A NOVEL HYBRID APPROACH FOR FEATURE EXTRACTION IN MALAYALAM HANDWRITTEN CHARACTER RECOGNITION

[1]AJAY JAMES, [2]SUJALA K, [3] CHANDRAN SARAVANAN

[1] Assistant Professor, Department of Computer Science and Engineering, Govt. Engineering College Thrissur, India

[2] Mtech Student, Department of Computer Science and Engineering, Govt. Engineering College Thrissur, India

[3] Associate Professor, Department of Computer Science and Engineering, National Institute of Technology West Bengal, India

E-mail: [1] ajayjames80@gmail.com, [2]sujala58@gmail.com, [3] dr.cs1973@gmail.com

ABSTRACT

Optical Character Recognition (OCR) is defined as the process of segregating textual scripts from a scanned document. To develop a digitally empowered society, information is made available in digital form. The OCR software assists in digitization of documents in different languages. Many researches are working on digitization of documents particularly to develop effective and error free character recognition models. To develop a digitally empowered society, information should be made digitally available. There arises the need for an OCR software in different languages. Malayalam handwritten character recognition precision is still inhibited around 90% due to the confrontations in Malayalam character set. The omnipresence of two different scripts old and new script, huge character set, ubiquity of similar shaped characters makes Malayalam handwritten character recognition more difficult. Feature extraction for each language may vary depending on various characteristics of the language. By observing the shape patterns in each language, different novel methods are developed to extract features and also to recognize the same. In this research, a novel hybrid approach is proposed which uses a combination of statistical and structural features (SSF). The statistical features are those derived from the statistical dissipating of pixels. Structural features are based on the topological and geometrical properties of the character. This study gives insight to the fact that combination of statistical and structural features gives more accuracy in Malayalam character recognition.

Keywords— *Optical Character Recognition, Binarization, Feature Extraction, Classification, Machine Recognition, Decision Tree.*

1. INTRODUCTION

Optical character recognition has many applications in day to day life. Nowadays, digitization is in high priority and machine recognition of characters is part of it. Researches in this area helps to develop a digitally empowered society. Many old documents in written or printed form are easily converted to editable text by means of digitization. OCR is a technology that enables us to convert different types of documents, such as scanned paper documents into editable and searchable data. For example, a written or printed document to be sent to our partner via email or other electronic medium by digitizing as an image other file forms. OCR helps to scan the document and store it as editable document for modifying the scanned document. In order to extract and repurpose data from scanned documents, OCR software is required that would single out letters on the image, put them into words and then words into sente`1nces, thus enabling accessing and editing the content of the original document.

Handwritten character extraction and recognition can ameliorate the human computer interaction and better integrate computers into human society. The handwritten text recognition

involves taking scanned text documents as inputs and transforming them into an editable text document format which can be stored in a computer. This transformation into digital format can be done either by online recognition or by offline recognition. The handwritten character recognition systems are broadly divided into two depending upon whether indited on touch sensitive surfaces or on a piece of paper.

- Offline
- Online

Online character recognition is the process of apperceiving handwritings, from digital surfaces like tabs, mobile phones etc. For this the pen pressure, movement of pen, direction of movement of pen are taken as attributes for character recognition. Online character recognition is much easier when compared to offline.

Offline Character Recognition is the process in which the scanned handwritten document is given as input to the system. The important steps in OCR are pre-processing, feature extraction, classification etc. [1]. Offline character recognition is complex because of the varying writing styles of individuals, age of the document, and amount of noise in the image. The different five phases in OCR are listed below:

i)   Pre-Processing

ii) Segmentation

iii) Feature Extraction

iv) Classification

v)   Post Processing

The following figure 1 shows an overview of OCR system. The first step is image acquisition in which scanned text documents are produced as inputs to this system. Then preprocessing steps are carried out like binarization [2], thinning, noise removal [3] etc. Then the image is normalized to a particular size. Segmentation steps are carried out like line, word and character segmentation [4]. Feature extraction is an important step in OCR which contributes a lot in character recognition accuracy. Feature extraction is the process of extracting the relevant features from objects or characters to form feature vectors. These feature vectors extracted from different training set images are given to the classifier. There are many different types of classifiers like

