Chapter 2

Written Language Input

2.1 Overview

Sargur N. Srihari & Rohini K. Srihari

State University of New York at Buffalo, New York, USA

The written form of language is contained in printed documents, such as newspapers, magazines and books, and in handwritten matter, such as found in notebooks and personal letters. Given the importance of written language in human transactions, its automatic recognition has practical significance. This overview describes the nature of written language, how written language is transduced into electronic data and the nature of written language recognition algorithms.

2.1.1 Written Language

Fundamental characteristics of writing are:

1. it consists of artificial graphical marks on a surface;
2. its purpose is to communicate something;
3. this purpose is achieved by virtue of the mark's conventional relation to language (Coulmas, 1989).

Although speech is a sign system that is more natural than writing to humans, writing is considered to have made possible much of culture and civilization.

Different writing systems, or scripts, represent linguistic units, words, syllables and phonemes, at different structural levels. In alphabetic writing systems, principal examples of which are the Latin, Greek and Russian scripts, alphabets are the primitive elements, or characters, which are used to represent words. Several languages such as English, Dutch, French, etc, share the Latin script. The Devanagari script, which represents syllables as well as alphabets, is used by several Indian languages including Hindi. The Chinese script, which consists of ideograms, is an alternative to alphabets. The Japanese script consists of the Chinese ideograms (Kanji) and syllables (Kana). There are roughly two dozen different scripts in use today (ignoring minor differences in orthography, as between English and French).

Each script has its own set of icons, which are known as characters or letters, that have certain basic shapes. Each script has its rules for combining the letters to represent the shapes of higher level linguistic units. For example, there are rules for combining the shapes of individual letters so as to form cursively written words in the Latin alphabet.

In addition to linguistic symbols, each script has a representation for numerals, such as the Arabic-Indic digits used in conjunction with the Latin alphabet. Also there are icons for special symbols found on keyboards.

2.1.2 Transducers

Since the invention of the printing press in the fifteenth century by Johannes (an invention whose principal elements included the movable type, an alloy for letter faces, printing mechanism and oil-based ink), most of archived written language has been in the form of printed paper documents. In such documents, text is presented as a visual image on a high contrast background, where the shapes of characters belong to families of type fonts.

Paper documents, which are an inherently analog medium, can be converted into digital form by a process of scanning and digitization. This process yields a digital image. For instance, a typical 8.5×11 inch page is scanned at a resolution of 300 dots per inch (dpi) to create a gray-scale image of 8.4 megabytes. The resolution is dependent on the smallest font size that needs reliable recognition, as well as the bandwidth needed for transmission and storage of the image. A typical fax image of a page is a binary image scanned at a resolution of 200 dpi along the scan line and 100 dpi along the paper feed direction.

More recently, it has become possible to store and view electronically prepared documents as formatted pages on a computer graphics screen, where the scanning and recognition process is eliminated. However, the elimination of printed paper documents is hardly in the offing due to the convenience and high-contrast offered by them in contrast to bulky computer screens of today.

Written language is also encountered in the form of handwriting inscribed on paper or registered on an electronically sensitive surface. Handwriting data is converted to digital form either by scanning the writing on paper or by writing with a special pen on an electronic surface such as a Liquid Crystal Display (LCD). The two approaches are distinguished as *off-line* and *on-line* handwriting. In the on-line case, the two-dimensional coordinates of successive points of the writing are stored in order– thus the order of strokes made by the writer are readily available. In the off-line case, only the completed writing is available as an image. The on-line case deals with a one-dimensional representation of the input, whereas the off-line case involves analysis of the two-dimensional image. The raw data storage requirements are widely different, e.g., the data requirements for an average cursively written word are: 230 bytes in the on-line case (sampling at 100 samples/sec), and 80 Kbytes in the off-line case (sampling at 300 dpi). The recognition rates reported are also much higher for the on-line case in comparison with the off-line case.

2.1.3 Recognition

Written language recognition is the task of transforming language represented in its spatial form of graphical marks into its symbolic representation. For English orthography, this symbolic representation is typically the ASCII representation of text. The characters of most written languages of the world are representable today in the form of the Unicode (Unicode Consortium, The, 1990).

We discuss here many of the issues in the recognition of English orthography, for printed text as well as handwriting. The central tasks are character recognition and word recognition. A necessary preprocessing step to recognizing written language is the spatial issue of locating and registering the appropriate text when there are complex two-dimensional spatial layouts employed. The latter task is referred to as document image analysis.

Character Recognition

The basic problem is to assign the digitized character into its symbolic class. In the case of a print image this is referred to as optical character recognition (OCR) (Srihari & Hull, 1992). In the case of handprint it is referred to as intelligent character recognition (ICR).

The typical classes are the upper- and lower-case characters, the ten digits, and special symbols such as the period, exclamation mark, brackets, dollar and pound signs, etc. A pattern recognition algorithm is used to extract shape features and assign the observed character into the appropriate class. Artificial neural networks have emerged as fast methods for implementing classifiers for OCR. Algorithms based on nearest-neighbor methods have higher accuracy, but are slower.

Recognition of characters from a single font family on a well-printed paper document can be done very accurately. Difficulties arise when there are many fonts to be handled, for example, decorative fonts, or the document is of poor quality. Some examples of poor quality machine-printed and handwritten characters are shown in Figure 2.1. In the difficult cases, it becomes necessary to use models to constrain the choices at the character and word levels. Such models are essential in handwriting recognition due to the wide variability of handprinting and cursive script.

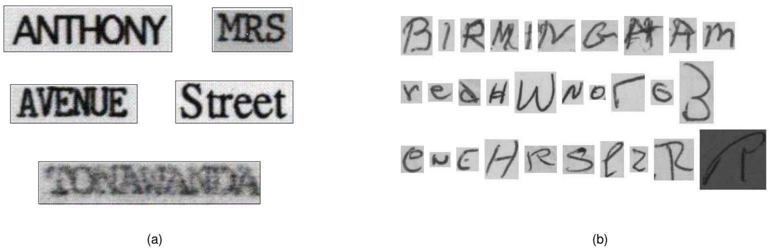

Figure 2.1: Examples of low-quality machine-printed characters involving segmentation difficulties (a) and handwritten characters (b).

A word recognition algorithm attempts to associate the word image to choices in a lexicon. Typically, a ranking is produced. This is done either by the *analytical* approach of recognizing the individual characters or by the *holistic* approach of dealing with the entire word image. The latter approach is useful in the case of touching printed characters and handwriting. A higher level of performance is observed by combining the results of both approaches. In the off-line unconstrained handwritten word recognition problem, recognition rates of 95%, 85% and 78% have been reported for the top choice for lexicon sizes of 10, 100 and 1,000 respectively (Govindaraju, Shekhawat, et al., 1993).

In the on-line case, larger lexicons are possible for the same accuracy; a top choice recognition rate of 80% with pure cursive words and a 21,000 word lexicon has been reported (Seni & Srihari, 1994).

Language Models

Language models are useful in recovering strings of words after they have been passed through a noisy channel, such as handwriting or print degradation. The most important model for written language recognition is the lexicon of words. The lexicon, in turn, is determined by linguistic constraints, e.g., in recognizing running text, the lexicon for each word is constrained by the syntax, semantics and pragmatics of the sentence.

The performance of a recognition system can be improved by incorporating statistical information at the word sequence level. The performance improvement derives from selection of lower-rank words from the word recognition output when the surrounding context indicates such selection makes the entire sentence more probable. Lexical techniques such as collocational analysis can be used to modify word neighbourhoods generated by a word recognizer. Modification includes re-ranking, deleting or proposing new word candidates. Collocations are word patterns that occur frequently in language; intuitively, if word A is present, there is a high probability that word B is also present.

Methods to apply linguistic knowledge include: n-gram word models, n-gram class (e.g., part-of-speech) models, context-free grammars, and stochastic context-free grammars. An example of a handwritten sentence together with recognition choices produced by a word recognizer and grammatically determined correct paths are shown in Figure 2.2. An increase in top choice word recognition rate from 80% to 95% is possible with the use of language models (Srihari & Baltus, 1993).

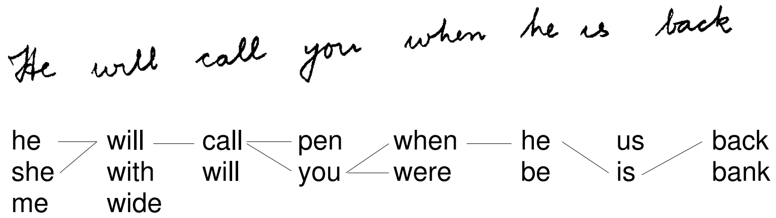

Figure 2.2: Handwritten Sentence Recognition. The path through top word choices is determined using part-of-speech tags.

