A cursive script-recognition system based on human reading models

Paul Edmund Bramall*, Colin Anthony Higgins**
Department of Computer Science, University of Nottingham, Nottingham, NG7 2RD, UK

Abstract. The human reading process is undoubtedly extremely complex; however, much work has been carried out in determining possible mechanisms behind it. A computer recognition system that makes use of some of the proposed models of human reading has been developed at the University of Nottingham. With it, we attempt to solve the problem of recognising handwriting on-line. The system, called NuScript, is based on the blackboard paradigm of artificial intelligence (AI). It initially uses easily extracted features to reduce a large lexicon to a smaller list of candidate words. Later stages use increasingly sophisticated knowledge sources, based on a diverse set of AI paradigms and other pattern-recognition techniques, to determine and subsequently refine a confidence value for each candidate. A description of the elements of the human recognition models on which the system is based is followed by a general description of the computer recognition system as a whole.

Key words: Handwriting recognition – Cursive script recognition – On-line – Reading models – Blackboard system

1 Introduction

The recognition of a written word involves the processing of the word representation into a more useful and understandable form. In humans, this corresponds to the reading process in which an image of the word, for example house, is processed in order that its meaning, i.e. a building for human habitation, becomes available. Word recognition by computer need not progress to this level; it is essentially a labelling task, that of transcribing the written word into a computer-readable form.

The recognition of handwriting by machine can be divided into two broad areas, off-line and on-line recognition. In the former, an image of the handwritten text to be recognised is presented to the machine for transcription. On-line recognition requires a digitiser to be present when the text is written. Its input data is a recording of the pen movements rather than a bit-mapped image of the text. The on-line recognition problem is generally considered easier than the off-line case because information regarding the order in which strokes are formed is present. The recognition system called NuScript, which has been developed at Nottingham, tackles the on-line problem of cursive script recognition (CSR) transforming time-ordered lists of \( x, y \) coordinates describing the motion of a stylus into a ranked set of candidate words.

In developing NuScript, much attention has been paid to keeping the processing requirements as low as possible to make the system readily applicable to the emerging field of portable pen-based computing and personal digital assistants (PDAs) [12].

2 Review of cursive script recognition (CSR)

Much work has been done concerning the on-line recognition of cursive handwriting. A common approach to the problem is to reduce the written word into segments [15], and match them against predetermined templates stored for each letter. In practice, unless the identity of the word is already known, robust segmentation at letter boundaries alone is impossible. The problem is further complicated as separate templates must be stored for each different way of forming a particular letter (allograph). For example, the letter ‘z’ may be formed with or without a crossbar, and with or without a descender. The segmentation approach can therefore be computationally expensive; also any previously unseen templates must usually be identified by hand, requiring specialist knowledge during configuration. As well as being undesirable from a practical and aesthetic viewpoint, human intervention such as this may also introduce inconsistencies into the system.

A variation on the segmentation approach is to divide the input into a sequence of small vectors, which are then coded according to their direction (e.g. Freeman coding [6]). The sequence of codes can then be processed in several ways, for example with hidden markov models (HMMs) [8, 20] or by elastic matching [27]. Although these techniques have some merits, they generally operate only on local information at each stage of the process.
At the other extreme to the segmentation approach is the 'whole word' approach to handwriting recognition. Here, instead of attempting to synthesise letters (and subsequently words) from small portions of the writing, the word is considered wholistically, and is compared with prerecorded features of known words. This approach has the advantage that the segmentation problem need not be addressed directly, but the disadvantage that a user may be required to supply many examples of every word in the lexicon to train the system. The volume of training data would be reduced if, by analysing a relatively small amount of writing, it were possible to extract the features that constitute each letter or ligature. Whole word templates could then be constructed by joining these features appropriately. Rapid lookup based on a large set of features is difficult and inefficient, and the problem of extracting features for each letter automatically is also far from trivial. To extract all letter features for every letter in every word, it would be necessary to segment at letter boundaries during training. This problem is a simplified variant of the segmentation problem, as the identity of the written word is known. However, a robust segmentation algorithm would still be required during recognition to make use of such templates.

