Handwritten Marathi Character Recognition on an Android Device

Tanvi Zunjarrao[1], Uday Joshi[2]

[1]MTech Student, Computer Engineering, KJ Somaiya College of Engineering,Vidyavihar,India
[2]Associate Professor, Somaiya College of Engineering,Vidyavihar,India

---***---

Abstract - *Android, a Linux based Operating System's popularity has been rising over the past few years. There are more than 60000 premium application built over this platform which meets user requirement. We focus on developing handwriting recognition system on Android platform that is acceptably fast, having a easy user-friendly interface, able to pre-process the given input to suppress the background, detect text regions present in the image. We present our work using OCR on Indian scripts, mainly on Devanagari, the most popular script in India. Recognition of Character is done using Nearest Neighbour, sampling techniques and geometric transformation. There are many research paper presented a new perspective which improve the current system along with experimental results. The proper selection of the technique is most probable for achieving high performance with a better accuracy.*

Key Words: **Devanagari, Gesture Library, Offline Character Recognition, Online Character Recognition,**

1. INTRODUCTION

In this paper, it is explored the efficiency of various stroke based handwriting analysis strategies in classifying Devanagari handwritten characters by using a template-based approach. Writing units are variable from time to time, even within the drawings of a specific character from the same user. Writing units include the properties of stroke such as, number, shape and size, order and writing speed. It is proposed to use structural properties of writing samples having such variability in writing units. This work employs Nearest Neighbour Algorithm to estimate the similarity.

The purpose of this project is to take Online handwritten Marathi characters as input, process the character, train using an algorithm, to recognize the pattern and modify the character to digital form. The project is aimed at developing an application which will be helpful in recognizing characters of different languages. One of the major applications is, it should be useful to the peoples having poor eyesight. During design we also wish to provide gestures in our application that will be appropriate for a blind user. Although blind people may use the same hardware as their sighted peers, it is possible that they will prefer to use different gestures.

Fig-1: Some printed Marathi Characters

2. TYPE OF CHARACTER RECOGNITION

Online and Offline character recognition entail different modes of input, representation, processing and recognition strategies.

2.1. Offline Character Recognition [1]:

In offline approach the image is scanned and loaded in the frame buffer. The writing is usually captured optically by a scanner and the writing is made available as an image. It is also called as static approach. Offline character recognition become more challenging due to shape of characters, great variation of character symbol and document quality.

2.2. Online Character Recognition [2]:

Online handwriting recognition is referred to as real-time recognition because here the characters are recognized as they are written. It refers to the recognition mode in which the machine recognizes the handwriting while the user writes on the surface of a digitizing tablet with finger tip. It is also called as dynamic approach. Online handwriting recognition is known to be based on trajectory and is represented as function of time. Features such as stroke pressure, velocity, trajectory are considered.

3. CHARACTERSTICS OF HANDWRITING SYSTEM

India is a multi lingual country with twenty two official languages written in twelve scripts. Different languages are used in different parts of the country. In India, more than 300 million people use Devanagari script for documentation. Devanagari was originally developed to write Sanskrit language but later on its use was extended to develop other Indian languages. There are many languages in India like Marathi, Sanskrit, Nepali and Hindi, written in Devanagari script. Devanagari script originated from ancient Brahmi script. Character set of Devanagari script consists of large number of basic characters called aksharas. The words in Devanagari not only consist of aksharas but also consist of vowels, consonants and conjunct characters called samyuktaksharas. It also includes compound characters which consist of two or more basic characters. Devanagari script consists of 34 consonants and 13 vowels. Besides consonants and vowels, it also contains vowel modifiers called matra's which are placed at left or right part of the character. The complexity of the script increases with the presence of half characters[3][4].

4. ISSUES IN INDIAN WRITING SYSTEM

4.1 Natural variability associated with handwriting

The recognition system should be able to distinguish between structural variations across similar characters and the natural variability that exists when the same character is written by different persons or at different times. It becomes difficult to identify the handwriting of different person.

4.2 Number of stroke classes

The presence of composite characters in Indian writing systems contributes to a large number of stroke classes, which could represent consonants, vowels and modifiers or combinations of consonants and vowels. The large number of stroke classes and the inherent structural complexity of various strokes increase the complexity of the recognition system

4.3 Directionality of writing

There exists a large variation in the directionality of writing strokes and stroke segments which could affect the uniformity in stroke representation using certain features. All such variations in writing may not be encountered in the examples captured for training the system. It is necessary to identify writing direction invariant features for representing the stroke.

