Handwritten Script Recognition

Mr. Awasare Bhushan[1], Ms. Patil Kalyani[2], Ms. Jolapure Supriya[3]

[1,2,3] *(Computer Science and Engg/Shivaji University, India)*

ABSTRACT: *We can use the handwriting recognition process for a quick . Handwriting recognition is in research for over four decades and has attracted many researchers across the world. Handwriting recognition involves the automatic conversion of text as it is written on an application or a writing pad, where hand-movements play a vital role. The obtained signal is converted into letter codes which are usable within computer and text-processing applications. The elements of an on-line handwriting recognition interface typically include: a pen or stylus for the user to write with. a touch sensitive surface, which may be integrated with, or adjacent to, an output display. a software application which interprets the movements of the stylus across the writing surface, translating the resulting strokes into digital text.*

Keywords – *Digital pen, Handwritten script, PDA, Stylus, Writer-Independent*

I. Introduction

Computers are increasingly becoming easy to use (easy to communicate with) and provide variety types of functionalities human beings need. Hence computers are now being involved in almost all activities of people.

Computers have gone through many improvements since their introduction in parallel with their involvement in people's activities. These improvements can be described in terms of functionality, ease of use and size. In terms of size, they have improved from a large villa sized early computers to pocket-sized devices.

Computers are influencing the lives of human beings and their usage is increasing at a tremendous rate. The ease with which we can exchange information between user and computer is of immense importance today because input devices such as keyboard and mouse have limitations vis-à-vis input through natural handwriting.

Automatic identification of handwritten script facilitates many important applications such as automatic transcription of multilingual documents and search for documents on the Web containing a particular script. The increase in usage of handheld devices which accept handwritten input has created a growing demand for algorithms that can efficiently analyze and retrieve handwritten data.

II. Areas of Handwriting Recognition

As already mentioned, Handwriting recognition is a technique by which a computer system can recognize characters and other symbols written by hand in natural handwriting. At the highest level, handwriting recognition can be broken into two categories on the basis of how the raw data is acquired [1]. Therefore, on the basis of raw data acquisition and the nature of handwritten data, handwriting recognition is divided into two distinct areas as follows: 1. Online Handwriting Recognition and 2.Offline Handwriting Recognition.[3,4,5]
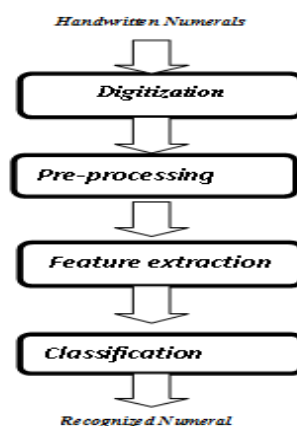

Fig 1: Block diagram of Handwritten Recognition System

2.1 Offline Handwriting Recognition

In this type of recognition, the text is not recognized at the same time as it is produced. Instead it is recognized after the user has finished writing i.e. in this case, the text is originally written on a surface such as paper and from there on it is recognized by the computer by scanning the surface. The scanned handwriting is first stored digitally in grey scale format e.g. bitmap image, and then further processing is done on it to have a good recognition accuracy.

Features for recognition are enhanced and extracted from the stored bitmap image by using digital image processing. This type of recognition is also sometimes called as "Optical Character Recognition" [2]. Recognition of machine printed characters is also a part of Optical Character Recognition. Offline methods are less suitable for man-machine communication because no real time interactivity is present. They are suitable for automatic conversion of paper documents to electric documents which then may be interpreted by computers.

2.2   Online Handwriting Recognition

In contrast to the offline method of handwriting recognition, online handwriting recognition is done in real time i.e. at the same time as the handwriting is produced. The surface used for handwriting is usually a digitized tablet and it is used along with a digital pen also sometimes called "Stylus", in order to write on the surface. As the pen moves across the surface, the two-dimensional co-ordinates of successive points are collected and stored as a function of time.

III.   Proposed System Overview

This system is useful where the online script is to be recognized. An evaluation of the system shows the features and model selected to be adequate for the task of writer-independent handwriting recognition at a high rate of word recognition. The system's robustness can be increased to handle cases where delayed strokes are written before the completion of a word part. Hence the number of errors can be reduced.

