# MALAYALAM RECOGNIZER: A LEARNING TO WRITE COLLABORATOR

[1]V. P. Deepa , [2]T.Navya, [3]Roopa Sree Mohan, [4]Soumya Muraleedhara Menon
[1]Student, Sreepathy Institute of Management And Technology / APJ Abdul Kalam Technological University, Koottanad, Palakad, Kerala, vpdeepa75@gmail.com
[2]Student, Sreepathy Institute of Management And Technology / APJ Abdul Kalam Technological University, Koottanad, Palakad, Kerala, navyat2015@gmail.com
[3]Student, Sreepathy Institute of Management And Technology / APJ Abdul Kalam Technological University, Koottanad, Palakad, Kerala, roopachandrasree@gmail.com
[4]Student, Sreepathy Institute of Management And Technology / APJ Abdul Kalam Technological University, Koottanad, Palakad, Kerala, soumyamuraleedharamenon@gmail.com

**Abstract**
**Handwriting recognition is an area under machine learning. Malayalam (Keralite language) gesture recognition is a challenging process because of alphabet written in different ways which is more complex to write among Indian languages and the recognition task is quite difficult due to wide intra-personal and inter-personal variation in human handwriting. Also recognition task on Malayalam language become multiplex since there are large number of classes with high similarities. Previous efforts of making Malayalam recognition more accessible have been through the inclusion of gesture recognizers through image processing. In this work, we propose a new model for handwriting gesture recognition in real time. The input of this model is a Malayalam alphabet. The aim of handwriting is to identify input gesture correctly then analysed to many process. This Malayalam recognizer averts image processing and make use of a real time recognition technology. Nowadays this technology has more relevance in devices like mobile phones, for giving input by hand and does the recognition process on writing itself. Since development of learning to write is a sophisticated procedure in Malayalam, so we introduce our project as an application which breaks this berg. To enable novice programmers to incorporate gestures into their UI prototypes, we present a "Malayalam letter recognizer" that is easy, cheap, and usable, in which the user found to make interactions through a built in canvas.**
**Keywords: $1 recognizer, unistroke, gesture, indicative angle, optimal cosine distance.**

I.

## II. INTRODUCTION

Malayalam is one of the 22 official languages and 14 regional languages of India. It is spoken by 38 million people primarily in the state of Kerala and in the Lakshadweep Islands in southern India. The Malayalam script, known as kolezhuthu (Rod-Script), is derived from the ancient Grandha script. The language includes 53 characters with 37 consonants and 16 long and short vowels. However, a new style of writing was introduced in 1981, which helped reduce the number of characters radically. As with many other world languages, Malayalam borrows some of its vocabulary from other languages. Its vocabulary has several words borrowed from Sanskrit, English and Portuguese.

Pen, finger, and wand gestures are increasingly relevant to many new user interfaces for mobile, tablet, large display, and tabletop computers. Even some desktop applications support mouse gestures. The Opera Web Browser, for example, uses mouse gestures to navigate and manage windows. As new computing platforms and new user interface concepts are explored, the opportunity for using gestures made by pens, fingers, wands, or other path-making instruments is likely to grow, and

with it, interest from user interface designers and rapid prototypers in using gestures in their projects.$1 recognizer that is easy, cheap, and usable almost anywhere. The recognizer is very simple, involving only basic geometry and trigonometry. It requires about 100 lines of code for both gesture definition and recognition. It supports configurable rotation, scale, and position invariance, does not require feature selection or training examples. Although $1 has limitations as a result of its simplicity, it offers excellent recognition rates for the types of symbols and strokes that can be useful in user interfaces.

The real time or dynamic has been used in place of online. Online handwriting recognition requires a transducer that captures the writing as it is written. The most common of these devices is the electronic tablet or digitizer.

The various approaches for handwritten character recognition are string machine matching schemes, structural approach, template matching, using neural networks, etc. The central objective is demonstrating how Malayalam characters are recognized by using Artificial Neural Networks. Such networks can be fed the data from graphic analysis of the input data. And also can be trained to output characters on one or another form. Multi-layer Perception model is one such network. It uses Delta learning rule for adjusting weights. It will force the output to one of nearby values if a variation of input is fed into the network.

