# Mobile Interactive Support System for Time-Critical Document Exploitation*

**George Nagy**
Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180
nagy@ecse.rpi.edu

**Daniel Lopresti**
Department of Computer Science
Lehigh University
Bethlehem, PA 18015
lopresti@cse.lehigh.edu

## Abstract

*Computer Assisted Visual Interactive Recognition, or CAVIAR, has proven to be an effective methodology for integrating human and machine expertise in addressing challenging pattern recognition tasks. In this paper, we examine the application of the CAVIAR paradigm to problems arising in real-time document analysis in the field. We identify hurdles we expect to encounter, propose potential solutions, and describe an evaluation framework to help determine whether this is a fruitful line of research.*

## 1 Introduction

Research aimed at fully automating the processing of document images has received a tremendous amount of attention over the past 40 years. As a quick perusal of any of the dozens of surveys to date will reveal, however, progress in automatic recognition and interpretation has been slower than predicted. We expect that further improvement in accuracy in domains such as cursive handwriting and degraded documents will be even more protracted because the challenges that remain are much harder. As in speech recognition, bridging the gap between machine and human knowledge to allow the former to draw even with the latter appears problematic. The context brought by humans to any classification task is much greater than what can be obtained from even the largest collections of training samples available to our community. Endowing fully automated systems with broad knowledge remains a far-off goal.

There exists, however, a significant body of applications where it is not necessary to fully automate the task of document analysis. Rather, the focus is on a relatively small number of high-value documents (perhaps just one) – say from a cache discovered in the field as part of an ongoing criminal justice, military, or intelligence operation – where the computer plays the role of an assistant to help the user acquire information that would otherwise remain inaccessible. While such documents could be collected and returned to a central repository for scanning and batch processing in the traditional manner, there is often a significant advantage in being able to exploit the information in real-time, immediately and in situ.

Recently, Nagy and Zou, et al. have begun exploring a concept known as CAVIAR: Computer Assisted Visual Interactive Recognition [3, 6, 24, 23]. Over the last few years this work has led to the development of a successful interactive system for recognizing faces and flowers, both problems of a level of difficulty (i.e., automated current accuracy) comparable in certain ways to document recognition in the field. Experiments on sizable databases of faces and flowers indicate that interactive recognition is more than twice as fast as the unaided human, and yields an error rate ten times lower than state-of-the-art automated classifiers. The benefit margin of interactive recognition increases with improved automated classification. Parsimonious human interaction throughout the interpretation process is much better than operator intervention only at the beginning and the end, e.g., framing the objects to be recognized or dealing with rejects. Furthermore, this interactive architecture has been shown to scale up: it can start with only a single sample of each class, and it improves as recognized samples are added to the reference database.

The notions embodied in CAVIAR differ in fundamental ways from past efforts at mobile and/or interactive recognition. Whether such an approach can be equally effective in the domain of documents as it is for flowers and faces is unproven. In this paper, we discuss the application of this paradigm to document analysis in the field, identify the hurdles we expect to encounter, propose potential solutions, and describe an evaluation framework to help determine whether this is a fruitful line of research.

* To be presented at the *Symposium on Document Image Understanding Technology*, November 2005, College Park, MD.

## 2 Related Work

The questions we plan to examine arise from adapting CAVIAR to tasks from document analysis. There are, however, other projects that share similar goals and assumptions; below we briefly cite a few of these. The Army Research Laboratory's Forward Area Language Converter (FALCon) system provides mobile optical character recognition (OCR) and translation capabilities [10, 21], but, so far as we know, employs a traditional user interface. A growing amount of research is being conducted on camera-based document acquisition (e.g., [7, 11]). This work employs the camera as a new form of capture device, but, as with FALCon, treats the later processing stages as though they will be fully automated

Camera-based systems for locating and recognizing text in traffic signs and providing translation services for non-native visitors in foreign lands are perhaps more similar to what we have in mind [8, 22]. Still, we have yet to encounter such a system with an interaction paradigm as integral and as sophisticated as CAVIAR's. Reading services for the vision-impaired are likewise focused on page-at-a-time processing, but employ an auditory user interface as opposed to a visual one for obvious reasons [9, 19]. A somewhat similar notion is recent work on developing tools to support forensic document analysis [20]. We note, however, that such systems are designed to solve a different, much more specific problem, are intended for off-line use by domain experts (as opposed to occasional users whose primary jobs lie elsewhere), and have no need for mobility.

