

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



Volume 28, issue 16

1 December 2007

ISSN 0167-8655

# Pattern Recognition Letters

An official publication of the  
International Association for Pattern Recognition



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Pattern Recognition Letters 28 (2007) 2335–2342

---



---

Pattern Recognition  
Letters

---



---

[www.elsevier.com/locate/patrec](http://www.elsevier.com/locate/patrec)

# Visible models for interactive pattern recognition

Jie Zou <sup>a,1</sup>, George Nagy <sup>b,\*</sup>

<sup>a</sup> National Library of Medicine, Bldg. 38A, Rm. 10S1000, 8600 Rockville Pike, Bethesda, MD 20894, USA

<sup>b</sup> Department of Electrical, Computer and Systems Engineering, DocLab, Rensselaer Polytechnic Institute, JEC 6020, 110 8th street, Troy, NY 12180, USA

Received 15 March 2007; received in revised form 8 June 2007

Available online 19 August 2007

Communicated by M.-J. Li

---

## 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.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Pattern recognition; Human–computer interaction; Visible model; Face recognition; Flower recognition

---

## 1. Introduction

We address the problem of many-class object recognition in a narrow domain, such as flowers or faces, where samples of different classes may bear a strong resemblance to each other. We adopt a parametric model-based approach for extracting classification features through weak segmentation. Failures of the automated model-fitting are corrected by operator intervention.

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 tech-

niques” (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).

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 vision 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.

## 2. Prior work

Interaction in image acquisition and in reject handling is well documented. In many OCR applications, every

---

\* Corresponding author. Tel.: +1 518 276 6078; fax: +1 518 276 6261.  
E-mail addresses: [jzou@mail.nlm.nih.gov](mailto:jzou@mail.nlm.nih.gov) (J. Zou), [nagy@ecse.rpi.edu](mailto:nagy@ecse.rpi.edu) (G. Nagy).

<sup>1</sup> He was with Rensselaer Polytechnic Institute, Troy, NY 12180 USA.

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).

We use landmarks for image registration. Homologous landmarks are 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. Chen et al. give examples of the use of landmarks for classifying species of the fish genus *Carpoides* (Chen et al., 2005).

In the broad domains of content-based image retrieval (CBIR), *relevance feedback* has been found effective (Rui et al., 1998). Interaction has been 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 designers of “Blobworld” (Carson et al., 2002) suggested that the CBIR system should display its representation of the submitted and returned images. In the next section we comment on how CAVIAR differs from CBIR.

*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, unlike CAVIAR, without interacting with images of the unknown object (MacKay, 1992; Cohn et al., 1996).

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 similar to ours (Pentland et al., 1994; Wiskott et al., 1997) classify 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.

Fig. 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.

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.

We did not compare relevance feedback with automated, interactive, and unaided classification because CAVIAR was designed for objective classification rather than subjective similarity retrieval. In our narrow domains of application, the top candidates retrieved algorithmically are already visually similar, like the top two in Fig. 2 left (the difference is the red ring at the flower center), so little gain can be expected from relevance feedback. Our experiments demonstrate that accurate segmentation is not essential for correct classification (Zou, 2004). Catastrophic mis-segmentation that results in arbitrary ranking (as in Fig. 1, top left) cannot be corrected by relevance feedback because CAVIAR provides only a single segmentation after any interaction.