Optical Character Recognition plays an important role in Digital Image Processing and Pattern Recognition. Even though ambient study had been performed on foreign languages like Chinese and Japanese, effort on Indian script is still immature. OCR in Malayalam language is more complex as it is enriched with largest number of characters among all Indian languages. The challenge of recognition of characters is even high in handwritten domain, due to the varying writing style of each individual. This method uses Chain code and Image Centroid for the purpose of extracting features and a two layer feed forward network with scaled conjugate gradient for classification.

Content Based Image Retrieval is one of the prominent areas in Computer Vision and Image Processing. Recognition of handwritten characters has been a popular area of research for many years and still remains an open problem. This uses visual image queries for retrieving similar images from database of Malayalam handwritten characters. Local Binary Pattern (LBP) descriptors of the query images are extracted and those features are compared with the features of the images in database for retrieving desired characters.

Apart from these, our system allows only characters to be drawn by unistroke, as $1 recognizer. How well does $1 perform on user interface gestures compared to two more complex algorithms used in HCI? How does recognition improve as the number of templates or training examples increases? How do gesture articulation speeds affect recognition? How do recognizers scores degrade as when moved down their N best lists? Which gestures do users prefer? These are answered in this.

Character recognition is a fundamental, but most challenging in the field of pattern recognition with large number of useful applications. The technique by which a computer system can recognize characters and other symbols written by hand in natural handwriting is called handwriting recognition system.

Handwriting recognition is classified into offline handwriting recognition and online handwriting recognition. If handwriting is scanned and then understood by the computer, it is called offline handwriting recognition. In case, the handwriting is recognized while writing through touch pad using stylus pen, it is called online handwriting recognition. Here we are concentrate more on online recognizers. On-line handwriting recognition requires a transducer that captures the writing as it is written. The most common of these devices is the electronic tablet or digitizer [1].

Handwritten recognition is divided into five phases, which are pre-processing, segmentation, feature extraction, classification and post processing [2].

An intelligent system for free hand entry of characters and words using light pen model is described. The developed system recognize the characters and words. The various approaches for handwritten character recognition are string machine matching schemes, structural approach, template matching, using neural networks, etc. The central objective is demonstrating how Malayalam characters are recognized by using Artificial Neural Networks. Network employs

learning rules to update the weights between the nodes. Such networks can be fed the data from graphic analysis of the input data. And also can be trained to output characters on one or another form. Multi-layer Perception model is one such network. It uses Delta learning rule for adjusting weights. It will force the output to one of nearby values if a variation of input is fed into the network. The word is finally recognized by checking the database trained for, and the proximity issue is solved [3].

Optical Character Recognition plays an important role in Digital Image Processing and Pattern Recognition. Even though ambient study had been performed on foreign languages like Chinese and Japanese, effort on Indian script is still immature. OCR in Malayalam language is more complex as it is enriched with largest number of characters among all Indian languages. The challenge of recognition of characters is even high in handwritten domain, due to the varying writing style of each individual. Here the proposed method uses Chain code and Image Centroid for the purpose of extracting features and a two layer feed forward network with scaled conjugate gradient for classification [4].

Content Based Image Retrieval is one of the prominent areas in Computer Vision and Image Processing. Recognition of handwritten characters has been a popular area of research for many years and still remains an open problem. The proposed system uses visual image queries for retrieving similar images from database of Malayalam handwritten characters. Local Binary Pattern (LBP) descriptors of the query images are extracted and those features are compared with the features of the images in database for retrieving desired characters. This system with local binary pattern gives excellent retrieval performance [5].

## 2. OBJECTIVE

The world is turning to be a digitalized one. Nowadays, no one depends on any kind of books for some reference. So here, we are implementing a web application where anyone can easily learn to read and write Malayalam language. The application is simple and user-friendly. With each of the Malayalam alphabets, an audio button is provided where the learner can click to hear the pronunciation. There will be a canvas provided where the learner can write on it using unistroke and then the corresponding alphabet will be identified.

## 3. PROJECT AREA

The area of the project is Machine Learning. Machine Learning is the field of study that provides the system the ability to learn automatically and improve from experience without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available.

Machine learning algorithms are often categorized as supervised or unsupervised. Supervised algorithms require a data scientist or data analyst with machine learning skills to provide both input and desired output, in addition to furnishing feedback about the accuracy of predictions during algorithm training. Data scientists determine which variables, or features, the model should analyze and use to develop predictions. Once training is complete, the algorithm will apply what was learned to new data. Unsupervised algorithms do not need to be trained with desired outcome data. Instead, they use an iterative approach called deep learning to review data and arrive at the conclusion.

