Interactive, Mobile, Distributed Pattern Recognition

George Nagy

DocLab, Rensselaer Polytechnic Institute, Troy, NY USA 12180
nagy@ecse.rpi.edu

Abstract. As the accuracy of conventional classifiers, based only on a static partitioning of feature space, appears to be approaching a limit, it may be useful to consider alternative approaches. Interactive classification is often more accurate then algorithmic classification, and requires less time than the unaided human. It is more suitable for the recognition of natural patterns in a narrow domain like trees, weeds or faces than for symbolic patterns like letters and phonemes. On the other hand, symbolic patterns lend themselves better to using context and style to recognize entire fields instead of individual patterns. Algorithmic learning and adaptation is facilitated by accurate statistics gleaned from large samples in the case of symbolic patterns, and by skilled human judgment in the case of natural patterns. Recent technological advances like pocket computers, camera phones and wireless networks will have greater influence on mobile, distributed, interactive recognition of natural patterns than on conventional high-volume applications like mail sorting , check reading or forms processing.

1. Introduction

I am grateful for this wonderful opportunity to proselytize for some heretical notions. First, I will suggest classifying pattern recognition applications into types A, B, AB, and O, according to the pattern recognition methodology that suits each best. Type A consists of *symbolic patterns*, the glyphs and sounds used for encoding messages. Type B includes *natural objects* like flowers and faces that are not used primarily for communication. I will try to substantiate the claim that interactive computer vision, where the tasks leading to object recognition are assigned according to the relative competence of human and machine, is particularly appropriate for Type B applications. On the other hand, context and style based classification seems better suited to Type A applications. Learning and adaptation benefit every type.

   Section 2 outlines the considerations that led to the proposed taxonomy of recognition problems. Section 3 summarizes our recent results on interactive classification of flowers and faces. Section 4 and 5 present the corollaries of interactive classification: mobile and networked recognition. In Section 6 we recapture the notion of style, show that it can lead to more accurate classification of multi-source patterns when the test samples are partitioned by source, and contrast it to the better established methods based on language context. In the last section I list some areas where rapid progress may be possible. This is not a survey: however, the cited publications contain extensive references to invaluable prior work by others.

2. Symbolic and Natural Patterns

Prototypical examples of Type A applications are character, hand print, and speech recognition (OCR, ICR and ASR). The first operational OCR system was installed at Readers' Digest in 1955. Eleven years later, at the 1966 IEEE Pattern Recognition Workshop in Puerto Rico, postal address readers, form processing, and spoken word recognition were among the most popular topics, and they remain so today. The early OCR systems were so expensive that they could not be justified unless they displaced 10-20 keypunch operators. Target rates were 1000-2000 characters per second. High throughput was necessary because our garrulous species spawns endless streams of print that we yearn to preserve for posterity. Type A applications share the following characteristics:

- they deal with symbolic patterns that represent natural or formal languages;
- any reader or speaker of the particular language can perform the classification manually;
- they require high throughput because every message consists of many patterns;

- many (millions) of samples are available for training;
- formal models of *context* (morphological, lexical, syntactic, pragmatic [1])
  and of *style* (typefaces, hand print, dialects [2, 3, 4]) have been developed;
- the error/reject tradeoffs are well understood [5];
- the classes are well defined: there are exactly ten digits and, in Italian, 21 letters of the alphabet;
- in feature space, the class centroids are located at the vertices of a regular simplex [6, 7, 8].

Examples of Type B include the recognition of birds, flowers and trees, many biometric applications, and biomedical pattern classification where the cost of preparing the samples often dominates the cost of recognition. Type B applications

- deal with natural patterns which may have developed without the discriminability of symbolic patterns;
- must be classified on demand rather than as part of a work-flow;
- are recognized only by relatively few, highly trained experts (bird-watchers, foresters, physicians);
- often have only small training sets because of the high cost of labeling;
- seldom have established models of context or style;
- because of the unpredictable cost of errors, require every decision to be checked by a human;
- exhibit a soft, hierarchical class structure, subject to change.

