Off-line Handwritten Malayalam Character Recognition Using Gabor Filters

Nusaibath C[#1], Ameera Mol P M[*2]

[#1] *M.Tech student(CSE), MES College of Engineering, Calicut University, India*

[*2] *Assistant Professor (CSE), MES College of Engineering, Calicut University, India*

***Abstract*— Handwritten character recognition is the ability of a Computer to receive and interpret handwritten input. Computers may find difficultly in deciphering many kinds of printed characters which is of different fonts and styles or handwritten characters. This paper focuses on the recognition of handwritten Malayalam characters. The proposed system consists of image acquisition, preprocessing, segmentation, feature extraction, classification & recognition. Here spatial Gabor filters are used to extract the Spatial Gabor based features and also ratios of grid values in horizontal and vertical directions are considered as additional features. The proposed system was tested with different handwritings and better recognition rates obtained. These results are comparable and better than published work in the same area. The experimental results indicate the effectiveness of this method for practical handwritten Malayalam character recognition.**

Keywords— **Character recognition, Pre-processing, Binarization, Feature extraction, Gabor filter**

I. Introduction

Handwriting recognition has been a popular area of research since few decades under the purview of image processing and pattern recognition. A major goal of pattern recognition research is to create human perception capabilities in artificial systems. The task of automatically reading handwriting with close to human performance is still an open problem and the central issue of an active field of research. This may be due to the large degree of variability of human writing. Handwritten character recognition (HCR) systems have to address issues such as infinite variety of character shapes, similarity between characters, and distorted and broken characters. The system having intelligence in recognizing the natural handwriting for all possible scripts around the world is the need of the current era. As a first step of document understanding, a digital image of the document to be analyzed needs to be captured by a scanner or digital camera. Then the segmented characters are subject to a number of preprocessing steps that aim at reducing complexity of character recognition. Malayalam scripts are rich in patterns because of their complex curved form, larger number of basic elements and the presence of conjuncts. The combinations of such patterns make the recognition of characters much complex and these patterns should be exploited to arrive at the solution. Here an image of handwritten Malayalam characters is given as the input and an editable document of Malayalam characters in a predefined format is produced as output. Handwritten character recognition provides significant benefit to man-machine communication.

II. Literature Survey

It is hard to say that a complete Malayalam OCR exits which meets all conditions. Malayalam OCR lacks an efficient algorithm. Even in the field of printed characters there are little advancements for this language. Even though the administrative language of Kerala state in India is Malayalam, only a few works were reported in this area. Government of Kerala has now taken initiative for the development of this language and scope of development in this area is promising. Some of the innovative approaches to handwritten Malayalam character recognition are described.

Jomy John et al [1] proposed a method, which is based on the application of wavelet processing in the domain of handwritten character recognition. The method consists of two stages: a feature extraction stage, which is based on Haar wavelet transform and a classification stage that uses support vector machine classifier.

Abdul Rahiman M et al. [2] proposed a method, in which the OCR process is presented in three modules: Preprocessing, Skeletonization and Recognition. In Preprocessing, we scan the input image and separate each character from it. In Skeletonization, we obtain one pixel thick skeleton of the character. In Recognition, classify the characters based on their features. The features of the characters are extracted based on the analysis of position and count of the horizontal and vertical lines.

Abdul Rahiman M et al. [3] modified this method by classifying the connected characters into 3 categories. Here propose an algorithm which uses the inveterate characteristic features to recognize these characters with perceptive accuracy by utilizing the intensity variations in the way in which they may be written. This algorithm recognizes Malayalam characters which are connected in nature.

Bindu S Moni et al. [4] proposed a character recognition system using run length count (RLC). In this method, implemented a feature extraction method based on RLC for the offline recognition of Hand written Malayalam Characters. RLC is the count of contiguous group of 1s encountered in a left to right / top to bottom scan of a character image or block of an image. For classification implemented Modified Quadratic Discriminate function (MQDF), which is a successful statistical approach for Handwritten Character Recognition.

Bindu S Moni et al. [5] proposed another character recognition system using directional features and modified quadratic classifier. Gradient of images is an effective discriminative feature, widely used in pattern recognition applications. In this method the twelve directional codes

depending on the gradient direction is coupled with a statistical classifier for designing an offline recognition system for handwritten isolated Malayalam characters. Preprocessed Character images are decomposed into sub-images using the Fixed Meshing strategy and the twelve directional codes are extracted to form the feature vector. Classification has been carried out by implementing MQDF.

Binu P. Chacko et al. [6] proposed a method deals with the recognition of handwritten Malayalam character using wavelet energy feature (WEF) and extreme learning machine (ELM). The wavelet energy (WE) is a new and robust parameter, and is derived using wavelet transform. This method used an extremely fast leaning algorithm called ELM for single hidden layer feed forward networks (SLFN), which randomly chooses the input weights and analytically determines the output weights of SLFN.

