ICFHR 2008 Panel Discussion

Handwriting recognition research: Twenty years of achievement… and beyond

Mohamed Cheriet*, Mounim El Yacoubi, Hiromichi Fujisawa, Daniel Lopresti, Guy Lorette

Synchromedia Lab, ETS, 1100, Notre Dame West Street, Montreal, Canada H3C 1K3

This panel discussion was a follow-up to the first edition of the ICFHR series, which was promoted as an international conference after 10 intensive and productive editions of the International Workshop on Frontiers in Handwriting Recognition initiated in 1990 in Montreal. It is essential, however, to recall that research on handwritten character recognition began some 40 years ago, and was expanded about 20 years ago to the more general notion of handwriting recognition (no longer constrained to isolated characters) that we work with today. The organizers of ICFHR 2008 wanted to take the opportunity to organize the panel discussion, chaired by M. Cheriet, to highlight progress in the field over the long course of its development. Panelists M. El Yacoubi, H. Fujisawa, D. Lopresti, and G. Lorette talked about the achievements of two decades of handwriting recognition and about its future. Following these outstanding presentations, conference attendees had the opportunity to air their questions and concerns during open and fruitful exchanges.

Each section of this summary is drawn from the original presentation of a panelist. In Section 1, Dr. Hiromichi Fujisawa of Hitachi (Japan) discusses the impact of handwriting recognition technologies on the industrial sector. In Section 2, Dr. Mounim A. El Yacoubi of Institut Telecom, T&M SudParis (France) discusses the need to restructure the efforts of the handwriting recognition community. In Section 3, Dr. Guy Lorette of IRISA—Université de Rennes (France) highlights the recent achievements in handwriting recognition and others on the horizon. In Section 4, Dr. Daniel Lopresti of Lehigh University (USA) discusses the future of handwriting recognition research as a "Grand Challenge" problem. Section 5 concludes this summary by highlighting the salient points that will drive research on handwriting recognition in the future.

1. Impact on industrial sectors: niches to develop—*the long tail era*

1.1. The current market for HR

The commercial and industrial application of handwriting recognition began in the 1960s, and included commercial OCRs produced by computer manufacturers and postal (zip) code readers for mail sorting machines. Handwritten numerals were constrained at the time, in the sense that the writers were asked to follow a guiding pattern in the writing box preprinted in a dropout color. To the best of our knowledge, unconstrained numeral recognition was introduced in the late 1970s, with the objective of reading bank check amounts,

including spelled-out numerals. In the late 1980s and early 1990s, unconstrained handwritten word recognition algorithms, including those for Chinese characters, were developed and applied to mail sorting systems in the US, Europe, and Japan.

The three commercial and industrial applications mentioned above, namely, business form reading (commercial OCRs), bank check reading, and full postal address reading, have constituted the majority of the market for handwriting recognition technology. It seems, at least in Japan, that this market has been saturated, and that the absolute size of the market is just sufficient for a few successful business players for each field of application. As a result of this saturation, the number of players does not seem to increase, especially in some lucrative businesses.

As one of those business players in Japan, Hitachi has been trying to expand the number of application fields. Our attempt to do this is briefly described below.

1.2. The "long tail" of the market

Recently, the author looked at the concept of the "long tail", and compared it to the situation of the OCR market, which illustrates the current situation of the market vividly. If we were to draw a graph of the number of OCR systems, or the number of transactions handled by such systems, as a function of the applications sorted in decreasing order of their value, then we would have what we could call a "long tail" distribution. In order to have enough applications to make the curve sufficiently long, we simply define "application" arbitrarily to include as many examples as possible. For instance, applications can be categorized based on the kind of (paper) forms being handled by OCR systems all over the world.

This, of course, is just a hypothetical exercise without actual statistics, performed for the purposes of analogical discussion only. What is expected, however, is that the three major applications described above would be positioned on the left-hand side of the *x*-coordinate of a Cartesian graph, and other miscellaneous applications would be distributed along the right-hand side of the *x*-coordinate, possibly forming a long tail. The question then becomes, what can we conclude from this analogy?

