International Journal of Engineering & Technology, 7 (4) (2018) 4629-4636

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.12551
Research paper

Handwritten Malayalam Character Recognition using Regional Zoning and Structural Features

Ajay James[1*], Raveena P.V[2] and Chandran Saravanan[3]

[1]*Assistant Professor*
[2]*M.tech student*
[3]*Associate Professor*
[*]*ajayjames80@gmail.com,raveenapv94@gmail.com, dr.cs.1973@gmail.com*

Abstract

Optical Character Recognition (OCR) extracts features from an image of script and converts it to machine-readable code. OCR comprises of line segmentation, word segmentation, character segmentation and character Recognition. Printed documents are efficiently converted to the editable text format with almost 100% accuracy. Handwritten character recognition places difficulties in identifying and translating scripts because of the wide variation in human handwriting. The writing styles like line spacing, word spacing, character sizes and shape of each character varies from person to person. Feature extraction and character recognition are different for different languages and become the most complicated task among the phases of OCR. By language characteristics, feature extraction can differ for each language. The Malayalam characters are characterized by their curved and non-cursive nature. The handwritten character recognition for the Malayalam language that proposed here uses a regional zone based method with structural feature extraction.

Keywords: *Feature Extraction, Classification, Regional Zoning, Optical Character Recognition, Structural Features.*

1. Introduction

Digitization of character images is getting improved and used for many applications such as human-computer interaction. Old documents which are currently preserved in their original handwritten form need to be converted to electronic format for better preservation, editability and to prevent tampering with the original copy. Digitization of documents without Optical Character Recognition (OCR) requires typing the entire document by using any documentation software. The emergence of OCR made this difficulty to disappear to a considerable extent.

OCR tries to recognize characters from an image of text, to encode the text in a particular format, which is more convenient to edit. The image of texts will be either in printed or in handwritten form. In printed format, the characters will be uniform throughout the document. Printed characters will have same appearance irrespective of the writers. But, for a handwritten document, the character varies from writer to writer making it challenging to process. The handwritten character recognition systems are broadly divided into two types, online and offline, depending upon whether written on touch-sensitive surfaces or a piece of paper. In the off-line recognition, the handwritten document is captured by a scanner and corresponding image file is created. But in the online system, the two-dimensional coordinates of successive points are represented as a function of time.

The output of an OCR system is either ASCII encoded or Unicode type which is post-processed easily by the machine. It is challenging to achieve 100% efficiency since the character recognition is different for different languages. When a definite static figure is known, it is possible to take common features correctly. That is, typical recognition rates for machine-printed characters exceed 99%. However, OCR is prone to errors when dealing with handwritten characters, resulting in reduced recognition rates.

Malayalam is an Indian language predominantly used in the state of Kerala. More than 95% people of Kerala use Malayalam for verbal and written communication. Malayalam script is curved in nature and is not cursive. The character types of Malayalam script consist of independent vowels, dependent vowel signs, consonant letters, consonant signs, consonant conjuncts and chillu as shown in Figure 1. Independent vowels are used to write syllables, which starts with a vowel. Dependent vowel signs occur only in combination with a base consisting of a symbol for a single consonant or a consonant cluster.

The following Figure 2 shows the overall process of character recognition. The method of Optical Character Recognition is divided into two phases, namely training phase and testing phase. Training data which are in image form are used to train the machine so that it will predict the unknown input. In the training phase, labeled character images are used as inputs and in testing phase test data are used as inputs. The character images are pre-processed to remove unwanted regions and noise and thereby increase the accuracy. The processed image is then fed to the feature extraction module to obtain a feature vector. The feature vector is a set of features which classifies each character uniquely. Feature extraction is the process of analyzing the whole image for extracting features vector. The feature vector thus obtained is sent to the classifier. There are many classifiers such

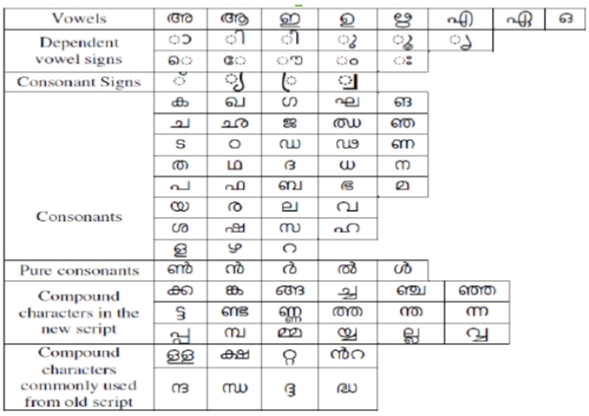


Figure 1: Malayalam Characters

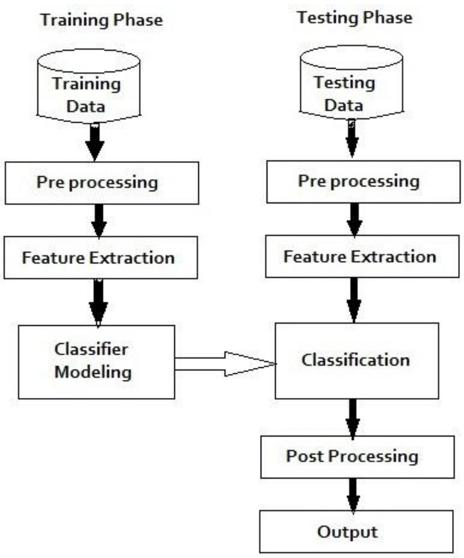


Figure 2: Overview of system

as Support Vector Machine (SVM) [1], k Nearest Neighbor (kNN) [2], Hidden Markov Model (HMM) [3], etc. Classifiers take these feature vectors and their corresponding labels as input and produce a classifier model. In the testing phase, pre-processed test image undergo feature extraction process and this feature vector along with classifier model is used to predict the classifier output. The output can be displayed or saved to a file.