## 3 Rationale for a Human-Machine Collaborative Approach to Document Analysis

A divide-and-conquer strategy for visual recognition should partition difficult domains into components that are relatively easier for both human and machine. There are pronounced differences between human and machine cognitive abilities. Humans excel in gestalt tasks, like object-background separation. We apply to recognition a rich set of contextual constraints and superior noise-filtering abilities. Computer vision systems, on the other hand, still have difficulty in recognizing "obvious" differences and generalizing from limited training sets. We can also easily read degraded text (e.g., CAPTCHA's [2]) on which the best optical character recognition systems produce only gibberish.

Computers, however, can perform many tasks faster and more accurately. Computers can store thousands of images and the associations between them, and never forget a name or a label. They can compute geometrical properties like higher-order moments whereas a human is challenged to determine even the centroid of a complex figure. Spatial frequency and other kernel transforms can be easily computed to differentiate similar textures. Computers can count thousands of connected components and sort them according to various criteria (size, aspect ratio, convexity). They can quickly measure lengths and areas. They can flawlessly evaluate multivariate conditional probabilities, decision functions, logic rules, and grammars. On the other hand, the study of psychophysics revealed that humans have limited memory and poor absolute judgment [16]. A detailed comparison of these differences appears in Table 1.

## 4 Technological Issues

Here we briefly summarize some of the technological issues that must be addressed in the implementation of a CAVIAR-like system to support document analysis and exploitation:

1. The rapid development of high-quality, low-cost digital cameras suitable for full-page imaging in color or a broad range of grays. Smaller versions of these cameras are now available as plug-ins for Personal Digital Assistants (PDA's) and within a year sufficient resolution should be available for camera phones. These devices make document-acquisition in the field simpler than with conventional page-width scanners.

2. Displays are approaching the resolution-limit of the human eye (discounting head movement). In addition to the small displays built into PDA's, cell phones, and helmets, larger flexible displays that can be incorporated into a user's clothing are on the verge of becoming available.

3. Interaction with a direct-action device like a stylus, and especially a thumb, is faster than with a mouse. Touch-sensitive screens are just now becoming available for cell phones.

4. The storage capacity of mobile devices is already sufficient for most DIA and OCR tasks.

5. Nagy, et al. have implemented interactive recognition on both stand-alone and wireless networked mobile platforms. Both are adequate for PDA's, but neither is yet acceptable for cell-phone based systems because these do not have enough computational resources for stand-alone operation, and their bandwidth is too low for acceptable interactive response time on image computations. However, according to technology forecasts, both of these shortcomings will disappear in a year or two.

Table 1: Comparison of relative strengths of human vs. machine in visual pattern recognition.

| Human | Machine |
|---|---|
| • dichotomies | • multi category classification |
| • figure-ground separation | • nonlinear, high-dimensional classification boundaries |
| • part-whole relationships | • store and recall many labeled reference patterns |
| • salience | • accurate estimation of statistical parameters |
| • extrapolation from limited training samples | • application of Markovian properties |
| • broad context | • estimation of decision functions from training samples |
| • gauging relative size and intensity | • evaluation of complex sets of rules |
| • detection of significant differences between objects | • precise measurement of individual features |
| • colored noise, texture | • enumeration |
| • non-linear feature dependence | • computation of geometric moments |
| • global optima in low dimensions | • orthogonal spatial transforms (e.g., wavelets) |
| | • connected component analysis |
| | • sorting and searching |
| | • rank-ordering items according to a criterion |
| | • additive white noise |
| | • salt & pepper noise |
| | • determination of local extrema in high-D spaces |

6. The need for network connectivity depends greatly on the targeted application domain: civilian, military, or covert. While it is clearly desirable for such a system to be operable completely autonomously, there may be substantial value in networking document acquisition and exploitation activities.

## 5 Data Entry on Hand-Held Devices

It is clear that an important constituent of mobile interactive document analysis is manual text entry. The alternatives seem to be restricted to virtual keyboard on a touch sensitive screen activated by a handheld stylus, finger-operated keyboards incorporated in the operators clothing (on arm or thigh), and automatic speech recognition. We believe that the stylus is the most appropriate solution, because in addition to text entry it can also mediate the graphical communication essential in some aspects of document image analysis.