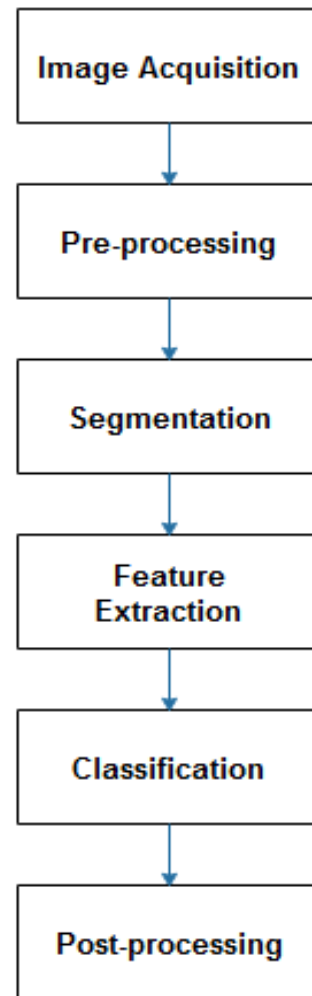SVM [5], Decision tree [6], KNN [7], Hidden Markov model [8] etc.

Fig 1. Overview Of OCR System

The selection of feature classifier pair plays an important role in the character recognition accuracy. Classifiers will take this feature vectors and their corresponding labels as input and produces a model. In testing phase, test image is pre-processed and placed for feature extraction and this feature vector along with classifier model is used to produce desirable output. The output of an OCR system can be converted to ASCII or Unicode type which is post processed facilely by the machine. It is very difficult to achieve 100% efficiency because the character

recognition systems are different for different languages.

The techniques used for feature extraction and feature selection approaches also affect the output of the OCR. The recognition rates of printed

2. MALAYALAM CHARACTERS

Malayalam is a language spoken by the people of Kerala State, India. The Malayalam script consists of a set of vowels and consonants. Malayalam language is enriched with immense number of characters and also large number of similar shaped characters. The Malayalam script consists of 15 vowels, 36 consonants and 5 pure consonants as shown in following figure 2.
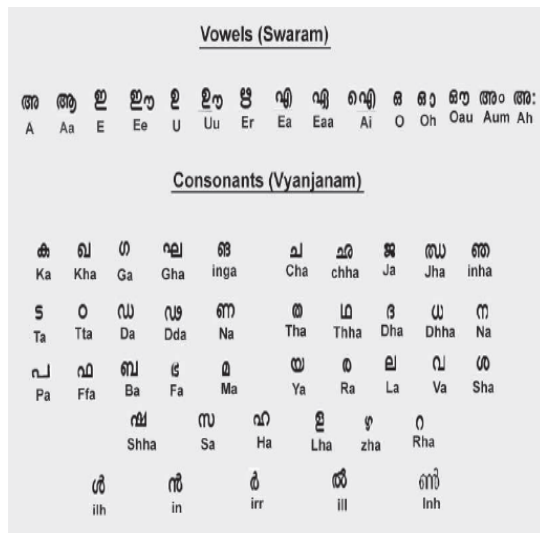

Fig 2. Malayalam Character Set

3. LITERATURE SURVEY

Feature Extraction is the process of extracting or identifying unique features from the segmented characters in such a way that it helps to recognize each and every character of the language effectively. A single feature or multiple features are extracted and combined to help in classification. Features are extracted in a way so that misclassifications between similar shaped characters are reduced. This section presents overview of different feature extraction methods followed by some of the works that have been implemented.

The feature extraction methods for handwritten character recognition are basically divided into two types: statistical and structural.

characters are above 99% because there are no style or shape variations in that. However, OCR is prone to errors when dealing with handwritten characters. Also, the recognition rates are poor in the case of handwritten character recognition. The statistical features are extracted from the statistical distributions of pixels, such as zoning [9], projection histograms [10].

Structural features are those based on the topological and geometrical properties of the characters like, end-points, intersection of lines and loops. There are also other methods like Gabor filter [11] and Fourier based approaches [12]. Assimilation of the statistical and structural features is a better elucidation to accommodate the large divergence recognized on the handwritten character images and focus different character properties.

Bindu S Moni et al. [13] suggested a feature extraction method based on extracting the count of contiguous one's in rows and columns, for the offline recognition of Handwritten Malayalam Characters. In this approach horizontal and vertical count of contiguous on pixels were used as features. The features were extracted after dividing the images into fixed sized meshes. Fixed meshes were constructed by dividing the image into N equal sized blocks. For example, a 27 X 27 image can be divided into nine 3 X 3 blocks. The proposed method quantified the transitions from 0 to 1 and 1 to 0 using the above defined feature extraction method. For classification, Modified Quadratic Discriminant function (MQDF) was implemented which was a good statistical approach for handwritten character recognition.