Document Image Analysis

Interactive with written language recognition is the task of document image analysis. It is the task of determining the physical (spatial) and logical structure of document content. There is wide variability in the structure of documents, as in the case of newspapers, magazines, books, forms, letters and handwritten notes. In the case of a newspaper page, the objective of document analysis is to:

1. determine spatial extent of document segments and to associate appropriate labels with them, e.g., half-tone photographs, text, graphics, separating lines, etc.,

2. group image parts into meaningful units, e.g., figure and caption, heading, subheading, etc.,

3. determine reading order of blocks of text.

Document image analysis involves traditional image processing operations to printed text, such as enhancement, gray-scale image binarization, texture analysis, segmentation, etc. Additional difficult problems in the case of handwriting are: separation of lines of text, separation of words within a line, and the separation of touching characters.

2.1.4 Future Directions

Research on automated written language recognition dates back several decades. Today, cleanly printed text in documents with simple layouts can be recognized reliably by off-the-shelf OCR software. There is also some success with handwriting recognition, particularly for isolated handprinted characters and words, e.g., in the on-line case, the recently introduced personal digital assistants have practical value. Most of the off-line successes have come in constrained domains such as postal addresses (Cohen, Hull, et al., 1991), bank checks, and census forms. The analysis of documents with complex layouts, recognition of degraded printed text, and the recognition of running handwriting continue to remain largely in the research arena. Some of the major research challenges in recognizing handwriting are in: word and line separation, segmentation of words into characters, recognition of words when lexicons are large and use of language models in aiding preprocessing and recognition.

2.2 Document Image Analysis

Richard G. Casey

IBM Almaden Research Center, San Jose, California, USA

Document analysis or more precisely, document image analysis, is the process that performs the overall interpretation of document images. This process is the answer to the question, "How is everything that is known about language, document formatting, image processing and character recognition combined in order to deal with a particular application?" Thus document analysis is concerned with the global issues involved in recognition of written language in images. It adds to OCR a superstructure that establishes the organization of the document and applies outside knowledge in interpreting it.

The process of determining document structure may be viewed as guided by a model, explicit or implicit, of the class of documents of interest. The model describes the physical appearance and the relationships between the entities that make up the document. OCR is often at the final level of this process, i.e., it provides a final encoding of the symbols contained in a logical entity such as *paragraph* or *table*, once the latter has been isolated by other stages. However, it is important to realize that OCR can also participate in determining document layout. For example, as part of the process of extracting a newspaper article the system may have to recognize the character string, *continued on page 5*, at the bottom of a page image, in order to locate the entire text.

In practice then, a document analysis system performs the basic tasks of image segmentation, layout understanding, symbol recognition and application of contextual rules in an integrated manner (Wong, Casey, et al., 1982; Nagy, Seth, et al., 1985). Current work in this area can be summarized under four main classes of applications.

2.2.1 Text Documents

The ultimate goal for text systems can be termed *inverse formatting* or completion of the *Gutenberg loop*, meaning that a scanned printed document is translated back into a document description language from which it could be accurately reprinted if desired. At the research level this has been pursued in domains such as technical papers, business letters and chemical structure diagrams (Tsujimoto & Asada, 1992; Schürmann et al., 1992; Nagy, Seth, et al., 1985). Some commercial OCR systems provide limited inverse formatting, producing codes for elementary structures such as paragraphs, columns, and tables (Bokser, 1992). Current OCRs will detect, but not encode, halftones and line drawings.

In certain applications less than total interpretation of the document is required. A system for indexing and retrieving text documents may perform only a partial recognition. For example, a commercially available retrieval system for technical articles contains a model of various journal styles, assisting it to locate and recognize the title, author, and abstract of each article, and to extract keywords. Users conduct searches using the encoded material, but retrieve the scanned image of desired articles for reading.

2.2.2 Forms

Forms are the printed counterparts of relations in a data base. A typical form consists of an n-tuple of data items each of which can be represented as an ordered pair (item name, item value). OCR is used to recognize the item value; more general document analysis operations may be needed in order to identify the item name (Casey et al., 1992).

The capability for locating items on a form, establishing their name class, and encoding the accompanying data values has many applications in business and government. Form documents within a single enterprise and single application are highly repetitive in structure from one example to the next. In such a case the model for the document can consist largely of physical parameters whose values are estimated from sample documents. Such systems for gathering form data are commercially available. The Internal Revenue Service of the U.S. has recently granted a large contract to automate processing of scanned income tax forms. This will require extraction of data from a large variety of forms, as well as adaptation to perturbations of a single form resulting from different printing systems.

2.2.3 Postal Addresses and Check Reading

These applications are characterized by a well-defined logical format (but a highly variable physical layout), and a high degree of contextual constraint on the symbolic data (Srihari, 1992). The latter is potentially very useful in the attainment of high accuracy. Contextual rules can modify OCR results to force agreement of city names and postal codes, for example, or to reconcile numeric dollar amounts on checks with the written entry in the legal amount field. Contextual constraints can also assist in the detection of misrecognized documents, so that these can be handled by manual or other processes. While mailpieces and checks are actually a subclass of form documents, the large amount of effort invested in these problems justifies listing them separately.

Current equipment in use for these applications make limited use of contextual information, and is limited to reading postal codes in the case of handwritten addresses, or numeric amounts for checks. Postal machines now in development will read the complete address field and obtain greater accuracy by applying contextual constraints. At the same time they will provide a higher granularity in the sorting of mail. In the U.S., for example, new machines are planned to arrange mailpieces into delivery order for the route of individual postmen.

2.2.4 Line Drawings

Much of the activity in this area centers on entry of engineering drawings to Computer-Assisted Design / Computer-Assisted Manufacture (CADCAM) systems (Kasturi, Sira, et al., 1990; Vaxiviere & Tombre, 1992). A project for input of integrated circuit diagrams has reported cost-efficient conversion of drawings compared with conventional manual input. This project offers evidence that new circuits can most efficiently be created on paper and then encoded by recognition processes. The claim is that this is better than direct input at a terminal, due to the small screen sizes on present-day equipment. A commercial version of such a system is available. Other research in progress aims at obtaining 3-D models for multiple views in drawings of manufactured parts. Research progress has also been reported in conversion of land-use maps.

2.2.5 Future Directions

One source of motivation for work in document analysis has been the great increase in image systems for business and government. These systems provide fast storage, recall and distribution of documents in workflow processing and other applications. Document analysis can help with the indexing for storage and recall, and can partition the image into subregions of interest for convenient access by users.

In the near future such capabilities will be extended to the creation of electronic libraries which will likewise benefit from automatic indexing and formatting services. In the longer range, efforts will increase to interpret more of the information represented in the stored images, in order to provide more flexible retrieval and manipulation facilities (Dengel et al., 1992).

How will document analysis capabilities have to improve to meet future needs? There is a strong need to incorporate context, particularly language context, into the models that govern document analysis systems. Over 35 years of research and development have still not been able to produce OCR based on shape that has the accuracy of human vision. Contextual knowledge must be invoked in order both to minimize errors and to reject documents that can not be interpreted automatically. An important research issue here is how to define such constraints in a generic way, such that they can easily be redefined for different applications. Beyond this, how are such rules to be converted to software that integrates with recognition processes in order to optimize performance?

Linguistic analysis may not simply be a postprocessing stage in future document analysis systems. Modern recognition processes often perform trial segmentation of character images and choose the best segmentation from a set of alternatives using recognition confidence as a guide. Such an operation might be performed most reliably if it were implemented as a sequential process, with contextual rules governing the choice of the sequence.

In order to facilitate future progress in document analysis, there is a need for a number of scanned document data bases, each representative of a different class of documents: text, engineering drawings, addresses, forms, handwritten manuscripts, etc. Currently such collections are limited to text-oriented documents. With access to common research material, different researchers will be able to compare results and gain greater benefit from each other's efforts.

2.3 OCR: Print

Abdel Belaïd

CRIN/CNRS & INRIA, Nancy, Lorraine, France

Nowadays, there is much motivation to provide computerized document analysis systems. Giant steps have been made in the last decade, both in terms of technological supports and in software products. Character recognition (OCR) contributes to this progress by providing techniques to convert large volumes of data automatically. There are so many papers and patents advertising recognition rates as high as 99.99%; this gives the impression that automation problems seem to have been solved. However, the failure of some real applications show that performance problems subsist on composite and degraded documents (i.e., noisy characters, tilt, mixing of fonts, etc.) and that there is still room for progress. Various methods have been proposed to increase the accuracy of optical character recognizers. In fact, at various research laboratories, the challenge is to develop robust methods that remove as much as possible the typographical and noise restrictions while maintaining rates similar to those provided by limited-font commercial machines.