The implementation of Malayalam handwritten character recognition is difficult to achieve better accuracy. One of the main problems is the difficulty in recognizing visibly similar characters. To improve the efficiency, the features need to be enhanced.

2. Related Work

Panyam Narahari Sastry et al. [4] proposed a zonal based feature extraction scheme for Telugu handwritten character recognition. In the Zoning feature extraction method, a character is divided into partitions of predefined size and those partitions are called zones. This method is simple and can be applied for optical character recognition. There are 18 vowels and 36 consonants in the Telugu language. From each zone, the pixel density was collected as a feature. The 100 zones result in 100 features in a vector form.

M. Blumenstein et al. [5] proposed a character recognition method using some features. Starting point and intersection point locations, distinguish individual line segments, labeling line segment information and line type normalization were discussed for direction feature extraction. These features were obtained from different equal sized windows or zones. Additionally, Transition feature was used which is obtained by computing transition from background to foreground.

Gayathri P et al. [6] solved Malayalam handwritten character recognition using hidden Markov model. The paper proposed some statistical features including some black pixels in various regions and zones, aspect ratio and angle baseline and global center of gravity. The features in their consideration are not language dependent; merely counts black pixels in different regions. The method achieved 81.38% accuracy for 13 Malayalam vowel characters.

Nisha Sharma et al. [7] developed a hybrid feature for handwritten character recognition for English lowercase and uppercase alphabets, numerals and special symbols. Statistical, Geometric and Directional Feature Extraction techniques were applied to each character image. Support Vector Machine was used for the classification. Four types of features were used in this method. The feature vector is large. The statistical feature was of size 69 feature values, geometric features of size 18 and directional feature yields around 20 features. The accuracy achieved was 92.16

Rajashekararadhya S. V et al. [8] suggested an isolated handwritten Kannada and Tamil numeral recognition approach. Projection distance and zoning based scheme was used for the numeral recognition. The nearest neighbor classifier was used for testing. The whole image was divided into 25 equal parts and these parts formed the zones. Pixel distance for grid column is computed from image centroid. If there was more than one pixel in a grid column, average pixel distance was calculated. The process was repeated for the entire grid to get 250 features (10 features for a zone). An average of 95% accuracy was achieved.

Giridharan. R et al. [9] proposed an information retrieval from temple Epigraphy using image zoning. The image was decomposed in several ways. Decompositions to vertical zones, equal horizontal zones, right diagonals, left diagonals, octants, diagonal quadrants, quadrants, etc. were used as zoning. A total of 54 zones were obtained. The density of foreground pixel was calculated in each zone. Perceptron was used for the recognition purpose. It creates a single output from multiple real-valued inputs using a linear combination according to the data.

Kalyan Sourav Dash et al. [10] presented a handwritten digit recognition using optimal zoning. Additionally, the system used non-redundant Stockwell transform. The selection of optimal zone from a test image was shown in the paper. Training phase and testing phases are isolated and are difficult to improve feature selection from error analysis adaptively.

Ganpat Singh G Rajput et al. [11] proposed a novel zone based method for recognition of Kannada handwritten characters. Simple traditional zoning was applied to the normalized image producing 64 equal zones of size 8x8 pixels each. Crack codes from left to right and top to bottom was calculated from each zone, which represented the line between the object pixel and the background when traversing in an anti-clockwise direction. With the SVM classifier this method achieved around 87.24% accuracy for Kannada characters. Giuseppe Pirlo et al. [12] developed an adaptive membership function for handwritten character recognition by Voronoi-based image zoning for handwritten numeral recognition. A membership function was selected according to each zone which best suit to covey the characteristics of that zone. The use of Voronoi based zoning over simple traditional zoning along with adaptive zone-based membership functions provided an excellent research platform.