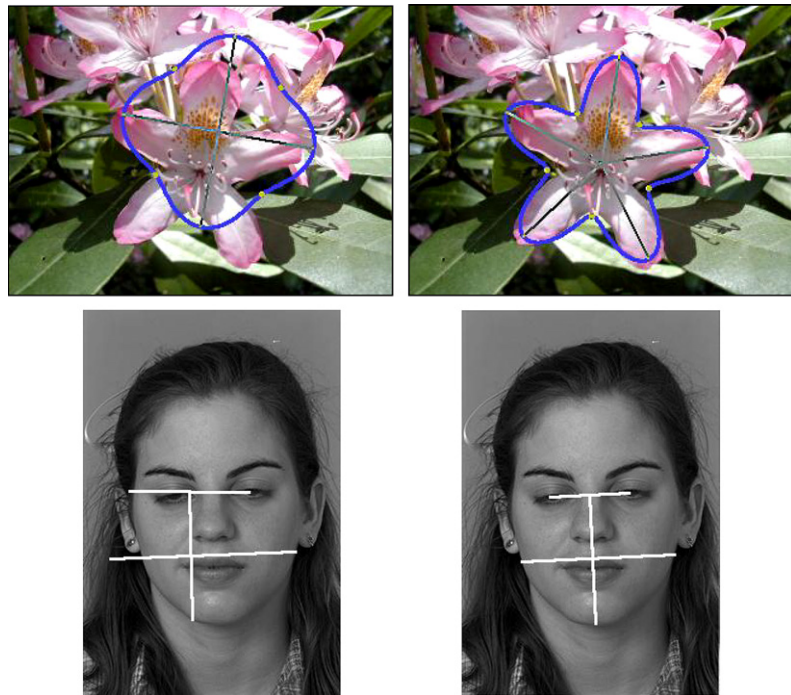


Fig. 1. Examples of CAVIAR-Flower (top) and CAVIAR-Face (bottom) models, before and after human adjustment. Here automatic model construction failed because of overlapping flowers and partially closed eyes, respectively.

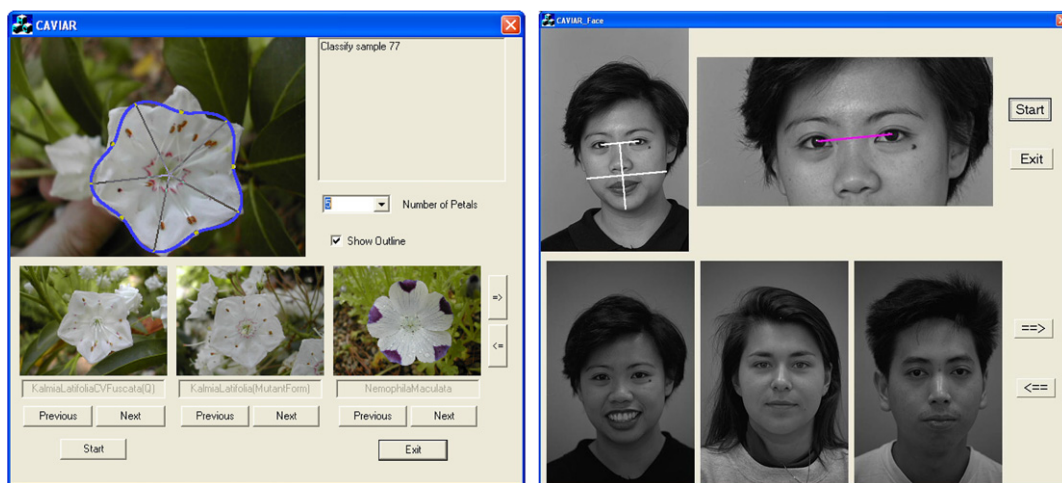


Fig. 2. CAVIAR-Flower (left) and CAVIAR-Face (right) graphic user interface. In CAVIAR-Flower, the top two candidates are often quite similar perceptually. It is difficult to imagine how relevance feedback, based on a prior display of other, quite dissimilar species, would promote retrieval of the correct candidate and rejection of the incorrect candidate. In CAVIAR-Face, because accurate pupil location is important, an enlarged view is provided for adjustment.

#### 4. CAVIAR-Flower and CAVIAR-Face systems

##### 4.1. Model building

The visible model of CAVIAR-Flower is a *rose curve* with six parameters (Fig. 1, top), which are estimated with prior probabilities learned from a training set (Zou, 2005). The visible model of CAVIAR-Face (Fig. 1, bottom) consists of pixels at the centers of the eyes (*pupils*), at the bottom of the chin (*chin*), and the points defined by the

junction of the bottom of the ear lobes with the contour of the face. 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.

We also conducted experiments with Intelligent Scissors (Mortensen and Barrett, 1998), and with multiple

propagating boundaries from multiple seeds. Although we did not conduct detailed evaluations, we are convinced that model building with guided general-purpose tools like the above is slower than interacting directly with the model. We did not see a clean way to combine the rose-curve model for flowers with guided segmentation based on other heuristics. Regardless of the specific segmentation scheme, we believe that the novelty here is interleaving human-aided segmentation with machine classification until the segmentation is just good enough (but no better).

#### 4.2. Feature extraction

In CAVIAR-Flower, the eight 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., in press).

#### 4.3. 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 the 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.

It is essential to show the candidates on the same screen as the unknown pattern, and the latter must be large enough to allow interaction. We display only the Top 3 candidates, because any further reduction in size (especially on a PDA or cell phone display) would preclude visual distinction between very similar flowers, such as those in Fig. 2 left. Furthermore, even if the human could select the correct candidate from the Top 10 candidates, it would result in little increase in average throughput because gross mis-segmentation usually demotes the correct candidate below the tenth rank.