There is a parallel analogy between the various stages of evolution of OCR systems and those of pattern recognition. To overcome the recognition deficiency, the classical approach focusing on isolated characters has been replaced with more contextual techniques. The opening of OCR domain to document recognition leads to combination of many strategies such as document layout handling, dictionary checking, font identification, word recognition, integration of several recognition approaches with consensual voting, etc.

The rest of this section is devoted to a summary of the state of the art in the domain of printed OCR (similar to the presentations in Impedovo, Ottaviano, et al., 1991; Govindan & Shivaprasad, 1990; Nadler, 1984; Mantas, 1986), by focussing attention essentially on the new orientations of OCR in the document recognition area.

2.3.1 Document Image Analysis Aspects

Characters are arranged in document lines following some typesetting conventions which we can use to locate characters and find their style. Typesetting rules can help in distinguishing such characters as *s* from *5*, *h* from *n*, and *g* from *9*, which can be often confused in multifont context (Kahan, Pavlidis, et al., 1987). They can also limit the search area according to characters' relative positions and heights with respect to the baseline (Luca & Gisotti, 1991a; Luca & Gisotti, 1991b; Kanai, 1990). The role of typesetting cues to aid document understanding is discussed by Holstege, Inn, et al. (1991).

Layout Segmentation

Location of characters in a document is always preceded by a layout analysis of the document image. The layout analysis involves several operations such as determining the skew, separating picture from text, and partitioning the text into columns, lines, words, and connected components. The portioning of text is effected through a process known as segmentation. A survey of segmentation techniques is given in Nadler (1984).

Character Building

In building character images, one is often confronted with touching or broken characters that occur in degraded documents (such as fax, photocopy, etc.). It is still challenging to develop techniques for properly segmentating words into their characters. Kahan, Pavlidis, et al. (1987) detected touching characters by evaluation of vertical pixel projection. They executed a branch-and-bound search of alternative splittings and merges of symbols pruned by word-confidence scores derived from symbol confidence. Tsujimoto and Asada (1991) used a decision tree for resolving ambiguities. Casey and Nagy (1982) proposed a recursive segmentation algorithm. Liang, Ahmadi, et al. (1993) added to this algorithm contextual information and a spelling checker to correct errors caused by incorrect segmentation. Bayer (1987) proposed a hypothesis approach for merging and splitting characters. The hypotheses are tested by several experts to see whether they represent a valid character. The search is controlled by the A^* algorithm resolving backtracking processing. The experts comprise the character classifier and a set of algorithms for context processing.

Font Consideration

A document reader must cope with many sources of variations, notably that of font and size of the text. In commercial devices, the multifont aspect was for a long time neglected for the benefit of speed and accuracy, and substitution solutions were proposed. At first, to cater for some institutions, the solution was to work on customized fonts (such as OCR-A and OCR-B) or on a selected font from a trained library to minimize the confusion between similar looking characters. The accuracy was quite good, even on degraded images on the condition that the font is carefully selected. However, recognition scores drop rapidly when fonts or sizes are changed. This is due to the fact that the limitation to one font naturally promotes the use of simple and sensitive pattern recognition algorithms, such as template matching (Duda & Hart, 1973).

In parallel with commercial investigations, the literature proposed multifont recognition systems that are based on typographical features. Font information is inherent in the constituent characters (Rubinstein, 1988) and feature-based methods are less font sensitive (Srihari, 1984; Ullman, 1973; Kahan, Pavlidis, et al., 1987). Two research paths were taken with multifont machines. One gears towards the office environment. This introduced systems which can be trained by the user to read any given font (Schurmann, 1978; Shlien, 1988; Belaïd & Anigbogu, 1991; Anigbogu & Belaïd, 1991a; Anigbogu & Belaïd, 1991b). The system is only able to recognize a font from among those learned. The others try to be font independent. The training is based on pattern differentiation rather than on font differentiation (Lam & Baird, 1987; Baird, Kahan, et al., 1986; Baird & Fossey, 1991).

2.3.2 Character Recognition

Feature Extraction

This step is crucial in the context of document analysis where several variations may be caused by a number of different sources: geometric transformation because of low data quality, slant and stroke width variation because of font changing, etc. It seems reasonable to look for features which are invariant and which capture the characteristics of the character by filtering out all attributes which make the same character assume different appearances. The classifier could store a single prototype per character. Schurmann, Bartneck, et al. (1992) applies normalizing transformations to reduce certain well-defined variations as far as possible. The inevitably remaining variations are left for learning by statistical adaptation of the classifier.

Character Learning

The keys of printed character learning are essentially training set and classification adaptation to new characters and new fonts. The training set can be given either by user or extracted directly from document samples. In the first case, the user selects the fonts and the samples to represent each character in each font and then guides the system to create models as in Anigbogu and Belaïd (1991b). Here, the user must use sufficient number of samples in each font according to the difficulty of its recognition. However, it is difficult in an omnifont context to collect a training set of characters having the expected distribution of noise and pitch size. Baird (1990) suggested parameterized models for imaging defects, based on a variety of theoretical arguments and empirical evidence. In the second case, the idea is to generate the training set directly from document images chosen from a wide variety of fonts and image quality and to reflect the variability expected by the system (Bokser, 1992). The problem here is that one is not sure that all valid characters are present.

Contextual Processing

Contextual processing attempts to overcome the shortcoming of decisions made on the basis of local properties and to extend the perception on relationships between characters into word. Most of the techniques try to combine geometric information, as well as linguistic information. See Srihari and Hull (1985) for an overview of these techniques. Anigbogu and Belaïd (1991a); Anigbogu and Belaïd (1991b); Belaïd and Anigbogu (1991) used hidden Markov models for character and word modeling. Characters are merged into groups which are matched against words in a dictionary using Ratcliff/Obershelp pattern matching method. In the situation where no acceptable words are found, the list of confused characters is passed through a Viterbi net and the output is taken as the most likely word. The bigram and character position-dependent probabilities used for this purpose were constructed from a French dictionary of some 190,000 words. The word-level recognition stands at over 98%.

2.3.3 Commercial Products

Commercial OCR machines came in practically at the beginning of 1950s and have evolved in parallel with research investigations. The first series of products heavily relied on customized fonts, good printing quality and very restricted document layout. Nowadays, we can find a vast range of products, more powerful than the previous ones. Among these are certain hand-held scanners, page readers, and integrated flat-bed and document readers. The tendency is to use the fax machine as an image sensor. Instead of printing the fax message on paper, it is taken directly as input to an OCR system. It is to be noted that the obtained images are of a poor quality. The challenge in this area is the development of high performing tools to treat degraded text with results as good as those of classical OCRs.

OCR is used in three main domains: the banking environment for data entry and checking, office automation for text entry, and the post office for mail sorting. We can find many surveys on commercial products in Mori, Suen, et al. (1992); Mantas (1986); Bokser (1992); Nagy (1992). Recently, the Information Science Research Institute had the charge to test technologies for OCR from machine printed documents.

A complete review has been published (Nartker, Rice, et al., 1994) giving a benchmark of different products in use in the U.S. market.

2.3.4 Future Directions

We have attempted to show that OCR is an essential part of the document analysis domain. Character recognition cannot be achieved without typesetting cues to help the segmentation in a multifont environment. We have also shown the unavoidable recourse to linguistic context; the analysis must be extended to this domain. The training still remains the weak side of OCR for now, as it is difficult to generate a training set of characters which includes all the variability the system will be expected to handle. Finally, it would appear more and more that in real-world OCR many different techniques must be combined to yield high recognition scores (Anigbogu & Belaïd, 1991b; Ho, 1992). For this reason, the tendency is to combine the results of many OCR systems in order to obtain the best possible performance.

2.4 OCR: Handwriting

Claudie Faure & Eric Lecolinet

Télécom Paris, Paris, France

2.4.1 The Domain

For more than thirty years, researchers have been working on handwriting recognition. As in the case of speech processing, they aimed at designing systems able to understand personal encoding of natural language.

Over the last few years, the number of academic laboratories and companies involved in research on handwriting recognition has continually increased. Simultaneously, commercial products have become available. This new stage in the evolution of handwriting processing results from a combination of several elements: improvements in recognition rates, the use of complex systems integrating several kinds of information, the choice of relevant application domains, and new technologies such as high quality high speed scanners and inexpensive powerful CPUs. A selection of recent publications on this topic is: Impedovo (1994); IWFHR (1993); Plamondon (1993); Pavlidis and Mori (1992); Impedovo and Simon (1992); Wang (1991).