Type AB applications have some characteristics of both Type A and Type B. An example is genetic sequence decoding. We defer consideration of Type O.

3. Interaction

Almost all operational pattern recognition systems require some human interaction, at least at the beginning or end. We focus here on systems where human and machine take turns to reach a decision that assigns a particular object (a flower, tree, face, or skin lesion) to a particular class.

There are essential differences between human and machine cognitive abilities. Humans excel in gestalt tasks, like object-background separation. They apply to recognition a rich set of contextual constraints gleaned from previous experience rather than from specific training and test sets. They have superior noise-filtering abilities (particularly with respect to colored noise.) They can easily read degraded text on which the best OCR systems produce only gibberish. Computer vision systems, on the other hand, still have difficulty in recognizing "obvious" differences and "generalizing" from limited training sets.

Computers, however, can perform many tasks faster and better. Computers can store thousands of images and the associations between them, and never forget a name or a label. They can evaluate geometrical properties like high-order moments whereas a human is challenged to determine even the centroid of a convoluted figure. They can compute wavelets and other kernel transforms to differentiate textures. Computers can quickly measure lengths and areas, count thousands of connected components and sort them according to various criteria (size, aspect ratio, convexity). They can flawlessly evaluate multivariate conditional probabilities, decision functions, logic rules, and grammars. In contrast, George Miller's psychophysical experiments in the 1950's revealed that humans have limited short-term memory (only ±7 items) and poor absolute judgment.

We believe that the key to effective interactive recognition is a *visible model* to mediate human-computer communication. The model is a geometric and topological abstraction of an object represented in a picture. It guides the machine to extract discriminative, localized intensity, color and texture features. The model mediates only a restricted set of information. It does not tell the computer anything about the rich perceptions that lead the operator to correct or approve the model, and it does not tell the human about the configuration of the resulting feature vectors in high-dimensional feature space.

We have developed CAVIAR (*Computer Assisted Visual Interactive Recognition*) systems for some Type B applications [9 - 12]. Figure 1 shows examples of our flower and face models. These models are constructed automatically, and corrected interactively only when necessary. The line drawing superimposed on the picture lets the operator judge whether the computer-suggested model fits the unknown object. We restrict the model to isolated points and curves, because color and intensity are difficult to modify with a mouse or stylus. In effect, the user can point and drag, but not paint or shade. The model is only an abstraction: by itself, it is not sufficient for classification by either human or computer. Both must
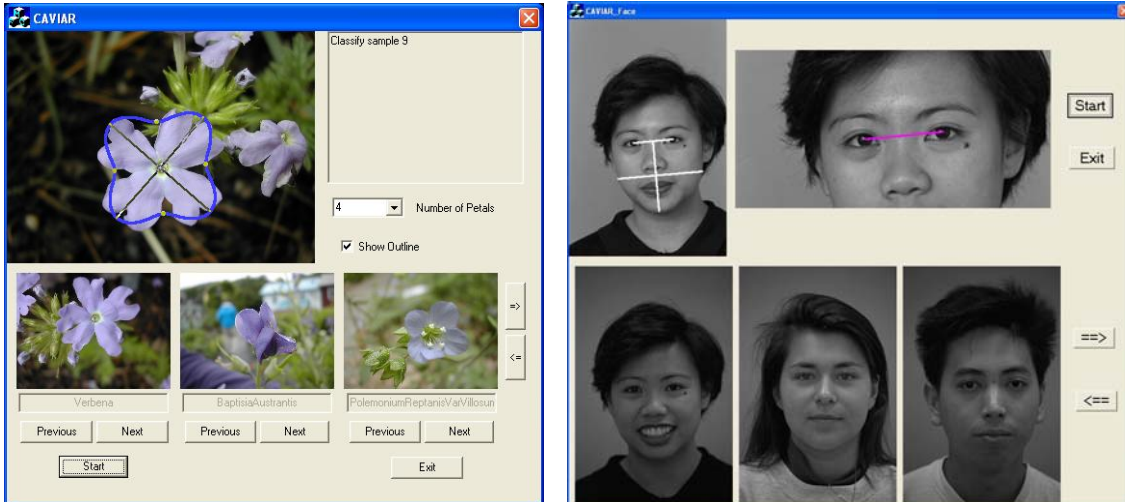
Fig. 1. CAVIAR-flower (left) and CAVIAR-face (right) visible models and graphic user interfaces. Because pupil localization is so important in face recognition, an enlarged view is provided for this part of the visible model.