## 4. SCOPE AND APPLICATIONS

A Malayalam learning platform for those who are likely to study the most tedious Malayalam language. Makes good user interface. Further, will move onto an Android app as a Malayalam learning path on your fingertip. As Malayalam is a Dravidian language and it has different scripting and style of writing ,it is very difficult to have a knowledge of writing. So for those who are in need to learn how to write each characters in Malayalam can use this as it includes user friendly nature as well as the world is moving more onto digitized, books will be a so called story later.

## 5. PROBLEM STATEMENT

To develop a gesture recognition system. The application of system comes in different areas like learning, deaf people interface, etc. Here we elaborate the recognition system for Malayalam learning, which will recognize the Malayalam alphabets written on a canvas. Our primary objective in solving this problem is to have a minimal set of training data in order to

quickly build a prototype system. Many number of coordinate points are taken and stored in the training set. When the alphabet is written on the canvas, it is compared with the corresponding points that are already stored.

## 6. IMPLEMENTATION

Consists of 2 modules:

- Learning Phase
- Recognition Phase

System architecture of Malayalam recognizer is shown in Fig. 6.1.

- ***Learning Phase***

Learning Phase is the first module and it is divided into various sub-modules, such as:

- Learn to read.
  - ➢ Virtual Keyboard.
  - ➢ Tune-in.
- Learn to write.
  - ➢ Scribble.

 Virtual Keyboard

This portraits the effect of having a virtual keyboard. This keyboard consist of Malayalam letters as keys. Malayalam letters which includes both 24 Swarakasharam and Vyanjhanaksharam. These letters are placed as buttons as found in our normal Keyboard.

 Tune-in

When buttons or keys of virtual keyboard are pressed an audio is generated such that it sounds the respective letter. In short, this sub-module pronounces each letter when pressed.

 Scribble

As the name indicates, user can scribble over the letter as much as he/she need to. After clicking the keys in virtual keyboard, a gif image, scribbling area, audio icon and a refresh button. Gif shows the user how to draw the particular letter, and at the same time the user can scribble it over the image displayed next to the gif. Also the image contains the path to draw. If he/she forgets how to pronounce the letter, then user can click on the audio icon displayed above the scribbling area. And also provided with refresh button, to refresh the page when scribbled roughly and to try again until the user is ready to claim.

- ***Testing Phase***

Here, in this module, a canvas is built with default size. User can draw gestures (alphabet) over that canvas using single stroke only and provides the result on the top of the canvas. And if the user needs to try again, he /she can.

❑ **Input**

There is a built in canvas in to which the gestures are drawn. User can give the gestures using mouse. When the user draw a gesture on the canvas ,the system detects the gesture and givens for next stages of recognition procedure.

❑ **Resample the Point Path**

To make gesture paths directly comparable even at different movement speeds, we first resample gestures such that the path defined by their original M points is defined by N equidistantly spaced points. To resample, we first calculate the total length of the M- point path. Dividing this length by (N1) gives the length of each increment, I, between N new points. Then the path is stepped through such that when the distance covered exceeds I, a new point is added through linear interpolation.

❑ **Rotate Once Based on the Indicative Angle**

The indicative angle is the angle formed between the centroid of the gesture (x,y) and the gestures first point. After finding the indicative angle we rotate the gesture so that this angle is at 0.

❑ **Scale and Translate**

After rotation, the gesture is scaled to a reference square. By scaling to a square, we are scaling non- uniformly. This will allow us to rotate the candidate about its centroid and safely assume that changes in pair-wise point-distances between C and T*i* are due only to rotation, not to aspect ratio. After scaling, the gesture is translated to a reference point. For simplicity, we choose to translate the gesture so that its centroid (x,y) is at (0,0).