Apart from Zoning and structural features, many other techniques were widely used in OCR systems. Manoj Kumar Mahto et al. [13] used vertical and horizontal projection for Gurmukhi characters and obtained 98% accuracy. Rajesh Gopakumar et al. [14] proposed a zoning based system to recognize multilingual characters from south Indian languages along with Hindi and English. Kalyan S Dash et al. [15] used a hybrid approach with Kirsch gradient operator and

curvature properties of Odia handwritten numerals. Principal Component Analysis (PCA) was used for feature dimension reduction and Modified Quadratic Discriminant Function (MQDF), Discriminative Learning Quadratic Discriminant Function (DLQDF) were used as classifiers
Aditya Raj et al. [16] proposed a conventional feature extraction with Structural Features for Oriya character recognition. The structural feature included a vertical line in a character and open space in a lower zone of a character. Kartar Singh Siddharth et al. [17] suggested Gurmukhi character recognition technique using background direction distribution feature and statistical features. Rafael M. O. Cruz et al. [18] presented different feature extraction algorithms for cursive character recognition with a classifier ensemble. The outputs from each feature set were combined using the group to C-Cube database. U. Pal et al. [19] solved curve-like Oriya handwritten character recognition system with curvature feature using bi-quadratic interpolation and direction of the gradient, etc.
Shanjana C et al. [20] presented Malayalam handwritten text recognition which comprised of all the OCR stages from segmentation of lines to the character classification. Curvature feature and direction changes were used as features and Support Vector Machine for classification. The gradient was calculated for finding the direction of writing. Primekumar K.P and Sumam Mary Idiculla et al. [21] presented an online Malayalam character recognition system using Time domain features such as input coordinates and the angular features. HMM and SVM classifier was used for classification and achieved maximum accuracy of 97.97% for SVM and 95.24% for HMM. Steffy Maria Joseph et al. [22] proposed an online character recognition system for Malayalam script using SVM classifier. Sk Md Obaidullah et al. [23] developed OCR for several Indian scripts including Malayalam using transform-based textual and statistical approach to get 34-dimensional features. Finally, 20 features were selected by applying Greedy Attribute Selection algorithm. Pawan Kumar Singh et al. [24] proposed a texture based method for word-level script identification with coefficients of Discrete Cosine Transform (DCT) and Moment invariants as features. Ashita T et al. [24] used N-gram segmentation, geometric features, and SVM classifier.

3. Proposed Work

The proposed method for Malayalam handwritten character recognition is based on regional zoning and structural features. Malayalam character images are given as input to the system, and it recognizes the character and produces the interpreted Unicode to the user. The following Figure 3 shows the overall system architecture. This system comprises of four major stages, 1) pre-processing, 2) feature extraction, 3) classification, and 4) post-processing. The method is discussed in detail in the following sections. Feature Extraction and classification are the major phases in this model.

3.1. Pre-processing

In the pre-processing stage, the character image is processed for removing noise. First, binarization is done on the input image and then the binarized image is cropped to delete the unwanted white area. A good recognition rate can be achieved by normalizing the image into 100x100 size. In this cropped image, noise removal technique is applied to avoid misclassifications and incorrect results. Thinning operation is performed to make a skeleton of the character. The character image now consists of a character represented by single pixel thickness in a white background. The various steps are as follows and and an example is shown in the following Figure 4.

1: Read the image
2: Convert into a binary image.
3: Edge detection and cropping the image

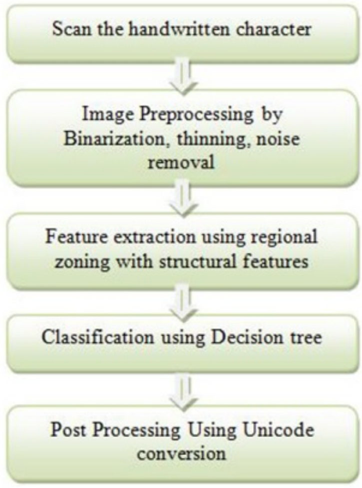


Figure 3: System Architecture

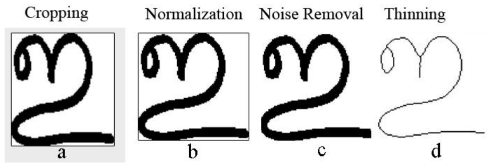


Figure 4: Preprocessing (a) Cropping, b) Normalized to 100 x 100 pixel, c) Noise Removed image, d) Thinned image)

4: Normalize to 100x100 pixels
5: Perform Noise removal
6: Apply the thinning algorithm

3.2. Feature Extraction

In the feature extraction phase, extended zoning feature extraction is used in this system. The image is divided into partitions, called zones and the various features are collected. The following structural and geometric features are captured from each region.

(1) Length of character in horizontal
(2) Length of character in vertical
(3) Number of endpoints
(4) Number of intersections in Horizontal
(5) Number of intersections in Vertical
(6) Number of loops
(7) Direction of writing
(8) Number of horizontal lines
(9) Number of vertical lines

The direction feature of a character is the direction of writing. Some characters start from bottom and end at top or bottom; other characters start from the top and end at the top or bottom. A total of 12 direction groups are there for the whole Malayalam character set. The character like ”ra” starts from the bottom and ends at the bottom, whereas those like ”va” starts at the bottom and end at the top, ”ga” starts at the top and ends at the bottom and ”dha” starts and ends at the top. The printed characters can be classified to a particular direction group precisely whereas the handwritten characters have different possibilities of direction due to varying writing styles. Thus, for a handwritten Malayalam character there is a possibility that it may belong to 2 different direction groups.
Additionally, the system comprises of regional zoning which indirectly performs statistical feature extraction. The image is divided into zones and then these zones are grouped to form different regions. The structural features for each of the region as listed before

is calculated and the feature vector is obtained. The listed structural features for each of the regions are found out by changing the position of the region each time and the obtained feature vector is stored. The different regions formed from regionalization are shown in the following Figure 5.