have unlimited access to, and make use of, the entire pixel array. The purpose of the model is only to guide the machine in extracting good features for classification.

A model instance need not depict faithfully intensity, color, or texture edges. A poorly fitting model may suffice to classify an "easy" object. Conversely, even an accurate model may result in ambiguous features. (One consequence of the role of the model in our system is that there can be no "ground truth" for it. Several models, or none, may lead to features that cause the correct candidate to be ranked on top.) The computer displays, in addition to the visible model, a set of reference pictures ranked according to the posterior class probabilities of the unknown object. The operator can correct the model if the top-ranked classes are implausible and if there are obvious mismatches between the current model and the unknown object. The operator can also scroll the reference patterns ("browse") to inspect candidates ranked below the top three. When an acceptable reference candidate appears on the display, the operator clicks on it to assign its class to the unknown object. Two CAVIAR graphic user interfaces (GUIs) are shown in Figure 1.

We compared CAVIAR to machine alone and to human alone in experiments conducted on a 612-flower database [10] and on the FERET face recognition benchmark [12]. Table 1 summarizes the results.

Table 1. CAVIAR compared to machine alone and to human alone

	Flowers		Faces	
	Accuracy	Time per flower	Accuracy	Time per face
Interactive	93%	12 sec	99.7%	7.6 sec
Machine alone	32%	**---**	48.0%	----
Human alone	93%	26 sec	~100.0%	66.3 sec

From this table, it appears that on some Type B applications interactive classification is more than twice as accurate than automated classification (at least with our classification algorithms), and more than twice as fast than unaided naïve human subjects.

In our CAVIAR-face experiments, 50 faces ("probes") were classified to one of 200 classes ("gallery") by naive subjects. Each subject classified a randomly selected set of 50 faces. The fraction of faces recognized correctly after each adjustment, and the time required, are shown in Table 2. For example, it takes 10.6 seconds to classify pictures that require two adjustments. Such pictures represent 15% of the total number of test pictures.

It is seen that the automated rank-ordering algorithm ranks the correct reference picture in the top three about 50% of the time. These faces are classified very quickly. Only about 1% of the faces require more than five model adjustments. Subjects seldom use the browsing option because it is slow. (A larger display would speed up the process further by displaying, with sufficient resolution for easy identification, more than the top three candidates.)

Table 2. Cumulative record of face recognition experiments. At each iteration, the operator can make an immediate decision (SELECT), modify the model (ADJUST), or look at more reference pictures (BROWSE). The times shown include the sequence of adjustments or browsing prior to classification. For example, after two adjustments, 85% of the faces are classified, in 5.0 seconds per face on average.

Iteration #	ADJUST %	SELECT %	Ave. time sec	BROWSE %	Ave. time sec	Classified %
0	48.7	50.3	2.3 sec	1.0	7.7	51.3
1	28.3	19.7	7.7 sec	0.7	16.1	71.7
2	15.0	13.3	10.6 sec	0.0	--	85.0
3	9.3	4.7	14.4 sec	1.0	42.6	90.7
4	3.7	5.3	16.6 sec	0.3	23.2	96.3
5	1.0	2.0	19.6 sec	0.7	33.2	99.0
6	0.7	0.3	42.0 sec	0.0	--	99.3
7	0.3	0.3	34.7 sec	0.0	--	99.6
8	0.0	0.0	--	0.4	49.8	100.0
Total		95.9		4.1		

