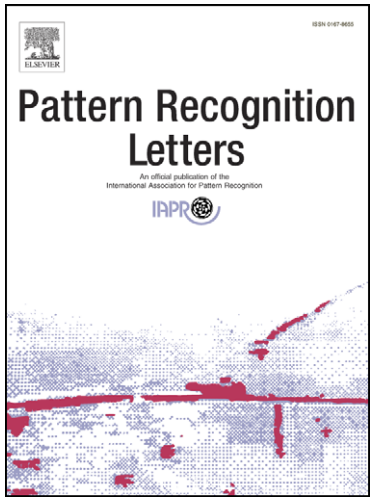

Visible Models for Interactive Pattern Recognition

Jie Zou, George Nagy

PII: S0167-8655(07)00240-1
DOI: 10.1016/j.patrec.2007.08.005
Reference: PATREC 4220

To appear in: *Pattern Recognition Letters*

Received Date: 15 March 2007
Revised Date: 8 June 2007
Accepted Date: 1 August 2007

Please cite this article as: Zou, J., Nagy, G., Visible Models for Interactive Pattern Recognition, *Pattern Recognition Letters* (2007), doi: 10.1016/j.patrec.2007.08.005

Visible Models for Interactive Pattern Recognition

Jie Zou, National Library of Medicine, Bethesda, MD

*George Nagy, DocLab, Rensselaer Polytechnic Institute, Troy, NY

***Abstract*—** The exchange of information between human and machine has been a bottleneck in interactive visual classification. The *visible model* of an object to be recognized is an abstraction of the object superimposed on its picture. It is constructed by the machine but it can be modified by the operator. The model guides the extraction of features from the picture. The classes are rank ordered according to the similarities (in the hidden high-dimensional feature space) between the unknown picture and a set of labeled reference pictures. The operator can either accept one of the top three candidates by clicking on a displayed reference picture, or modify the model. Model adjustment results in the extraction of new features, and a new rank ordering. The model and feature extraction parameters are re-estimated after each classified object, with its model and label, is added to the reference database. Pilot experiments show that interactive recognition of flowers and faces is more accurate than automated classification, faster than unaided human classification, and that both machine and human performance improve with use.

Keywords — Pattern Recognition, Human-Computer Interaction, Visible Model, Face Recognition, Flower Recognition.

* Corresponding author. E-mail: nagy@ecse.rpi.edu.
J. Zou, was with Rensselaer Polytechnic Institute, Troy, NY 12180 USA. He is now with the National Library of Medicine, Bldg. 38A, Rm. 10S1000, 8600 Rockville Pike, Bethesda, MD 20894 USA (telephone: 301-496-7086, E-mail: jzou@mail.nlm.nih.gov).
G. Nagy is with the Department of Electrical, Computer and Systems Engineering, JEC 6020, 110 8th street, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (telephone: 518-276-6078, FAX: 518-276-6261, E-mail: nagy@ecse.rpi.edu).

1. INTRODUCTION

CAVIAR (Computer Assisted Visual InterActive Recognition) is a paradigm for interaction in narrow domains where higher accuracy is required than is currently achievable by automated systems, but where there is enough time for limited human interaction. The key to efficient interaction in CAVIAR is a *visible model*, overlaid on the unknown picture, which provides two-way communication between human and machine.

As early as 1992, a workshop organized by the National Science Foundation in Redwood, California, recommended that "computer vision researchers should identify features required for *interactive image understanding*, rather than their discipline's current emphasis on automatic techniques" (Jain, 1992). A panel discussion at the 27th AIPR Workshop asserted "... the needs for Computer-Assisted Imagery Recognition Technology" (Mericsko, 1998). Kak's ICPR'02 keynote emphasized the difficulties facing fully automated model-based vision (Kak and Desouza, 2002).

2. PRIOR WORK

In the broad domains of Content-Based Image Retrieval (CBIR), *relevance feedback* has been found effective (Rui et al. 1998). Interaction has been, however, necessarily limited to the initial query formulation and the selection of acceptable and unacceptable responses (Harman, 1992; Cox et al. 2000; Carson et al., 2002). The designer of "Blobworld" (Carson et al., 2002) suggested that the CBIR system should display its representation of the submitted and returned images.

Active Learning makes use of human intervention to reduce the number of training samples that the classifier needs to achieve a target error rate, but without interacting with images of the unknown object (MacKay 1992; Cohn et al., 1996).

In many OCR applications, every scanned page is checked by the operator before further processing (Bradford, 1991; Dickey, 1991; Klein and Dengel, 2004). In camera-based text recognition, the operator defines a bounding box (Haritaoglu, 2001; Zhang et al., 2002). In face recognition, the operator sets the pupil-to-pupil baseline (Yang et al., 2002). In all three tasks, the operator intervenes again at the end to classify *rejects* (Sarkar et al., 2002). In contrast, we propose that the human and the machine take turns throughout the classification process, each doing what they do best. A survey of the relevant psychophysical literature appears in (Zou and Nagy, 2006).