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,

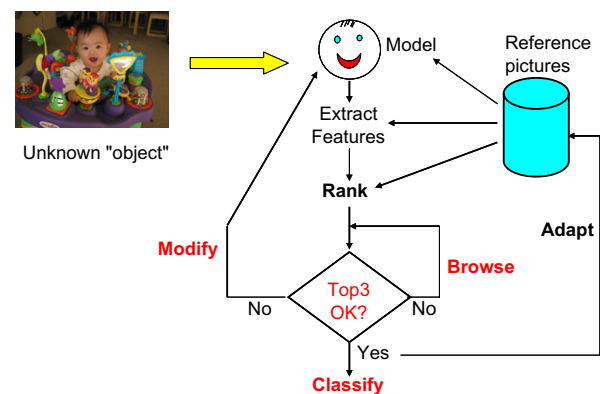


Fig. 3. CAVIAR system architecture. Operator interactions shown in red.

and the number of available reference samples per class (Nagy, 2004). Only the interactive recognition system architecture that we propose (Fig. 3) is general across different domains.

## 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  $320 \times 240$  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>.

### 5.1. 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), where classified samples are automatically added to the training sample. (We call samples with user-assigned labels and models *pseudo-training samples*.) The initial segmentation parameters are recalculated after each pattern is added, using the current segmentation as though it were ground truth. Adding the newly classified samples also provides a larger and more representative reference set for improved rank ordering. Table 1 shows the experimental design, based on 612 distinct flower images.

5.2. 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.

5.3. Machine learning

Table 3 shows the median values of the classification accuracy and the human time for T2, T3, T4, and T5.

Table 1  
CAVIAR-Flower recognition experiments

Task	Purpose	Classification	Training set composition	
			No. of labeled samples	No. 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

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

Table 3  
Machine learning

	No. of labeled samples per class	No. 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 median time spent on each interactive recognition task decreased from 16.4 to 10.7 s, 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.

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.

5.4. Human recognition strategy and learning

Fig. 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. Fig. 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

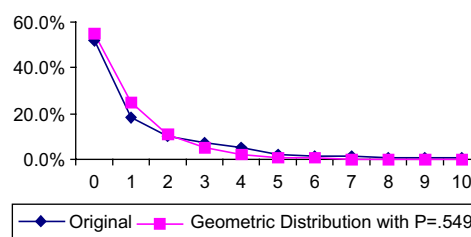


Fig. 4. Successive rose curve adjustments.

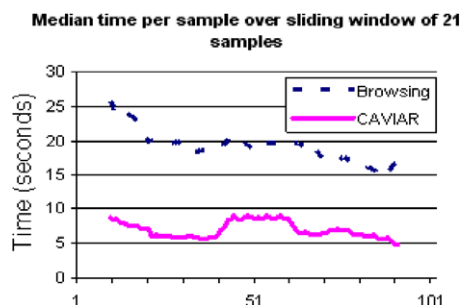


Fig. 5. Recognition time as a function of experience with the system.

26 to 17 s per flower as the subjects become familiar with the database. With CAVIAR, the time decreased from 9 to 5 s.

### 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. We could not test decision-directed machine learning here because one of the two samples of each face available in the database was used for the initial training, and the second was used for testing. Therefore there were no independent samples left for evaluating any improvement due to decision-directed learning.

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

Fig. 6 and Table 4 summarize the experimental results. 50.3% of the photos were classified without adjustment, in 2.3 s on average. The accuracy was 99.7%, and the average recognition time, including adjustments and browsing, was 7.6 s 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.

Early presentations of our results on wildflower classification often raised the question whether the CAVIAR method was restricted to this domain. We were convinced that it was not by the results of small experiments on fruit, vegetable, and Han character images that we had conducted before collecting the flower database. When we became involved in automated face recognition after completing the wildflower project, it was appealing to test interactive face classification.