The rank-ordering mechanism is chosen according to the number of classes and the number of reference samples per class. For CAVIAR-flower, where we ran experiments with 1, or 3, or 5 reference samples per class, we used nearest-neighbors with 8 features [10]. For CAVIAR-face, where we had only one reference sample per class, we used the Borda Count on small patch features near the eyes and mouth. The test samples are added to the reference set after they are classified, thereby gradually improving the statistical estimates for automated model construction and for rank ordering. Despite the occurrence of misidentified samples and the lack of adequate theoretical justification, we have always found this type of adaptive learning extremely effective [14 - 17, 10]

In summary, CAVIAR combines the following aspects that contribute to its accuracy and speed:
- interaction throughout the classification process, rather than only at the beginning or end;
- automatic correction and interactive modification of a simple, domain-specific visible model which guides automated feature extraction;
- pruning the reference pictures according to their similarity with the unknown, and displaying the top candidates for selection by the operator;
- adaptive re-estimation of internal parameters;
- leaving the human wholly in charge throughout the process by letting him or her decide when to classify an object, when to modify the model, and when to browse lower-ranked reference pictures.

4. Mobile Recognition

Mobile recognition systems with hand-held cameras offer obvious advantages for recognizing objects outside the office or home. The interaction takes place with stylus or thumb on the photo display screen. The computation can be carried out on the same platform or through a wireless link to a nearby laptop or to the Internet. Figure 2 shows our Mobile-CAVIAR interface on a Toshiba pocket computer with a plug-in camera and a Wi-Fi card [18].

One of the most interesting aspects of mobile systems is their potential capability to improve classification accuracy because the operator can immediately take additional pictures of a difficult object. Classification can then be based on several pictures by merging relevant information. However, we don't yet know just how to do this effectively. Note that this is neither classifier combination nor sensor fusion: it is more akin to 3-D model construction from multiple views, except that the objective here is classification rather than representation. Some of the problems awaiting solution are listed below [18].

- When is a single picture of the object not enough, and another required?
- What viewpoint, scale, or illumination is best for a second or third picture?
- Should the computer or the human decide? If the former, how should it specify the desired picture?
- At what level(s) (pixel, feature, classifier) should the information from multiple pictures be combined?

- Several views of an object would require these views to be represented in the reference set. What would be an appropriate sampling scheme to ensure this during enrollment of training/reference samples?

Although PDA cameras still lag stand-alone digital cameras in terms of optical and digital resolution (and convenience features), the greatest limitation of handheld systems compared to PC recognition platforms is their limited screen size, which prevents simultaneous display of several objects in adequate detail (multiple zooms are disorienting). It is clear that camera-phones will soon have enough storage and computing capacity, as well as appropriate operating systems, for interactive recognition. However, the display size limitation will be even more stringent – current cell phones have display sizes ranging from 96x64 to 128x160 pixels. Some expect that it will be overcome by wireless access to large public displays [19], but we can hardly expect a public display to pop up whenever we wish to recognize a flower or a face. Furthermore, cell phones lack a direct pointing device: fitting a visible model with arrow keys is clumsy

Fig. 2. M-CAVIAR graphic user interface. The model does not have to be perfect to rank the unknown in the top 3

5. Pattern recognition networks

Interactive pattern recognition may be ready to benefit from the kind of distributed computing envisioned by the creators of the ARPANET 30 years ago. Experts anywhere on the Internet can, in principle, interact with any image through its visible model. Just as bird watchers and wildflower enthusiasts band together for collective judgment, dermatologists at dispersed locations will be able to pool their expertise to diagnose difficult skin lesions. The interaction may take place in parallel, or in a hierarchy where the more difficult pictures percolate to the more qualified experts.

As mentioned already, one advantage of human interaction is that newly labeled specimens can be safely added to the training or reference database, thus improving the performance of the automated model construction, feature extraction, and rank ordering. We don't yet know whether a democratic CAVIAR will be more accurate than a CAVIAR operated by a Designated Expert. In any case, the larger sample size resulting from many persons collecting pictures, correcting models, and assigning class labels should benefit overall performance [20].