The two wider categories of online and offline handwriting recognition systems have their own areas of application. If data entry is needed at the time of writing, online handwriting is the best choice. To facilitate online handwritten text entry, special devices and/or supporting applications must be attached to the computer.

3.1   Data collection

Online handwritten recognition requires a transducer that captures writing as it is written. The most common of these devices is the electronic tablet or digitizer. These devices uses a pen that is digital in nature.Data collection is the first phase in online handwriting recognition that collects the sequence of coordinate points of the moving pen. A typical pen includes two actions, namely, PenDown and PenUp[6]. The connected parts of the pen trace between PenDown and PenUp is called a stroke. These pen traces are sampled at constant rate, therefore these pen traces are evenly distributed in time and not in space. The common names of electronic tablet or digitizer are personal digital assistant (PDA), cross pad (or pen tablet) and tablet PC. The appearances of personal digital assistant, cross pad and tablet PC are shown in Figure 2.

Personal Digital Assistant   **Cross pad or pen tablet**   **Tablet PC**

Fig. 2:Commonly used hardware devices for capturing handwriting.

3.2   Preprocessing

Preprocessing phase in handwriting recognition is applied to remove noise or distortions present in input text due to hardware and software limitations vis-à-vis smooth handwriting. These noise or distortions include irregular size of text, missing points during pen movement collections, jitter present in text, left or right bend in handwriting and uneven distances of points from neighboring positions.
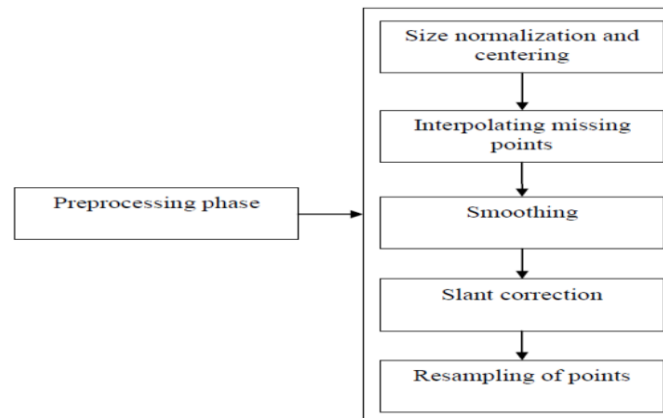
Fig.3:Common steps in preprocessing phase.

he goal of preprocessing is to reduce or eliminate these variances in order to decrease the dissimilarity between a to be recognized entity and its correct prototype[9]. The noises introduced in the data are caused by the inaccuracy of the data capturing device, the erratic movement of the hand or fingers while writing and the varying writing speed of the writer.

3.3 Feature Extraction

Once the data has been preprocessed and appropriate segmentation has been done on it (if any), the data will be ready enough to apply classification. But, it is sensible that all the sampled data points are not equally useful for recognition while the possibility to do classification using the whole raw data is still possible[12,14]. Systems that rely on raw data generally have lesser recognition rate than those that rely on only the important part of the data or attribute(s) derived from the raw data[12]. Hence, the appropriateness of using selected features.

3.4 Character Recognition

The term 'recognition' is a typical word used in many pattern recognition systems. Recognition systems are in general devoted to classify input patterns to corresponding entities. These entities vary from one system to another. Recognition systems, in which characters are expected to be classified, are referred as character recognition systems.

One major distinguishing feature of character recognition systems is the type of characters, which are supposed to be recognized. There are systems for recognition of printed characters, typewritten characters or written characters. Due to the varieties occurring to handwritten characters, development of the handwriting recognition systems is the most challenging.

The character recognition refers to the procedure of correcting misclassified results by applying linguistic knowledge. All the possible outcomes of an individual character are studied in terms of graph and the best suitable nature of character is depicted. Language information can increase the accuracy obtained by feature extraction.

3.4.1 Handwriting Recognition

Handwriting recognition is the task of transforming a language represented in its spatial form of graphical marks into its symbolic representation[3]. Symbolic representation refers to the digital representation of characters like in the case of 8-bit ASCII character set. Handwriting has been a basic form of communication and still a good way of expressing ones' ideas. In relation to this fact, handwriting recognition systems are useful and are used to realize ideal applications of computers such as pen computing.