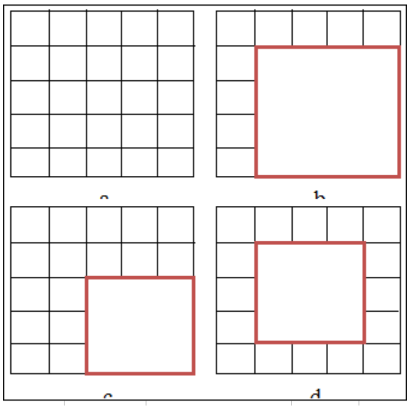


Figure 5: Forming Regions; a) 25 equal zones, b) Forming a region of 4 x 4 grids, c) Reducing the size of the region, d) Moving the region to the middle.

In the feature extraction phase, nine structural features of the whole image are collected and are stored as a feature vector. The image is then divided into 25 zones each of 20x20 pixels. From each of the region under consideration, the same set of features is collected repeatedly. Thus, features from 13 such regions and also of the whole image is taken, to form a feature vector of size 126. The methods to capture the different features are explained below.

Length of character in horizontal and vertical (1,2)

The length of the character in the horizontal and the vertical is computed before the pre-processing is done. To find this length, the entire image is scanned row-wise until a black pixel is located which will be the upper boundary and a lower limit is also found out in a similar manner. Figure 6 shows an example and the various steps involved are as follows.

1: Scan the image row wise
2: Stop scanning if a black pixel is found and store the row number
3: Scan the image until a row with all white pixels is located and store the row number
4: Subtract the row numbers to find the height of the character
5: Perform the same algorithm in each column for detecting the width of the character

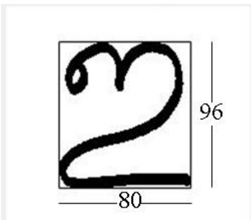


Figure 6: Length of character in horizontal and vertical

Number of endpoints (3)

A pixel is considered as an endpoint, if it is surrounded by only one black pixel. The image is scanned one pixel at a time and its surrounding pixel status are checked. Finally, the count of such points are added. The following Figure 7 shows the endpoints of some of the characters.

Number of intersections in horizontal & vertical (4,5)

A line is drawn in the middle horizontally and vertically and the black pixels that touches the lines are counted. The following Figure

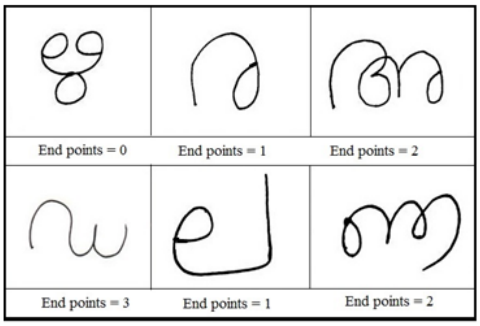


Figure 7: Number of endpoints

8 and Figure 9 shows the number of intersections in horizontal and vertical respectively.

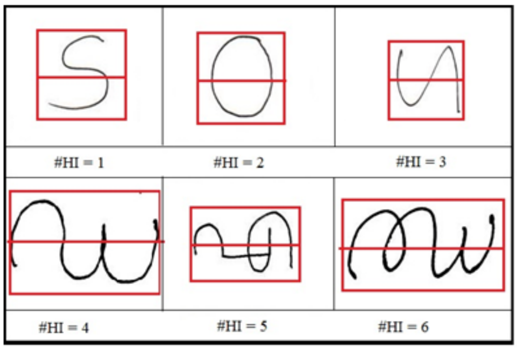


Figure 8: Number of intersections in horizontal

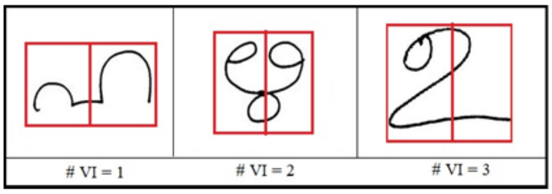


Figure 9: Number of intersections in vertical

Number of loops (6)

The number of loops ranges from zero to three for standard printed Malayalam character set. But, as explained in the endpoints, number of loops differ based on the writing styles. Some partially completed loops make the character loop free. Euler number method is used to find the number of loops in a character image.

$e = \text{No. of objects} - \text{No. of holes}$

$\text{No. of loops} = 1 - e$

Some example characters with the identified loops are shown in the following Figure 10.

Direction of writing (7)

There are 12 classes based on the direction of writing and are listed below.