A typical network interaction session of our wireless ethernet (802.11b) link through TCP/IP sockets is shown in Figure 3. Whenever the PDA sends data, it blocks execution flow till it receives a response from the server. After interpreting the required type of service according to the header data of the byte stream, the appropriate server routine is invoked. Each user interaction requires a response from the server to update the display, but the only long message is uploading a new picture to the host computer. With current GSM, TDMA or CDMA cell phone links, this would take several seconds even if the picture were first compressed. However, the third generation cellular networks already being deployed will have more than sufficient bandwidth.
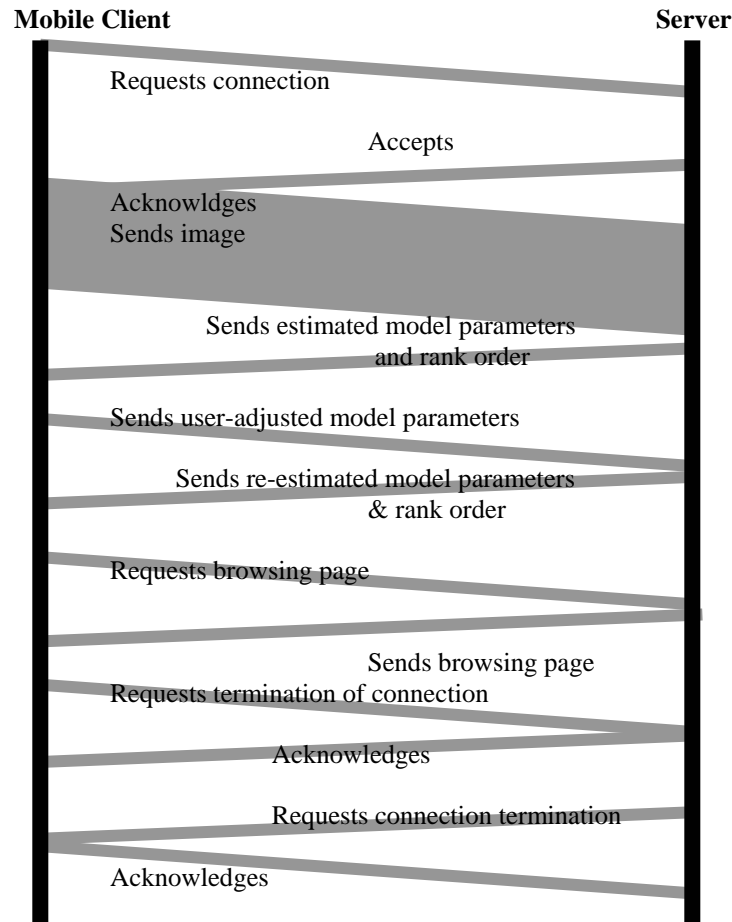
Mobile Client   **Server**

Requests connection

Accepts

Acknowldges
Sends image

Sends estimated model parameters
and rank order

Sends user-adjusted model parameters

Sends re-estimated model parameters
& rank order

Requests browsing page

Sends browsing page
Requests termination of connection

Acknowledges

Requests connection termination

Acknowledges

Fig. 3. Communication sequence between the PDA and the server for identifying a test sample {adapted from [18])

6. Style and Context

Interactive classification is too slow for Type A applications, so we must resort to other means to improve the conventional classification paradigm. Because they normally convey messages, symbolic patterns tend to appear together in groups (fields) that have a common origin. A printed message is usually a field of character patterns printed in the same font. Each font makes use of only a few well-matched typographic components—bowls, stems, bars, finials, and serifs. Printers, copiers and scanners leave individual imprints that differentiate documents from different sources. Hand printing and cursive writing are characterized by a certain writer-dependent uniformity of strokes and spacing. Therefore the feature-space representation of any single postal address, bureaucratic form, or printed article displays a measure of homogeneity due to isogeny (common origin). We say that isogeny induces *style* in features measured on patterns (cf. Fig 4a).