Methods and recognition rates depend on the level of constraints on handwriting. The constraints are mainly characterized by the types of handwriting, the number of scriptors, the size of the vocabulary and the spatial layout. Obviously, recognition becomes more difficult when the constraints decrease. Considering the types of roman script (roughly classified as hand printed, discrete script and cursive script), the difficulty is lower for handwriting produced as a sequence of separate characters than for cursive script which has much in common with continuous speech recognition. For other writing systems, character recognition is hard to achieve, as in the case of Kanji which is characterized by complex shapes and a huge number of symbols.

The characteristics which constrain hand writting may be combined in order to define handwriting categories for which the results of automatic processing are satisfactory. The trade-off between constraints and error rates give rise to applications in several domains. The resulting commercial products have proved that handwriting processing can be integrated into working environments. Most efforts have been devoted to mail sorting, bank check reading, forms processing in administration and insurance. These applications are of great economic interest, each of them concerning millions of documents.

Mail sorting is a good illustration of the evolution in the domain. In this case, the number of writers is unconstrained. In the early stages, only ZIP code was recognized. Then cities (and states as in the U.S.) were processed, implying the recognition of several types of handwriting: hand printed, cursive, or a mixture of both. The use of the redundancy between the ZIP code and the city name, as well as redundancy between numeral and literal amounts in bank checks, shows that combining several sources of information improves the recognition rates. Today, the goal is to read the full address, down to the level of the information used by the individual carrier. This necessitates precisely extracting the writing lines, manipulating a very large vocabulary and using contextual knowledge as the syntax of addresses (such as in the case of reading the literal amount of checks, the use of syntactic rules improves the recognition).

These new challenges bring the ongoing studies closer to unconstrained handwritten language processing which is the ultimate aim. The reading of all of the handwritten and printed information present on a document is necessary to process it automatically, to use content dependent criteria to store, access and transmit it and to check its content. Automatic handwritten language processing will also allow one to convert and to handle manuscripts produced over several centuries within a computer environment.

2.4.2 Methods and Strategies

Recognition strategies heavily depends on the nature of the data to be recognized. In the cursive case, the problem is made complex by the fact that the writing is fundamentally ambiguous as the letters in the word are generally linked together, poorly written and may even be missing. On the contrary, hand printed word recognition is more related to printed word recognition, the individual letters composing the word being usually much easier to isolate and to identify. As a consequence of this, methods working on a letter basis (i.e., based on character segmentation and recognition) are well suited to hand printed word recognition while cursive scripts require more specific and/or sophisticated techniques. Inherent ambiguity must then be compensated by the use of contextual information.

Intense activity was devoted to the character recognition problem during the seventies and the eighties and pretty good results have been achieved (Mori, Suen, et al., 1992). Current research is rather focusing on large character sets like Kanji and on the recognition of handwritten roman words. The recognition of handwritten characters being much related to printed character recognition, we will mainly focus on cursive word recognition.

Character Recognition

Character Recognition techniques can be classified according to two criteria: the way preprocessing is performed on the data and the type of the decision algorithm.

Preprocessing techniques include three main categories: the use of global transforms (correlation, Fourier descriptors, etc.), local comparison (local densities, intersections with straight lines, variable masks, characteristic loci, etc.) and geometrical or topological characteristics (strokes, loops, openings, diacritical marks, skeleton, etc.).

Depending on the type of preprocessing stage, various kinds of decision methods have been used such as: various statistical methods, neural networks, structural matchings (on trees, chains, etc.) and stochastic processing (Markov chains, etc.). Many recent methods mix several techniques together in order to provide a better reliability to compensate the great variability of handwriting.

Handwritten Word Recognition

As pointed out in the chapter overview, two main types of strategies have been applied to this problem since the beginning of research in this field: the holistic approach and the analytical approach (Lecolinet & Baret, 1994; Lorette & Lecourtier, 1993; Hull, Ho, et al., 1992; Simon, Baret, et al., 1994). In the first case recognition is globally performed on the whole representation of words and there is no attempt to identify characters individually.

The main advantage of holistic methods is that they avoid word segmentation (Rocha & Pavlidis, 1993). Their main drawback is that they are related to a fixed lexicon of word descriptions: as these methods do not rely on letters, words are directly described by means of features and adding new words to the lexicon require human training or the automatic generation of word descriptions from ASCII words. These methods are generally based on dynamic programming (DP) (edit distance, DP-matching, etc.) or model-discriminant hidden Markov models.

Analytical strategies deal with several levels of representation corresponding to increasing levels of abstraction (usually the feature level, the grapheme or pseudo-letter level and the word level). Words are not considered as a whole, but as sequences of smaller size units which must be easily related to characters in order to make recognition independent from a specific vocabulary.

These methods are themselves subclassed into two categories: analytical methods with explicit (or *external*) segmentation where grapheme or pseudo-letter segmentation takes place before recognition (Lecolinet & Crettez, 1991) and analytical methods with implicit (or *internal*) segmentation (Burges, Matan, et al., 1992; Chen, Kundu, et al., 1992) which perform segmentation and recognition simultaneously (segmentation is then a *by-product* of recognition). In both cases, lexical knowledge is heavily used to help recognition. This lexical knowledge can either be described by means of a lexicon of ASCII words (which is often represented by means of a lexical tree) or by statistical information on letter co-occurrence (n-grams, transitional probabilities, etc.). The advantage of letter-based recognition methods is that the vocabulary can be dynamically defined and modified without the need of word training.

Many techniques initially designed for character recognition (like neural networks Burges, Matan, et al., 1992) have been incorporated to analytical methods for recognizing tentative letters or graphemes. The contextual phase is generally based on dynamic programming and/or Markov chains (edit distance, Viterbi algorithm, etc.). Fruitful research has been realized in recent years in the field of analytic recognition with implicit segmentation using various kinds of hidden Markov models (Chen, Kundu, et al., 1992).

2.4.3 Future Directions

Exploitable results can already be obtained when the data is sufficiently constrained. Commercial products are already available for hand printed characters recognition in forms and recent research projects have shown that cursive word recognition is feasible for small lexicons and/or when strong sentence syntax is provided. For instance recognition rates of 95% (respectively 90%) or more have been obtained for lexicons of American city names whose size varies between 10 and 100 (respectively 1000) words (Kimura, Shridhar, et al., 1993).

Recent studies show the emergence of two promising tendencies:

1. hybrid systems that combine several recognition techniques; and
2. the use of contextual analysis at word, sentence or text level to predict or confirm word recognition.

This is already the direction that several major research teams have decided to follow (Hull, 1994) and there is no doubt that contextual analysis will be a field of intense research and achievements in the next few years.

2.5 Handwriting as Computer Interface

Isabelle Guyon & Colin Warwick

AT&T Bell Laboratories, Holmdel, New Jersey, USA

2.5.1 Pen Computers: Dream and Reality

Pen computers (Forman & Zahorjan, 1994) offer an interesting alternative to paper. One can write directly on a Liquid Crystal Display (LCD) screen with a stylus or *pen*. The screen has an invisible sensitive matrix which records the position of the pen on the surface. The trajectory of the pen appears almost instantaneously on the screen giving the illusion of ink (electronic ink). Handwriting recognition allows text and computer commands to be entered.

While nothing opposes the idea of a computer that would use multiple input modalities including speech, keyboard and pen, some applications call for a pen-only computer interface: in a social environment, speech does not provide enough privacy; for small hand-held devices and for large alphabet (e.g., Chinese), the keyboard is cumbersome. Applications are numerous: personal organizer, personal communicator, notebook, data acquisition device for order entries, inspections, inventories, surveys, etc.

The dream is to have a computer that looks like paper, feels like paper but is better than paper. Currently, paper is the most popular medium for sketching, note taking and form filling, because it offers a unique combination of features: light, cheap, reliable, available almost everywhere any time, easy to use, flexible, foldable, pleasing to the eye and to the touch, silent. But paper has also its drawbacks: in large quantities it is no longer light and cheap, it is hard to reuse and recycle, difficult to edit, expensive to copy and to mail, inefficient to transform into computer files. With rapid technology progress, electronic ink could become cheaper and more convenient than paper, if only handwriting recognition worked!

As of today, the mediocre quality of handwriting recognition has been a major obstacle to the success of pen computers. Users report that it is "too inaccurate, too slow and too demanding for user attention" (Chang & Scott MacKenzie, 1994). The entire pen computing industry is turning its back on handwriting and reverting to *popup keyboards*. On small surfaces, keypad tapping is difficults and slow: 10–21 words per minute, compared to 15–18 wpm for handprint and 20–32 wpm for a full touch screen keyboard. However, it remains the preferred entry mode because of its low error rate: less than 1% for the speed quoted, compared to 5–6% with a state-of-the-art recognizer (CIC) (MacQueen, Scott MacKenzie, et al., 1994; Chang & Scott MacKenzie, 1994). In one of our recent studies, we discovered that a good typist tolerates only up to 1% error using a special keyboard that introduced random typing errors at a software-controllable rate; 0.5% error is unnoticeable; 2% error is intolerable! (Warwick, 1995) Human subjects make 4–8% error for isolated letters read in the absence of context and 1.5% error with the context of the neighboring letters (Wilkinson, Geist, et al., 1992; Geist et al., 1994). Therefore, the task of designing usable handwriting recognizers for pen computing applications is tremendously hard. Human recognition rates must be reached and even outperformed.