Chen et al. give examples of the use of landmarks for classifying species of the fish genus Carpiodes (Chen et al., 2005). Homologous landmarks are also well established in aerial photography (Drewniok and Rohr, 1997), in radiography (Yue et al., 2005), and in dactylography (Nilsson and Bigun, 2002). The homologies or correspondences of interest range from translation, rotation, and scale invariance to affine and projective transformations.

The recognition of flowers has been investigated in (Das et al., 1999; Saitoh et al., 2004), while face recognition is a growth industry with entire conferences devoted to it (Zhao et al., 2004). Local matching methods (Pentland et al., 1994; Wiskott et al., 1997) classify the faces by comparing the local statistics of the corresponding facial features.

3. THE CAVIAR MODEL

The visible model consists of a minimal set of perceptually salient landmark points (pixels) that establish a homology between two images. The desired homology is the structural correspondence between imaged objects of the same class, which allows mapping a region (a compact set of pixels) from one image into another. In CAVIAR, unlike in numerical taxonomy and some face classification methods, the landmarks serve only to define the similarity

transformation required for *registering* (juxtaposing) pairs of images rather than as condensed object descriptors. The homology specified by the visible model ensures that the features extracted from images of the same class are commensurable. What distinguishes CAVIAR from previous work is the interactive refinement of the visible model – the landmarks – *according to the results of the classification.*

The model mediates only a restricted set of information. It does not tell the computer anything about the image-based perceptions that lead the operator to correct or approve the model, and it does not reveal to the operator the configuration of the resulting feature vectors in high-dimensional feature space. Rich contextual knowledge and superior noise-filtering abilities render the operator superior in tasks like object-background separation (Palmer, 1999), but the machine can faultlessly compute geometric and histogram moments, posterior probability distributions, and rank orders. It also stores all the reference images, labels, feature vectors and the associations between them. The interaction itself can be modeled by a simple finite-state machine.

The above formulation of the visible model leads to two evaluation criteria: (1) the error rate of interactive classification based on accurately instantiated visible models, and (2) the human time required to refine the automatically generated visible models as necessary.

Figure 1 shows examples of our flower and face models. The interaction is restricted to isolated points: the user can point and drag, but not paint or shade. A line drawing is superimposed on the picture to let the operator judge whether a computer-suggested model fits the unknown object. These models are constructed automatically, and corrected interactively only when necessary.