It is possible to model "style" mathematically and thereby develop a basis for more accurate classification of a group (field) of digitized characters from the same source. The features of patterns co-occurring in a field are statistically dependent because they share the same style. Effects of style

consistency on the distributions of field-features (concatenation of pattern features) can be modeled by discrete, hierarchical, or continuous mixtures of Gaussian variables. Based on such a model, a style-constrained classifier can be constructed to recognize entire fields of patterns rendered in a consistent but unknown style. In experiments on printed alphabets and on National Institute of Standards and Technology (NIST) hand-printed test sets, style constrained classification reduced errors on fields of digits by nearly 25 percent over singlet classifiers [2, 3]. We are currently trying to develop style-constrained SVMs that should yield even lower error rates when the underlying densities are not Gaussian.

1 k 4 d 7 h 1	Isabella
1 k 4 d 7 h l	l40 mm long
(a)	(b)
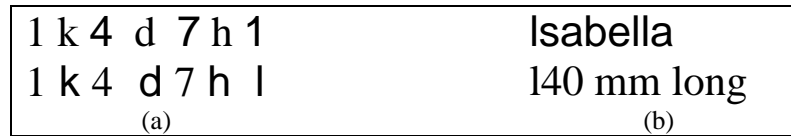
Fig. 4. (a) Style: in the part numbers on the left, the first pattern in the top row must be the letter l, and the identical first pattern in the bottom row must be the numeral 1. (b) Language context: here the surrounding text disambiguates the identical glyphs in either row

Language context has been exploited in OCR and ASR far longer than style. Unlike style, textual context depends on the reading order of the patterns. Instead of modeling the dependence between the *features* of patterns in a same-source field, it models the dependence between their *labels* (Fig. 4b). For use in classification, morphological, lexical and syntactic conventions are converted into letter n-gram frequencies, word frequencies, and word transition probabilities. A vocabulary of 60,000-100,000 words provides very thorough coverage of English text, but three or four times more entries are needed for highly-inflected languages such as Italian or Hungarian. Specialized lexicons are needed for mail sorting, telephone directories, technical handbooks, part-number catalogs, and other non-narrative compendia. Nowadays such information can be readily obtained from the Web.

Specialized syntactic conventions dictate the placement of punctuation marks and the construction of abbreviations, citations, legal and courtesy amounts on checks, and postal addresses. Mathematical formulas, chemical structure diagrams, and arithmetic redundancy in financial documents also have their own rules. These rules have been compiled in style guides ranging from a few pages for technical journals and mailing instructions to manuals of hundreds of (web) pages. Markov models provide an efficient and trainable alternative to formal parsing. Although several books have been published on the appropriate design and layout (i.e., syntax) of tables, tables have proved singularly resistant to correct interpretation.

In the last decade, segmentation, context, and shape based classification have been successfully integrated into algorithms that search a trellis of trial segmentations with transitions dictated by context. The weight given to shape information vs. contextual information depends on the application. In the limit, with no prior shape information but perfect shape consistency, the recognition problem is equivalent to decoding a substitution cipher [21 - 24].

The distinction between the various types of statistical dependence underlying correlated features, styles, and language context is clarified when they are modeled with graphical Bayesian networks [25].

7. Applications

Type A Applications. The manifold applications of character and speech recognition are already well known. Further research on style based recognition of fields of characters or even entire documents is much more convenient than on interactive recognition, because it does not require interaction with human subjects. We believe, however, that it too is inspired by human recognition. Indeed, the automatic recognition of *isolated* printed and hand-printed letters or digits is at least as accurate as human recognition. It is only when we are faced with a coherent sequence of patterns designed to deliver a message that human recognition is still far superior. Style and context can narrow the gap.