Anitha Mary M.O. Chacko et al. [14] proposed a multiple classifier system for the recognition of handwritten Malayalam characters. The proposed technique used gradient and density based features. At each pixel position of a character image, gradient feature points in the direction where there is a greatest change of intensity. The pixel density feature was computed as the ratio of number of foreground pixels in each zone to the total number of pixels in that zone. Here two feed forward neural networks used for classification.

M Abdul Rahiman et al. [15] Proposed an efficient algorithm for recognizing the handwritten Malayalam characters by extracting structural features like count of horizontal and vertical lines. Using this count, characters were classified into Ra type, Pa type and special type characters. But for some characters these line

count will be similar making it difficult to recognize and thus diminishing the accuracy.

Rajashekararadhya, S. V et al. [16] proposed a zone based feature distillation method for the handwritten numeral recognition in the south Indian scripts. In the proposed method the image centroid was computed first and then the image was divided into 25 zones. In each zone the distance between the pixel and image centroid was computed. The average distance from each zone was used as features.

Shanjana C et al. [17] developed a method for the offline recognition of Malayalam handwritten text. Since majority of the characters in Malayalam are curved in nature, here structural features like curvature and directional features are used as features for recognizing each character. Bi-quadratic interpolation method [18] was used for extracting curvature features. And Robert's gradient diagonal operator [19] was used for extracting the gradient feature.

Steffy Maria Joseph et al. [20] suggested a feature extraction method which used a combination of time domain features, directional and frequency domain features. Time domain features like writing direction, curvature features etc. were used. Frequency domain features include discrete cosine and Fourier transform [21].

Ashitta T Jia et al. [22] proposed a Malayalam OCR combining n-gram segmentation along with geometric feature extraction methodology. Geometric features were used to train a support vector machine in order to obtain a better recognizing accuracy. Structural features like count of loops, edges were used in feature extraction stage.

K. Khan et al. [23] suggested an Urdu character recognition system which used three different types of features like Hu moments, Zernike moments [24] and PCA [25].

The current recognition accuracy rate for Malayalam handwritten characters shows that there is a need for further improvement in a language dependent feature extraction approach. A hundred percent accurate HCR system in Malayalam is not yet available. The aim of this proposed system is to employ a feature extraction approach such that misclassifications are reduced.

The article is organized in the remaining portion as, section IV describes the proposed new character recognition method, section VI shows the experimental results, and finally, section VII concludes the research article.

4. SSF METHOD

The proposed system is developed for Malayalam handwritten character recognition. MATLAB is used to implement the proposed OCR system. Segmented Malayalam character images are given as input to the system.

Then these images are pre-processed and unique features are extracted from each image. The extracted features are used to train the classifier and to effectively recognize each character. Feature extraction and classification are the phases in consideration.

Every language has some features which are utilized for the effective recognition of the characters. Thus, to develop a character recognition system for Malayalam, the features of the Malayalam characters are observed. Also, common shapes of the Malayalam character set are classified as upward curve, downward curve, left curve and right curve. Another important shape is a closed loop. These shapes forms an important part of Malayalam characters. By making use of these shape features a hierarchical classification for Malayalam characters are proposed.

Many characters are easily identified by using these shapes. But, recognizing similar shaped characters are challenging. Thus, identifying basic common features to recognize such characters are essential. The following figures 3 and 4 shows the common curves in Malayalam characters. These curves are named as top, bottom, left and right reservoirs, because of its resemblance to a reservoir. Even though these shapes have important features, there are few other characters have similar curve, challenging to recognize few other characters.
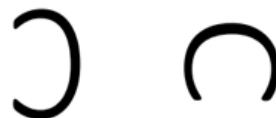
Fig 3. Left and Down Curve

Fig 4. Up And Right Curve

The following table 1 shows the curve count of some characters in Malayalam. The table 1 shows the curve features are not sufficient to recognize a character accurately. Also, other important structural features in Malayalam like count of horizontal lines, vertical lines etc., are

not sufficient for recognizing similar shaped characters. Further, these feature values are varying according to the different writing styles.

Table I. Curve Count Of Characters

Character	Curve Count			
	Up	Down	Left	Right
○		1		
ല	1	1		
വ	1	1		
�സ	1	1		
ട			1	1

Considering these varying writing styles it is evident that structural features like count of loops, horizontal lines, end points etc. may vary person to person. Thus, structural features are not adequate to recognize a Malayalam character.