The virtual keyboard appeared in the seventies to accelerate the digitization of maps and line-drawings by avoiding having the operator shift constantly between pointing device and keyboard. It consisted of a picture of a keyboard on a piece of paper that could be shifted to the area of the drawing being vectorized. The current virtual keyboard usually appears in a fixed partition of the touch-sensitive screen of a handheld device. The text being entered appears on another partition. Edwards surveyed input interface issues in mobile devices in 1997 [4]. Data input is usually a local operation, so it makes little difference whether the device is networked or not.

The key consideration for stylus data entry are speed, (perceived) operator comfort, and ramp-up time. The first two factors are influenced by the amount of space allocated to the keyboard, to the recognized or entered text, and to control functions. The third factor depends heavily on the keyboard layout. The QWERTY layout, developed to prevent binding of type bars in mechanical typewriters, is suboptimal even for typing, and even more so for one-handed stylus entry.

There has been much work on evaluating alternative keyboards and word-completion algorithms (for a recent overview, see [1]). An upper bound on the speed of individual character entry is imposed by Fitts' Law, which is a nonlinear relationship between pointing time and the distance and size of the target. The relevant distance is that between the screen areas ("keys") corresponding to consecutive letters. The letter transition frequency is given by a language model. As an example of the compromises necessary in keyboard design, it is possible to reduce the average distance by having several space keys, but this decreases the size of the keys. Several researchers have optimized keyboard designs according to various language models [14, 13, 15, 12]. The computed speeds hover about 40 words per minute, but actual text entry is much slower.

The speed increase obtainable by word completion also depends on the language model. Ancona [1] demonstrates a keyboard with separate keys for the ten most common words (with a cumulative word frequency of 28%). After each tap on the screen, the ten most likely words appear in the selection area of the screen. If the correct word is included, it can be selected with one additional tap. If not, another letter is tapped, which brings up ten new words. With a vocabulary of 13,000 words, the expected number of taps per word is claimed to be 3.3. Some of the notions here are clearly similar to those incorporated in CAVIAR.

The performance of word-completion systems depends on how well the stored lexicon is matched to the user input. Multiple lexicons – for different languages and applications – can be either stored on board, or downloaded via a wireless connection.

Another important source of ideas is the technology developed for vectorizing maps and engineering drawings [17, 18]. Manual vectorization was first conducted from hardcopy with a digitizing table or tablet. The operator traced the lines with cross-hairs under a magnifying glass with a MARK button. After the advent of large-size roller-feed scanners and bitmapped displays, all service bureaus and in-house operations converted to on-screen vectorization from scanned copy. One advantage was that already vectorized lines could be displayed with a different color, and departures between manually entered lines and original were clearly visible. The operator could zoom in on dense portions of the drawing. However, the aspects of interest here are the algorithms developed for semi-automated data entry.

For colored maps, different color layers were first separated according to RGB values. Vectorizing algorithms were manually initialized to a line segment or curve, and then could automatically follow that line at least to the next intersection point. The systems would also attempt to automatically recognize map and drawing symbols (for schools or resistors). If it failed, the operator would override it. The character recognition software recognized cleanly lettered labels (elevations, part numbers, resistor values), but left labels confused by overlaid line art or poor lettering to the operator.

If most of the labels cannot be recognized by OCR because of poor document quality or unusual character shapes, it is still possible to rapidly mark their location and orientation, rotate them to horizontal, and move them to a single area of the screen. This accelerates manual label entry. A single E-sized drawing may contain 3,000 alphanumeric symbols, which is more than a densely printed page of text.

Most such data-entry systems are part of larger GIS or CAD software, and are typically designed for standard workstations. All graphical operator interaction is therefore mediated by the mouse. As demonstrated by Engelbart and colleagues at SRI long ago, direct-action devices, like a touch-sensitive stylus, allow faster and more accurate interaction [5]. Indeed, we found that to be the case in comparing desk-top CAVIAR systems with handheld CAVIAR.

Like CAVIAR, these interactive systems exhibit clear speed advantages over completely manual data entry, and are robust enough (unlike automated systems) for operational application. Although some of these systems are laboriously trainable, one key difference compared to CAVIAR is that no commercial system that we are aware of incorporates adaptive algorithms that take advantage of routine operator input. Unlike CAVIAR, they also fail to provide visible models for the entry of complex 2-D patterns.

