

CHAPTER 1

INTRODUCTION

1.1 Background

One of the most common and natural medium of communication for human beings is through handwritings. Even with the introduction of new technologies, handwriting persists as an effective means of communication and information recording in our daily life.

On the other hand, the computer, which was created in the mid nineteenth century, with the initial intention to help performing complex mathematical calculations, has now emerged as a popular and important tool in most of the corporations and organizations. It has been used extensively to perform repetitive and routine tasks to increase productivity and efficiency of an organization. Although current computer technology has gone so far in terms of processing speed, it still offers an unnatural way of communication between human and computer. We still need to learn a set of input methods to adapt ourselves before we can use the computer efficiently. Therefore, having a computer with the ability to interact with human in a more natural way, such as the ability to recognize our speech and handwritings, might further increase the usage of the computers in the area where current computer technology are not accessible.

Although research in optical character recognition (OCR) has started quite some time, research in unconstrained handwriting recognition has only gain much

attention since the early 80 [Nagy, 2000]. Many of the successfully commercialized of handwriting recognition systems are still limited in application which can be applied to small and specific vocabularies, such as bank cheque reading and postal address interpretation [Augustin, et al, 1998] [Guillevic, et al, 1995] [Knerr, et al, 1997]. Usage in other applications especially those required large lexicons, such as recognizing handwritten notes, are, however, still not available.

1.2 Problems

Figure 1.1 shows various styles of handwriting for the word “*six*”. For word recognition purpose, all the examples shown belong to only one class out of all classes available in our known vocabulary. The examples shown in the figure only unveil very small portion of the handwriting styles for the word “*six*”. By only looking at the variability of each example of this particular class, we shall realize that to design a system that can recognize handwritings is indeed an extremely challenging task. From this example, we also realize the incredible recognition capability that humans have. Albeit its slow processing speed, our brain can still perform real-time handwriting recognition task that even the fastest computer can hardly achieve.

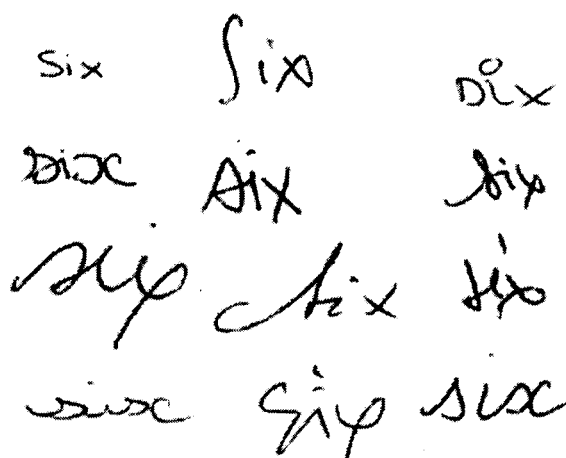


Figure 1.1. Various handwriting styles for the word “*six*”

The problems and difficulties of handwriting recognition task can be generally summarized into five categories:

- Nature of the handwriting signals
- Handwriting styles
- Writer dependency
- Size of vocabulary
- Language

The following subsections elaborate in detail each of the problems mentioned.

1.2.1 Nature of the Handwriting Signals

Depending on the nature of the applications, there are two different types of handwriting signals that can be retrieved from the input sources: offline handwriting signal and online handwriting signals.

Offline handwriting recognition deals with the problem of reading the handwritings at some point in time after they were written. The handwriting is typically captured and digitized from a paper by a scanner or camera. Thus, it is in the form of two-dimensional set of pixels with binary, gray-scale or color value.

On the other hand, an online system uses a graphic tablet to enter the handwriting. The data from the tablet is one-dimensional temporal signals $\{x(t), y(t)\}$, i.e. coordinates sampled at a constant interval in time, t . Information that can be retrieved by the system is the relative position of each point, velocity of the pen along its trajectory, and whether the pen is currently lifted or touching the tablet.

As the online signals record the temporal sequence of handwriting, we can easily transform the online signals into offline signals. However, it is quite difficult, if not impossible, to generate online signals from offline signals. In other words,

online handwriting signals contain more information, and thus, online handwriting recognition is regarded easier to solve in terms of recognition accuracy than its offline counterpart.

Nevertheless, handwriting recognition system for offline and online signals represent different perspective of challenges and applications. As online handwriting recognition is usually applied in direct interface with the user, it often requires real-time recognition response, immediately after the user finished writing a word. Apart from that, it also needs to handle recognition with very large vocabulary. This represents a challenge to the system especially when the recognition engine is built on a less powerful computational platform with stringent memory constraint.