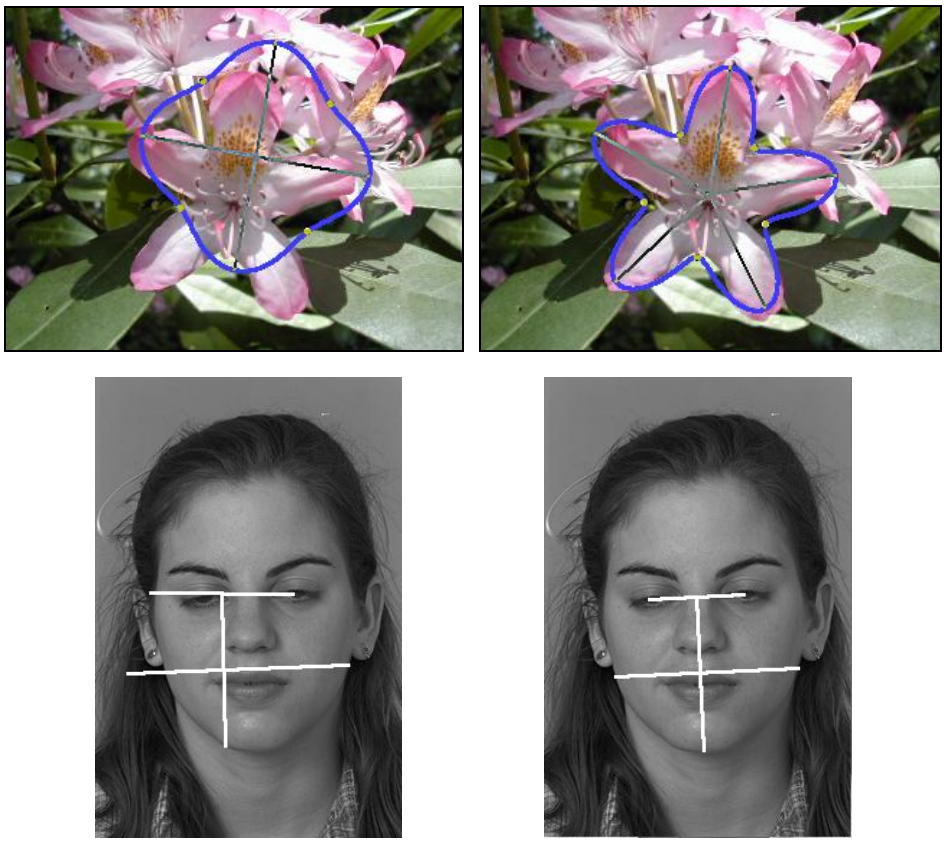

Figure 1. Examples of CAVIAR-Flower (top) and CAVIAR-Face (bottom) models, before and after human adjustment. Here automatic model construction failed because of overlapping flowers and partially closed eyes, respectively.

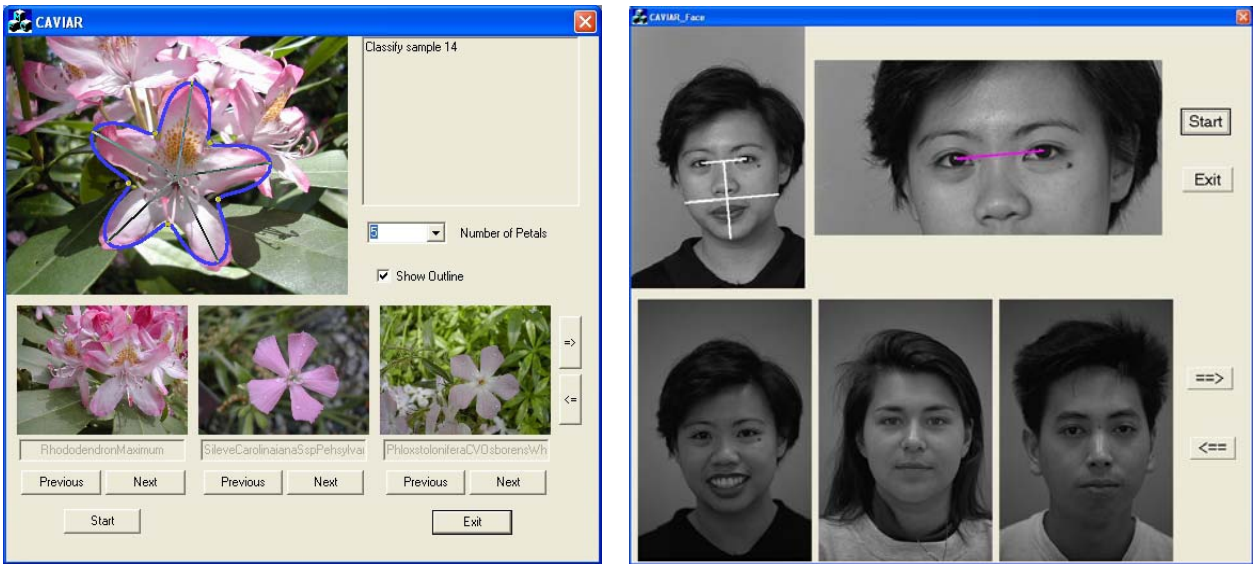

Figure 2. CAVIAR-Flower (left) and CAVIAR-Face (right) graphic user interface. In CAVIAR-Face, because accurate pupil location is important, an enlarged view is provided for adjustment.

A model instance need not depict faithfully intensity, color, or texture edges. An ill-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 visual 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 (Fig. 2). The operator can then either correct the model if none of the top three candidates match, or consult more reference images to find a better match.

4. CAVIAR-FLOWER AND CAVIAR-FACE SYSTEMS

Model building

The visible model of CAVIAR-Flower (Figure 1, top) is a *rose curve* with 6 parameters, which are estimated with prior probabilities learned from a training set (Zou, 2005). The visible model of CAVIAR-Face (Figure 1, bottom) consists of pixels at the centers of the eyes (*pupils*), at the bottom of the chin (*chin*), and below the ears (*jowls*, for lack of a better word). These characteristic points are located by hierarchic template matching. In both CAVIAR-Flower and CAVIAR-Face, interactive model correction requires positioning the cursor to "acquire" some landmark, and then dragging it to a preferred location. The machine then re-estimates the posterior probabilities and the resulting rank order according to the adjusted model.

Feature extraction

In CAVIAR-Flower, the 8 features are the two similarity-invariant parameters of the rose curve, and the first three moments of the hue and saturation histograms of the region enclosed by the curve (Zou, 2004). In CAVIAR-Face, the face is aligned based on the five landmarks, and

then divided into a large number of local regions. The pixel configurations of these local regions serve as features (Zou et al., 2006).