## 6 Test Data Collection

Any comparison of the proposed mobile interactive document exploitation system with existing and forthcoming automated and manual data entry systems requires a test database. It is, of course, desirable that the test database reflect the characteristics of the documents that are to be processed. For the purposes of this discussion, we assume that the domain of interest is handwritten Arabic documents; however, it should be clear how the the same issues and potential solutions generalize to other genres and languages.

We believe that existing handwritten Arabic databases cannot fulfill the objectives we have in mind for CAVIAR research because they are too small, they were designed specifically for testing a particular class of recognition algorithms, and they do not reflect the characteristics of the target data. For the sake of concreteness, we make the following assumptions:

1. The proposed system is to be applied to transcription and further processing of documents similar to those in a growing collection of handwritten Arabic documents, referred to here as the *Target Database*. Although we propose a mobile, personal device operated by a non-specialist, the test database will also be used for evaluating batch-mode DIA and OCR systems.

2. The amount of data and corresponding meta-data, including ground truth (GT), will be comparable to an existing database used for evaluating typeset Arabic documents, i.e., 400,000 words. We take this as a fixed point for now. Handwritten documents have fewer words than

printed text: say 100 words/document on average. The documents are not restricted to a single page, but each document is created by a single author.

3. Handwritten documents have statistically different textual content than typeset documents; therefore we cannot mirror the existing database. Handwritten material is less likely to form a grammatical narrative: it may have unusual abbreviations, short lists of phrases, unlabeled strings of digits, underlines, corrections, cross-outs and erasures, and unstylized layouts.

4. Character formation in free writing in any script is markedly different from copying or taking dictation from tape or a person. The conditions under which the document is composed, including the degree of stress or haste, will perceptibly alter even the same person's writing. We believe that this must be taken into consideration to avoid disappointment when the system is deployed.

5. The most important components of systematic variability are (a) country of education, (b) level of education (years of schooling), and (c) age of the writer.

6. Digitization and character recognition are affected by the medium: pen or pencil, $n^{th}$ generation copy, fax, paper quality, and physical conditions (writing on a desk is different from writing in a hand-held notebook).

7. The chosen approach for (semi-)automated transcription should depend on the distribution of the amount of writing among individuals. If there are many long passages or multiple documents by the same writer, the system ought to take advantage of this. Language context and writer/style adaptation are powerful aids to recognition.

8. A service bureau with Arabic software can perform scanning and ground-truthing more efficiently than students. Such a service bureau could be located in a country with a significant Arab-speaking population (US, Canada, UK), or in a friendly Arab country (Jordan, Egypt) with an advanced computer infrastructure.

In order to plan data collection, it would be desirable to have the following information. The relevant document statistics could either be estimated (guessed) by a designated expert, or obtained from the agency responsible for processing the target documents.

Ideally, a sizeable random sample of the original documents (or very good copies) with translations from the Target Database, with accompanying metadata, would be made available. Most of the necessary information could be derived from such a sample. If, as is likely, such a sample cannot be released, then specification of the following distributions (in the form of histograms) would be useful:

- Number of *words per document*.
- Number of *documents per writer* and of *words per writer*.
- *Educational profile* of the writer.
- *Country of schooling* of the writer.
- *Vocabulary:* a lexicon indexed by frequency and, if possible, by writer.
- *Media:* pen, pencil, copy, fax, photo, etc.

In order to gauge feasibility, in Table 2 we provide a rough estimate (perhaps accurate to ±50%) of the cost of deriving the required statistics from a hardcopy database of documents. These costs do not include the significant cost of collecting the documents, nor planning/management costs.

Table 2: Estimated costs of a hardcopy document ground-truthing activity.

| Category | Cost |
|---|---|
| Scanning (200 or 300 dpi, 8-bit grayscale), for 4,000 mostly single-page documents, with quality control, at $1.00 per document | $4,000 |
| Transcription, with verification and correction, at $2.00 per document | $8,000 |
| System acquisition or customization for detailed (word-box level) zoning | $10,000 |
| Interactive box location, verification and correction | $8,000 |
| Database creation (accession numbers, bitmaps, boxes, GT) | $12,000 |
| Estimated total | $42,000 |

The handwriting data could consist of either: existing originals (or high-quality copies thereof), or handwritten documents generated for this purpose by about 1,000 subjects.