Type B applications. I believe that the time is right for developing interactive, mobile pattern recognition applications. I am excited about their potential impact on education, even though none of the new waves of

technology – 16mm movies, radio, television, time-sharing, personal computers, laptops, web-teach – has lived up to its educational expectations. The necessary mobile wireless computing platforms (PDAs and cell phones) are already widespread among the school-age population. Education – from elementary to graduate school – involves many visual recognition tasks. At various stages of our academic careers, we learn to identify flowers, trees, rocks, insects, clouds, paintings, statues, architectural styles. Applications to industrial training – recognizing electrical and mechanical components and assemblies – may also be worth exploring. Interactive computer-aided recognition could enhance learning almost as much as visiting a zoological or botanical garden or a museum with an expert personal tutor. Group dynamics can be added by networked recognition, either with nearby classmates or students and teachers at the antipodes. CAVIAR-like instructional systems will be considered successful only if, after a period of use, they can be discarded.

Visitors who saw a demonstration of CAVIAR have suggested several applications: searching for new medicinal plants in the jungle; identifying endangered cryptic cats (like jaguars) caught in photo-traps in Costa Rica; fruit and vegetable checkout in supermarkets; assisting farmers and foresters to identify pests and crop diseases; helping to learn the artistic style of painters in art history classes; assessing the value of collectibles (spoons, coins, stamps, porcelain dolls) using home web cams (to draw customers to advertising websites).

Among medical applications, the recognition of skin lesions seems particularly appropriate, because they are often diagnosed by inspection. Large dermatological atlases are available on the Web, but the most of the posted photographs were obtained at high resolution and with specialized lighting, and these databases contain only one or two examples of each disease. Collecting appropriate data with a mobile platform entails restrictive ethical considerations and will require close collaboration with medical researchers. Anyone of the following types of visible skin conditions appears ripe for experimentation:

- Cosmetic dermatology, scar assessment, beauty-aids;
- Infectious and contagious diseases with spots: measles, chickenpox, rubella;
- Rashes: hives, eczemas, psoriasis;
- Burns; cuts, frostbite;
- Sexually transmitted diseases;
- Poisonous plants and bugs: poison ivy, insect bites;
- Bio-terrorism agents: cutaneous anthrax, smallpox, plague, tularemia.

A personal diagnostic system may be appropriate when the patient is too embarrassed to seek help, or when medical personnel is not available (on battlefields or expeditions, or in impoverished rural areas). The same platform can help collect ancillary information in complete privacy: *Does it itch? How long have you had it? Do you feel lethargic? Did you eat fish recently?* Close-ups and pictures of healthy areas of the skin can be taken with different sources of illumination for comparison. If the severity of the condition warrants it, or if a confident diagnosis cannot be reached, the system may also request its owner to forward electronically both the pictures and the ancillary information to the appropriate health maintenance network. Paramedical personnel or physicians not specialized in dermatology could use CAVIAR-derma for continuing health care education.