Rank ordering (classification)

In CAVIAR-Flower, the classes are ordered according to the Euclidian distance of the unknown features from the nearest feature vector of each class. In CAVIAR-Face, the local regions from the unknown image are compared against corresponding local regions of every reference face. The classes are then ordered by their total rank, i.e., the *Borda Count* (Ho et al., 1994), computed over all local regions.

In both CAVIAR-Flower and CAVIAR-Face, when the reference pictures of top three candidates are displayed, the operator decides whether to (1) accept one of the displayed classes by clicking on it, or (2) modify the model superimposed on the picture of the unknown object, or (3) inspect lower-ranked candidates ("browse") until a good match is found. The operator need not be able to classify the unknown object, but only to decide whether it matches one of the displayed reference pictures.

The *model* must be based on visible and readily discernible vertices and edges. The *features* must address properties of the objects that differentiate the classes. The choice of *classifier* is dictated by the number of classes, the number of features, the range and distribution of feature values, and the number of available reference samples per class (Nagy, 2004). Only the interactive recognition system architecture that we propose (Figure 3) is general across different domains.

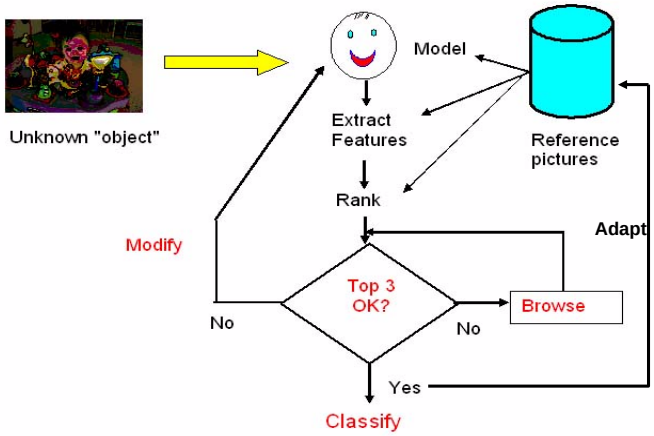

Figure 3. CAVIAR system architecture. Operator interactions shown in red.

5. EVALUATION: CAVIAR-FLOWER

We could not use pictures from any of the many excellent flower sites on the web because none have more than one or two samples per specie, and labeling conventions, background, and resolution differ from site to site. We therefore collected a database of 1078 flowers from 113 species, mostly from the New England Wildflower Garden (http://www.newfs.org/). Our system was developed on a subset of 216 flowers with 29 classes (Nagy and Zou, 2002) and evaluated on a new subset of 612 flowers, consisting of 102 classes with 6 samples per class (Zou and Nagy, 2004). For classification, the photos, taken at the lowest resolution of a Canon Coolpix camera, were further reduced to 320 x 240 pixels.

The flowers were photographed against complex backgrounds (dirt, weeds, and other flowers of the same or other species), under highly variable illumination (sharp shadows on foreground or background, specular reflections, saturation of some of the color channels), and poor imaging conditions (blur, incomplete framing), without necessarily a clear view of the camera viewfinder screen. The color distribution is not uniform, but most of our flowers are yellow, white, red, or blue. Several pictures contain multiple, tiny, overlapping flowers. Our database, including labels and segmentation, is freely available on *http://www.ecse.rpi.edu/doclab/flowers*.

Experimental protocol

We asked 30 naive subjects (male and female adults without any connection to our department) to classify, as rapidly and as accurately as possible, one flower of each of 102 different categories (this took about one hour per subject). The order of the 102 flower images was randomized for each subject. None of these test images was used in training the CAVIAR system used by that subject.

The 30 sessions addressed five different tasks (T1-T5). Each task was replicated by 6 subjects, with images of different instances of the same 102 species. In baseline Task 1, neither the subjects nor the machine made use of a model: the subject just browsed the reference set (which was kept in an arbitrary fixed order) to find an acceptable candidate class. In Task 2, all 5 of the available reference pictures were used to train the system. The remaining tasks (Task 3-5) explored semi-supervised learning based on decision directed approximation (Nagy and Shelton, 1966; Baird and Nagy, 1994; Veeramachaneni and Nagy, 2004) and differed only by letting the users add the samples they classified to the training sample (we call samples with user-assigned labels and models *pseudo-training samples*). Table 1 shows the experimental design, based on 612 distinct flower images.

Table 1. CAVIAR-Flower recognition experiments

Task	**Purpose**	**Classification**	**Training set composition**	
			# of labeled samples	# of pseudo-training samples
T1	Unaided classification	Browsing only	None	None
T2	Interactive classification	Rose curve adjustment + browsing	510	None
T3			102	None
T4			102	204
T5			102	408

Interactive accuracy and time compared to human alone and to machine alone

Table 2 shows the median performance of six subjects for human-alone (T1), machine-alone (T2 initial auto), and CAVIAR (T2). There is no significant difference between CAVIAR and human-alone in accuracy. However, CAVIAR reduced the time spent on each test sample to less than half.