First, the applications on the "head" (left) of the distribution are associated with rather lucrative businesses. Demands for the processing of paper documents, such as pieces of mail, can be huge, resulting in a formidable task for humans to accomplish. For instance,

every mail sorting machine in Japan processes more than 10 million New Year's greeting cards in the two-week period at the end of the year. The contents of the documents in those successful applications are well defined and often semantically rich. In other words, the semantics, syntax, and vocabulary are confined to a certain limit and well defined, which makes language and knowledge processing so effective that handwriting recognition systems are able to recognize unconstrained handwritten addresses with a sufficiently high degree of accuracy in these cases. More important is the fact that the manufacturers at the head of the OCR market may be able to invest in advanced next-generation technologies, yielding positive feedback in terms of the value created by the products. This would lead to what we refer to as the *virtuous cycle*, according to which more advanced technology leads to more advanced products and systems, with the possibility of more profits, and so on.

However, the real question to be asked is whether or not this market head has been saturated. The answer seems to be that it has, at least in Japan. So, we now look at the "tail" of the OCR market to see what it is like.

The tail distribution contains small niche markets representing many different needs. Unlike the head, the profits from the tail market applications are generally not large enough for an investment to be made in a technology leap to generate growth. In such markets, there is a tendency for the focus to shift from industrial applications to personal use applications, like the PDA, in the attempt to find new areas for growth. Owing to their nature, these new applications demand a much less constrained user environment, and technical difficulties associated with it make a sufficiently high degree of accuracy a challenge in such an environment. For example, digital pen applications would seem to be a promising area, but we should ensure that the users of such applications are allowed to write anything they want from the unlimited choices available, especially when taking notes or annotating paper documents. So far, the only successful digital pen applications have involved the data entry of numerals, names, and addresses.

One way to remedy this limiting situation would be to make the recognition technology more *versatile* and at the same time enable it to cover a wider area of application. More versatile products and software may lead to bigger returns and another cycle of investment. In the next section, we would like to address the technical challenges inherent in the quest for versatility.

1.3. Technical challenges for versatility

As discussed above, the key to future growth will be *versatility* of handwriting recognition systems, which will enable them to cover wider areas of application. Listed below are the technical challenges that are considered crucial to overcome in order to achieve versatility:

- Rapid and sustained adaptation to writers with different writing characteristics
- Contextual constraint utilization with incomplete knowledge and/or unlimited scope
- Immunity to erroneous or incomplete inputs
- Language-independent recognition infrastructure
- Search engines for handwriting

Rapid and sustained adaptation is almost completely lacking at the present time, and most commercial and industrial OCRs do not have this feature. The recognition engines are usually customized at an engineering site prior to shipment. Customization, the fine tuning of parameters, and retraining are all carefully performed by the manufacturer's engineers. It is the lack of a method that theoretically guarantees the convergence of the adaptation process, or at least

that the process will not diverge. How to provide class information for retraining is also a problem, but more important is the fact that there are many other parameters that have to be adapted than those for classifiers, for which no single algorithm for direct adaptation is known. These include parameters for layout analysis, word segmentation, character segmentation, and so on.

The integration of language processing into word recognition has been successful in the limited domains of check amount and postal address reading. Such algorithms have been referred to as lexicon-directed methods or knowledge-based methods. However, even in the successful applications, incomplete knowledge has been a serious problem, although an effort is always made to minimize its effect. Address information, for example, is always subject to change, and so, without maintenance, the lexicon for an address reader is incomplete. Incompleteness of the knowledge base may also be the result of human error introduced into data capture process. This issue is most significant when the domain of application is unlimited, as discussed in the context of the long tail.

Immunity to erroneous and incomplete inputs is another problem which has not yet been addressed. In our experience of developing an address reader, we found that capturing common error patterns and integrating them into a knowledge base in advance, which is a kind of customization, is effective. The versatility we are looking for now is the capacity of the system to automatically identify errors on its own and even to correctly recognize the input.

The creation of a language-independent recognition infrastructure which can be shared by many applications is an engineering issue. The aim is to drastically minimize the cost of developing multilingual OCR systems internationally. Such an infrastructure would include tools for lexicon development, pattern matching algorithms, an application interface, and so on.

