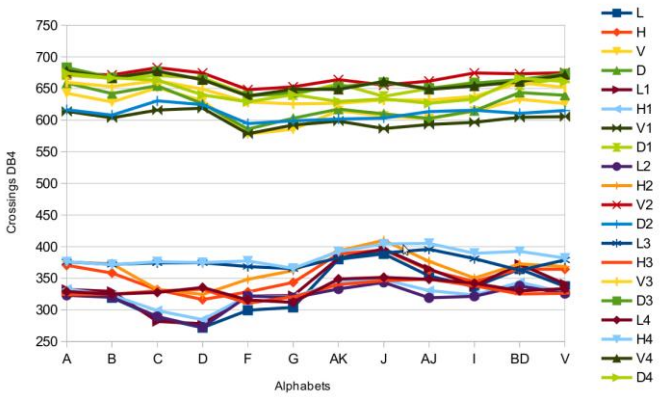

Fig. 3. Plot of number of zero crossings.

Code	Character	Code	Character
A		AJ	
B		I	
C		BD	
D		V	
F		AH	
G		AQ	
AK		AN	

Fig. 4. Code given to a few characters.

VI. CONCLUSION

This work is aimed to investigate the effectiveness of the feature vector created by combining the number of zero crossings in wavelet transform coefficients with gradient directions for offline recognition of unconstrained handwritten Malayalam characters. A set of 71 Malayalam alphabets has been considered in the study which include combined characters. All of the previous work done in this area have limited their studies to a subset of the Malayalam characters and no attempt was made to recognize combined characters. The output obtained is promising with 98.8% recognition accuracy with Multilayer Perceptron classifier. The study has to be analyzed further with more number of sample data to assess the strength of the proposed feature vector. Also the size of the feature vector need be reduced to improve the time complexity.

REFERENCES

[1] Govindaraju, V. and S. Setlur, "Guide to OCR for Indic Scripts: Document Recognition and Retrieval". 2009: Springer Publishing Company, 325.

[2] Ghosh, D., T. Dube, and A.P. Shivaprasad, "Script Recognition;A Review", Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. 32(12): p. 2142-2161.

[3] O Matan et. al. ," Handwritten Character Recognition Using Neural Network Architectures", Proceedings of the 4th USPS Advanced technology Conference, pp 1003 – 1011, november 1990.

[4] Lajish V. L., "Handwritten character recognition using perpetual fuzzy zoning and class modular neural networks", Proc. 4th Int. National conf. on Innovations in IT, 2007, 188 – 192.

[5] G. Raju and K. Revathy, "Wavepackets in the Recognition of Isolated Handwritten Characters", Proceedings of the World Congress on Engineering, Vol I, July 2 - 4, 2007, London, U.K

[6] R. John, G. Raju and D. S. Guru, "1D Wavelet transform of projection profiles for isolated handwritten character recognition", Proc. ofICCIMA07, Sivakasi, 2007, 481-485, Dec 13-15.

[7] Moni, B.S. and Raju, G. "Modified quadratic classifier for Handwritten Malayalam Character recognition using Run Length Count", International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), pp 600-604, March 2011.

[8] Abdul Rahiman M and Rajasree M S, "An Efficient Character Recognition System for Handwritten Malayalam Characters Based on Intensity Variations", International Journal of Computer Theory and Engineering, Vol. 3, No. 3, June 2011.

[9] Jomy John, Kannan Balakrishnan, Pramod K. V, "A System for Offline Recognition of Handwritten Characters in Malayalam Script", International Journal of Image, Graphics and Signal Processing, Vol 4, pp 53-59, 2013.

[10] John, J., K.V. Pramod, and K. Balakrishnan,"Unconstrained Handwritten Malayalam Character Recognition using Wavelet Transform and Support vector Machine Classifier", Procedia Engineering, 30(0): pp. 598-605, 2012.

[11] Binu P. Chacko, Babu Anto P, "Discrete Curve Evolution Based Skeleton Pruning for Character Recognition", Seventh International Conference on Advances in Pattern Recognition, 2009.

[12] Richard G. Casey, and Eric Lecolinet, "A survey of Methods and Strategies in Character Segmentation",IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 7, July 1996.

[13] Lajish V. L., "Handwritten character recognition using gray scale based state space parameters and class modular NN",Proc. 4th Int. National conf. on Innovations in IT, 2007, 374 – 379.

[14] Abdul Rahiman M et al ,"Recognition of Handwritten Malayalam Characters using Vertical & Horizontal Line Positional Analyzer Algorithm", Third International Conference on Machine Learning and Computing (ICMLC 2011) 2011.

[15] Plamondon R and Srihari S N, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, pp 63 – 84, January 2000.

[16] Lajish V L,Suneesh T K and Narayanan N K, " Recognition of Isolated handwritten images using Kolmogorov-Smirnov Statistical classifier and K–nearest neighbour classifier", Proc. of International Conf. on Cognition and Recognition, December, 2005.