4.4 Variation of the number and writing order of strokes in multi-stroke characters

The variation in the number of strokes is addressed using appropriate rules for character identification. The writing order of strokes is addressed by making the application of rules independent of the order of strokes in the character.

4.5 Arbitrary pen-lifts in the course of writing a stroke and stray marks

These add to confusability and adversely affect the performance of the recognition system. Substroke based approaches have been proposed to handle arbitrary pen lifts.

5. PROPOSED WORK

The screen is a source of input in android. Pre-processing helps to reduce unwanted noise and distortion. It helps to remove variations and normalized character. Feature extraction helps us to discriminate and extract the exact source. Classification and Recognition helps to match input images with the trained set. It looks for an exact match with the existing pattern and displays the output. Gesture Library helps to implement the work.

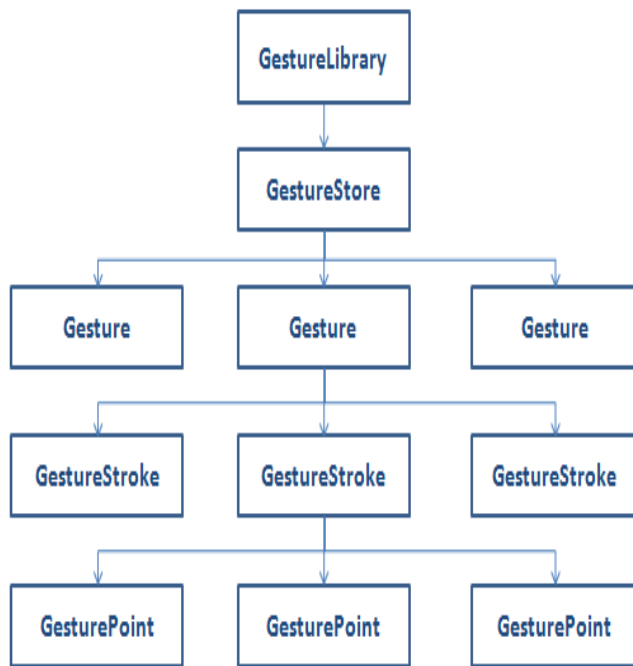

Fig- 2: Flow of proposed work

Fig-3: Hierarchical representation of gesture library

Fig -4: Gesture Utility Methods

6. THE HANDWRITTING RECOGNITION SYSTEM

Our Recognition system consists of following Interfaces and Classes used for development the application:

1. The front end of the application in prepared in XML (Xtended Markup Language). XML is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. The design goals of XML emphasize simplicity, generality, and usability over the Internet.

2. To start recognizing gestures in our application, we have to add a Gesture Overlay View to our XML layout. Gesture overlay view i.e. the input window which acts as a simple drawing board on which the user can draw his gestures[5].

3. The user interface is built using View and ViewGroup objects. The ViewGroup class serves as the base for subclasses called "layouts". A View object is a data structure whose properties store the layout parameters and content for a specific rectangular area of the screen. A View object handles its own measurement, layout, drawing, focus change, scrolling, and gesture interactions for the rectangular area of the screen in which it resides[6].

4. Gesture store maintains gesture examples and makes predictions on a new gesture.

5. Gesture class helps to recognize hand-drawn shape on a touch screen. It can have one or multiple strokes. Each stroke is a sequence of timed points. A user-defined gesture can be recognized by a GestureLibrary.

6. A gesture stroke started on a touchdown and ended on a touch up[5].

7. GestureUtility functions for gesture processing & analysis. It includes feature extraction using sampling and geometric transformation - translation, rotation and scaling, gesture similarity comparison - calculating Euclidean or Cosine distances between two gestures[6].
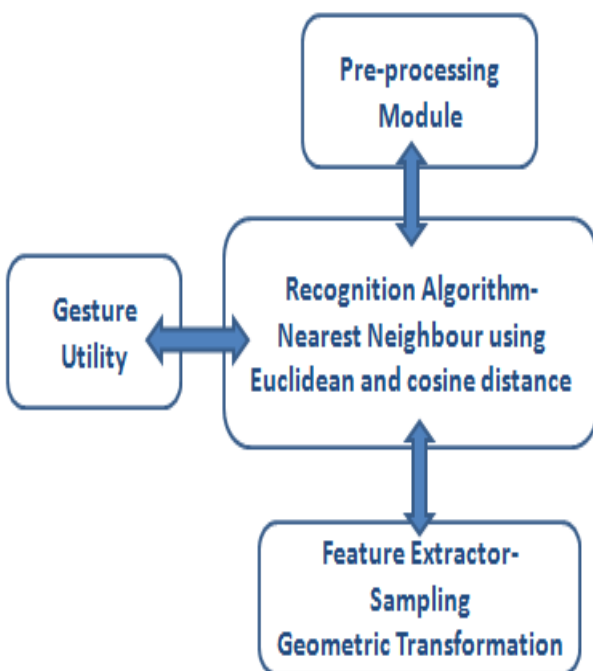
7. TESTING AND RESULTS