Although the issue of search engines for handwriting cannot be addressed by the recognition algorithm by itself, it is listed here to emphasize those new applications, like digital pens for example, may have greater need for a search function than for a recognition function. Of course, if we had a perfect recognition engine, it would not be at all difficult to build a search engine for handwriting. One recommendation is to develop search algorithms that do not require perfect recognition engines, and then to develop applications that fully exploit the search capability. An annotation system using a digital pen would be an application of that kind.

Finally, one our expectations is that a scientific study will be conducted to clarify the relationship between handwriting and brain activities. The hope is to answer the question of whether or not electronic systems, such as a keyboard and mouse, will totally supersede the physical act of handwriting. Handwriting appears to be the most direct, natural way of expressing the ideas in our brain. At the same time, the use of the hand to write words and draw illustrations seems to stimulate the brain more than the use of electronic means, as well as to enhance the memory. This is conjecture, but, if true, then the seamless integration of handwriting and electronic information processing would be an ideal form of human information processing, and a digital pen solution would work well for this ideal system.

2. Is there a need to restructure the HR community effort?

Effective handwriting recognition systems exist for constrained application domains, and two of the most successful so far have been automatic mail sorting and automatic check processing. One of the main features of these tasks is that both involve the processing of inputs exhibiting rich context and high redundancy. In other words, zip code recognition allows for a dramatic and dynamic reduction in the vocabulary associated with cities. Furthermore, the combination and cross validation of the output zip, state, and city scores usually

leads to an effective and reliable decision at the definitive *destination* (city–state–zip) level. The same goes for the decision at the *delivery* level, which efficiently makes use of the recognition of the street (or post office box) number and the recognition of the street (post office box) name. The destination and delivery lines can also interact to produce an optimal decision at the address level. As far as check recognition is concerned, the courtesy amount and the legal amount constitute the input information for cross validation, which plays a similar role in decreasing recognition complexity and in providing a reliable decision at the check level. The above observations do not mean that mail sorting and check processing are easy, because they involve other non trivial tasks, such as the automatic location of the fields of interest (address bloc, individual address fields, courtesy and legal amount fields), and background removal (especially in the context of real checks associated with different financial institutions, each exhibiting a different and complex background). These tasks can be challenging on their own and require effective pre-processing techniques to make subsequent recognition feasible. The main point, however, is that impressive commercial systems could become a reality, despite insufficient handwriting recognition accuracy. The read rate of US handwritten mail pieces, for instance, is currently about 95%, with an error rate below 3%.

Speech recognition research, in comparison, took another direction from the outset. The speech community benefited early on from the support of large-scale programs like DARPA, which organized and funded several competitions which, on each occasion, took speech recognition maturity to the next level. Large datasets from various sources had been made available and sound evaluation protocols defined to encourage researchers to compare their results on widely agreed-upon criteria. One of the immediate effects was that, unlike the handwriting community, the speech community quickly moved from *isolated* speech recognition to focus on the more challenging task of *continuous* speech recognition. The other major consequence was that most research labs quickly adopted state-of-the-art techniques on some speech recognition phases, like the widespread adoption of a common front end (roughly speaking, sampling at about 10 ms, Mel-frequency Cepstral Coefficients (MFCCs), and working under the log scale) and focused on open issues like discriminative and adaptive methods (beginning in the 1990s) and more robust language models.

If we look now at handwriting recognition, there are some questions that are surprisingly hard to answer, despite the huge amount of effort dedicated to them over the last 20 years: What is the best feature extraction method available today? Should we over segment handwritten words? If the answer to this latter question is no, how should the sliding window width be optimally determined? Is the labeling of uppercase and lowercase letters necessary for an optimal performance of an unconstrained handwriting recognizer? The point is, it is hard to answer these questions because there has been no systematic comparison of state-of-the-art systems on common datasets under unique and well defined evaluation protocols. Moreover, because of their varied experimental settings (size of datasets, vocabulary used, etc.), these systems cannot benefit from one another in an optimal way from a development perspective.