Following figure 5 shows the varying styles of hand writing of few characters in Malayalam.
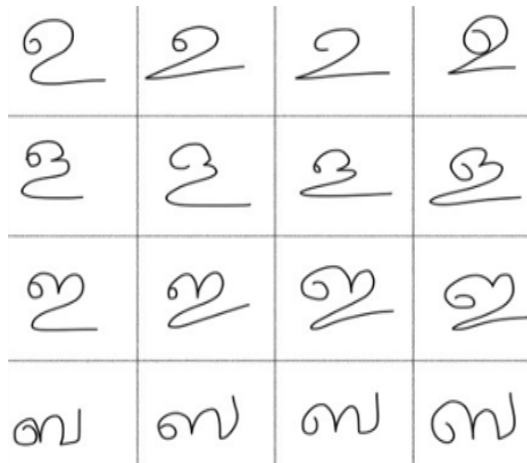

Fig 5. Varying Writing Styles Of Different Characters

The reason of misclassifications of characters in Malayalam is mainly due to the varying hand writing styles. Thus, it is noticed that recognizing a Malayalam character with the help of structural features is challenging. Thus, a novel hybrid

approach combining statistical and structural features [SSF] will recognize Malayalam character accurately. Thus, a novel hybrid approach extracting 12 different features from each character which accounts for a feature vector of size 50 is developed.

A total of 12 features are proposed in this approach. The proposed features are:-

1. Count of number of horizontal lines
2. Count of number of vertical lines
3. Number of endpoints
4. Number of top, bottom, left, right curves
5. Simple zoning
6. Region based zoning horizontally
7. Region based zoning vertically
8. Region based zoning diagonally
9. Region based zoning anti-diagonally
10. Extended diagonal based zoning
11. Count of contiguous black pixels in all rows
12. Count of contiguous black pixels in all columns

Classification phase is another important part of this work. The Malayalam handwritten character recognition system proposed here uses Decision tree for classification. Training images are used to train the machine in order to create a model. Such a model will be used to classify the unlabeled test image. Feature vectors are extracted for each of the training images. This feature vector is stored in a matrix. Feature matrix along with manually created labels is used to construct a training model. The proposed system architecture is shown in figure 6.

The proposed system consists of four vital stages.