Acknowledgments

References

1. G. Nagy, Teaching a computer to read, *Procs. Int'l Conf. Pattern Recognition (ICPR XI),* vol. 2, pp. 225-229, The Hague, August 1992.
2. P. Sarkar, G. Nagy, Style consistent classification of isogenous patterns, *IEEE Trans. PAMI,* vol. 27, no. 1, pp. 88-98, Jan. 2005,
3. S. Veeramachaneni, G. Nagy, Style context with second order statistics, *IEEE Trans. PAMI,* vol. 27, no. 1, pp. 14-22, Jan. 2005.
4. N. Nagy, X. Zhang, G. Nagy, E.W. Schneider, A quantitative categorization of phonemic dialect features in context, *Procs. International Conference on Context,* Paris, July 2005.
5. G. Nagy, P. Sarkar, Document style census for OCR*, Procs. First International Workshop on Document Image Analysis for Libraries* (DIAL04), Palo Alto, CA, IEEE Computer Society Press, pp. 134-147, January 2004.
6. G. Nagy and S. Veeramachaneni, A Ptolemaic model for OCR, *Procs. ICDAR-03*, Edinburgh, pp. 1060-1064, August 2003.
7. G. Nagy, Visual pattern recognition in the years ahead, *Procs. ICPR XVII*, vol. IV, pp. 7-10, Cambridge, UK, August 2004.
8. G. Nagy, X. Zhang, Simple statistics for complex features spaces, to appear in Data Complexity in Pattern Recognition, Springer Verlag, Editors: Mitra Basu & Tin Kam Ho, Publication Date: December 2005.
9. G. Nagy, J. Zou, Interactive visual pattern recognition, *Procs. ICPR XVI,* IEEE Computer Society Press, Aug. 2002, Vol. III, pp. 478-481.
10. J. Zou, G. Nagy, Evaluation of model-based interactive pattern recognition, *Procs. ICPR XVII*, Vol. II, p. 311-314, Cambridge, UK, August 2004.
11. J. Zou, G. Nagy, Human-computer interaction for complex pattern recognition problems, to appear in Data Complexity in Pattern Recognition, Springer Verlag, Editors: Mitra Basu & Tin Kam Ho, December 2005.
12. A. Evans, J. Sikorski, P. Thomas, S-H Cha, C. Tappert, J. Zou, A. Gattani, G. Nagy, Computer Assisted Visual Interactive Recognition (CAVIAR) Technology, *Procs. IEEE International Conference on Electro-Information Technology,* Lincoln, NE May 2005.
13. J. Zou, A Model-Based Interactive Image Segmentation Procedure, *IEEE Workshop on Applications of Computer Vision (WACV), Breckenridge CO, January* 2005.
14. G. Nagy, G.L. Shelton, Self-Corrective Character Recognition System, *IEEE Transactions on Information Theory IT-12,* #2, pp. 215-222, April 1966.
15. H.S. Baird, G. Nagy, A Self-correcting 100-font Classifier, *Proc. SPIE Conference on Document Recognition, Volume SPIE-2181,* pp. 106-115, San Jose, CA, February 1994.
16. S. Veeramachaneni, G. Nagy, Adaptive classifiers for multi-source OCR, *IJDAR* vol 6*,* no. 3, pp. 154-166, March 2004.
17. G. Nagy, Classifiers that improve with use, *Procs. Conference on Pattern Recognition and Multimedia ,* IEICE, Tokyo, pp. 79-86, February 2004.
18. A. Gattani, Mobile Interactive Visual Pattern Recognition, MS thesis, Rensselaer Polytechnic Institute, December 2004.
19. M. Raghunath, C. Narayanaswami, and C. Pinhanez, Fostering a Symbiotic Handheld Environment, *IEEE Computer,* pp. 56-65, Sept. 2004.
20. J. Zou, A. Gattani, Computer Assisted Visual InterActive Recognition and Its Prospects of Implementation Over the Internet, *IS&T/SPIE 17th Annual Symposium Electronic Imaging, Internet Imaging VI*, 2005.
21. R.G. Casey, G. Nagy, Autonomous Reading Machine, *IEEE Transactions on Computers C-17*, #5, pp. 492-503, May 1968.
22. R.G. Casey, G. Nagy, Advances in Pattern Recognition, *Scientific American 224,* #4, pp. 56-71, 1971.
23. G. Nagy, S. Seth, K. Einspahr, Decoding Substitution Ciphers by means of Word Matching with Application to OCR, *IEEE Transactions on Pattern Analysis and Machine Intelligence 9,* #5, pp. 710-715, September 1987.
24. T.K. Ho, G. Nagy, OCR with no shape training, *Procs. ICPR-XV,* vol. 4, pp. 27-30, Barcelona, September 2000.
25. S. Veeramachaneni, P. Sarkar, G. Nagy, Modeling context as statistical dependence, *Procs. International Conference on Context,* Paris, July 2005.

Most of these papers are available as PDF files at: http://www.ecse.rpi.edu/homepages/nagy/