In general, most of the methods discussed, and particularly the more popular segmentation methods, are effectively trying to analyse the process of writing to determine what was written, i.e. they are based on models of human writing. Since writing is the operation of creating input data for the reading process, it seems that a recognition system based on the human reading process may be more appropriate.

There is much psychological evidence that in fluent reading, humans do not necessarily process every letter [24]. Several studies have shown that the appearance of the word as a whole is important in reading contextually related words [9, 11]. It seems, therefore, that a wholistic approach to the recognition problem may resemble human reading more closely than more traditional segmentation methods. Using this as a basis, it seems appropriate to take inspiration from other aspects of the human reading process in constructing a CSR system [13].

3 The human reading process

Many suggestions have been made as to possible mechanisms on which the human reading process may be based [1, 17, 28]. The data on which these are founded range from watching the physical aspects of reading, such as eye movements, to psychological tests on readers to gain understanding of more abstract elements of the process. Theories have proposed a basic reading unit, and speculated that fluent reading involves the successful recognition of these units, and subsequent synthesis to form meaningful text [7, 19]. Much work has been done on identifying the basic reading unit; suggestions include letters, letter clusters, syllables, and words. The evidence suggests that as a reader becomes more fluent, more importance is attached to the larger reading units, and less to the smaller ones [16, 23].

Most of the investigations into the human reading process have used printed (typewritten) text, either on paper or on a VDU. There are several reasons why experiments with handwritten text have not been pursued in the same depth. These range from the ease of ensuring consistency with typewritten text, to the purely pragmatic reason that typewritten text is easier to generate for psychological tests in which readers are required to respond to stimuli on a computer screen. Inspiration for NuScript is taken from the human reading of printed text for the parts of the system for which insufficient information about reading handwritten text is available. NuScript has been designed as a flexible and easily expandable system that allows new models of cursive script reading, based on the readers knowledge of the dynamics of script formation, to be incorporated. Indeed, any appropriate mechanisms that improve recognition rates can be incorporated irrespective of whether they model part of the human reading process.

It is important to note that context undoubtedly plays a large role in human reading. This is especially true in reading connected handwriting in which letter separations are often ambiguous. It has been suggested that handwriting in context can be read with little more than the initial letter and word shape. Taylor and Taylor illustrate this with a diagram prepared at Bouma’s suggestion [28]. However, it seems that more legible handwriting preserves more of the features traditionally associated with the printed letters; for example, the letter 'e' contains a loop, etc.

The areas of syntactic and semantic context have been covered extensively in the literature [4, 5, 29], and are beyond the scope of this paper. The role of context is also unclear in reading especially terse text, such as that written while taking notes. It seems likely that the relative importances of exact feature detection and of contextual information may depend on the quality and type of the script, (i.e. from high-quality printed words to scrawled handwriting), and the proficiency of the human reader.

When learning to read, children often name each letter of a word as it is encountered. The recognised word is then pronounced as the concatenation, or modified concatenation, of these constituent letters. Early reading is therefore likely to be on a letter-by-letter basis, the successful reading of a word depending on the correct identification of its letters. It has been suggested that there is a phonological stage (the phonological bridge) in which the written word must be converted before being recognised by the same mechanism as a spoken word. This approach is known as reading by ear. On encountering an unfamiliar or difficult word, a fluent reader may exhibit behaviour similar to that of an inexperienced reader, slowing from the normal reading speed to process the problematic word on a letter-by-letter basis.

This process of intermediate conversion to a phonological representation fails to explain why fluent readers are not confused by homophones such as 'sea' and 'see', which would receive different phonological representations. Also, fluent readers can decide correctly on a pronunciation for words such as 'lead' when they encounter it in context, implying that the meaning is available before the phonological representation. Other criticisms have been levelled at reading based on the phonological bridge including the reading performance of 'phonological dyslexics' who can read words successfully, but cannot articulate individual letters or letter clusters [3], and that of speed readers who can read as many