1. Pre-processing
2. Feature Extraction
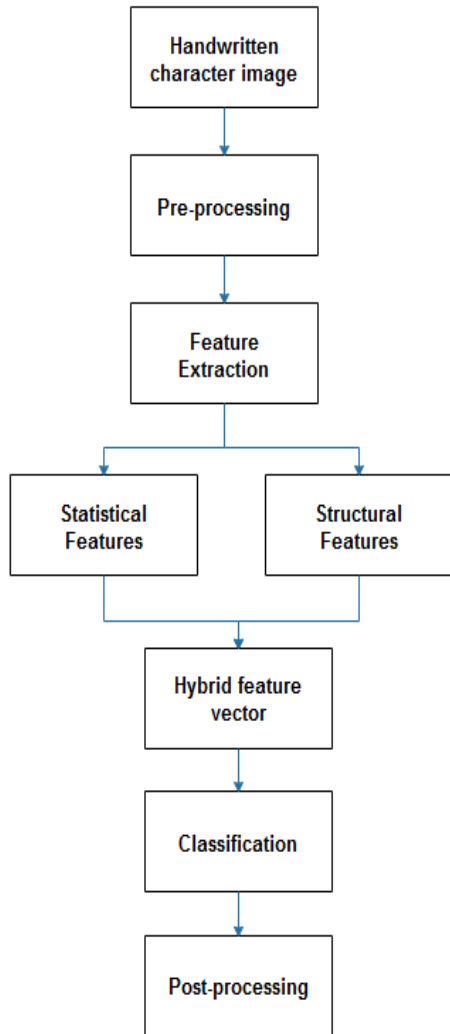3. Classification
4. Post processing

Fig 6. Proposed SSF System

4.1 Pre-processing

In the pre-processing stage, the character image is first binarized by means of Otsu's method [26]. To remove the unwanted excess white areas, the image is cropped. Noise is removed using median filtering. Thinning [27] operation is applied to get the skeletonized image. Following figure 7 shows three pre-processed images applied to a Malayalam character image.
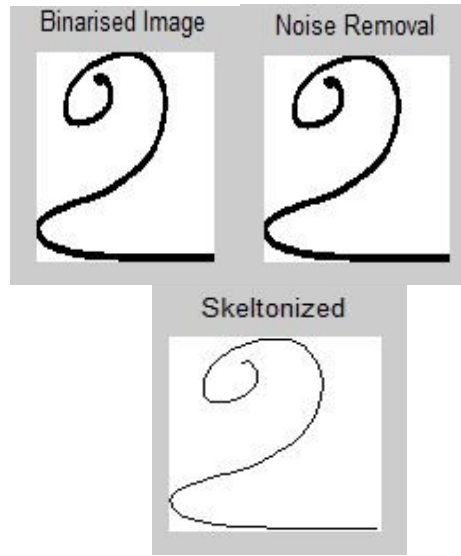

Fig 7. Pre-processing Steps

4.2. Feature Extraction

The image is first normalized into 100x100 pixels to attain uniformity. From each segmented character image the above described 12 features are extracted. A feature vector of size 50 is collected from each character image and stored in an array. Number of H lines and V lines of few Malayalam characters are shown in the following table 2.

Table II. Number of H And V Lines

Character	Number of H lines	Number of V lines
	2	3
	2	3
	2	1
	1	3
	3	4

4.2.1   No.of lines in Horizontal and Vertical

Search the whole image and set a threshold t. If there are t consecutive points in horizontal, that is a horizontal line. If there are t consecutive points in vertical, that is a vertical line.

4.2.2 Number of top, bottom, left and right curves

The count of up, bottom, left and right curves formed in the character images can be used as the curve features.

For up curve:-

- Scan image horizontally from top
- If a black pixel is encountered then find the next black pixel in that row.
- If distance between 2 pixels is greater than a threshold, find midpoint and perform vertical scan along midpoint.
- If black pixel is found in downward direction and no black in upward, then it's an upward curve.

Similarly scanning horizontally from bottom lead to down curve and left scan and right scan leads to corresponding curves.
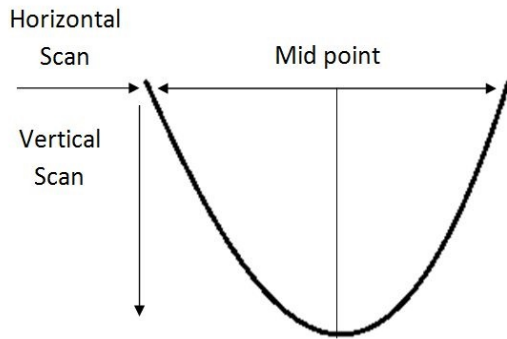

Fig 8. Upward Curve Detection

4.2.3 Number of end points

Scan each pixel of thin image and its surrounding 8 pixel. If there is only one on pixel in 8 pixel surrounding count it as an endpoint.

4.2.4 Region based zoning horizontally

- Divide the image horizontally into two halves
- Count the number of black pixels in corresponding zones.

4.2.5 Region based zoning vertically

- Divide the image vertically into two halves
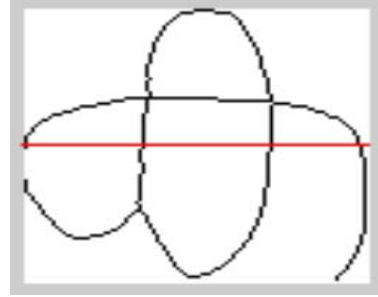- Count the number of black pixels in the two zones.
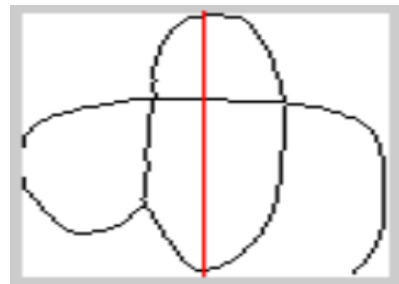

Fig 9. Horizontal Zoning

Fig 10.Vertical Zoning

4.2.6 Region based zoning diagonally

- Divide the image diagonally into two halves
- Count the number of black pixels in the respective zones.

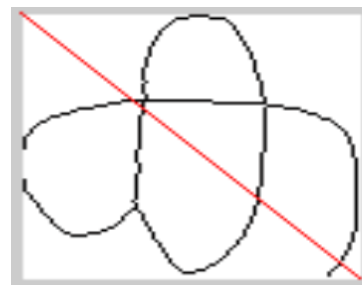Diagonal zoning implementation in a character image is shown in Figure 11.