1. Starting at top and ending at the top
2. Starting at top and ending at the bottom
3. Starting at bottom and ending at the top
4. Starting at bottom and ending at the bottom
5. Starting at left and ending at the right
6. Starting at top
7. Starting at bottom
8. Ending at top
9. Ending at bottom
10. Ending at right

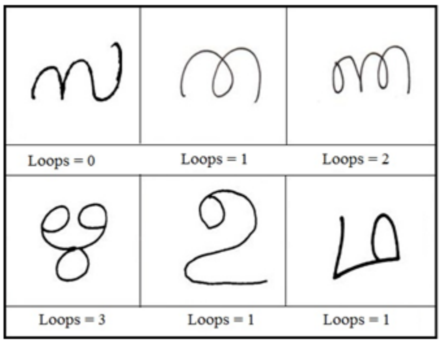


Figure 10: Number of loops

11. Ending at left
12. Others

The whole Malayalam character set is included in any one of these 12 classes. To find the direction, the image is split in the middle vertically. From the two halves the endpoints are determined. The endpoint is scanned and the direction is found out using the direction of the neighboring pixel. If the adjacent pixel is above the end point, then the direction is bottom. An example is shown in the following Figure 11. Similarly, the remaining directions are also determined.

Figure 11: Starting and ending of a character

The algorithm for determining the direction of writing is as follows.

1: Divide the image (showing endpoints as white and all other pixels as black) vertically into 2
2: Consider the first half and start scanning
3: Find the leftmost endpoint and its position; this endpoint is taken as the beginning of the character
4: Find the direction of the beginning pixel by considering the direction of its neighboring pixel.
5: Consider right half of the image
6: Find rightmost endpoint which is considered as ending of that character.
7: Perform the same steps to find the direction of ending pixel.

After finding the start and end of the character, a character is classified into any one of the listed 12 classes. The output of this feature should range from 1 to 12. If there are no endpoints in both the halves of the image, then that character is classified into group 12. The following Figure 12 shows the various Malayalam characters classified into different classes based on the position of endpoints

Number of lines in horizontal and vertical (8,9)

The whole image is searched for finding horizontal and vertical lines. A threshold value *t* is set. If there are *t* consecutive points in horizontal, then it is a horizontal line. Similarly, if there are t successive points in vertical, then it is a vertical line. The following Figure 13 and Figure 14 shows examples of the number of horizontal and vertical lines respectively.

The following Figure 15 gives an example of features extracted from a Malayalam character. These features are collected from the whole image. Now, for each zone region, the same features are calculated and stored in a vector for classification.

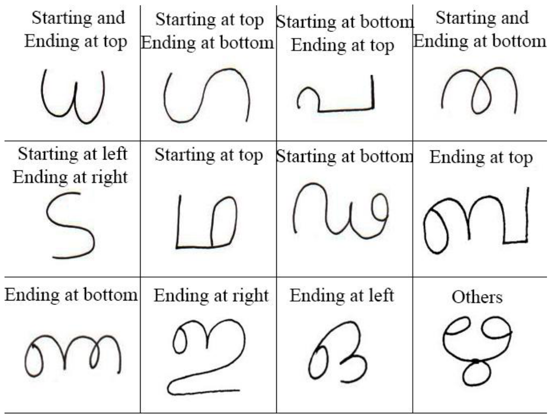


Figure 12: Malayalam characters classified into Direction classes

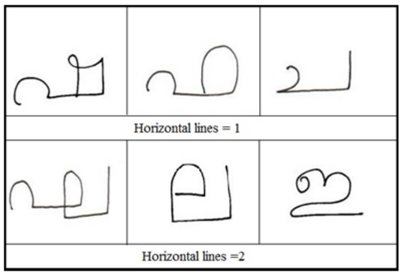


Figure 13: Number of horizontal lines

3.2.1. Regional Zoning

The nine features are collected from the whole image (Region R1) and the results are stored in a feature vector. Further, some smaller regions are considered from the image and the same features are again obtained. First, the image is divided into 25 equal parts as shown in Figure 3 (a). Then, 4 x 4 grid is considered as a region R2 as shown in Figure 3 (b). The structural features are collected from the part of the image. The region is moved above, below, and left to get R3, R4, and R5. A total of 36 features are collected from this zoning. Similarly, 3 x 3 grid region is considered as shown in figure 3 (c). Thus, regions R6 through R14 are produced. These 14 regions form a feature vector of size 126 and is shown in Table 1. The following Figure 16 and Figure 17 shows example of Regionalization.

Table 1: Regionalization

Zoning	Regions	Pixels
Zoning 1	R1	100 x 100
Zoning 2	R2, R3, R4, R5	80 x 80
Zoning 3	R6, R7, R8, R9, R10, R11, R12, R13, R14	60 x 60

The following Figure 18 shows some example characters and their feature values. Six characters and their nine features are shown in the table. The following Figure 19 shows an example of the various steps done on a Malayalam character. Figure 19 (a) is after cropping the character image; Figure 19 (b) after normalizing the image size; Figure 19 (c) after performing noise removal for the character; Figure 19 (d) after thinning. After preprocessing, the feature extraction phase displays some of its output namely number of endpoints as in Figure 19 (e) and horizontal and vertical intersections as in Figure 19 (f). The detected features are written in a file as in Figure 19 (g) and the Malayalam character recognized is written into a text editor as shown in Figure 19 (h).