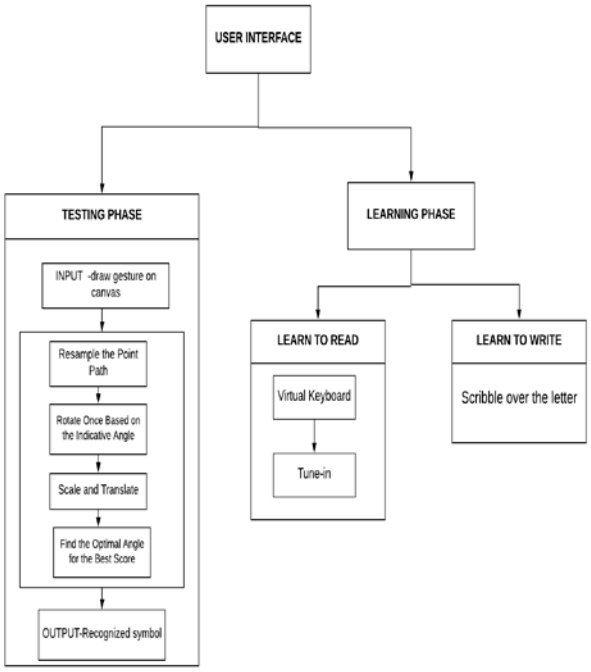



Fig. 6.1: System Architecture of Malayalam Recognizer.

- **Find the Optimal Angle for the Best Score**

A candidate C is compared to each stored template T i to find the average distance $d_i$ between corresponding points:

$$d_i = \sum_{k=1}^{n} \sqrt{(C[k]_x - T_i[k]_x)x^2 + (C[k]_y - T_i[k]_y)x^2}$$

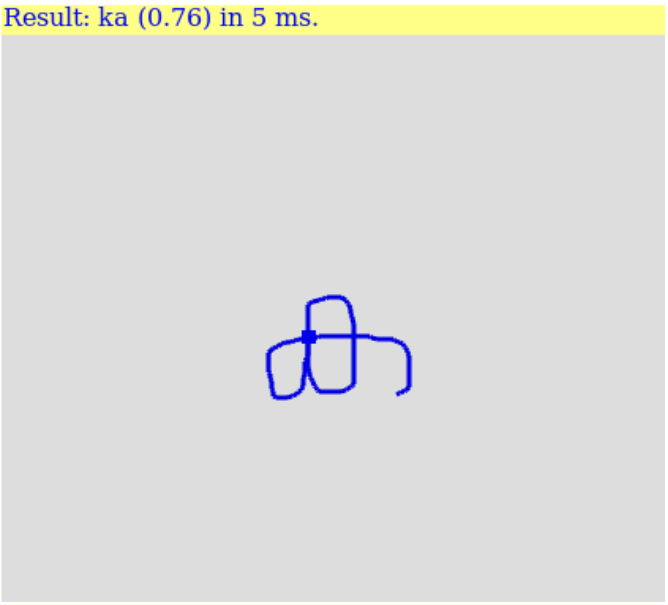
the path-distance between C and $T_i$ . The template $T_i$ with the least path-distance to C is the result of the recognition. This minimum path-distance $d_i$* is converted to a [0..1] score using:

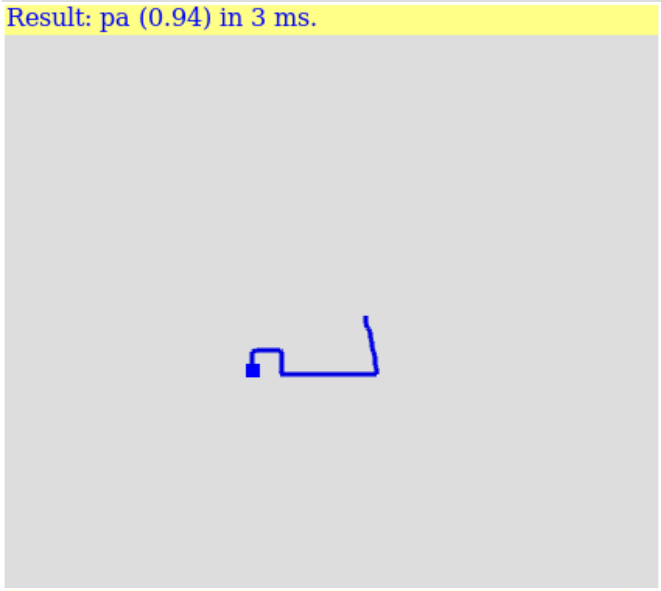
$$score = 1 - \frac{d_i}{\frac{1}{2}\sqrt{x^2} + x^2}$$