Conversely, offline handwriting recognition systems normally perform the task at the back-end. Although the recognition speed is an important issue, it can be overcome by high-speed computers. Nonetheless, in most offline scenarios, there is no control on the type of writing medium and instrument used. The artifacts of the complex interactions between mediums, instruments and subsequent operations such as scanning and binarization present additional challenges to algorithms for offline handwriting recognition.

1.2.2 Handwriting Styles

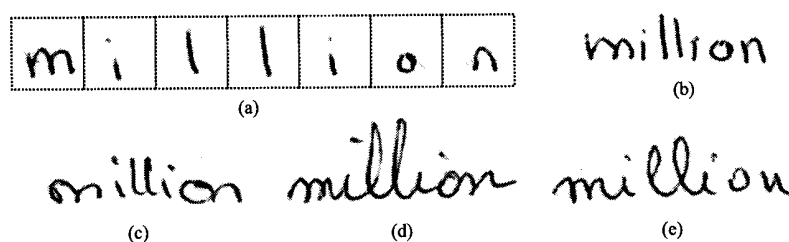


Figure 1.2. Different types of handwriting styles. (a) Box discrete characters. (b) Spaced discrete characters. (c) Run-on discretely written characters. (d) Pure cursive handwriting. (e) Unconstraint handwriting.

Handwriting styles vary depending on the constraint applied on applications. Figure 1.2 displays five main types of handwriting styles:

- **Boxed discrete characters.**
For applications like automatic form processing, normally users are requested to write each isolated capital letter in the pre-printed rectangles. These “boxed discrete characters” do not require character segmentation as the position of each isolated character is known. Handwritten characters can be extracted from the rectangles and can be recognized by a character recognizer.
- **Spaced discrete characters.**
The “spaced discrete characters” are separately written. Each character does not have ink traces to link with its neighboring characters. Only simple character segmentation algorithm is needed to retrieve each isolated character.
- **Run-on discretely written characters.**
For this type of handwriting, each character is written one after another. There may be ink traces that link between neighboring characters due to the writing speed.
- **Pure cursive handwriting.**
This is the handwriting style that is usually being taught in the school. Each character has to be nicely link with its neighboring characters. In the cursive handwriting, diacritic marks such as i-dots and t-bar are written after the main part of the word.
- **Unconstraint handwriting.**
This is the most frequently encountered handwriting style in our daily life. It is a mixture of spaced discrete, run-on discrete and cursive handwriting.

1.2.3 Writer Dependency

Handwriting styles are extremely diverse for individuals, depending on the geographical region, native handwriting and so on. It is much more difficult to design a system to recognize many people's handwritings than that of a single author. There are three levels of difficulties:

- **Mono-scriptor.**
For mono-scriptor system, the handwriting recognition system is trained and tested on the same scriptor. As the handwriting of the same writer is normally similar, the handwriting recognition accuracy is usually higher.
- **Multi-scriptor.**
Another tougher problem is the multi-scriptor system, where the handwriting recognition system has to deal with different person's handwriting. The training and testing data sets are from the same group of scriptors.
- **Omni-scriptor/Scriptor Independent.**
The toughest problem is the omni-scriptor system, where the handwriting recognition system has to deal with handwriting styles that it has never seen during the design or training stage. During design stage, the system is trained with onset of available database from one group of scriptors. However, it will be tested on a fresh database from an entirely different group of scriptors. This posed a great difficulty to the system. This is usually the problem for a real commercial handwriting recognition system.

1.2.4 Vocabulary Size

Table 1-1. Different categories of problems based on size of vocabulary

Category	Vocabulary Size	Applications
Small and specific	<100	Legal Amount Recognition
Limited but dynamic	<1000	Address Recognition
Large	>1000	General Applications

According to [Steinherz et al, 1999], one can classify the nature of the vocabulary or lexicon size into three main categories as shown in Table 1-1. The size of vocabulary has significant influence on the design of a handwriting recognition system. As the size of vocabulary grows, the chances to have more words that are similar increase, thus, more confusion occur. Therefore, as the vocabulary size increased, we would expect performance and speed of a particular recognizer to degrade.

1.2.5 Language

There are many different languages used in this world. Three most widely used handwritten languages are Latin-based languages, e.g. English or French, Chinese language and Arabic language. English alphabet has 52 letters (lower and upper case). French has all English alphabets plus 24 more alphabets with diacritical marks (lower and upper case). In English or French, the position and size of the letters is important. Upper case letters sit on the baseline and are full size. Lower case letters are smaller and most are about half the height of upper case letters. Handwritten English or French words are normally separated spatially. Letters within a word, however, are not usually separated spatially. For letters like “t”, “i”, “j”, etc., when written cursively, crosses (for “t”) and dots (for “i”, “j”, etc) tend to be delayed and thus they are not written exactly above of the main part of the characters.