The following Table 2 shows the different feature vectors used in this method. In this method, a feature vector of size 9 (structural

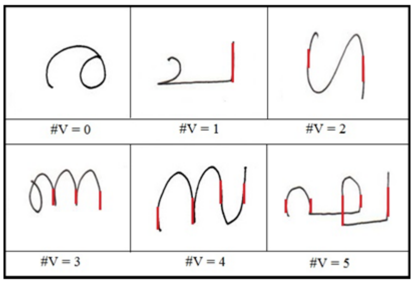


Figure 14: Number of vertical lines

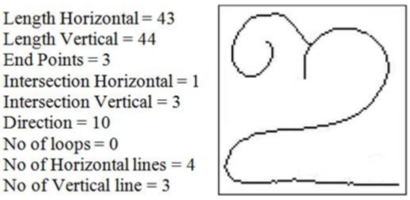


Figure 15: Example features for Malayalam letter ”e”

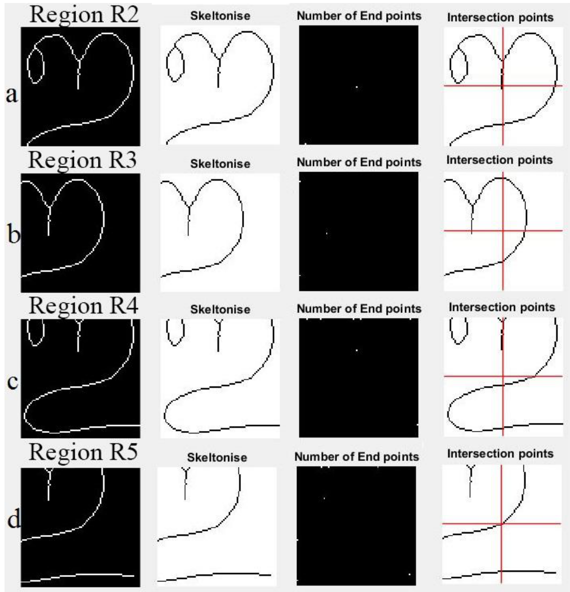


Figure 16: Regional zoning showing thinning, Number of endpoints and intersection points for (a) R2, b) R3, c) R4 and d) R5)

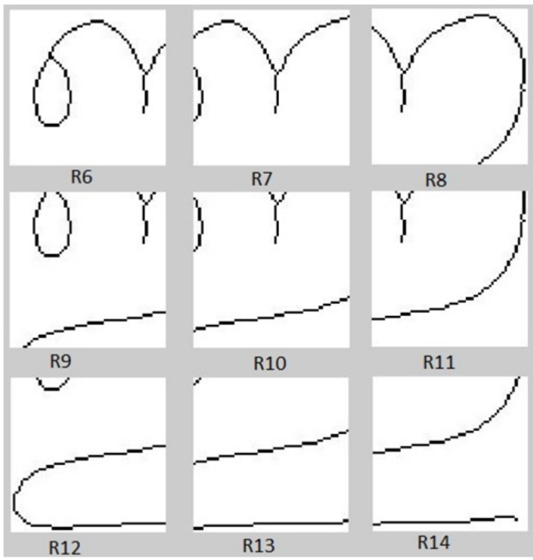


Figure 17: Regional zoning R6 to R14

features) collected from the whole image is named feature set 1 (f1). A second set consists of a feature vector of size 45 collected from first five regions and is considered as the feature set 2 (f2). The third category is a feature vector of size 126, collected from 14 regions forming a feature set 3 (f3).

Table 2: Feature Vector Specification

	No of Regions	No of features
The feature set 1 (f1)	1	9
The feature set 2 (f2)	5	45
The feature set 3 (f3)	14	126

3.3. Classification

The decision tree classifier is used to predict the unlabelled sample. It creates a tree with leaf nodes as output to predict the class. The decisions are followed in each node starting from the root, down to a single leaf node. Classification trees produce a response that is nominal, such as ’true’ or ’false.’ The decision tree is formed using the features obtained from the training images. The testing module deals with the test images. It first pre-processes the input image and collects the feature vector. This feature vector along with the classifier model created by the training module is used to classify the unlabeled test data.

3.4. Post-Processing

Finally, the output of the classifier is mapped to the Malayalam character Unicode. Classifier outputs any of the integer labels corresponding to the character class. This integer label is compared against a statically created look-up table and retrieves the Unicode of interest. The end user is not familiar with the Unicode representation of the Malayalam character. Thus, the Unicode character is written to a text editor for viewing it by the end user.

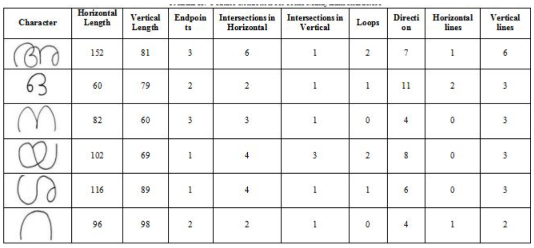
4. RESULTS

4.1. Data Set

A standard dataset for Malayalam handwritten character was not available at the time when the work started. So, a 2000 samples of each of the 57 characters including vowels, consonants, dependent vowels, dependent consonants, etc. were collected, thus making a dataset of size 116,000 available online as P-ARTS Kayyezhuth [25]. 200 samples of each character were collected from different individuals in a paper. This sample was produced by both ball point pen and sketch pen. The individual characters are cropped and stored in *jpg* format. Thus, collected 200 samples were processed by applying different morphological changes producing 2000 samples of each character. These characters were organized into corresponding classes named from class 1 to class 57. 80% of this data is used for training and the remaining is used for testing. Individual components in compound characters are collected separately.