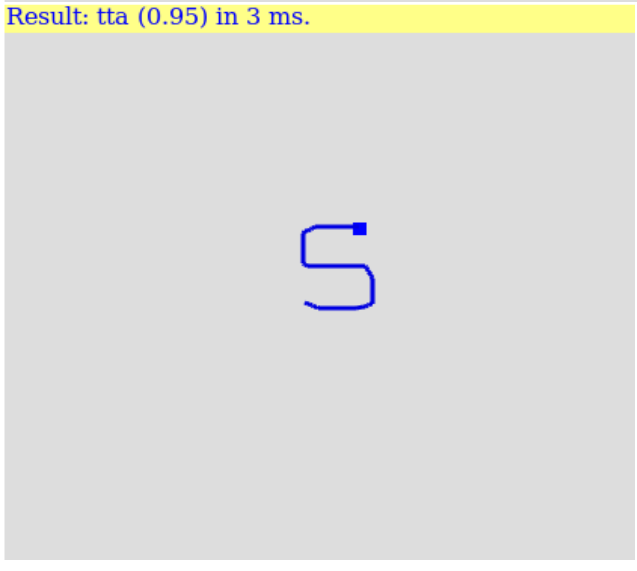
When comparing C to each $T_i$, the result of each comparison must be made using the best angular alignment of C and $T_i$. Rotating C and $T_i$ once using their indicative angles only approximated their best angular alignment. The exact recognition is performed in this stage and finally we derive the result which is the gesture inputted by the user.

## 7. RESULTS

For our problem, the solution deals with geometric and trigonometric functions throughout the procedure like distance calculation, score value calculations, etc.Some recognized letters are shown in Fig.4.1.



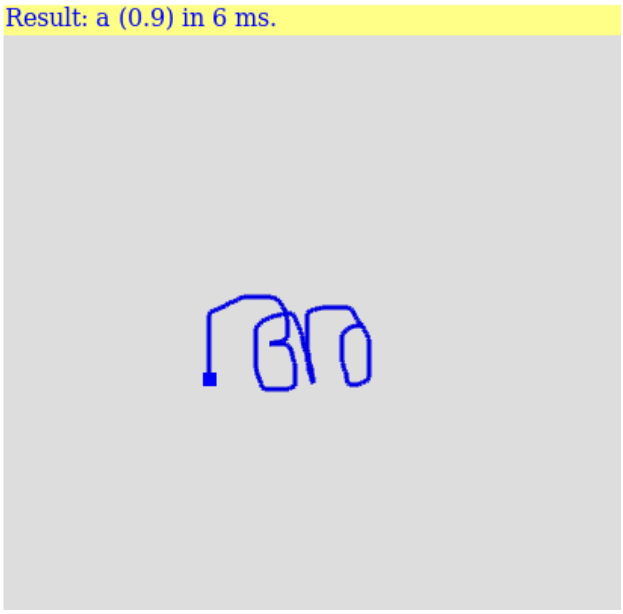



Fig.7.1: Recognized image on canvas of “Ka”, “pa”, “tta” and “ a” respectively.

## 8. CONCLUSION

Used a $1 recognizer that is easy, cheap, and usable as it includes only 100 lines of code. Involves only basic geometry and trigonometry. Offers excellent recognition rates for the types of symbols and strokes that can be useful in user interfaces. If any new features are required to involve then will move onto more tools and frameworks. And finally, to Android app using Android Studio.

## 9. REFERENCES


1. Charles C. Tappert, Ching Y. Suen, Toru Wakahara, "*The state of the art in online handwriting recognition*" , IEEE Trans. Pattern Analysis & Machine Int. 12 (8), 787-808, 1990.
2. Ayush Purohit , Shardul Singh Chauhan, ”*A Literature Survey on Handwritten Character Recognition*”, International Journal of Computer Science and Information Technologies, Vol. 7 (1) , 2016.
3. Manoj Kumar, Sandeep Chandran, ”*Handwritten Malayalam Word Recognition System using Neural Networks*”, International Journal Of Engineering Research and Technology, Volume 4, Issue 05, April-2015.
4. Jomy John, Pramod K. V, Kannan Balakrishnan, “*Offline Handwritten Malayalam Character Recognition Based on Chain Code Histogram*”, 2011
5. Simily Joseph, Jomy John, Kannan Balakrishnan, K. Pramod Vijayaraghavan, “*Content Based Image Retrieval System for Malayalam Handwritten Characters*”, 2011.