Regardless of the size of the gallery, interactive face classification is much more accurate than wildflower classification because subjects have a lifetime of face recognition experience. Therefore any statistically valid comparison of the small error rates of interactive and human-only classification would require very lengthy experiments. The data presented here is intended to show only that there exist applications very different from wildflowers where CAVIAR may prove appropriate. It is also clear that higher initial (i.e., automatic) accuracy, using better images, better domain-specific features, and better classifiers, can only

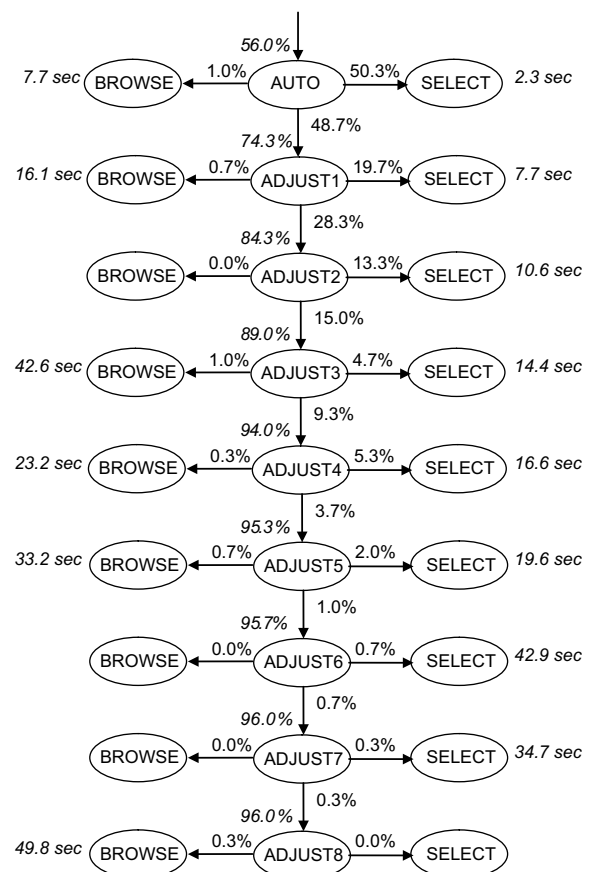


Fig. 6. Interactions in CAVIAR-Face (6 subjects). SELECT means choosing one of the displayed candidates. BROWSE means looking at more than 3 displayed faces before selecting a winner. Both SELECT and BROWSE terminate the interactive classification. ADJUST prompts automated rank ordering, which results 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.

Table 4  
CAVIAR-Face compared to human alone and to machine alone

	Accuracy (%)	Time per face (s)
Interactive	99.7	7.6
Machine alone	48.0	–
Human alone	~100.0	66.3

boost the speed and accuracy of CAVIAR. While it is possible that algorithms will be developed eventually to determine more accurately than by human inspection whether two photographs represent the same person, or the same species of flower, that is far from the case at present.

### 7. Mobile CAVIAR

An early version of CAVIAR-Flower was reprogrammed in Java on a camera-equipped Sharp Zaurus Per-



Fig. 7. M-CAVIAR graphic user interface.

sonal 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).

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.

Our visual models and graphic interfaces benefited from the advice of three botanists (including the author of a CD-ROM guide for the classification of ferns), a graphic artist who teaches portrait painting and digital photography, and a perceptual psychologist. In collaboration with biomedical researchers we are attempting to extend CAVIAR to skin lesions (e.g., insect bites, poison ivy) photographed by the afflicted person, and to an existing NIH database of cervix images.