Table 2. CAVIAR compared to human alone and to machine alone

	Time (s)	**Top-1 accuracy (%)**	**Top-3 accuracy (%)**	**Rank Order**
T1 (human alone)	26.4	94	N/A	51.0
T2 initial auto	0	39	55	6.6
T2 (CAVIAR)	10.7	93	N/A	N/A

Machine Learning

Table 3 shows the median values of the classification accuracy and the human time for T2, T3, T4, and T5. The median time spent on each interactive recognition task decreased from 16.4 to 10.7 seconds, which is the same as the median time of T2. The speed-up is due to an increase in the initial machine accuracy of about 10% resulting from the addition of the (not-necessarily correctly) classified flowers to the database.

Table 3. Machine learning.

	# of labeled samples per class	# of pseudo-training samples per class	Human time	Accuracy (%)
T3	1	0	16.4	90
T4	1	2	12.7	95
T5	1	4	10.7	92
T2	5	0	10.7	93

The CAVIAR system can achieve high accuracy even when initialized with only a single training sample per class. Adding pseudo-labeled training samples improved automatic recognition, which in turn helped the subjects to identify the flowers faster.

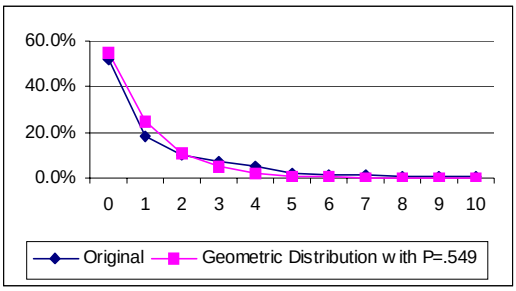

Figure 4. Successive rose curve adjustments

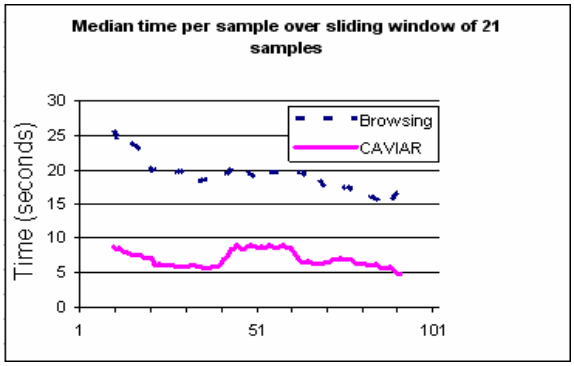

Figure 5. Recognition time as a function of experience with the system.

Human recognition strategy and learning

Figure 4 shows the average percentage of successive rose curve adjustments. A geometric distribution with p=0.55 fits the curve well: the probability of success on each adjustment is just over one half. On average, each sample requires 1.3 adjustments. Figure 5 shows the human time as a function of experience with the system, i.e., the number of samples that have been classified, for T1 and T2. Even without the machine's help, human time decreased from 26 to 17 seconds per flower as the subjects become familiar with the database. With CAVIAR, the time decreased from 9 to 5 seconds.

6. EVALUATION: CAVIAR-FACE

We downloaded the FERET face database from the National Institute of Standards and Technology (NIST) (Phillips et al., 1998; Phillips et al., 2000). Series BK was used for the "gallery" (reference) images. Part of the BA series was reserved for training, which requires pairs (BA and BK) of images of the same individual. Each of six subjects classified 50 randomly selected BA test pictures (different from the training set) against the same gallery of 200 BK pictures (taken on the same day as the test pictures but with a different camera and lighting). The faces vary in size by about 50%, and horizontal and vertical head rotations of up to 15° can be observed. Although the subjects had been asked to keep a neutral expression and look at the

camera, some blinked, smiled, frowned, or moved their head. Because we had only two samples per face, here we could not test decision-directed machine learning.

Earlier experiments showed that human-alone (browsing only) required an average of 66 seconds per photo, and most subjects did not misclassify any photos (Zou, 2004).