IV Merits of Proposed System

The basic input to character-based handwriting recognition systems is a pattern that represents a character. In fact, this pattern should be digitized before it is available to the system. The way the input is provided along with the digitization technique are taken as the two factors to classify handwriting recognition systems as online or offline.

The basic differences between these online and offline handwriting recognition systems as stated in a number of literatures [3, 4, 7] are:

➢ Input Method:

In online handwriting recognition systems, the handwriting data is obtained with the help of a transducer such as an electronic or tablet digitizer. Systems that simulate a pen-paper like arrangement for a writer are the sources of online handwriting data. Such systems record the pen-tip information as a sequence of (x,y) coordinates of data points sampled over time. Whereas, on the other hand, in offline handwriting recognition systems, the data is captured optically by a scanner in the form of image. This necessitates the development of a method to identify the pixels which are part of the handwriting data pattern from pixels lie on the background [8].

➢   Kind of available information:

For online system, the coordinates of successive points as a function of time in order is available where as in the offline case, only the completed writing is available. Additionally the speed of the writer, stroke number and order, and the Pen-Up/PenDown states are detected during the collection of online handwriting data.

➢   Time of processing:

Clearly seen fact is that offline systems run after their data have been collected. The handwriting ought to be put on a media such as paper exhaustively and brought to the scanner, which in turn digitizes it. The case for online handwriting recognition systems is quite opposite. An online system receives handwriting data and recognizes it in real or nearly real time. This situation adds one useful feature to online systems that is interactivity. Being interactive, online handwriting recognition systems allow users to edit what they have written and recognition errors can be corrected immediately.

➢   Speed Determination:

The speed of online systems depends on the writing speed of the user since the handwriting is analyzed and recognized in real or near real time. Conversely, in offline systems, the speed depends on the specification of the system in words or character per second.

➢   Adaptation:

Adaptation of the writer to machine and machine to the writer is possible in online handwriting recognition systems. When the writer sees that some of her characters are not being classified correctly, she will alter the characters' shapes to improve accuracy and adapt to the system from time to time. On the other hand, some recognition engines are made to be capable of adapting to the user, basically by storing samples of the writer's characters for subsequent recognition. This two-side adaptation could never be made to happen to offline systems.

V.   Conclusion

This paper discussed an approach that involved the use of Online Handwriting Recognition technique to promote the user to have performance beyond the expectations. In broader sense, online handwriting character recognition is the process of finding out what character is represented by a given online handwritten character data. Online handwriting character data is, in general, a set of the coordinate values of data points that are sampled along the trajectory of the character as it is written.

The difficulty of identifying what character is actually represented by the data set start from the fact that identical set of data points will not be obtained for different occurrences of the same character. Thus, pattern representation scheme that produces similar representations for different occurrences of the same character solve the problem of character classification greatly.

Another future area of work will address using segments at the word level for identifying the script. This will allow us to identify multi-script documents where different portions of the document can contain different scripts.

REFERENCES

[1] Tappert, C. C., Suen, C. Y., and Wakahara, T., 1988. Online Handwriting Recognition- A Survey. International Conference on Pattern recognition, vol. 2, pp. 1123-1132.

[2]   Tappert, C. C., Suen, C. Y., and Wakahara, T., 1990. The State of the Art in Online Handwriting Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 8, pp. 787-808.

[3] Palmondon, S.N. Srihari: "*Online and Offline Handwriting Recognition: A Comprehensive Survey*", IEEE Transactions on pattern Analysis and Machine Intelligence, Vol. 22, No-1, 2000.

[4] A. Drissman, "*Handwriting Recognition Systems: An Overview*", availablefromwww.drissman.com/avi/school/HandwritingRecognition.pdf

[5] C.C. Tappert, C.Y. Suen, T.Wakahara, "*The State of the Art in Online Handwriting Recognition*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 8, August, 1990.

[6] "*Pen Technologies*", IBM Pen Technologies, Available from www.research.ibm.com/electricInk/

[7] H. Shu, "*On-line Handwriting Recognition Using Hidden Markov Models*", (Master Thesis), Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1997.

[8] Scott D.Connell, "*Online Handwritten Recognition Using Multiple Pattern Class Models*", (Doctor of Philosophy Dissertation), Department of Computer Science and Engineering, Michigan State University, 2000.