2.5.2 The State of the Art in On-line Handwriting Recognition

The problem of recognizing handwriting recorded with a digitizer as a time sequence of pen coordinates is known as *on-line handwriting recognition*. In contrast, *off-line handwriting recognition* refers to the recognition of handwritten paper documents which are optically scanned.

The difficulty of recognition varies with a number of factors:

- Restrictions on the number of writers.
- Constraints on the writer: entering characters in boxes or in combs, lifting the pen between characters, observing a certain stroke order, entering strokes with a specific shape.
- Constraints on the language: limiting the number of symbols to be recognized, limiting the size of the vocabulary, limiting the syntax and/or the semantics.

Until the beginning of the nineties, on-line handwriting recognition research was mainly academic and most results were reported in the open literature (Tappert, Suen, et al., 1990). The situation has changed in the past few years with the rapid growth of the pen computing industry. Because of the very harsh competition, many companies do no longer publish in the peer reviewed literature and no recent general survey is available.

In the last few years, academic research has focussed on cursive script recognition (Plamondon, 1995c; Lecolinet & Baret, 1994). Performances are reported on different databases and are difficult to compare. It can be said, with caution, that the state of the art for writer independent recognition of isolated English cursive words, with an alphabet of 26 letters, and with a vocabulary of 5,000–10,000 words, is between 5% and 10% character error rate and between 15% and 20% word error rate.

Most commercial recognizers do writer independent recognition and can recognize characters, words or sentences, with either characters written in boxes or combs, or in run-on mode with pen-lifts between characters (e.g., CIC, AT&T-EO, Grid, IBM, Microsoft, Nestor). In addition, those systems recognize a set of gestures and can be trained with handwriting samples provided by the user. Some companies provide recognizers for both Latin and Kanji alphabets (e.g., CIC). Companies like Paragraph International and Lexicus offer cursive recognition. Palm Computing recently introduced a recognizer for a simplified alphabet (similarly as Goldberg & Richardson, 1993). It presumably reaches below 1% error, but no controlled benchmark has been performed yet.

AT&T-GIS anonymously tested seven Latin alphabet recognizers, including five commercial recognizers, using an alphabet of 68 symbols (uppercase, lowercase, digits and six punctuation symbols) on two different tasks (Allen, Hunter, et al., 1994):

- The recognition of isolated characters written in boxes;
- The recognition of American addresses written in run-on mode on a baseline, without the help of boxes or combs, but with pen-lifts between characters. The vocabulary list was not disclosed.

The first task imposes constraints on the writer, but not on the language. Without any contextual information given by neighboring characters which are part of a same word or sentence, it is impossible to distinguish between the digit "0," the letter "O," and the letter "o." Even humans make errors on such cases which we call *legitimate* errors. If all errors are counted, including the legitimate errors, the best recognizer has a 19% error rate. This error-rate is reduced by more than half if legitimate errors are removed. On such a data set, humans still make approximately half as many errors. Much higher recognition rates are obtained on subsets of the characters set which do not contain intrinsic ambiguities. For instance, less than 2% error can be obtained on digits only, which is close to the human performance on the same task.

The second task imposes less constraints on the writer, thus characters are harder to segment. However, the recognizers can use neighboring letters to determine relative character positions and relative sizes, which is helpful to discriminate between uppercase and lowercase letters. Using only such limited contextual information, the best recognizer has a 30% character error rate (including insertions, substitutions and deletions). Use can also be made of a model of language to help correcting recognition mistakes. The performance of the best recognizer using an English lexicon and a letter trigram model was 20% character error. Humans perform considerably better than machines on this task and make only a few percent error.

2.5.3 A Brief Review of On-line Handwriting Recognition Techniques

Considerably more effort has been put in developing algorithms for Optical Character Recognition (OCR) and speech recognition than for on-line handwriting recognition. Consequently, on-line handwriting recognition, which bears similarity to both, has been borrowing a lot of techniques from them.

There is a natural temptation to convert pen trajectory data to pixel images and process them with an OCR recognizer. But, the on-line handwriting recognition problem has a number of distinguishing features which must be exploited to get best results:

- **preprocessing** operations such as smoothing, deslanting and deskewing and dehooking and **feature extraction** operations such as the detection of line orientations, corners, loops and cusps are easier and faster with the pen trajectory data than on pixel images.
- **discrimination** between optically ambiguous characters (for example, "j" and ";") may be facilitated with the pen trajectory information.
- **segmentation** operations are facilitated by using the pen-lift information, particularly for handprinted characters.
- **immediate feed-back** is given by the writer whose corrections can be used to further train the recognizer.

Another temptation is to use the pen trajectory as a temporal signal and process it with a speech recognizer. Other problems arise:

- **stroke reordering** is usually necessary, to get rid of stroke order variability and of the problem of delayed strokes.
- **data unfolding** in a purely one-dimensional representation may result in loosing direct reference to the two-dimensional structure of the data.

Classically, on-line recognizers consist of a preprocessor, a classifier which provides estimates of probabilities for the different categories of characters (or other subword units) and dynamic programming postprocessor (often a Hidden Markov Model) which eventually incorporates a language model (ICDAR, 1993; Hanson, Cowan, et al., 1993; ICASSP, 1994). The system has usually adjustable parameters which values are determined during a training session. The Expectation Maximization (EM) algorithm (or its *K-means* approximation) is used to globally optimize all parameters.

While all postprocessors are very similar, a wide variety of classifiers have been used, including statistical classifiers, Bayesian classifiers, decision trees, neural networks and fuzzy systems. They present different speed/accuracy/memory tradeoffs but none of them significantly outperforms all others in every respects. On-line systems also differ from one another in data representations which range from 2-dimensional maps of pixels or features to temporal sequences of features, and from local low level features to the encoding of entire strokes.

2.5.4 Future Directions

Only a few years ago, cursive handwriting recognition seemed out of reach. Today the dream has become reality. Yet, recognizers currently available are still disappointing to users. There is a wide margin for improvement which should challenge researchers and developers.

Because of the lack of success of the first generation of pen computers, the industry is currently focusing two kinds of products:

- Data acquisition devices for form filling applications requiring only a limited alphabet and allowing very constrained grammars or language models. Users such as commercial agents would be willing to print characters in boxes or combs.
- Personal Digital Assistants combining agenda, address book and telecommunications facilities (phone, fax and mail). Users would want to use natural unconstrained handwriting, cursive or handprinted.

In the short term, to meet the accuracy requirements of industry applications, it is important to focus on simplified recognition tasks such as limited vocabulary handprinted character recognition. In the long term, however, research should be challenged by harder tasks such as large vocabulary cursive recognition.

Hardware constraints presently limit commercial recognizers but the rapid evolution of computer hardware ensures that within two to three years discrepancies between the processing power of portable units and today's workstations will disappear. Therefore, it seems reasonable to use as a metric the processing power of today's workstations and concentrate most of the research effort on improving recognition accuracy rather than optimizing algorithms to fulfill today's speed and memory requirements.

To be able to read cursive writing, humans make use of sources of information that are still seldom taken into account in today's systems:

- elaborate language model;
- writing style model.

The success of incorporating both kind of models in speech recognition systems is an encouragement for handwriting recognition researchers to pursue in that direction.

Finally, there is often a large discrepancy between the error rate obtained in laboratory experiments and those obtained on the field. Recognizers should be tested, as far as possible, in realistic conditions of utilizations or at least on realistic test data. With projects such as UNIPEN (Guyon, Schomaker, et al., 1994), it will be possible to exchange a wide variety of data and organize public competitions.

2.6 Handwriting Analysis

Rejean Plamondon

Ecole Polytechnique de Montréal, Montréal, Québéc, Canada

2.6.1 Problem Statement

As in many well-mastered tasks, human subjects generally work at the highest and most efficient level of abstraction possible when reading a handwritten document. When difficulties are encountered in decyphering a part of the message using one level of interpretation, they often switch to a lower level of representation to resolve ambiguities. In this perspective, the lower levels of knowledge, although generally used in the background, constitute a cornerstone on which a large part of the higher and more abstract process levels relies. For example, according to motor theories of perception, it is assumed that motor processes enter into genesis of percepts and that handwriting generation and perception tasks interact and share sensorimotor information. Cursive script recognition or signature verfication tasks therefore require, directly or indirectly, an understanding of the handwriting generation processes.