Possible sources of existing data include schools (assignments, term papers, course notes, administrative records), businesses (orders, reports, correspondence), and government agencies (dead files of forms and free-form correspondence). Privacy concerns may preclude certain of these approaches, however.

If new data must be generated, a possible scenario for distributed data collection is:

1. Prepare from five to eight printed single or multi-page protocols on various topics, e.g., economics, politics, technology, military, religion. The topics should correspond to their proportions in the target database. The protocols should include questions that require a narrative answer, including requests for summarization, organization, decision, rebuttal, comment or memory-aids in the subjects' own words. It is essential to avoid having the subjects copy fixed material, and to give rise to varied vocabulary, syntax, spelling, and document length.

2. Recruit ~1000 Arab-literate subjects according to specified demographics, possibly in different countries, and request each to respond to one or more forms. In U.S. university settings, this would require permission for the cognizant institutional review board on human factors experimentation, and obtaining signed consent forms.

3. Provide each writer a writing surface, paper and writing instrument according to the randomized experimental design.

4. Collect the forms, possibly with the demographic data attached to each sheet. Scan the image data and organize the metadata in a widely readable format (ASCII perhaps). The size of the database after digitization at 200 dpi (8-bit depth) should be about 16 GB. Because of the prevalence of white space, after compression it may fit on a single CD.

For DIA and OCR research and development, this method would far superior to the customary collection of copies from prescribed forms. The latter, however, would be less expensive because the handwritten data need not be transcribed but only proofread to catch copying errors.

Part of the resulting database should be sequestered for independent testing. The remainder could be released to the research and development community. Used in this manner, the proposed database would have a relatively long shelf-life and provide performance measures far more representative of operational systems than the customary greenhouse test data.

## 7 Examples of Interactive Document Image Analysis

We mention some DIA tasks where automated algorithms work accurately only on exceptionally clean documents, but where a little interaction can quickly produce acceptable results.

Most OCR algorithms, especially for handwriting, are designed for binarized images, because scripts generally avoid discrimination based only on shades of gray or color. Instead of using thresholding hardware built into the scanner, today documents are usually digitized to 8-bit gray scale or RGB, and subsequently converted to binary images. Global binarization algorithms work only if the foreground and background reflectance are uniform throughout the document, which may not be the case if, for example, part of a folded documents suffers prolonged exposure to sunlight, or if there are dark areas around the edges of a photocopy. Local binarization algorithms measure the distribution of reflectance in a window translated through the page, and set the local threshold between foreground and background reflectance peaks according to the estimated dispersions of the components of the mixture distribution. The window size and reflectance distribution estimates invariably depend on explicit or implicit assumptions about the relative density and configuration (strokes) of the foreground (ink) and background. These assumptions generally hold only for a narrow class of documents. Fortunately, the binarization algorithms are simple, therefore an operator can easy set the appropriate window size and foreground density either for the whole document, or for selected areas. This still allows local algorithmic thresholding, and therefore requires much less interaction than setting the threshold manually everywhere, and is much more robust than fully automated local thresholding.

Line finding is another instance where interaction may be effective. The first step is usually estimating global document skew, i.e., the angle of the written lines with respect to the paper or digitizer axes. While very accurate skew estimation and correction algorithms have been developed for printed matter, they do not work well on handwriting because the orientation of individual lines varies, the margins are not straight, there may be only a few words on a page, and there may be several columns of words or phrases at different angles. Humans can, however, judge skew remarkably well, and convey this information to the computer by a few well chosen stylus taps or by rotating a superimposed grid. After the computer-proposed skew correction and line finding is corrected, the occasional merged pair of lines – due to overlapping ascenders and descenders – can be likewise rapidly separated.

Word segmentation is relatively easy for printed text, except for extremely tightly-set, micro-justified print. In handwriting, however, large spaces often appear within words and, towards the end of a line, words are often squeezed together. In Arabic and other scripts, some inter-letter spaces are mandatory. Underlines that link word sequences can further complicate the task. Again, humans can usually spot missed word boundaries even in unfamiliar languages and scripts. If the writing lines are already properly segmented, then a simple interface can be designed to correct linked and broken words.