Even though dozens of datasets exist for many languages, what is missing is the large-scale adoption of some of them as benchmarks for testing by the community. Furthermore, the size of many of these sets is not sufficient for this purpose. Studies may therefore be required to assess the needs of the community in terms of datasets not only containing isolated handwritten words, but also handwritten sentences and handwritten text pages. Unique evaluation protocols will be used to judge the effectiveness of a whole handwriting recognizer, one of its modules (e.g. preprocessing or feature extraction), or a language model.

To conclude, the following are some of the open issues and challenging tasks that need to be addressed in the near future: (1) the recognition of natural, unconstrained handwritten texts, which requires the development of a robust recognizer for both cursive and hand printed words, and even for machine printed text, as many handwritten texts might, in real-life tasks, be preceded or followed by machine printed data in headers or footers. Failure to do so might not only prevent us from recognizing some important contextual fields which could help guide subsequent handwriting recognition tasks, but even produce serious errors, owing to the fact that a recognizer conceived for and trained on handwritten data only may mistake some printed data for handwritten data with a relatively high confidence; (2) the development of language models (*LMs*) specific to handwriting (punctuation marks, lowercase vs. uppercase, accentuated letters), as well as speech-like *LMs*, and their dynamic adaptation to handwriting style and to *LMs* (by means of a dynamic topic detection method, for instance); (3) the development of reliable measures to assess the recognizer confidence and to prompt a manual operator transcription, as a 100% recognition rate is not possible; and, finally; (4) sound measures of image quality, so that less effort is spent on data which are easy to recognize and more effort and resources on data which are difficult to recognize.

Despite its current performances in unconstrained contexts, handwriting recognition will continue to attract commercial applications involving a wide range of domains, including healthcare and insurance, as well as large stores and administration facilities. However, the key to opening the door to far more businesses and new applications will be the maturity of context-independent natural handwriting recognition.

3. Recent achievements in HR and future prospects

The online counterparts of the offline HR techniques, which are implemented on small devices such as PDAs and smartphones, have not yet been very successful. We believe there are several reasons for this. First, the acquisition surfaces of such devices are generally considered by most users to be too small, resulting in the limited resolution of the digitizer and in too little space on the screen in which to write more than a few words. Second, a virtual keyboard is only acceptable for inputting very short pieces of information or messages, and HR is only useful for the input of entire pages of notes, annotated diagrams, etc. Third, most existing pen-based man-machine interfaces (MMI) are still not very user friendly, and some are very cumbersome or tedious to use. Fourth, the initial handwriting recognition rate is generally not sufficiently high to immediately convince the user to change to a new terminal. Moreover, adapting the system to the user's handwriting style is usually too time-consuming and too tedious a process.

3.1. Challenges for the near future

Except for some very successful industrial applications described previously, neither the online nor the offline HR techniques are in widespread use on a day-to-day basis, and the problems posed by HR are far from being solved. Further investigation is needed, and the HR domain should be explored both in greater depth and greater breadth.

Investigation of the domain in greater depth will consist, for example, of studying the recognition of distorted, incomplete, or noisy handwriting data, developing methods to enable handwriting styles to be learned from very small samples, developing algorithms for fast adaptation to a new personal handwriting style, designing very user friendly pen-based MMIs, and designing systems capable of recognizing and perceiving basic handwriting primitives, letters, and words in their respective contexts, and capable of understanding the mean-

ing of large pieces of handwritten text (at least one subsection). To do so will probably require the use of various knowledge sources, AI techniques, Gestalt theory, and linguistic techniques, among others.

Investigation of HR in greater breadth will consist, for example, of developing tools for the input of new composite document online systems to annotate or correct pre-existing digitized online documents. It could also serve to design multimodal man–machine interfaces combining speech and handwriting recognition techniques or to design automatic indexation systems for archiving document images according to their semantic content, etc.

3.2. A grand challenge problem for HR

On the one hand, the offline HR problem has not yet been totally solved, and cannot be considered as an off-the-shelf tool. There exist huge amounts of archived documents all around the world, which cannot be easily and immediately consulted at a distance by everyone on the Web. A new and major challenge might be to digitize massive quantities of these documents in order to create databases that are Web-accessible. This procedure would probably also require some specific search tools to help users find specific documents. From this point of view, it would be very useful to use offline HR for word spotting techniques in order to be able to automatically index the content of archived documents. This will require writer-independent offline HR systems, based on a great deal of linguistic knowledge.