Consistent with these hypotheses, some design methodologies incorporate this theoretical framework in the development of automatic handwriting processing systems. So far, numerous models have been proposed to study and analyze handwriting (Plamondon & Maarse, 1989; Plamondon, Suen, et al., 1989; Galen & Stelmach, 1993; Faure, Lorette, et al., 1994). Depending on the emphasis placed on the symbolic information or on the connectionist architecture, two complementary approaches have been followed: top-down and bottom-up. The top-down approach has been developed mainly by those researchers interested in the study and application of the various aspects of the high-level motor processes: fundamental unit of movement coding, command sequencing and retrieval, movement control and learning, task complexity, etc. The bottom-up approach has been used by those interested in the analysis and synthesis of the low-level neuromuscular processes. For this latter approach to be of interest in the study of the perceptivomotor strategies involved in the generation and perception of handwriting, two criteria must be met. On the one hand, a model should be realistic enough to reproduce specific pentip trajectories almost perfectly and, on the other, its descriptive power should be such that it provides consistent explanations of the basic properties of single strokes (asymmetric bell-shaped velocity profiles, speed accuracy trade-offs, etc.). In the other words, the most interesting bottom-up models should allow the link to be made between the symbolic and connectionist approaches.

2.6.2 A Model of the Handwriting Generation System

A serious candidate model for a basic theory of human movement generation, in the sense that it addresses some of the key problems related to handwriting generation and perception, is based on two basic assumptions. First, it supposes that fast handwriting, like any other highly skilled motor process, is partially planned in advance (Lashley, 1987; van der Gon & Thuring, 1965), with no extra control during execution of a continuous trace of handwritten text, hereafter called a string (Plamondon, 1989b). Second, it assumes some form of rotation invariance in movement representation and used differential geometry to describe a handwritten string by its change of line curvature as a function of the curvilinear abscissa (Plamondon, 1989a).

In this context, a string can be described by a sequence of virtual targets that have to be reached within a certain spatial precision to guarantee the message legibility. Each individual stroke can be seen as a way to map these targets together in a specific two dimensional space. To produce a continuous and fluent movement, it is necessary to superimpose these discrete movement units in time, that is to start a new stroke, described by its own set of parameters, before the end of the previous one. This superimposition process is done vectorially in a 2D space. A complex velocity pattern, representing a word, thus emerges from the vectorial addition of curvilinear strokes.

A general way to look at the impulse response of a specific controller, say the module controller, is to consider the overall sets of neural and muscle networks involved in the production of a single stroke as a synergetic linear system producing a curvilinear velocity profile from an impulse command of amplitude D occurring at t_0 (Plamondon, 1992). The curvilinear velocity profile thus directly reflects the impulse response H_{t-t_0} of neuromuscular synergy.

The mathematical description of this impulse response can be specified by considering each controller as composed of two systems that represent the sets of neural and muscular networks involved in the generation of the agonist and antagonist activities resulting in a specific movement (Plamondon, 1995b). Although various forms of interaction and coupling between these two systems probably exist throughout the process, we assume that their global effect can be taken into account at the very end of the process by subtracting the two outputs. If so, each of the systems constituting a controller can be considered as a linear time-invariant system and the output of a controller as the difference between the impulse responses of the agonist and antagonist systems, weighed by the respective amplitude of their input commands. The mathematical description of an agonist or antagonist impulse response can be specified if the sequential aspects of the various processing steps occurring within a system are taken into account. Indeed, as soon as an activation command is given, a sequence of processes goes into action. The activation command is propagated and a series of neuromuscular networks react appropriately to it. Due to the internal coupling between each of the subprocesses, one stage is activated before the activation of the previous one is completed. Within one synergy, the coupling between the various subprocesses can thus be taken into account by linking the time delays of each subprocess.

Using a specific coupling function and making an analogy between this function and the predictions of the central-limit theorem, as applied to the convolution of a large number of positive functions, it is predicted that the impulse response of a system under the coupling hypothesis will converge toward a log-normal curve (Plamondon, 1995b) provided that the individual impulse response of each subsystem meets some very general conditions (real, normalized and non-negative, with a finite third moment and scaled dispersion). So, under these conditions, the output of the module or the direction controller will be described by the weighted difference of two lognormals, hereafter called a delta lognormal equation (Plamondon, 1995b).

In this context, the control of the velocity module can now be seen as resulting from the simultaneous activation (at $t = t_0$) of a controller made up of two antagonistic neuromuscular systems, with a command of amplitude D_1 and D_2 respectively. Both systems react to their specific commands with an impulse response described by a lognormal function, whose parameters characterize the time delay and the response time of each process (Plamondon, 1995b).

One of the most stringent conclusion of this model, apart from its consistency with the major psychophyusical phenomena regularly reported in studies dealing with speed/accuracy tradeoffs, is that the angular component of the velocity vector just emerges from this superimposition process and is not controlled independently by a specific delta lognormal generator (Plamondon, 1995a). Each string is thus made up of a combination of curvilinear strokes, that is, curvilinear displacements characterized by delta-lognormal velocity profiles. Strokes can be described in terms of nine different parameters: t_0, the time occurence of a synchronous pair of input command; D_1 and D_2, the amplitude of agonist and antagonist commands respectively; μ_1, μ_2 and σ_1, σ_2, the logtime delays and the logresponse times of the agonist and the antagonist systems; θ_0 and C_0, the initial postural conditions, that is, stroke orientation and curvature. In this general context, a curvilinear stroke is thus defined as a portion of the pentip trajectory that corresponds to the curvilinear displacement resulting from the production of a delta-lognormal velocity profile, produced by a specific generator in response to a specific pair of impulse commands fed into it. These strokes are assumed to be the fundamental units of human handwriting movement and serve as the coding elements of the motor plan used in trajectory generation.

2.6.3 Testing the Model

Several comparative studies have been conducted to test and validate this model (Plamondon, Alimi, et al., 1993; Alimi & Plamondon, 1994; Alimi & Plamondon, 1993). Without entering into the details of each study, let us simply point out that it was concluded that the delta equation was the most powerful in reconstructing curvilinear velocity profiles and that its parameters were consistent with the hierarchical organization of the movement generation system. Computer simulations have also demonstrated that the delta lognormal model predicts the majority of phenomena consistently reported by many research groups studying the velocity profiles of simple movements (Plamondon, 1995b).

2.6.4 Conclusion

Further, the delta lognormal model provides a realistic and meaningful way to analyze and describe handwriting generation and provides information that can be used, in a perceptivomotor context to tackle recognition problems. Its first practical application has been the development of a model-based segmentation framework for the partitioning of handwriting (Plamondon, 1992) and its use in the development of an automatic signature verification system (Plamondon, 1994b). Based on this model, a multilevel signature verification system was developed (Plamondon, 1994a), which uses three types of representations based on global parameters and two other based on functions. The overall verification is performed using a step wise process at three distinct levels, using personalized decision thresholds.

2.6.5 Future Directions

As long as we partially succeed in processing handwriting automatically by computer, we will see on-line tools designed to help children learn to write appearing on the market, as well as intelligent electronic notebooks, signature verification, and recognition systems, not to mention the many automated off-line systems for processing written documents.

In order to see these newest inventions (all of which are dedicated to the popularization of handwriting) take shape, become a reality, and not be relegated to the status of laboratory curios, a great deal of research will be required, and numerous theoretical and technological breakthroughs must occur. Specifically, much more time and money must be spent on careful research and development, but with less of the fervor that currently prevails. False advertising must be avoided at all costs when technological breakthroughs are made, when development is still far from complete and any undue optimism arising from too many premature expectations risks compromising the scientific achievement.

In this perspective, multidisciplinarity will play a key role in the future developments. Handwriting is a very complex human task that involves emotional, rational, linguistic and neuromuscular functions. Implementing any pen-based system requires us to take a few of these aspects into account. To do so, we have to understand how we control movements and how we perceive line images. Any breakthrough in the field will come from a better modeling of these underlying processes at different levels with various points of view. The intelligent integration of these models into functional systems will require the cooperation of scientists from numerous complementary disciplines. It is a real challenge for patient theoreticians.

2.7 Chapter References

Alimi, A. and Plamondon, R. (1993). Performance analysis of handwritten strokes generation models. In *Proceedings of the Third International Workshop on Frontiers in Handwriting Recognition*, pages 272–283, Buffalo, New York.

Alimi, A. and Plamondon, R. (1994). Analysis of the parameter dependence of handwriting generation models on movements characteristics. In Faure, C., Lorette, G., Vinter, A., and Keuss, P., editors, *Advances in Handwriting and Drawing: A multidisciplinary Approach*, Paris. Presse de l'Ecole Nationale Superieure de Telecommunication.