Fig 11. Diagonal Zoning

4.2.7 Region based zoning anti-diagonally

- Divide the image anti-diagonally into two halves4
- Count the number of black pixels in the respective zones.

4.2.8 Extended diagonal based zoning

Number of black pixels in the region above and below the diagonals in the four zones thus forming eight features as shown in Figure 12.
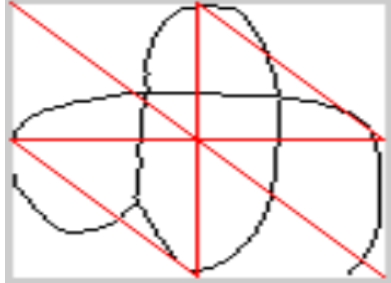

Fig 12. Extended Diagonal Based Zoning

4.2.9   Number of black pixels in each zone by dividing the   image into 5x5 zones.
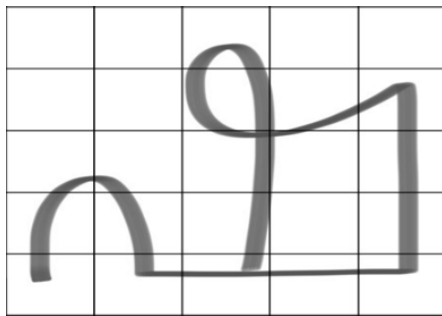

Fig 13. Simple Zoning

4.2.10   Count of contiguous on pixels in all rows and columns by scanning the image horizontally and vertically.

4.3. Classification

In this system Decision tree classifier is used for classification. It follows a top-down approach. Each internal node represents a test using a splitting attribute. Each leaf node represents the character label. Splitting attributes are chosen using different methods like information gain, Gini index etc. Existing different decision tree algorithms are CART, ID3, and C4.5 [28] etc. and in this SSF system CART algorithm is used.

The test images will first pre-process the input image and collect the feature vector. This feature vector along with the classifier model that created by the training module are used to classify the unlabeled test data. Unlabeled test data are recognized by the SSF model.

4.4.   Post Processing

In the post processing stage of the proposed system the classifier output is mapped to the character Unicode. The outputs of the classifier are integer labels. This integer label is converted into corresponding character Unicode. The Unicode is written in a text file.

5 DATASET

In this research work a dataset "P-ARTS KAYYEZHUTHU   MALAYALAM HANDWRITTEN DATASET" [29] is used, which contains 2000 samples per character of 56 Malayalam characters. Character samples are collected from people belonging to different age groups. 80% of the samples are used for training and remaining for testing.

6 EXPERIMENTAL RESULTS

The proposed system is implemented using MATLAB R2014b. The experiment is conducted on 56 selected characters of Malayalam which includes 15 vowels, 36 consonants and 5 pure consonants. The feature set consists of a combination of statistical and structural features. So statistical features alone are first passed to the classifier and compared the accuracy. Then structural features followed by hybrid feature set are separately passed to the classifier and compared the accuracy. The following table 3 shows the feature vector size and the accuracy obtained in the proposed SSF approach.

Table III. Feature Vector Size and Accuracy of SSF

Feature	Feature Size	Accuracy
Structural	9	70.37%
Statistical	41	92.83%

Hybrid	50	97.87%

Results shows that the hybrid feature vector is producing more accurate results compared to Structural and Statistical techniques. An evaluation of accuracy in three feature sets is shown in following figure 14.

The proposed system also resulted in some misclassifications. And they are shown in figure 15.
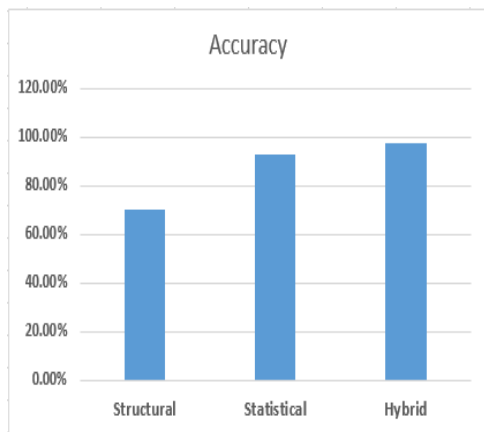

Fig 14. Evaluation Of Accuracy Of Different Feature Sets

The proposed hybrid feature vector also played a major role in reducing misclassifications among similar shaped characters. Also no misclassifications were shown in the pure consonants and dependent vowels section of the character set. Correctly recognized similar shaped characters are shown in figure 16.