Chinese has a much larger set of characters (alphabets). A complete vocabulary of Chinese characters is about 50,000. A more basic vocabulary that is sufficient for normal reading purpose is around 3,000 to 5,000 characters [Tappert et al, 1990]. A Chinese character has an average of 8-10 strokes: the simplest having one stroke and the most complicated contains more than 30 strokes. If written cursively, strokes are connected and shapes may deform from the basic style. Japanese language use Hiragana, Katakana, and Kanji. The first two groups are phonetic alphabets, and each has 46 full-size characters. The strokes of these characters are less than the one in Chinese character. Kanji are subset of Chinese characters. It has 6,349 characters as defined by Japanese Industry Standard but

daily usage limited to around 2,000. Chinese and Japanese characters are usually written in block; thus, there is less character segmentation problem.

Arabic alphabet consists of 29 letters. However, due to the dependency of the shape of a character on its position in a word, the number of character shapes increases from 29 to about 60. Arabic is written from right to left. Handwritten Arabic words usually include vertical combinations of characters called ligatures. This feature makes it difficult to determine the boundaries of the characters. [Amin, 1997]

Different languages of handwriting pose different kind of recognition problems. For this thesis, only the English and French language will be looked into detail.

1.3 Applications of Handwriting Recognition

There have been significant growth in the application of handwriting recognition systems in both online and offline domain during the past decade. The most important of these has been in bank cheque processing and handwritten address interpretation, whereas popularity of cellular phones and personal digital assistants (PDAs) lately has contributed to the online handwriting recognition on electronic appliances. The following sections describe the major applications in details.

1.3.1 Bank Cheque Processing

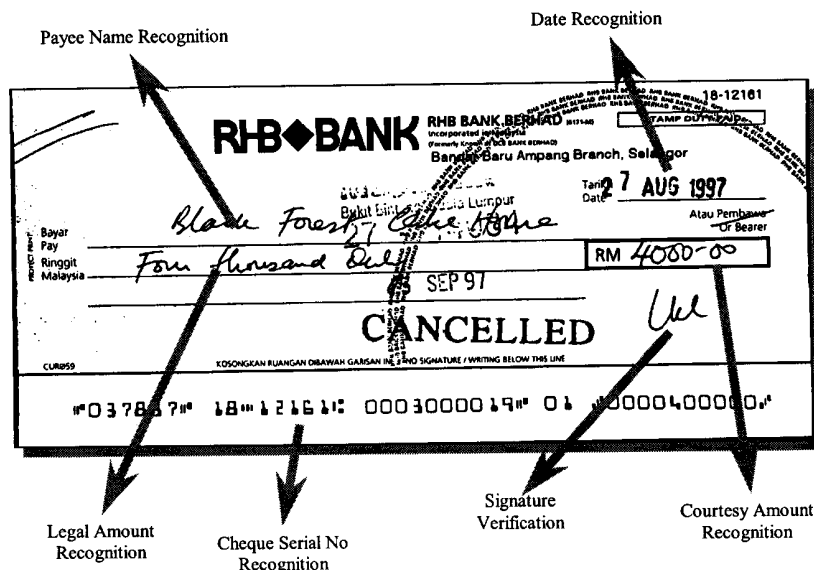


Figure 1.3. Various recognition processes to be performed to automate back cheque processing.

An automated bank cheque processing system is one that uses a computer system to replace the human operator for processing bank cheques automatically in an efficient way. The system is able to recognize the amount of cheques based on the courtesy amount, as well as the legal amount written on them. Apart from that, it is also able to recognize the date written on cheque. An automated bank cheque recognition system reduces laborious and repetitive human data entry, thus increase the productivity and efficiency of a bank.

Figure 1.3 illustrates an example of a cheque. In order to automate the cheque reading, various recognition processes are needed. The most important process is the courtesy amount recognition to recognize the amount of payment. To further verify the amount written, text written on the legal amount field is recognized. Date, signature, serial number and the payee name, are also need to be recognized or verified to make sure that it is a valid cheque.

There are substantial research challenges in the area of document image analysis and recognition for bank cheque processing. Due to the complex backgrounds and stamping noises, extraction of completely clean handwriting for

recognition is impossible. For this, we need to rely on the robustness of the handwriting recognition system to perform recognition with noisy inputs or to reject if it is too noisy. Due to limited vocabulary in legal amount, it has been getting rather promising in the development of commercial cheque-reading systems [Knerr, et al, 1998] [LeCun et al, 1998].