Allen, T., Hunter, W., Jacobson, M., and Miller, M. (1994). Comparing several discrete handwriting recognition algorithms. Technical report, AT&T GIS, Human Interface Technology Center, New User Interface Group.

Anigbogu, J. C. and Belaïd, A. (1991a). Application of hidden Markov models to multifont text recognition. In *Proceedings of the First International Conference on Document Analysis and Recognition*, volume 2, pages 785–793, St. Malo, France. AIPR-IEEE, IAPR.

Anigbogu, J. C. and Belaïd, A. (1991b). Recognition of Multifont Text Using Markov Models. In *7^{th} Scandinavian Conference on Image Analysis*, volume 1, pages 469–476.

Baird, H. S. (1990). Document Image Defect Models. In *Proceedings of the Workshop on Syntactical and Structural Pattern Recognition*, pages 38–47.

Baird, H. S. and Fossey, R. (1991). A 100-font classifier. In *Proceedings of the First International Conference on Document Analysis and Recognition*, volume 1, pages 332–340, St. Malo, France. AIPR-IEEE, IAPR.

Baird, H. S., Kahan, S., and Pavlidis, T. (1986). Components of a Omnifont Page Reader. In *Proceedings of the 8th International Conference on Pattern Recognition*, pages 344–348, Paris.

Bayer, T. (1987). Segmentation of Merged Character Patterns with Artificial Intelligence Techniques. In *Scandinavian Conference on Image Analysis*, pages 49–55, Stockholm.

Belaïd, A. and Anigbogu, J. C. (1991). Text recognition using stochastic models. In Gutiérrez, R. and Valderrama, M. J., editors, *5^{th} International Symposium on ASMDA*, pages 87–98. World Scientific.

Bokser, M. (1992). Omnidocument Technologies. *Proceedings of the IEEE*, 80(7):1066–1078.

Burges, C. J. C., Matan, O., Le Cun, Y., Denker, J. S., Jackel, L. D., Stenard, C. E., Nohl, C. R., and Ben, J. I. (1992). Shortest path segmentation: A method for training a neural network to recognize character strings. In *Proceedings of the 1992 International Joint Conference on Neural Networks*, volume 3, page 165, Baltimore, Maryland.

Casey, R. G. et al. (1992). Intelligent forms processing system. *Machine Vision Applications*, 5:143–155.

Casey, R. G. and Nagy, G. (1982). Recursive Segmentation and Classification of Composite Patterns. In *Proceedings of the 6th International Conference on Pattern Recognition*, pages 1023–1026.

Chang, L. and Scott MacKenzie, I. (1994). A comparison of two handwriting recognizers for pen-based computers. In *Proceedings of CANSON'94*, Canada.

Chen, M. Y., Kundu, A., Zhou, J., and Srihari, S. N. (1992). Off-line handwritten word recognition using hidden Markov model. In *Proceedings of the USPS Advanced Technology Conference*, page 563.

Cohen, E., Hull, J. J., and Srihari, S. N. (1991). Understanding handwritten text in a structured environment: determining zip codes from addresses. *International Journal of Pattern Recognition and Artificial Intelligence*, 5(1-2):221–264.

Coulmas, F. C. (1989). *The Writing Systems of the World*. Blackwell, New York.

Dengel, A. et al. (1992). From paper to office document standard representation. *IEEE Computer*, 25(7):63–67.

Duda, R. O. and Hart, P. E. (1973). *Pattern Recognition and Scene Analysis*. John Wiley, New York.

Faure, C., Lorette, G., Vinter, A., and Keuss, P. (1994). *Advances in Handwriting and Drawing: A Multidisciplinary Approach*. Europia, Paris.

Forman, G. and Zahorjan, J. (1994). The challenges of mobile computing. *IEEE Computer*, pages 38–47.

Galen, V. and Stelmach, G. E. (1993). Handwriting, issue of psychomotor control and cognitive models. *Acta Psychologica*, 82(1-3).

Geist, J. et al. (1994). The second census optical character recognition systems conference. Technical Report NISTIR-5452, National Institute of Standards and Technology, U.S. Department of Commerce.

Goldberg, D. and Richardson, D. (1993). Touch-typing with a stylus. In *INTERCHI'93 conference on Human factors in computer systems*, New York. ACM.

Govindan, V. K. and Shivaprasad, A. P. (1990). Character recognition—a review. *Pattern Recognition*, 23(7):671–683.

Govindaraju, V., Shekhawat, A., and Srihari, S. N. (1993). Interpretation of handwritten addresses in U.S. mail stream. In *Proceedings of the Second International Conference on Document Analysis and Recognition*, Tsukuba Science City, Japan. AIPR-IEEE, IAPR.

Guyon, I., Schomaker, L., Plamondon, R., Liberman, M., and Janet, S. (1994). UNIPEN project of on-line data exchange and recognizer benchmarks. In *Proceedings of the 12th International Conference on Pattern Recognition*, Jerusalem.

Hanson, S. J., Cowan, J. D., and Giles, C. L., editors (1993). *Advances in Neural Information Processing Systems 5*. Morgan Kaufmann.

Ho, T. K. (1992). *A Theory of Multiple Classifier Systems and Its Application to Visual Word Recognition*. PhD thesis, State University of New York at Buffalo. Also technical report number 92-12.

Holstege, M., Inn, Y. J., and Tokuda, L. (1991). Visual Parsing: An Aid to Text Understanding. In *Recherche d'Informations Assistée par Ordinateur 1991*, pages 175–193, France. RIAO.

Hull, J. (1994). Language-level syntactic and semantic constraints applied to visual word recognition. In Impedovo, S., editor, *Fundamentals in Handwriting Recognition*, NATO-Advanced Study Institute Series F, page 289. Springer-Verlag.

Hull, J., Ho, T. K., Favata, J., Govindaraju, V., and Srihari, S. (1992). Combination of segmentation-based and wholistic handwritten word recognition algorithms. In Impedovo, S. and Simon, J. C., editors, *From Pixels to Features III*, page 261. Elsevier Science, Amsterdam.

ICASSP (1994). *Proceedings of the 1994 International Conference on Acoustics, Speech, and Signal Processing*, Adelaide, Australia. Institute of Electrical and Electronic Engineers.

ICDAR (1991). *Proceedings of the First International Conference on Document Analysis and Recognition*, St. Malo, France. AIPR-IEEE, IAPR.

ICDAR (1993). *Proceedings of the Second International Conference on Document Analysis and Recognition*, Tsukuba Science City, Japan. AIPR-IEEE, IAPR.

Impedovo, S., editor (1994). *Fundamentals in Handwriting Recognition.* NATO-Advanced Study Institute Series F. Springer-Verlag.

Impedovo, S., Ottaviano, L., and Occhinegro, S. (1991). Optical Character Recognition—A Survey. *International Journal on Pattern Recognition and Artificial Intelligence*, 5(1-2):1–24.

Impedovo, S. and Simon, J. C., editors (1992). *From Pixels to Features III.* Elsevier Science, Amsterdam.

IWFHR (1993). *Proceedings of the Third International Workshop on Frontiers in Handwriting Recognition*, Buffalo, New York.

Kahan, S., Pavlidis, T., and Baird, H. S. (1987). On the Recognition of Printed Characters of Any Font and Size. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2):274–288.

Kanai, J. (1990). Text Line Extraction and Baseline Detection. In *Recherche d'Informations Assistée par Ordinateur 1990*, pages 194–209, France. RIAO.

Kasturi, R., Sira, S., and O'Gorman, L. (1990). Techniques for line drawing interpretation: an overview. In *IAPR International Workshop on Machine Vision Applications*, pages 151–160, Tokyo.

Kimura, F., Shridhar, M., and Chen, Z. (1993). Improvements of a lexicon-directed algorithm for recognition off unconstrained handwritten words. In *Proceedings of the Second International Conference on Document Analysis and Recognition*, page 18, Tsukuba Science City, Japan. AIPR-IEEE, IAPR.

Lam, S. W. and Baird, H. S. (1987). Performance Testing of Mixed-font Variable-size Character Recognizers. In *Scandinavian Conference on Image Analysis*, pages 563–570, Stockholm.

Lashley, K. S. (1987). The accuracy of movements in the absence of excitation from the moving organ. *American journal of Physiology*, 20:169.

Lecolinet, E. and Baret, O. (1994). Cursive word recognition: Methods and strategies (survey on CSR). In Impedovo, S., editor, *Fundamentals in Handwriting*

Recognition, volume 124 of *NATO-Advanced Study Institute Series F*. Springer-Verlag.