At the character recognition level, there are also several opportunities for effective interaction. First, humans can often tell where perfect accuracy is important, as in telephone numbers, email addresses, and proper nouns. If the automated algorithms fails on important words, phrases or numbers, they can be either entered manually using the virtual keyboard, or selected by a stylus tap from the top recognition candidates.

The human can provide global assistance to the character recognition system. The operator may be able to recognize the language or script of a document even from a few words, perhaps by using common sense or contextual information related to the source of the document (e.g., indicate that the document probably contains a mixture of Korean and German.) He or she can indicate the average slant, and in Western scripts, the prevalent case (e.g., if a writer uses only capital letters). The operator may also decide which of the available lexicons would provide the best language model. (The lexicons will be automatically updated with entries from the processed documents that have been deemed correct.)

Most importantly, entering only part of a document may provide enough training data – to a recognition system designed with this in mind – for fine-tuning the classification algorithms. The underlying assumption is that if the remainder of the document (and perhaps also additional documents) is from the same source, the adjusted parameters will yield more accurate recognition. If that is not the case, the operator can easily separate the portions of the document written by different individuals (assuming that the second writer was not attempting to mimic the first). Of course, human handwritten data entry in an unfamiliar language and script is also problematic. However, it requires far less training than learning to speak, write and understand a language. Off-shore data-entry is sometimes carried out by operators who do not know English. Anecdotal evidence suggests that on printed matter at least, non-speakers are more accurate, because they must look at each letter.

# 8 Evaluation Plans and Issues

The requirements for evaluating a CAVIAR-like approach to document analysis differ from traditional, fully-automated techniques. In the latter case, the focus is entirely on classification accuracy, whereas in the former, the system is comprised of a human user and a machine working in collaboration: both time and accuracy are important measures. It is unlikely the system will be more accurate than a human working alone, it must be faster. Similarly, it is unlikely the system will be faster than a fully-automated algorithm, it must be more accurate. Note also that such work naturally demands the use of human test subjects, unlike most pattern recognition research.

One of the most intriguing aspects of the earlier CAVIAR work was its demonstration of an inherent potential to incorporate adaptation, yielding a system that improves with use. This also provides the opportunity to study human learning: the operator and the system learn together, a little like a blind person and her Seeing Eye dog. As was done in the case of CAVIAR for flowers, we plan to study this effect.

It is also evident that one of the most plausible targets for a CAVIAR-like system involves documents written in a language that the user is unable to read, perhaps using a script that he/she does not recognize. A typical end-to-end scenario, then, might including the following processing stages:

1. Categorization: source, age, type of document, size, quality, language, script.
2. Image-level preprocessing including various levels of segmentation.
3. Word recognition.
4. Translation.
5. Information importance assessment/transformation into actionable knowledge.

Our initial experiments will be conducted on workstations with simulated mobile-device window sizes.

Among other expected benefits, we believe that the development cycle for interactive recognition systems may be faster than for wholly automated systems. One reason to suspect this will be true is that it is no longer necessary to take care of rare contingencies: they can be dealt with by the human operator, who can always override every automated option. This will allow deployment of a whole family of support systems, specialized to different levels of operator familiarity with the target language(s) and script(s).

We intend to develop the interfaces necessary to accomplish the above interactive tasks in parallel with data collection. While this would have been a monumental endeavor only a few years ago, today excellent tools are available for the purpose. Perhaps the most difficult part of the project is devising and incorporating a customized logging system for timing and recording every interaction and its effect. This must be done in sufficient detail to allow replaying the whole interactive document information acquisition sequence in order to find out where and what goes wrong.

The individual files log must be compiled and aggregated for statistical analysis. This will also form the basis for evaluating the interactive system and for comparing it with entirely automated and entirely manual data entry.

We emphasize that we do not address the extraction of data from a large backlog of accumulated documents. Unlike much of the academic research on document analysis, our interest here is not in archival documents, but in those whose value decreases exponentially with time. We propose a system that may prevent future backlogs of nearly worthless documents by on-site, just-in-time information extraction.

## References

[1] M. Ancona, S. Locati, M. Mancini, A. Romagnoli, and G. Quercini. Comfortable textual data entry for PocketPC: the WTX system. In *Advances in Graphonomics, Proceedings of IGS 2005*, Salerno, Italy, June 2005.

[2] H. S. Baird and D. P. Lopresti, editors. *Human Interactive Proofs: Second International Workshop*, volume 3517 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, Germany, 2005.

