39

A new method of off-line text recognition

Susan Laflin

School of Computer Science, University of Birmingham.
Email: laflins@bham.ac.uk

39.1   Introduction

The problem of recognising handwritten text as a means of data input to a computer has been with us almost as long as computers themselves. Some areas have achieved a large measure of success, others very little. The main problem is the lack of consistency in human hand-writing – humans are generally better at perceiving the intended word lying behind the scrawl on the paper. Computer methods are less good at guessing what is intended and very few computer-based methods use anything like the amount of background information on context and likely meaning that the human reader applies without consciously realising what he or she is doing. Other problems arise because of the poor quality of many of the surviving texts and, to a lesser extent, the use of 'obvious' abbreviations which can usually be guessed by the human reader but must be expicitly defined before they can be handled automatically by the computer.

As usual, to tackle this large and ill-defined problem, the first stage is to subdivide it into smaller, more closely-defined problems. The first division is into *on-line* or *dynamic* methods contrasted with the *off-line* or *static* ones. With an *on-line* method, the data is captured *as it is written*. Letters and words are written on a digitising tablet or pad and successive coordinates of pen-position are recorded and the whole is deciphered by the computer (Tappert 1990). This records not only the pen path – giving information about both the shape of the letters and the order in which the strokes are made – but also the speed with which different parts of the word are drawn and recording any pauses made by the writer during the word. All this information is lost once the writing is complete. With this extra information, some allowance can be made for inconsistencies in human writing and this improves the success rate of recognition. Recently a spate of portable 'notebooks' allowing handwritten input have been released on the market and with most of these the recognition has reached an acceptable level. They require an initial training session to accustom the system to the style and idiosyncracies of the owner and after that may be used by this one writer with an acceptable level of success.

By contrast, the *off-line* methods have only an image of the completed document to work on. Typed or printed documents have a much greater consistency and once a system has been trained on that particular typeface, it will recognise most of the document. Problems still arise on faded or discoloured areas, but in general scanning and automatic recognition are quicker than re-typing the whole document into the computer. The next problem, in order of difficulty, is the recognition of handwritten characters, either letters or digits. This is of particular interest when these characters make up post-codes and their recognition is necessary for automatic sorting of mail. Some success has been achieved and the current state of the art is described in a special issue of *Pattern Recognition Letters* (Tominaga 1993).

The final and most difficult problem is that of cursive handwriting and there are two main approaches to this problem. One of them attempts to extend the experience gained with on-line methods to apply this to the off-line problem. The aim is to identify pen-strokes within the image, and then build these up into letters and then words. Much work has been done on this approach (e.g. Helsper & Schomaker 1993; Helsper *et al.* 1993) and success has been obtained in a number of cases. However many others still remain, cases where the human reader is much more successful than the computer system, and these have led me to devise an alternative approach, based on the methods used by the human reader.

39.2   Global method.

Let us for a moment consider how the human reader deals with a diffucult passage in a handwritten text. We have all met the problem – whether in our own or someone else's handwritng – of the illegible phrase. How do we tackle it? For myself, I am rarely successful in trying to spell it out letter by letter. Far more often I puzzle over it for some time and then, suddenly, solve the entire phrase not just a single word within it. Sometimes the whole problem arises because of the misreading of an earlier word and when this is corrected the rest falls into place. In all this process, it is a little difficult to watch myself and decide what I am really doing. I am aware of failing to recognise the phrase at the first attempt, of puzzling over it for some time and, during this time, of trying and rejecting certain possibilites and finally a feeling of jubilation when a successful match is found and the whole text falls into place. Now I need to produce a similar process, explicitly, on a computer system.

The main reason for the greater success of the human reader in deciphering these difficult cases come from knowlege of the context of the word. The human reader does not attempt to compare all the words in the language with the unknown word, but uses all experience of the writer and subject to restrict comparison to those which are believed to be reasonable. Sometimes a change in interpretation of another word or words in the phrase may extend the scope of what is reasonable and allow a fit

where one did not previously exist. To design a similar system for the computer, it is necessary to define the method explicitly. Instead of vague ideas about what is likely or unlikely from that particular correspondent, the computer system will require a list of words defining the vocabulary to be compared with the unknown word-images within the document. This vocabulary will have to be supplied by the users, who know that the manuscript to be recognised comes from a particular context and this context can be used to give a list of probable words and phrases. In addition, the manuscript is known to be written in a particular hand, and so the computer will need precise instructions for generating an image of the handwritten word from the ascii string. This image can be compared with the unknown word-images in the manuscript and when matches are found, the problem of recognition has been solved. The process can be repeated for every word in the vocabulary. This, very briefly, is the general idea behind my current research. I call it a 'global' method because it compares complete words (blocks within the image of the document) rather than attempting to analyse these into individual strokes and then build them up into letters and words.