On the other hand, because there is still a big demand for nomadic computing (computing everywhere, and standing up, for example), online HR is still of interest in this area too. Despite the fact that it exists in a number of devices, such as PDAs, smartphones, tablet PCs, and netbooks, there is still no light, highly user friendly, pen-based and cheap terminal device that can be used on a daily basis. Such a device would be very useful in the field or during lectures or meetings for taking notes, drawing schemas and diagrams, and so on. The study of such a device (both hardware and software) remains a challenge for the online scientific and industrial HR communities.

4. Future handwriting recognition research: a grand challenge problem

Since, to date, offline handwriting recognition research has achieved notable successes in certain focused application areas, including postal address reading and bank check and form processing, we might now wonder, why just here and not elsewhere? Certainly economics—i.e. business considerations—plays some part in this, but, as researchers, we can content ourselves with meeting the challenges posed by the technical side of the issue. What are barriers to solving the broader handwriting recognition problem?

One view that might help explain the present situation is to regard specific instances of handwriting recognition as points in a three-dimensional space, as illustrated in Fig. 1.

Here, the dimensions are style, language, and layout. The first of these represents the myriad ways that people express their individuality when writing, the second captures the complexities of human language (including the expression of visual concepts), and the third refers to the manner in which information is organized on the page. Of these three dimensions, substantial progress has been made in the first by developing techniques that can account for inter-writer variability, but we have seen much less movement along the other two dimensions. This is illustrated by the two data points in the figure and the dashed line showing the direction of our progress.

Hence, handwriting recognition for inputs that are fully unconstrained in terms of layout and language remains an elusive goal. In an interesting (and perhaps instructive) side note, language modeling has played a much more prevalent and productive role in the
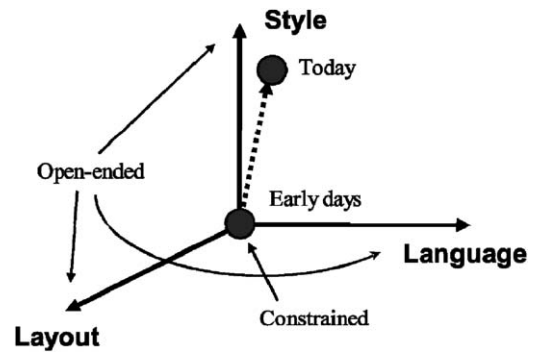

Fig. 1. Three dimensions of the HR problem.

field of speech recognition, a problem that is in some sense "easier" than document analysis, because it is only two-dimensional (missing the layout component).

A truly unconstrained handwriting recognition system would be able to handle any of the inputs shown in Fig. 2 below, which provides only a very small sampling of the range of ways that humans express themselves in writing. Without the economic burden of having to worry about funding or requiring a "killer" business application, what is the most general formulation of the handwriting recognition problem?

This is an interesting question in its own right and has the flavor of what have come to be known as "Grand Challenge Problems." In 1900, the famous mathematician David Hilbert proposed 23 unsolved problems which helped define the field of modern mathematics and drove it forward for almost a century. Nearly all these problems have now been solved, and so, in 2000, the Clay Foundation proposed seven new problems, known as Millennium Problems, each with a $1 million prize (http://www.claymath.org/millennium/).

In fields that are perhaps more similar to an experimental science, like pattern recognition research, one can point to the recent DARPA Urban Challenge, which had as its goal the building of a driverless vehicle that could navigate a 96-km (60-mile) urban area course in less than 6 h, while obeying all traffic regulations and dealing with other vehicles and obstacles (http://www.darpa.mil/grandchallenge/index.asp). Teams from around the world competed, generating tremendous media attention and widespread interest in the field of robotics.

In the area of artificial intelligence, at the website longbets.org, two well known founders in the field of computing, Mitch Kapor and Ray Kurzweil, have staked a $20,000 wager on the provocative statement, "By 2029 no computer—or 'machine intelligence'—will have passed the Turing Test" (http://www.longbets.org/1). That this debate is taking place in a very public forum will undoubtedly cause bright, highly motivated students and researchers to ponder whether they might have something to contribute to answering the question and, in doing so, advance the science.