III. Proposed Work

Feature extraction and classification are the two major steps in any HCR system. Performance of an HCR system relies more on the selected feature classifier pair. Feature extraction methods are generally based on statistical and structural features of the characters. Offline HCR systems are matured only in few languages like English and Chinese. Nature of the character set (the Language) has great influence on the performance of an HCR system. In Malayalam language, characters of the alphabet are rich in shape and they are subjected to many variations in terms of handwriting styles. Malayalam characters are isolated in nature; there is no cursive writing and no upper and lower case difference. Only few works are reported in Malayalam HCR. There is no bench marking data base available for comparison.

The proposed work focuses on the identification of Handwritten Malayalam characters. The proposed system consist of image acquisition, preprocessing, segmentation, feature extraction, classification & recognition. The proposed system architecture is shown in figure 1.

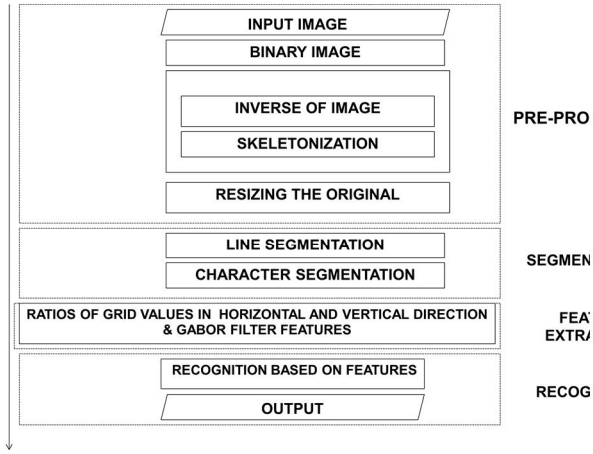

Fig.1 Proposed system architecture

A. Image Acquisition

In image acquisition the recognition system acquires a scanned image as an input image. This image is acquired through a scanner, digital camera or any other suitable digital input device. A sample training set is shown in figure 2.

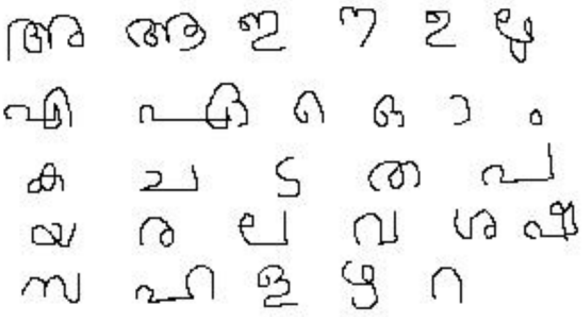

Fig. 2 Input image

B. Preprocessing

Preprocessing includes the steps that are required to shape the input image into a form suitable for segmentation. Preprocessing steps are binarization and skeletonization.

C. Segmentation

In segmentation stage the input image is segmented into separate lines and then each character is separated. Line separated images and character separated images are shown in figure 3 and figure 4 respectively.

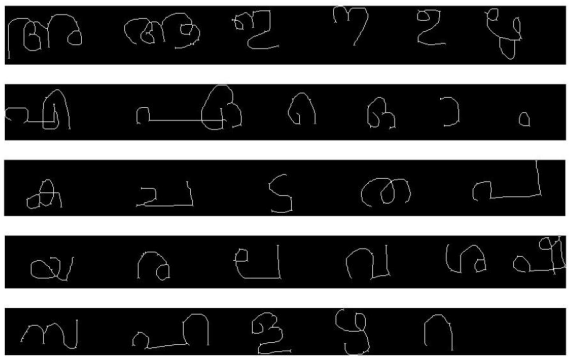

Fig. 3 Line separated image

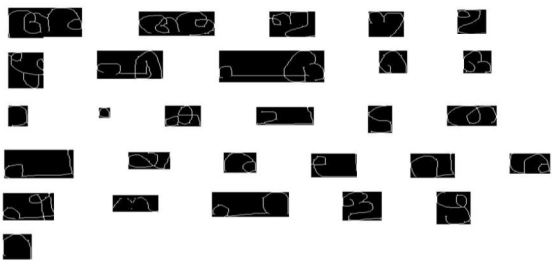

Fig. 4 Character separated image

D. Feature Extraction

After that a feature extraction method is applied to capture the most relevant characteristic of the character to recognize. Here the features are aspect ratio, ratios of grid values in horizontal and vertical directions and Gabor features.