A comparison of proposed SSF system with other different feature extraction approaches existing in Malayalam specified in the literature survey is shown in following table IV. The comparison reveals that most of the papers are using structural features resulting in a poor accuracy. The proposed SSF system with hybrid feature set is producing better accuracy.

Fig 16. Correctly Recognized Similar Shaped Characters

The above comparison reveals that different datasets is used in different approaches. The dataset used in these approaches are comparatively small. The proposed system is tested with a huge dataset including 56 letters of the Malayalam character set. Also, the results show that the proposed hybrid approach provides better accuracy than other approaches.

The existing works in Malayalam character recognition uses either statistical features alone or structural features alone. The table IV shows in detail the feature extraction methods used in prior works.

Table IV. Comparison of Proposed System With Other Systems

Paper	Feature	Accuracy	Dataset

Fig 15. Misclassifications

Proposed SSF System	Hybrid (Statistical and Structural)	97.87%	2000 samples per character of 56 characters in Malayalam
Modified Quadratic Classifier for Handwritten Malayalam Character Recognition [2]	Statistical	89.18%	500 samples each of the 30 selected character
Multiple Classifier System for Malayalam Character Recognition[3]	Structural	75.15%	33 selected characters, total 825 character samples
Isolated Handwritten Malayalam Character Recognition using HLH Intensity [4]	Structural	88.6%	A total of 661 character samples of selected 44 characters
Offline recognition of Malayalam handwritten text [5]	Structural	82%	122 samples of each of 53 Malayalam character
Malayalam OCR N-gram approach Using SVM Classifier[6]	Structural	95.6%	56 Malayalam alphabets, total of 7809 samples

7. CONCLUSION

Existing handwritten recognition systems in Malayalam always resulted in misclassifications among similar shaped characters. The proposed system implements a handwritten Malayalam character recognition system for 15 vowels, 36 consonants and 5 pure consonants. This hybrid approach (SSF Method) minimizes the chances of misclassifications between similar shaped characters by using a combination of statistical and structural features. Structural features include curve features in Malayalam. Curve features provide an important shape feature in Malayalam. Statistical features like zoning and region based zoning are used to improve the accuracy and reduce the misclassifications. The proposed

model obtained better results than existing character recognition systems in Malayalam (around 97%).

The proposed system also resulted in some misclassifications. And they are shown in Figure 14. Even though my aim was to build an OCR system without misclassifications, this system resulted in minor misclassifications among some characters. The main reason for this is the varying writing styles of individuals. This shows that if more training samples are used we can improve the accuracy. Thus this research shows the importance of dataset in character recognition accuracy. In this study around 2000 samples per character is used for comparing recognition accuracy. The dataset used in this research is much larger compared to the existing works in Malayalam character recognition. This infers that higher the number of character samples for training, higher the accuracy achieved.

This research also highlights one of the important limitations in Malayalam character recognition. Absence of a standard dataset is the main problem in this area of research. If there exists a standard dataset, we can compare the existing approaches with the new ones. In order to solve this problem we are sharing the dataset [29] used in this work online. So that researchers can use this to compare their works.

REFERENCES:

[1] Impedovo, S., L. Ottaviano, and S. Occhinegro. "Optical character recognition—a survey." *International Journal of Pattern Recognition and Artificial Intelligence* 5.01n02 (1991): 1-24.

[2] Karthika, M., and Ajay James. "A Novel Approach for Document Image Binarization Using Bit-plane Slicing." *Procedia Technology* 19 (2015): 758-765.

[3] Alginahi, Yasser. Preprocessing techniques in character recognition. *INTECH Open Access Publisher*, 2010.

[4] Pal, Umapada, Ramachandran Jayadevan, and Nabin Sharma. "Handwriting recognition in indian regional scripts: a survey of offline techniques." *ACM Transactions on Asian Language Information Processing* (TALIP) 11.1 (2012): 1.

[5] Shanthi, N., and K. Duraiswamy. "A novel SVM-based handwritten Tamil character recognition system." *Pattern Analysis and Applications 13.2 (2010)*: 173-180.

[6] Safavian, S. Rasoul, and David Landgrebe. "A survey of decision tree classifier methodology." *IEEE transactions on systems, man, and cybernetics 21.3* (1991): 660-674.

[7] Lee, Yuchun. "Handwritten digit recognition using k nearest-neighbor, radial-basis function, and backpropagation neural networks."*Neural computation3.3 (1991)*: 440-449.