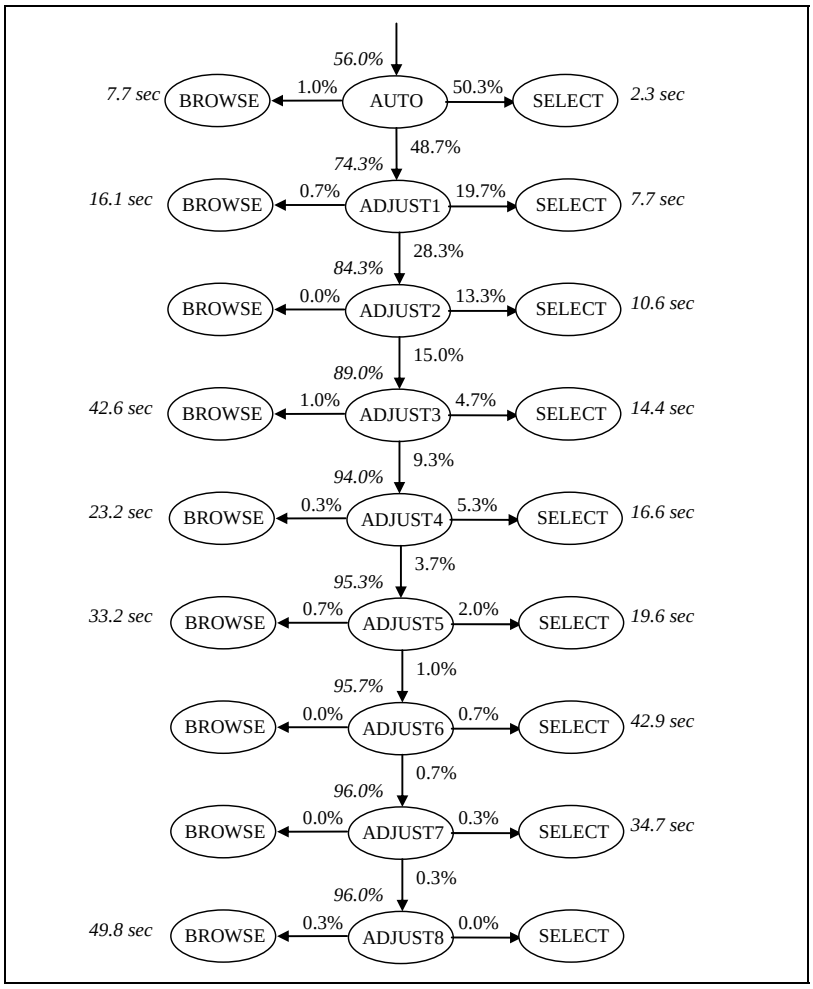

Figure 6. Interactions in CAVIAR-face (6 subjects). SELECT means choosing one of the displayed candidates. BROWSE means looking at more than 3 displayed faces before selecting a winner. Both SELECT and BROWSE terminate the interactive classification. ADJUST prompts automated rank ordering, which results in a new set of candidates for selection or browsing. Listed next to SELECT and BROWSE is the average (over subjects and pictures) human time required for final classification, including any adjustment or browsing. The percentage (in italic) indicated above AUTO or every ADJUST is the machine TOP-3 recognition accuracy.

Figure 6 and Table 5 summarize the experimental results. 50.3% of the photos were classified without adjustment, in 2.3 seconds on average. The accuracy was 99.7%, and the average recognition time, including adjustments and browsing, was 7.6 seconds per photo. Only 15% of the faces required more than two interactions. The top-3 accuracy rose from 56% for automatic recognition to 96% after interactive model modifications.

Table 5. CAVIAR-Face compared to human alone and to machine alone.

	Accuracy	Time per face
Interactive	99.7%	7.6 sec
Machine Alone	48.0%	----
Human Alone	~100.0%	66.3 sec

7. MOBILE CAVIAR

An early version of CAVIAR-Flower was reprogrammed in Java on a camera-equipped Sharp Zaurus Personal Digital Assistant (PDA) at Pace University (Evans et al., 2005). Subsequently it was ported in our laboratory to a camera and Wi-Fi equipped Toshiba e800 PDA dubbed M-CAVIAR (Fig. 7). The PDA forwards, via the wireless network interface, each newly acquired image to a host laptop computer. The host computes the initial visible model and rank order using its stored reference images, and returns the model parameters and index number of the top candidates to the PDA. The PDA then displays the top three candidates from its stored database of thumbnail reference pictures. If the user adjusts the model (using stylus or thumb), the adjusted model parameters are sent to the laptop and a new model and rank order is computed and communicated to the PDA (Gattani, 2004; Zou and Gattani, 2005).

Figure 7. M-CAVIAR graphic user interface.

We repeated on the PDA some of the earlier experiments with six new subjects. With this system, it was also possible to conduct field experiments to recognize flowers in situ. An additional 68 classes of flowers, with 10 samples of each, were collected with the new, lower-

quality PDA camera. Recognition time per flower was over 20% faster than using the desktop, mainly because model adjustment was faster with either stylus or thumb than with a mouse. Recognition accuracy was slightly lower, because some reference flowers could not be easily distinguished on the small PDA display. The networked computation did not impose any significant delay: except for uploading each new flower picture to the laptop, only very short messages (model coordinates and rank orders) are exchanged.