Lecolinet, E. and Crettez, J.-P. (1991). A grapheme-based segmentation technique for cursive script recognition. In *Proceedings of the First International Conference on Document Analysis and Recognition*, page 740, St. Malo, France. AIPR-IEEE, IAPR.

Liang, S., Ahmadi, M., and Shridhar, M. (1993). Segmentation of Touching Characters in Printed Document Recognition. In *Proceedings of the Second International Conference on Document Analysis and Recognition*, pages 569–572, Tsukuba Science City, Japan. AIPR-IEEE, IAPR.

Lorette, G. and Lecourtier, Y. (1993). Is recognition and interpretation of handwritten text a scene analysis problem? In *Proceedings of the Third International Workshop on Frontiers in Handwriting Recognition*, page 184.

Luca, P. G. and Gisotti, A. (1991a). How to Take Advantage of Word Structure in Printed Character Recognition. In *Recherche d'Informations Assistée par Ordinateur 1991*, pages 148–159, France. RIAO.

Luca, P. G. and Gisotti, A. (1991b). Printed Character Preclassification Based on Word Structure. *Pattern Recognition*, 24(7):609–615.

MacQueen, C., Scott MacKenzie, I., Nonnecke, B., Riddesma, S., and Meltz, M. (1994). A comparison of four methods of numeric entry on pen-based computers. In *Proceedings of Graphics Interface'94*, Toronto. Canadian Information Processing Society.

Mantas, J. (1986). An Overview of Character Recognition Methodologies. *Pattern Recognition*, 19(6):425–430.

Mori, S., Suen, C. Y., and Yamamoto, K. (1992). Historical review of OCR research and development. *Proceedings of the IEEE*, 80(7):1029–1058. Special Issue on Optical Character Recognition.

Nadler, M. (1984). A Survey of Document Segmentation and Coding Techniques. *Computer Vision and Image Processing*, 28:240–262.

Nagy, G. (1992). At the frontiers of OCR. *Proceedings of the IEEE*, 80(7):1093–1100.

Nagy, G., Seth, S., and Stoddard, S. (1985). Document analysis with an expert system. In *Proceedings of Pattern Recognition in Practice*, volume 2, Amsterdam.

Nartker, T. A., Rice, S. V., and Kanai, J. (1994). OCR Accuracy: UNLV's Second Annual Test. *Technical journal INFORM of the University of Nevada Las Vegas*.

Pavlidis, T. and Mori, S. (1992). Special issue on optical character recognition. *Proceedings of the IEEE*, 80(7).

Plamondon, R. (1989a). Handwriting control, a functional model. In Cotterill, R., editor, *Models of Brain Function*, pages 563–574. Academic Press, Cambridge, New York, Port Chester, Melbourne, Sydney.

Plamondon, R. (1989b). A handwriting model based on differential geometry,. In Plamondon, R., Suen, C. Y., and Simner, M., editors, *Computer Recognition and Human Production of Handwriting*, pages 179–192. World Scientific, Signapore, New Jersey, London, Hong Kong.

Plamondon, R. (1992). A model based segmentation framework for computer processing of handwriting. In *Proceedings of the 11th International Conference on Pattern Recognition*, pages 303–307.

Plamondon, R. (1993). Handwriting processing and recognition. *Pattern Recognition*, 26(3).

Plamondon, R. (1994a). The design of an on-line signature verification system: From theory to practice. In *Progress in Automatic Signature Verification*, volume 13 of *Machine Perception and Artificial Intelligence*, pages 795–811. World Scientific.

Plamondon, R. (1994b). A model-based dynamic signature verification system. In Impedovo, S., editor, *Fundamentals in Handwriting Recognition*, NATO-Advanced Study Institute Series F. Springer-Verlag.

Plamondon, R. (1995a). A delta-lognormal model for handwriting generation. In *Proceedings of the Seventh Biennal Conference of the International Graphonomics Society*, pages 126–127, London, Ontario.

Plamondon, R. (1995b). A kinematic theory of rapid human movements. *Biological Cybernetics*, 72:I:295–307 and II:309–320.

Plamondon, R. (1995c). Special issue on cursive script recognition. *Machine Vision and Applications*. In press.

Plamondon, R., Alimi, A., Yergeau, P., and Leclerc, F. (1993). Modelling velocity profiles of rapid movements: A comparative study. *Biological Cybernetics*, 69(2):119–128.

Plamondon, R. and Maarse, F. J. (1989). An evaluation of motor models of handwriting. *IEEE Trans. on Syst. Man Cybernetic*, 19(5):1060–1072.

Plamondon, R., Suen, C. Y., and Simner, M., editors (1989). *Computer Recognition and Human Production of Handwriting*. World Scientific.

RIAO (1991). *Recherche d'Informations Assistée par Ordinateur 1991*, France. RIAO.

Rocha, J. and Pavlidis, T. (1993). New method for word recognition without segmentation. In *Proceedings of SPIE*, volume 1906, page 76.

Rubinstein, R. (1988). *Digital Typography: An Introduction to Type and Composition for Computer System Design*. Addison-Wesley.

Schurmann, J. (1978). A Multifont Word Recognition System for Postal Address Reading. *IEEE Transactions on computers*, C-27(8):721–732.

Schurmann, J., Bartneck, N., Bayer, T., Franke, J., Mandler, E., and Oberlander, M. (1992). Document Analysis—From Pixels to Contents. *Proceedings of the IEEE*, 80(7):1101–1119.

Schürmann, J. et al. (1992). Document analysis—from pixels to contents. *Proceedings of the IEEE*, 80(7):1101–1119.

Seni, G. and Srihari, R. K. (1994). Hierarchical approach to on-line script recognition using a large vocabulary. In *Proceedings of the Fourth International Workshop on Frontiers in Handwriting Recognition*, Taipei, Taiwan.

Shlien, S. (1988). Multifont Character Recognition For Typeset Documents. *International journal of Pattern Recognition and Artificial Intelligence*, 2(4):603–620.

Simon, J. C., Baret, O., and Gorski, N. (1994). A system for the recognition of handwritten literal amounts of checks. In *Proceedings of DAS*, page 135, Kaiserslautern, Germany.

Srihari, R. and Baltus, C. M. (1993). Incorporating syntactic constraints in recognizing handwritten sentences. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, page 1262, Chambery, France.

Srihari, S. N. (1984). *Computer Text Recognition and Error Correction*. IEEE Computer Society Press, Silver Springs, Maryland.

Srihari, S. N. (1992). High-performance reading machines. *Proceedings of the IEEE*, 80(7):1120–1132.

Srihari, S. N. and Hull, J. J. (1985). Address Recognition Techniques in Mail Sorting: Research Directions. Technical Report 85-09, Department of Computer Science, SUNY at Buffalo.

Srihari, S. N. and Hull, J. J. (1992). Character recognition. In Shapiro, S. C., editor, *Encyclopedia of Artificial Intelligence, second edition*, pages 138–150. John Wiley.

Tappert, C. C., Suen, C. Y., and Wakahara, T. (1990). The state of the art in on-line handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(8):787–808.

Tsujimoto, S. and Asada, H. (1992). Major components of a complete text reading system. *Proceedings of the IEEE*, 80(7):1133–1149.

Tsujimoto, Y. and Asada, H. (1991). Resolving Ambiguity in Segmenting Touching Characters. In *Proceedings of the First International Conference on Document Analysis and Recognition*, pages 701–709, St. Malo, France. AIPR-IEEE, IAPR.

Ullman, J. R. (1973). *Pattern Recognition Techniques*. Crane-Russak, New York.

Unicode Consortium, The (1990). *The Unicode Standard, Worldwide Character Encoding*. Addison-Wesley.

van der Gon, J. J. D. and Thuring, J. P. (1965). The guiding of human writing movements. *Kybernetik*, 4(2):145–148.

Vaxiviere, P. and Tombre, K. (1992). Celesstin: CAD conversion of mechanical drawings. *IEEE Computer*, 25(7):46–54.

Wang, P. S. P. (1991). Character and handwriting recognition: Expanding frontiers. *International journal of Pattern Recognition and Artificial Intelligence*, 5.

Warwick, C. (1995). Trends and limits in the talk time of personal communicators. *Proceedings of the IEEE*, 83(4). In press.

Wilkinson, R. A., Geist, J., Janet, S., Grother, P. J., Burges, C. J. C., Creecy, R., Hammond, B., Hull, J. J., Larsen, N. J., Vogl, T. P., and Wilson, C. L. (1992). The first census optical character recognition systems conference. Technical Report NISTIR-4912, National Institute of Standards and Technology, U.S. Department of Commerce.

Wong, K. Y., Casey, R. G., and Wahl, F. M. (1982). A document analysis system. *IBM journal of Research and Development*, 26(6):647–656.