4.2. Recognition Results

The following Table 3 consolidates the recognition rates for the feature sets with the vowels and consonants separately. For the feature size of 126, the accuracy obtained is 95.4% and for a feature vector of size 9 is 78.67% of efficiency. Some classification labels affect the recognition in such a way that the 8 class vowel recognition attains



Character	Horizontal Length	Vertical Length	Endpoints	Intersections in Horizontal	Intersections in Vertical	Loops	Direction	Horizontal lines	Vertical lines
	152	81	3	6	1	2	7	1	6
	60	79	2	2	1	1	11	2	3
	82	60	3	3	1	0	4	0	3
	102	69	1	4	3	2	8	0	3
	116	89	1	4	1	1	6	0	3
	96	98	2	2	1	0	4	1	2

Figure 18: Feature extraction for some Malayalam characters

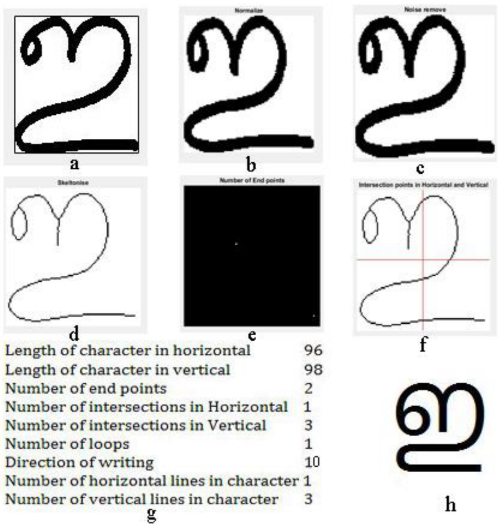


Figure 19: Output (a) Cropping b) Normalize c) Noise removal d)Thinning e) Endpoints f) Intersections in horizontal and vertical lines g) Feature file h) output character written in file

98.72% accuracy while the consonant recognition was obtained to be 95.68% for the feature vector of size 126.

Table 3: Accuracy for vowels and consonants

Feature set	Vowels (8)	Consonants (36)	Total characters (57)
f1 (9)	84.20	79.86	78.67
f2 (45)	95.88	92.10	91.88
f3 (126)	98.72	95.68	95.40

Figure 20 is a graph showing the recognition rate against the feature sets. For the handwritten character recognition, different feature sets and different results are obtained. The 20% of the dataset images for all the three categories are tested and the corresponding recognition rates are depicted in the graph. From this evaluation, its is found that the feature set 3 is more accurate than others, leading us to conclude that the recognition rate improves with the size of feature vectors. From Table 3, it is observed that the number of classes to be processed affects the recognition rates. Recognition of 8 vowels gave high accuracy when compared to 36 consonants. The accuracy is better than the total recognition rate of the system, that is the whole character set consisting of vowels, consonants, dependent consonants, and dependent vowels.

5. CONCLUSION

OCR is critical in digitization of documents and other related applications. This research work implements a handwritten Malayalam character recognition system using extended regional zone based features with structural features. A list of structural features is considered in combination with regions and zones, which produced a

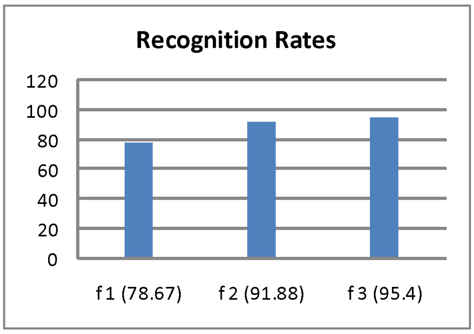


Figure 20: Graph of recognition rate to feature sets

feature vector of size 126. For the feature size of 126, the accuracy obtained is 95.4% and for a feature vector of size 9 is 78.67% of efficiency. Some classification labels affect the recognition in such a way that the 8 class vowel recognition attains 98.72% accuracy while the consonant recognition was obtained to be 95.68% for the feature vector of size 126. It can be concluded that as the size of feature vectors increase, the recognition rate also improves.

References

[1] A. T. Jia, Y. Ayappally and K. Syama, "Malayalam OCR: N-gram approach using SVM classifier," 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, 2013, pp. 1799-1803.

[2] S. A. Chaudhari and R. M. Gulati, "An OCR for separation and identification of mixed English — Gujarati digits using kNN classifier," 2013 International Conference on Intelligent Systems and Signal Processing (ISSP), Gujarat, 2013, pp. 190-193.

[3] K. A. Mohand, T. Paquet, N. Ragot and L. Heutte, "Structure Adaptation of HMM Applied to OCR," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 2877-2880.

[4] P. N. Sastry, T. R. V. Lakshmi, N. V. K. Rao, T. V. Rajinikanth and A. Wahab, "Telugu Handwritten Character Recognition Using Zoning Features," 2014 International Conference on IT Convergence and Security (ICITCS), Beijing, 2014, pp. 1-4.