8. SUMMARY

We presented a case for interaction *throughout* the recognition of visual objects, rather than only at the beginning or the end. The human retains the initiative at all times and, as final arbiter of correct matches (as opposed to merely proofreading already classified items), ensures high accuracy. The visible models formulated for flowers and for faces show that such models can mediate human-computer communication. In these applications an interactive system is more accurate than the machine alone and faster than the human alone. Furthermore, it improves with use.

The feature extraction and automated rank ordering can obviously be improved. Any improvement of the automated part of the system will further reduce interaction time. The network protocol of M-CAVIAR will require some changes for camera-phone applications. Careful interface and display design will be required to avoid disorienting the operator, but direct action manipulation will be faster with stylus and thumb than with a mouse.

Portable, wireless CAVIAR systems offer the possibility of Internet-wide reference data collection and collaborative interactive recognition, including some medical and educational applications. They may also prove valuable for constructing very large labeled training sets for

automated algorithms by “growing” training sets with interactive classification under operational conditions.

Acknowledgements

Hamei Jiang collected pictures of fruit, stamps, coins and Han characters for the early CAVIAR experiments. Greenie Cheng and Laura Derby photographed *many* flowers and helped build the database. Borjan Gagoski recruited subjects, conducted the 30 flower recognition experiments, and compiled the results. Rebecca Seth (City Naturalist, Lincoln, NE), Dr. Richard Mitchell (NY State Botanist) and Prof. Robert Ingalls (RPI CS Dept) gave us valuable advice about the classification of plants and flowers. Arthur Evans, John Sikorski, and Patricia Thomas, under the supervision of Professors Sung-Hyuk Cha and Charles Tappert at Pace University, ported CAVIAR to the Zaurus. Abhishek Gattani developed and tested M-CAVIAR as part of his MS thesis at Rensselaer. Prof. Qiang Ji (ECSE Dept) suggested that we apply CAVIAR to face recognition. We are indebted to the many enthusiastic volunteers who helped us test the various incarnations of CAVIAR under an agreement that their names would not be revealed. Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

References

Baird, H.S., Nagy, G., 1994. A Self-Correcting 100-font Classifier. Proc. SPIE Conference on Document Recognition. SPIE-2181, 106-115.

Bradford, R., 1991. Technical Factors in the Creation of Large Full-Text Databases. Proc. DOE Infotech Conference.

Carson, C., Belongie, S., Greenspan, H., Malik, J., 2002. Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying. IEEE Trans. Pattern Analysis and Machine Intelligence. 24(8), 1026-1038.

Chen, Y., Bart, H.L., Jr., Huang, S., Chen, H., 2005. A Computational Framework for Taxonomic Research: Diagnosing Body Shape within Fish Species Complexes. Proc. of the Fifth IEEE International Conference on Data Mining. 593-596.

Cohn, D.A., Ghahramani, Z., Jordan, M.I., 1996. Active Learning with Statistical Models. Journal of Artificial Intelligence Research. 4, 129-145.

Cox, I.J., Miller, M.L., Minka, T.P., Papathomas, T.V., Yianilos, P.N., 2000. The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments. IEEE Trans. Image Processing. 9(1), 20-37.

Das, M., Manmatha, R., Riseman, E.M., 1999. Indexing Flower Patent Images Using Domain Knowledge. IEEE Intelligent Systems. 14(5), 24-33.

Dickey, L.A., 1991. Operational Factors in the Creation of Large Full-Text Databases. Proc. DOE Infotech Conference.

Drewniok, C., Rohr, K., 1997. Model-Based Detection and Localization of Circular Landmarks in Aerial Images. International Journal of Computer Vision. 24(3), 187-217.

Evans, A., Sikorski, J., Thomas, P., Cha, S.-H., Tappert, C., Zou, J., Gattani, A., Nagy, G., 2005. Computer Assisted Visual Interactive Recognition (CAVIAR) Technology. IEEE Int. Conf. Electro Information Technology. 22-25.

Gattani, A., 2004. Mobile Interactive Visual Pattern Recognition, MS thesis, Rensselaer Polytechnic Institute.

Haritaoglu, I., 1994. Scene Text Extraction and Translation for Handheld Devices. IEEE conf. on Computer Vision and Pattern Recognition. 2, 408-413.

Harman, D., 1992. Relevance Feedback and Other Query Modification Techniques. Information Retrieval: Data Structures and Algorithms, W.B. Frakes and R. Baeza-Yates, eds., 241-263.

Ho, T.K., Hull, J.J., Srihari, S.N., 1994. Decision Combination in Multiple Classifier Systems. IEEE Trans. Pattern Analysis and Machine Intelligence.16(1), 66-75.

Jain, R., 1992. US NSF Workshop on Visual Information Management Systems.