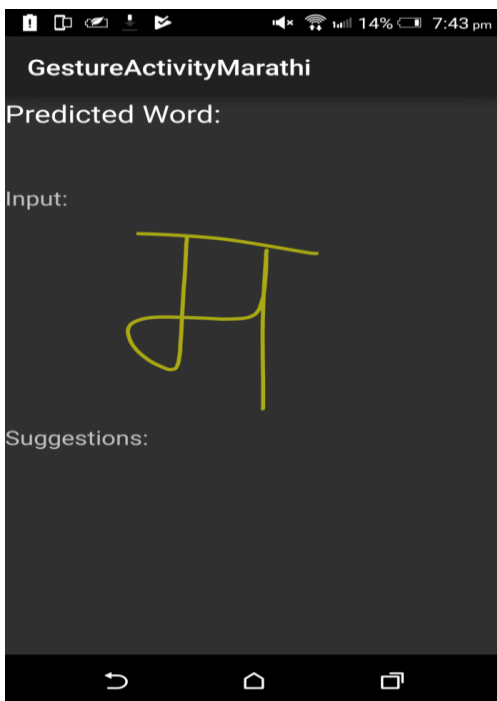Fig- 5: The interface for adding new samples to the library

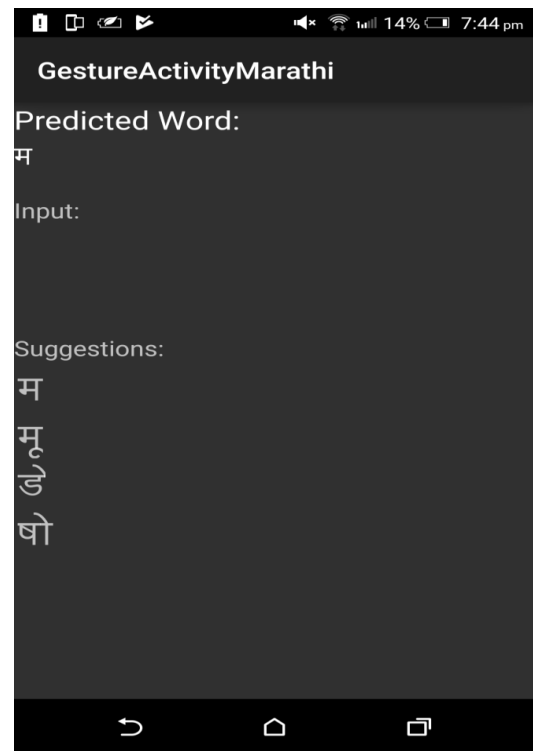Fig-6: A drawn gesture corresponding to Marathi Character

Fig-7: Predicted word

8. CONCLUSION AND FUTURE WORK

The Handwritten Character Recognition system recognizes the Marathi handwritten character and converts it into computerized font. The accuracy rate is more than 90%. For easy and fast integration into the existing user interface of Android, we use the platform's framework in some parts instead of defining our own data structures.

In future we can apply different algorithms, template matching techniques for recognition of characters. It can be helpful for recognition of different scripts. It can be further extended to recognize compounded characters. The recognition result will provides new benchmark for recent and future research.

REFERENCES

[1]Anisha Priya, Surbhi Mishra, Saloni Raj, Sudarshan Mandal and Sujoy Datta,"Online and Offline Character Recognition: A Survey", International Conference on Communication and Signal Processing, April 6-8, 2016, India.

[2]Bhushan C. Bhokse, Bhushan S.Thakare,"Devnagari Handwriting Recognition System using Dynamic Time Warping Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 52– No.9, August 2012.

[3]Dhanashree Joshi, Sarika Pansare, "Combination of multiple image features along with KNN classifier for classification of Marathi Barakhadi." 2015 International

Conference on Computing Communication Control and Automation.

[4]Satish Kumar, "A Study for Handwritten Devanagari Word Recognition", International Conference on Communication and Signal Processing, April 6-8, 2016, India.

[6] https://developer.android.com

[5]D. Dutta, A. Roy Chowdhury, U. Bhattacharya, S. K. Parui, "Building a Personal Handwriting Recognizer on an Android Device", 2012 International Conference on Frontiers in Handwriting Recognition.