What, then, would be a Grand Challenge Problem for handwriting recognition? How would the task be formulated? How would success be measured? These are intriguing questions, and taking a step back from our day-to-day research activities to consider such issues can be illuminating.

Thinking more deeply about the ground rules, it is possible to come up with a long list of related questions. Should all three dimensions be completely open-ended? Online or offline handwriting? Unilingual or multilingual? Text only, or anything a person can write or sketch? What level of performance is required to win? Is success measured by word error rate or a task-based evaluation? If the latter, what is the task? Should there be limits placed on computational resources or time used?
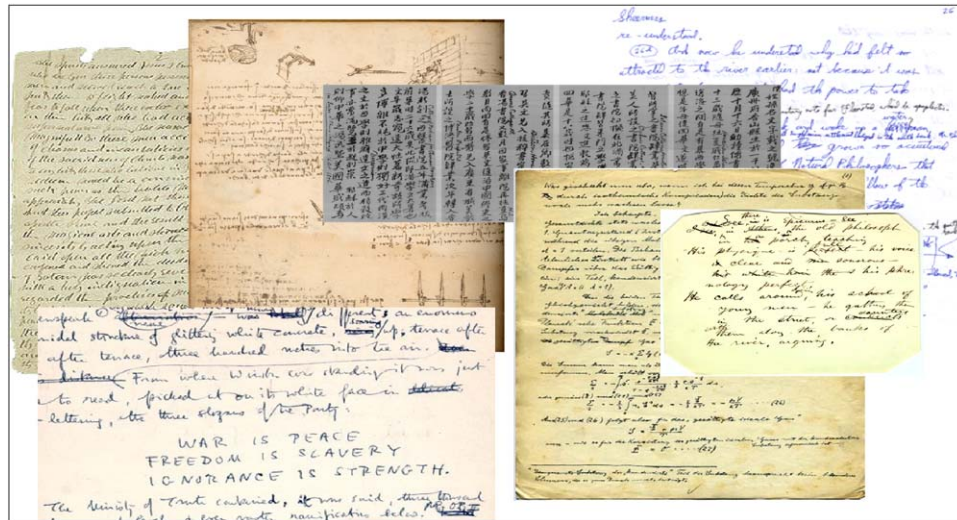
Fig. 2. Samples of challenging handwriting.

There are more questions than answers, and we will not argue here for one formulation over another. Rather, we hope to spur debate within the community. Having our own set of Grand Challenge Problems for handwriting recognition research would advance the field and draw well-deserved attention to the good work that is already being done.

5. Conclusions

In pondering the future of handwriting recognition research, certain key points come to the fore. The first of these are practical in nature, and include the need to develop business models where handwriting recognition is the technology of choice in applications that consumers find compelling. The requirements dictated by the commercial marketplace will help frame the problem in ways that are general enough to be useful, but not so overly ambitious as to be impractical in the near term.

Thinking about the issue from the other direction, it may be instructive to try to frame handwriting recognition as a Grand Challenge Problem, one worthy of thought and energy unconstrained by monetary concerns. As an intellectual exercise, there is some value in taking such a view.

In posing his famous test for machine intelligence, Alan Turing defined the question in such a way that few actually try to solve it, but everyone is aware of it. The goal of achieving performance which is indistinguishable from that of a human on a suitably broad range of inputs could be the starting point. People are not perfect, however, and this is a fact we sometimes forget. Given the myriad uses of handwriting and the sheer number of written languages in existence, it is clear that no single human could handle every possible input. Recognizing this limitation is also a human trait. So, how good would a machine have to be, to be declared the equal of a human?

Attempts to define the Grand Challenge Problem for handwriting recognition lead naturally to two key sub-goals: the need for comprehensive, versatile multilingual datasets and the search for evaluation paradigms that extend beyond current measures derived from basic word or character accuracy rates. We encourage the international community to ponder these points, as we explore the papers included in this special issue and shape the goals for our own future research.