From the above discussion, it should be clear that this very large problem can be broken down into stages and each one tackled separately. There will probably be several methods which may be applied to some of the stages and these will need to be compared and the most successful ones included in the system.

39.2.1 Text generation

This involves a study of the methods of writing the various hands in order to generate an image for any given ASCII string. Work has already started in this area and is described later in this paper.

39.2.2 Classification of manuscripts

This requires a study of different types of manuscripts at different periods to produce the relevant lists of words and phrases. The lack of any standardised spelling may make these lists very long, or may require their amendment during use.

39.2.3 Segmentation of document images

When a page of a document has been scanned, it will be necessary to subdivide the image into lines of text and the lines into individual words. It is likely that automatic methods are not entirely successful and will need to be checked and edited interactively.

39.2.4 Comparison of word images

The generated images of known words have to be compared with the images of unknown words in the manuscript and matches sought. The success or failure of this section is crucial to the whole approach. It is likely that the many possible methods will need to be compared and a choice of the more successful ones included in the package. Since human handwriting is frequently variable,

it will be necessary to try fuzzy matching techniques, various distortions of the geometry of these images, and neural net methods to find suitable methods for these comparisons.

39.2.5 Testing of prototype system

Once at least one method exists for each section, it will be possible to produce a prototype system and invite users to test it and comment on the facilities needed.

39.3 Text Generation

The problem may be defined very simply: *Given any ASCII string, generate an image of the corresponding handwritten word.* This requires an analysis of the way in which the handwriting was formed in order to produce an image for comparison. Modern texts on calligraphy (e.g. Hardy Wilson 1990) describe the production of a number of hands in terms of pen-path for each of a number of strokes, with the pen-width and angle to the horizontal assumed constant throughout. This suggested that first approximation should model this using these three variables. The description of the *gothic (blackletter) hand* required a broad pen held at an angle of about $35°$ to the horizontal. This was simulated by drawing a line of width $w/2$ at this angle on each side of the pen-position. The pen-path was produced by recording a series of (x, y) coordinates and joining them by a sequence of straight lines (polyline). This produced a first approximation and the gothic hand was chosen because it was well suited to this method. The individual strokes making up the letters are linear and the script is not truly cursive, so that words may be generated by the juxtaposition of the individual
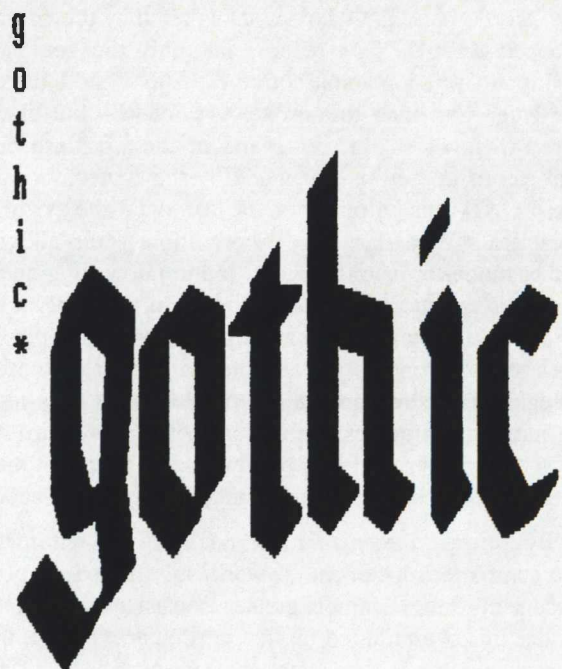

Figure 39.1: Example output from the experimental text generator.

letters without having to bother about a smooth join. Figure 39.1 shows an example of the output from this initial experiment, with the ascii characters down the left-hand side and the corresponding 'handwritten' output in the centre of the page.