[5] Michael Blumenstein, Brijesh Verma, and Hasan Basli. 2003 "A novel feature extraction technique for the recognition of segmented handwritten characters", In Document Analysis and Recognition, 2003. Proceedings, Seventh International Conference on, IEEE 2003, pp. 137-141.

[6] Gayathri, P., and Sonal Ayyappan. 2014 "Off-line handwritten character recognition using Hidden Markov Model", In Advances in Computing, Communications and Informatics ICACCI, 2014 International Conference on, IEEE, 2014, pp. 518-523.

[7] Nisha Sharma, Bhupendra Kumar, Vandita Singh, 2014 "Recognition of Offline Hand printed English Characters, Numerals and Special Symbols", 5th International Conference, 2014, pp.640-645.

[8] R. S. V, V. R. P and M. A. V. N, "Isolated Handwritten Kannada and Tamil Numeral Recognition: A Novel Approach," 2008 First International Conference on Emerging Trends in Engineering and Technology, Nagpur, Maharashtra, 2008, pp. 1192-1195.

[9] R. Giridharan, E. K. Vellingiriraj and P. Balasubramanie, "Identification of Tamil ancient characters and information retrieval from temple epigraphy using image zoning", 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, 2016, pp. 1-7.

[10] Dash, Kalyan Sourav, Niladri B. Puhan, and Ganapati Panda 2015 "Handwritten numeral recognition using non-redundant Stockwell transform and bio-inspired optimal zoning",IET Image Processing 9, no. 10, 2015, pp.874-882

[11] G. S. G. Rajput and R. Horakeri, "Zone based handwritten Kannada character recognition using crack code and SVM," 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, 2013, pp. 1817-1821.

[12] Pirlo, Giuseppe, and Donato Impedovo 2012 "Adaptive membership functions for handwritten character recognition by voronoi-based image zoning", IEEE Transactions on Image Processing,21, no. 9, 2012, pp. 3827-3837.

[13] M. K. Mahto, K. Bhatia and R. K. Sharma, "Combined horizontal and vertical projection feature extraction technique for Gurmukhi handwritten character recognition," 2015 International Conference on Advances in Computer Engineering and Applications, Ghaziabad, 2015, pp. 59-65.

[14] R. Gopakumar, N. V. Subbareddy, K. Makkithaya and U. Dinesh Acharya, ”Zone-based structural feature extraction for script identification from Indian documents,” 2010 5th International Conference on Industrial and Information Systems, Mangalore, 2010, pp. 420-425.

[15] K. S. Dash, N. B. Puhan and G. Panda, ”A hybrid feature and discriminant classifier for high accuracy handwritten Odia numeral recognition,” 2014 IEEE REGION 10 SYMPOSIUM, Kuala Lumpur, 2014, pp. 531-535.

[16] A. Raj, ”An optical character recognition of machine printed Oriya script,” 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015, pp. 543-547.

[17] Kartar Singh Siddharth, Mahesh Jangid, Renu Dhir, Rajneesh Rani, 2011”Handwritten Gurmukhi Character Recognition Using Statistical and Background Directional Distribution Features”, International Journal on Computer Science and Engineering, Vol. 3, No. 6 June 2011, pp. 0975-3397.

[18] R. M. O. Cruz, G. D. C. Cavalcanti and T. I. Ren, ”An ensemble classifier for offline cursive character recognition using multiple feature extraction techniques,” The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, 2010, pp. 1-8.

[19] U. Pal, T. Wakabayashi and F. Kimura, ”A System for Off-Line Oriya Handwritten Character Recognition Using Curvature Feature,” 10th International Conference on Information Technology (ICIT 2007), Orissa, 2007, pp. 227-229.

[20] Shanjana C, Ajay James, 2014 ”Offline Recognition of Malayalam Handwritten Text”, 8th International Conference Interdisciplinary in Engineering, INTER-ENG 2014, 9-10 October 2014. pp. 772-779.

[21] Primekumar K.P., Idiculla S.M, ”Online Malayalam handwritten character recognition using HMM and SVM”, IEEE 2013 International Conference on Signal Processing, Image Processing and Pattern Recognition (ICSIPR).

[22] S. M. Joseph, V. A. Rahiman and K. M. Abdul Hameed, ”SVM based feature set analysis in dynamic malayalam handwritten character recognition,” 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, 2015, pp. 238-243.

[23] S. M. Obaidullah, N. Das, C. Halder and K. Roy, ”Indic script identification from handwritten document images — An unconstrained block-level approach,” 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS), Kolkata, 2015, pp. 213-218.

[24] P. K. Singh, S. K. Dalal, R. Sarkar and M. Nasipuri, ”Page-level script identification from multi-script handwritten documents,” Proceedings of the 2015 Third International Conference on Computer, Communication, Control and Information Technology (C3IT), Hooghly, 2015, pp. 1-6.

[25] A. T. Jia, Y. Ayappally and K. Syama, ”Malayalam OCR: N-gram approach using SVM classifier,” 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, 2013, pp. 1799-1803.

[26] P-ARTS Kayyezhuthu - https://tinyurl.com/ydecjnfy