I. Gabor Filter

A two-dimensional Gabor filter can be viewed as a complex plane wave and an elliptical Gaussian. The even and odd Gabor filters in the 2-dimensional spatial domain cab be formulated as,

$$g_{e(x,y;\lambda,\theta)} = e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} \cos\left(2\pi\frac{x}{y}\right) \quad (1)$$

$$g_o(x,y;\lambda,\theta) = e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} \sin\left(2\pi\frac{x}{y}\right) \quad (2)$$

Where λ is Gabor filter wavelength in pixels, θ is the angle of filter in degrees, k_x , k_y scale factors specifying the filter sigma relative to the wavelength of the filter to make the shapes of the filter invariant to scale, k_x controls the sigma in the x direction which is along the filter, and hence controls the bandwidth of the filter. k_y controls the sigma across the filter and hence controls the orientation selectivity of the filter; $\sigma_x = \lambda k_x$ and $\sigma_y = \lambda k_y$

The response of the filter in (1 & 2) to an image i(x, y) can be calculated with the convolution

$$G_e(x,y;\lambda,\theta) = G_e(x,y;\lambda,\theta) \times i(x,y) \quad (3)$$

$$G_o(x,y;\lambda,\theta) = G_o(x,y;\lambda,\theta) \times i(x,y) \quad (4)$$

$$G_a(x,y;\lambda,\theta) = \sqrt{G_e^2(x,y;\lambda,\theta) + G_o^2(x,y;\lambda,\theta)} \quad (5)$$

E. classification & recognition

In the final step realize the character according to features extracted. Character recognition is shown in figure 5.

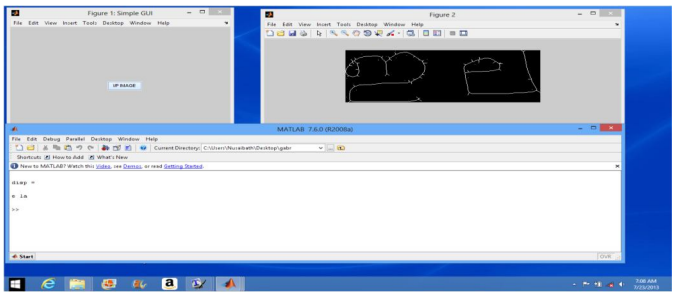

Fig. 5 character recognition

F. Implementation Details

Implementation of the system was done in MATLAB using Image Processing Toolbox (IPT). The images used fo training and testing of the proposed system are in JPEG format.

G. Experimental Results

Initially the system was trained with alphabets of different handwritings. Different size characters with different handwriting are tested. This recognition system achieves 96.80% recognition accuracy. The proposed system recognizes both isolated and connected characters. Performance comparisons of different methods are shown in table 1.

Methods	Character Types		RA(Recognition accuracy in %)
	Isolated	Connected	
WT&SVM classifier [1]	√	√	90.25
Vertical& horizontal line positional analyzer [2]	√	×	91.00
Intensity variations [3]	×	√	91.41
Intensity variations [3]	√	×	95.96
Run length count & MQDF[4]	√	×	94.18
Directional features &MQDF[5]	√	×	95.42
WE & ELM [6]	√	√	95.59
Proposed system	√	√	96.80

Table 1: Performance comparison

IV. Conclusions

The strength of the selected feature and the effectiveness of the classifier are the two key factors determining the performance of a handwritten Character Recognition System. The result of the proposed system is better than published work in the same area. It uses lesser number of features, which are highly uncorrelated; the computation time to extract the features will highly reduce. In future make the system suitable for recognition of noisy images and design the system using different feature classifier pair for better recognition accuracy.

REFERENCES

[1] Jomy John, Pramod K. V., Kannan Balakrishnan. Unconstrained Handwritten Malayalam Character Recognition using Wavelet Transform and Support vector Machine Classifier. In International Conference on communication Technology and System Design, ELSEVIER 2011.

[2] Abdul Rahiman M, M S Rajasree Masha N, Rema M , Meenakshi R, Manoj Kumar G. Recognition of Handwritten Malayalam Characters using Vertical Horizontal Line Positional Analyzer Algorithm. In International Conference IEEE, 2011.

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

[4] Bindu S Moni, G Raju. Modi_ed Quadratic Classifier for Handwritten Malayalam Character Recognition using Run length Count. In International Conference IEEE, 2011.

[5] Bindu S Moni G Raju. Modified Quadratic Classifier and Directional Features for Handwritten Malayalam Character Recognition. IJCA Special Issue on Computational Science - New Dimensions Perspectives NCCSE, 2011.

[6] Binu P. Chacko V. R. Vimal Krishnan G. Raju P. Babu Anto . Handwritten character recognition using wavelet energy and extreme learning machine. In Int. J. Mach. Learn. Cyber. (2012), Springer-Verlag 2011.