Handwriting as Computer Interface

Isabelle Guyon and Colin Warwick
AT&T Bell Laboratories
Holmdel, New Jersey, USA
isabelle@research.att.com
colin@spin.att.com

Pen Computers: Dream and Reality

"Pen computers" [4] offer an interesting alternative to paper. One can write directly on a LDC (Liquid Crystal Display) 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" [3]. 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) [11, 3]. 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! [?] 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 [14, 5]. Therefore, the task of designing usable handwriting recognizers for pen computing applications is tremendously hard. Human recognition rates must be reached and even outperformed.

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 [13]. 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 [12, 10]. 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 5000-10000 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 [6]). 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 [2]:

- 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.

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" an ";") 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 [1, 9, 8]. 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.

Suggested Research 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 [7], it will be possible to exchange a wide variety of data and organize public competitions.

References

[1] AIPR-IEEE. *International Conference on Document Analysis and Recognition*, Tsukuba, Japan, October 1993.

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

[3] L. Chang and I. Scott MacKenzie. A comparison of two handwriting recognizers for pen-based computers. In *Proc. of CANSON'94*, Canada, 1994.

[4] G. Forman and J. Zahorjan. The challenges of mobile computing. *IEEE Computer*, pages 38–47, April 1994.

[5] J. Geist et al. The second census optical character recognition systems conference. Technical Report NISTIR-5452, NIST, US Department of Commerce, 1994.

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

[7] I. Guyon, L. Schomaker, R. Plamondon, M. Liberman, and S. Janet. UNIPEN project of on-line data exchange and recognizer benchmarks. In 12^{th} *International Conference on Pattern Recognition*, Jerusalem, Israel, October 1994.

[8] IEEE. *International Conference on Acoustics Speech and Signal Processing*, Adelaide, Australia, April 1994.

[9] INNS-IEEE. *Advances in Neural Information Processing Systems 6*, Denver, Colorado, December 1993.

[10] E. Lecolinet and O. Baret. Cursive word recognition: Methods and strategies. In S. Impedovo, editor, *Fundamentals in Handwriting Recognition*, volume 124 of *series F*, New York, October 1994. Springer-Verlag.

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

[12] R. Plamondon, editor. *Special issue of Machine Vision and Applications on cursive script recognition.* 1994, to appear.

[13] C. C. Tappert, C. Y. Suen, and T. Wakahara. The state of the art in on-line handwriting recognition. *IEEE trans. on PAMI*, 12(8):787–808, 1990.

[14] R. A. Wilkinson et al. The first census optical character recognition systems conference. Technical Report NISTIR-4912, NIST, US Department of Commerce, 1992.