Kak, A.C., Desouza, G.N., 2002, Robotic Vision: What Happened to the Visions of Yesterday. Proc. Int. Conf. Pattern Recognition. 2, 839-847.

Klein, B., Dengel, A.R., 2004. Problem-Adaptable Document Analysis and Understanding for High-Volume Applications. International Journal of Document Analysis and Recognition. 6, 167-180.

MacKay, D.J.C., 1992. Information-Based Objective Functions for Active Data Selection. Neural Computation. 4(4), 590-604.

Mericsko, R. J., 1998. Introduction of 27th AIPR Workshop - Advances in Computer-Assisted Recognition. Proc. of SPIE. 3584.

Nagy, G., Shelton, G.L., 1966. Self-Corrective Character Recognition System. IEEE Transactions on Information Theory. 12(2), 215-222.

Nagy, G., Zou, J., 2002. Interactive Visual Pattern Recognition. Proc. International Conference on Pattern Recognition. 2, 478-481.

Nagy, G., 2004. Visual Pattern Recognition in the Years Ahead. Proc. International Conference on Pattern Recognition. 4, 7-10.

Nilsson, K., Bigun, J., 2002. Prominent Symmetry Points as Landmarks in Finger Print Images for Alignment. 16th International Conference on Pattern Recognition. 3, 395-398.

Palmer, S.E., 1999. Vision Science, Photons to Phenomenology, MIT Press.

Pentland, A., Moghaddam, B., Starner, T., 1994. View-Based and Modular Eigenspaces for Face Recognition. Proc. IEEE Conference on Computer Vision and Pattern Recognition, 84-91.

Phillips, P.J., Wechsler, H., Huang, J., Rauss, P., 1998. The FERET Database and Evaluation Procedure for Face Recognition Algorithms. Image and Vision Computing J. 16(5), 295-306.

Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J., 2000. The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Trans. Pattern Analysis and Machine Intelligence. 22, 1090-1104.

Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S., Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. IEEE Trans. Circuits and Systems for Video Technology. 8(5), 644-655.

Saitoh, T., Aoki, K., Kaneko, T., 2004. Automatic Recognition of Blooming Flowers. International Conference on Pattern Recognition, 27-30.

Sarkar, P., Baird, H.S., Henderson, J., 2002. Triage of OCR Output Using ‘Confidence’ Scores. Proceedings of the SPIE/IS&T 2003 Document Recognition and Retrieval IX Conf., 20-25.

Veeramachaneni, S., Nagy, G., 2004. Adaptive Classifiers for Multi-Source OCR. International Journal of Document Analysis and Recognition. 6(3), 154-166.

Wiskott, L., Fellous, J.-M., Kruger, N., von der Malsburg, C., 1997. Face Recognition by Elastic Bunch Graph Matching. IEEE Trans. Pattern Analysis and Machine Intelligence. 17(7), 775-779.

Yang, J., Chen, X., Kunz, W., 2002. A PDA-based Face Recognition System. Proc. the 6th IEEE Workshop on Applications of Computer Vision. 19-23.

Yue, W., Yin, D., Li, C., Wang, G., Xu, T., 2005. Locating Large-Scale Craniofacial Feature Points on X-ray Images for Automated Cephalometric Analysis. IEEE International Conference on Image Processing. 2, 1246–1249.

Zhang, J., Chen, X., Yang, J., Waibel, A., 2002. A PDA-based Sign Translator. Proc. the 4th IEEE Int. Conf. on Multimodal Interfaces. 217-222.

Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P., 2004. Face Recognition: A Literature Survey. ACM Computing Surveys. 35(4).

Zou, J., 2004. Computer Assisted Visual InterActive Recognition, Ph.D. thesis, ECSE Department, Rensselaer Polytechnic Institute.

Zou, J., Nagy, G., 2004. Evaluation of Model-Based Interactive Flower Recognition. International Conference on Pattern Recognition. 2, 311-314.

Zou, J., 2005. A Model-Based Interactive Image Segmentation Procedure. IEEE Workshop on Applications of Computer Vision. 1, 522-527.

Zou, J., Gattani, A., 2005. Computer Assisted Visual InterActive Recognition and Its Prospects of Implementation Over the Internet. IS&T/SPIE 17th Annual Symposium Electronic Imaging, Internet Imaging VI.

Zou, J., Nagy, G., 2006. Human-Computer Interaction for Complex Pattern Recognition Problems. Book chapter in Data Complexity in Pattern Recognition, Springer-Verlag London, Editor: M. Basu and T.K. Ho, 271-286.

Zou, J., Ji, Q., Nagy, G., 2006. A Comparative Study of Local Matching Approach for Face Recognition. Submitted to IEEE Trans. Image Processing, 2006.