## Acknowledgement

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. In: Proc. SPIE Conf. on Document Recognition. SPIE-2181, pp. 106–115.
- Bradford, R., 1991. Technical factors in the creation of large full-text databases. In: Proc. DOE Infotech Conf.
- Carson, C., Belongie, S., Greenspan, H., Malik, J., 2002. Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pattern Anal. Machine Intell.* 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. In: Proc. 5th IEEE Internat. Conf. on Data Mining, pp. 593–596.
- Cohn, D.A., Ghahramani, Z., Jordan, M.I., 1996. Active learning with statistical models. *J. Artif. Intell. Res.* 4, 129–145.
- Cox, I.J., Miller, M.L., Minka, T.P., Papatomas, T.V., Yianilos, P.N., 2000. The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. *IEEE Trans. Image Process.* 9 (1), 20–37.
- Das, M., Manmatha, R., Riseman, E.M., 1999. Indexing flower patent images using domain knowledge. *IEEE Intell. Syst.* 14 (5), 24–33.
- Dickey, L.A., 1991. Operational factors in the creation of large full-text databases. In: Proc. DOE Infotech Conf.
- Drewniok, C., Rohr, K., 1997. Model-based detection and localization of circular landmarks in aerial images. *Internat. J. Comput. 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. In: IEEE Internat. Conf. Electro Information Technology, pp. 22–25.
- Gattani, A., 2004. Mobile Interactive Visual Pattern Recognition, MS thesis, Rensselaer Polytechnic Institute.
- Haritaoglu, I., 2001. 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. In: Frakes, W.B., Baeza-Yates, R. (Eds.), *Information Retrieval: Data Structures and Algorithms*. Prentice Hall, Inc., Upper Saddle River, NJ USA, pp. 241–263.
- Ho, T.K., Hull, J.J., Srihari, S.N., 1994. Decision combination in multiple classifier systems. *IEEE Trans. Pattern Anal. Machine Intell.* 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. In: Proc. Internat. Conf. on Pattern Recognition, vol. 2, pp. 839–847.
- Klein, B., Dengel, A.R., 2004. Problem-adaptable document analysis and understanding for high-volume applications. *Internat. J. Document Anal. Recognition* 6, 167–180.
- MacKay, D.J.C., 1992. Information-based objective functions for active data selection. *Neural Comput.* 4 (4), 590–604.
- Mericsko, R.J., 1998. Introduction of 27th AIPR workshop – advances in computer-assisted recognition. In: Proc. SPIE. 3584.
- Mortensen, E.N., Barrett, W.A., 1998. Interactive segmentation with intelligent scissors. *Graph. Models Image Process.* 60 (5), 349–384.
- Nagy, G., 2004. Visual pattern recognition in the years ahead. In: Proc. Internat. Conf. on Pattern Recognition, vol. 4, pp. 7–10.
- Nagy, G., Shelton, G.L., 1966. Self-corrective character recognition system. *IEEE Trans. Inform. Theory* 12 (2), 215–222.
- Nagy, G., Zou, J., 2002. Interactive visual pattern recognition. In: Proc. Internat. Conf. on Pattern Recognition, vol. 2, pp. 478–481.
- Nilsson, K., Bigun, J., 2002. Prominent symmetry points as landmarks in finger print images for alignment. In: 16th Internat. Conf. on Pattern Recognition, vol. 3, pp. 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. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 84–91.
- Phillips, P.J., Wechsler, H., Huang, J., Rauss, P., 1998. The FERET database and evaluation procedure for face recognition algorithms. *Image Vision Comput. 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 Anal. Machine Intell.* 22, 1090–1104.
- Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S., 1998. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans. Circuits Syst. Video Technol.* 8 (5), 644–655.
- Saitoh, T., Aoki, K., Kaneko, T., 2004. Automatic recognition of blooming flowers. In: Internat. Conf. on Pattern Recognition, pp. 27–30.
- Sarkar, P., Baird, H.S., Henderson, J., 2002. Triage of OCR output using ‘Confidence’ scores. In: Proc. of the SPIE/IS&T 2003 Document Recognition and Retrieval IX Conf., pp. 20–25.
- Veeramachaneni, S., Nagy, G., 2004. Adaptive classifiers for multi-source OCR. *Internat. J. Document Anal. 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 Anal. Machine Intell.* 17 (7), 775–779.
- Yang, J., Chen, X., Kunz, W., 2002. A PDA-based face recognition system. In: Proc. 6th IEEE Workshop on Applications of Computer Vision, pp. 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 Int. Conf. on Image Process.* 2, 1246–1249.
- Zhang, J., Chen, X., Yang, J., Waibel, A., 2002. A PDA-based Sign Translator. In: Proc. 4th IEEE Int. Conf. on Multimodal Interfaces, pp. 217–222.
- Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P., 2004. Face recognition: a literature survey. *ACM Comput. Surveys* 35 (4).
- Zou, J., 2004. Computer Assisted Visual InterActive Recognition, Ph.D. thesis, ECSE Department, Rensselaer Polytechnic Institute.
- Zou, J., 2005. A model-based interactive image segmentation procedure. In: IEEE Workshop on Applications of Computer Vision, vol. 1, pp. 522–527.
- Zou, J., Gattani, A., 2005. Computer assisted visual interactive recognition and its prospects of implementation over the internet. In: IS& T/ SPIE 17th Symposium Electronic Imaging, Internet Imaging VI.
- Zou, J., Nagy, G., 2004. Evaluation of model-based interactive flower recognition. In: Internat. Conf. on Pattern Recognition, vol. 2, pp. 311–314.
- Zou, J., Nagy, G., 2006. Human–computer interaction for complex pattern recognition problems. In: Basu, M., Ho, T.K. (Eds.), *Data Complexity in Pattern Recognition*. Springer-Verlag, London, pp. 271–286.
- Zou, J., Ji, Q., Nagy, G., in press. A comparative study of local matching approach for face recognition. *IEEE Trans. Image Process.*