[8] Agazzi, Oscar E., and Shyh-shiaw Kuo. "Hidden Markov model based optical character recognition in the presence of deterministic transformations."*Pattern recognition 26.12 (1993)*: 1813-1826.

[9] Trier, Øivind Due, Anil K. Jain, and Torfinn Taxt. "Feature extraction methods for character recognition-a survey." *Pattern recognition 29.4* (1996): 641-662.

[10] Chacko A. M, Dhanya P. M, "A comparative study of different feature extraction techniques for offine Malayalam character recognition", *In Computational Intelligence in Data Mining, Volume* 2(pp. 9-18), Springer India, 2015

[11] Wang, Xuewen, Xiaoqing Ding, and Changsong Liu. "Gabor filters-based feature extraction for character recognition." Pattern recognition 38.3 (2005): 369-379.

[12] Hong, Zi-Quan. "Algebraic feature extraction of image for recognition." *Pattern recognition 24.3 (1991*): 211-219.

[13] Moni, Bindu S., and G. Raju, "Modified quadratic classifier for handwritten Malayalam character recognition using run length count", *Emerging Trends in Electrical and Computer Technology (ICETECT*), (pp. 600-604), 2011.

[14] Chacko, Anitha Mary MO, and P. M. Dhanya. "Multiple classifier system for offline malayalam character recognition." *Procedia Computer Science 46: 86-92,*(2015).

[15] Rahiman, M. Abdul, et al. "Isolated handwritten Malayalam character recognition using HLH intensity patterns", *Machine Learning and Computing (ICMLC), 2010 Second International Conference on IEEE*, 2010.

[16] Rajashekararadhya, S. V., and P. Vanaja Ranjan. "The Zone-Based Projection Distance Feature Extraction Method for Handwritten Numeral/Mixed Numerals Recognition of Indian Scripts." *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on. IEEE*, 2010.

[17] Shanjana, C., and Ajay James. "Character segmentation in Malayalam Handwritten documents" *,Advances in Engineering and Technology Research (ICAETR),* 2014 International Conference on IEEE, 2014.

[18] Pal, Umapada, Tetsushi Wakabayashi, and Fumitaka Kimura. "A system for off-line Oriya handwritten character recognition using curvature feature."*Information Technology,(ICIT 2007). 10th International Conference on. IEEE*, 2007.

[19] Haralick, Robert M. "Statistical and structural approaches to texture."*Proceedings of the IEEE 67.5 (1979):* 786-804.

[20] Joseph, Steffy Maria, V. Abdu Rahiman, and KM Abdul Hameed. "SVM based feature set analysis in dynamic malayalam handwritten character recognition." *Signal and Image Processing Applications (ICSIPA), 2015 IEEE International Conference on. IEEE*, 2015.

[21] Parisi, R., et al. "Car plate recognition by neural networks and image processing." Circuits and Systems, 1998. ISCAS'98. *Proceedings of the 1998 IEEE International Symposium on. Vol. 3. IEEE*, 1998.

[22] Parisi, R., et al. "Car plate recognition by neural networks and image processing." *Circuits and Systems, 1998. ISCAS'98. Proceedings of the 1998 IEEE International Symposium on. Vol. 3. IEEE*, 1998.

[23] Khan, K., et al. "Urdu text classification using decision trees." *High-Capacity Optical Networks and Enabling/Emerging Technologies (HONET), 2015 12th International Conference on. IEEE*, 2015.

[24] Khotanzad, A. L. I. R. E. Z. A., and Yaw Hua Hong. "Rotation invariant pattern recognition using Zernike moments." Pattern Recognition, 1988., *9th International Conference on. IEEE*, 1988.

[25] Deepu, V., Sriganesh Madhvanath, and A. G. Ramakrishnan. "Principal component analysis for online handwritten character recognition." *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. Vol. 2. IEEE*, 2004.

[26] Nina, Oliver, Bryan Morse, and William Barrett. "A recursive Otsu thresholding method for scanned document binarization." *Applications of Computer Vision (WACV), 2011 IEEE Workshop on. IEEE*, 2011.

[27] Parker, Jim R. *Algorithms for image processing and computer vision*. John Wiley & Sons, 2010.

[28] Javidi, Bahram. *Image recognition and classification: algorithms, systems, and applications*. CRC Press, 2002.

[29] https://drive.google.com/open?id=0B1eLyjU euERZWUVIOU9OZm40RHc.