[3] S-H. Cha, A. Evans, A. Gattani, G. Nagy, J. Sikorski, C. Tappert, P. Thomas, and J. Zou. Computer Assisted Visual Interactive Recognition (CAVIAR) technology. In *Proceedings of the IEEE Electro/Information Technology Conference*, May 2005. http://www.ecse.rpi.edu/homepages/nagy/PDF_files/EIT2005_Evans_etal.pdf.

[4] J. Edwards. New interfaces: Making computers more accessible. *IEEE Computer*, pages 12–14, December 1997.

[5] W. K. English, D. C. Engelbart, and M. L. Berman. Display-selection techniques for text manipulation. *IEEE Transactions on Human Factors in Electronics*, HFE-8(1):5–15, March 1967.

[6] A. Evans, J. Sikorski, P. Thomas, J. Zou, G. Nagy, S.-H. Cha, and C. C. Tappert. Interactive visual system. Technical Report CSIS No. 196, Pace University, 2003. http://csis.pace.edu/∼ctappert/ivs/ivs-nagyetal.pdf.

[7] F. Fisher. Digital camera for document acquisition. In *Symposium on Document Image Understanding Technology*, Columbia, MD, 2001. http://www.dtic.mil/matris/sbir/sbir022/a038.pdf.

[8] H. Fujisawa, H. Sako, Y. Okada, , and S. Lee. Information capturing camera and developmental issues. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition (ICDAR'99)*, pages 205–208, Bangalore, India, September 1999.

[9] T. Hedgpeth, M. Rush, J. Black, and S. Panchanathan. The iCare project reader. In *Sixth International ACM SIGACCESS Conference on Computers and Accessibility*, October 2004. http://cubic.asu.edu/grants/downloads/RDE/iCareReader.doc.

[10] M. Holland and C. Schlesiger. High-mobility machine translation for a battlefield environment. In *Proceedings of NATO/RTO Systems Concepts and Integration Symposium*, volume 15, pages 1–3, Monterey, CA, May 1998.

[11] C. Jacobs and P. Simard. Low resolution camera based OCR. *International Journal on Document Analysis and Recognition*. To appear.

[12] C. L. James and K. M. Reischel. Text input for mobile devices: Comparing model prediction to actual performance. In *Proceedings of the ACM Conference on Computer-Human Interaction*, pages 365–371, 2001.

[13] D. J. Langendorf. Textware solution's Fitaly keyboard v1.0 easing the burden of keyboard input. *WinCELair Review*, February 1998.

[14] S. Mackenzie and W. Soukoreff. Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction*, 17:147–198, 2002.

[15] T. Masui. An efficient text input method for pen-based computers. In *Proceedings of the ACM Conference on Computer-Human Interaction*, pages 328–335, 1998.

[16] G. Miller. The magical number seven plus or minus two; some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 1956.

[17] G. Nagy, L. Li, A. Samal, S. Seth, and Y. Xu. Cooperative text and line-art extraction from a topographic map. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition (ICDAR'99)*, pages 467–470, Bangalore, India, September 1999.

[18] G. Nagy, L. Li, A. Samal, S. Seth, and Y. Xu. Integrated text and line-art extraction from a topographic map. *International Journal on Document Analysis and Recognition*, 2(4):177–185, June 2000.

[19] J. P. Peters, C. Thillou, and S. Ferreira. Embedded reading device for blind people: a user-centred design. In *IEEE Emerging Technologies and Applications for Imagery Pattern Recognition (AIPR 2004)*, pages 217–222, 2004.

[20] S. Srihari and Z. Shi. Forensic handwritten document retrieval system. In *First International Workshop on Document Image Analysis for Libraries (DIAL'04)*, pages 188–194, 2004.

[21] K. Swan. FALCon: Evaluation of OCR and machine translation paradigms, August 1999. http://www.arl.army.mil/seap/reports/kreport.pdf.

[22] J. Yang, X. Chen, J. Zhang, Y. Zhang, and A. Waibel. Automatic detection and translation of text from natural scenes. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, May 2002.

[23] J. Zou. *Computer Assisted Visual InterActive Recognition*. PhD thesis, Rensselaer Polytechnic Institute, 2004.

[24] J. Zou and G. Nagy. Evaluation of model-based interactive flower recognition. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 2, pages 311–314, 2004.