1.3.2 Form Processing

Numerous government agencies or private institutions use forms to collect data from the public. The design of the forms can be as simple as a few check boxes for the user to select the answers, to as complex as accepting unconstrained handwritings for particular questions. Much of this raw data must be stored structurally in the computer to be manipulated to produce useful pieces of information. Manual data entry is currently the bottleneck in the process, and it is exposed to human errors. Replacing human operators in manual data entry with reliable handwriting recognizers can indeed help produce higher and consistent productivity. Although forms must usually be hand printed to keep the writing as legible as possible, for human as well as for machine processing, cursive recognition systems are very much useful in many instances.

1.3.3 Handwritten Address Interpretation

The main objective of interpreting handwritten addresses is to automate the sorting of mail pieces. An address consists of country, state, city, street, primary number (street number or post-office box number), secondary number (apartment or house number) and finally the receiver's name. The interpretation result is represented in the form of a barcode and printed at the bottom of the envelope so that subsequent stages of sorting can be made by a barcode reader.

Handwritten Address Interpretation can be seen as an ideal application for scriptor-independent handwriting recognition, since it has a wide variety of difficulty, from postcode written at predetermined locations on an envelope, up to complete determination of an address without postcode, from discretely written postcode up to unconstrained multiple-lines address written by a foreigner.

1.3.4 Input Method for Electronic Appliances

With a widespread of various electronic handheld devices such as cellular phones, personal digital assistants (PDAs), electronic appliances have inevitably entered into our personal life and household. Current user interface of these electronic appliances are still very crude. The most popular input method used in today's computing environment – keyboard, is not suitable as a convenient user input method due to its size and cost of manufacturing. Handwriting input methods, such as the Graffiti[®] on PDA powered by PalmOS[®], only use uni-strokes handwritten character recognition. Pocket PCs that come with its default Transcriber cursive handwriting recognition software, is still yet to meet the normal user requirement. Thus, large vocabulary online handwriting recognition system is very much in demand in order to provide better user experience with the electronic appliances.

1.4 Scope and Objectives

The main objective of this thesis is to construct a system that is able to recognize isolated cursive handwritten English and French words. Although offline handwriting recognition is used through out the thesis, the classification model can also be applied to the online handwriting recognition.

This thesis describes a Segmentation-by-Recognition (SegRec) approach in handling handwriting recognition. As will be mentioned in the later chapters, that it is quite difficult, if not impossible, to develop an intelligent character segmentation process without the knowledge of characters. Therefore, our approach employs a simple segmentation process to segment a word into a left to right sequence of sub-characters, and then propose various hypotheses of merge sub-characters into characters. The main purpose of the segmentation process here is to preserve as much handwriting information as possible at this level. The recognition process then

try to spot the most probable segmentation based on the recognition results of each of the character hypothesis.

For SegRec approach, apart from performing classification between each character class, the recognizer also needs to model the character hypotheses that do not belong to any character class, which we term it as *junk*. A detailed experimental analysis in handling the junk problem will be described in this thesis.

Two fundamental training schemes for the hybrid ANN/HMM are presented in this thesis. First is the character-level discriminant training that optimizes the ANN at the character level and further optimizes the HMM at the word level. A method is introduced to carefully select junk examples from the training database to train the ANN explicitly. Another scheme is called word-level discriminant training where the ANN is optimized at the word-level using the stochastic gradient-descent (back-propagation) algorithm.

1.5 Thesis Layout

The thesis is divided into six chapters:

- **Chapter 1** describes some background information, the challenges and applications of the handwriting recognition as well as the objectives and contents of the thesis.
- **Chapter 2** provides literature review on handwriting recognition and summarizes the achievements of other works in the field of handwriting recognition. An introduction of the statistical pattern recognition system, which is the basis of this thesis, is also presented.
- **Chapter 3** describes in details the offline handwritten cursive word recognition system that has been designed. This chapter is separated into

four main parts, i.e. image preprocessing, segmentation, feature extraction and recognition, which corresponds to different processing stages in the handwriting recognition system. The last part of the chapter discusses the combination of ANN and HMM to perform the recognition.

- **Chapter 4** explains the training of the hybrid ANN/HMM recognition system. Two approaches, namely Character-level Discriminant Training and Word-level Discriminant Training are presented. Both approaches use popular gradient descent algorithm to optimize the parameters in the system.
- **Chapter 5** presents the experiments carried out to evaluate the performance of the system based on different training approaches and configurations. Detailed Error Analysis on the wrongly recognized examples is also presented.
- **Chapter 6** concludes about the handwriting recognition system and summarizes what has been achieved in this research. Further works that could be carried on from this thesis is also suggested.