There are a number of modifications which must be included in the next version. In the book, the height of the script is described in terms of pen-width – in one place the gothic hand is said to have a height of five or six nib-widths and be written with an angle of $45°$ while later on the blackletter gothic is said to have a height of three to five nib-widths and an angle of $30–40°$. This suggests that the coordinates describing the pen-path should be normalised to fractions of a nib-width rather than millimetres on the page. Ascenders and descenders appear to be two nib-widths above or below the body of the text, and this ratio can be checked. Also there are a few thin flourishes, which are probably produced by the edge of the nib and need to be coded separately. These are minor additions, but they would improve the appearance of the text produced by this program.

The major problem encountered when using this book it that it assumes modern equipment, especially metal nibs which remain constant in shape for the majority of their lifetime and which can be replaced by another identical one when worn out. This was certainly not the case when medieval manuscripts were written – the quill pens then in use had to be re-sharpened at very frequent intervals and were unlikely to be of identical shape or thickness after each operation. Also a very slight change of pressure would have a large change in the thickness of the resulting line. All of this suggests that the 'constants' of modern calligraphy may have been far from constant in the ancient manuscripts with which the comparisons are desired. To produce a system which would fit more closely to the actual conditions of the time, a specimen sheet by the German writing master Gregory Bock ($c.$1510) (Friedman 1993, 70, Fig 3) showing how each letter was built up, stroke by stroke was used. This suggested that the angle of pen-nib to the horizontal was far from constant – indeed the angle between the pen-nib and the gradient of the pen-path is a more important factor and the script has both thick, black strokes (with this angle close to $90°$) and thin flourishes (with this angle close to $0°$). In-between angles seem rare at a first inspection of this page. It also appeared that some of the lines of the pen-path were in fact curved and so the polyline approximation would require many more points, whilst the number of strokes per letter should be increased from that shown in the modern text.

If the pen-angle is to be defined as the angle between the nib and the gradient of the pen path, this implies that gradient as well as coordinates must be recorded at the points defining the outline. This additional information immediately offers the alternative of representing the actual path as an interpolating cubic rather than a polyline. Such a cubic can also be used to provide smooth joins in a cursive script and thus to represent many other hands using the same techniques. The present software will be developed further during the summer of 1994, using the ideas discussed here, and incorporated in the full system in due course.

39.4 Conclusion.

The aim of this research is to provide an interactive tool which will help users to obtain a text copy of the contents of a manuscript (Laflin 1993; Laflin forthcoming). It assumes that as the cost of computer hardware continues to fall and as the demands on Public Record offices and others archives continue to grow, the use of computer systems will become more widespread. Instead of allowing users to handle the ever-more-fragile original documents, these archives will scan them and make the images available on the computer screen. Under such a system (and I realise it may be many years into the future), a tool such as this will be of great use. I expect the user to alternate between automatic comparison of the word-segments in the document and the images generated from the appropriate word-list and interactive intervention to identify a particular image as a particular ASCII string. When the session is finished, the user should be able to get hard-copy, either all text (if the transcription is complete) or all image or a mixture, indicating the state achieved at the end of the session.

References

Friedman, J. F. 1993. 'Computerized Script Analysis and Classification: Some Directions for Research' in *Optical Character Recognition in the Historical Discipline*', Halbgruae Reihe Band A18, Max Planck Institut.

Hardy Wilson, D. 1990. *The Encyclopaedia of Calligraphy Techniques.* Headline Books.

Helsper, E. L. & Schomaker, L. R. 1993. 'Off-line and On-line Handwritten Recognition' in *Optical Character Recognition in the Historical Discipline*', 39–51 Halbgruae Reihe Band A18, Max Planck Institut.

Helsper, E. L., Schomaker, L. R. & Teulings, H-L. 1993 'Tools for the Recognition of Handwritten Historical Documents', *History and Computing* 5, 88–93.

Laflin, S. 1993. 'An Interactive System for Off-line Text Recognition' in *Optical Character Recognition in the Historical Discipline*', 53–58, Halbgruae Reihe Band A18, Max Planck Institut.

Laflin, S. (forthcoming) 1994. 'Processing Historical Information with the Aid of Computers,' in F. Bocchi & P. Denley (eds.) *Storia & Multimedia.*

Tappert, C. C., Suen, C. Y. & Wakahara, T. 1990. 'The State of the Art in On-Line Handwriting Recognition' *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 12, 787–808.

Tominaga, H. (ed.) 1993. Special Issue on Postal Processing and Character Recognition *Pattern Recognition Letters*, 14, 257–354