Evaluation of Model-Based Interactive Flower Recognition

Jie Zou and George Nagy
Rensselaer Polytechnic Institute, DocLab, ECSE department
110 8th ST, Troy, New York, 12180
zouj@rpi.edu and nagy@ecse.rpi.edu

Abstract

We introduce the concept of Computer Assisted Visual InterActive Recognition (CAVIAR). In CAVIAR, a parameterized geometrical model serves as the human-computer communication channel. We implemented a flower recognition system and evaluated it on 30 inexperienced subjects. Major conclusions include: 1) the accuracy of the CAVIAR system is much higher than that of the machine alone; 2) its recognition time is much lower than that of the human alone; 3) it can be initialized with as few as one training sample per class and still achieve high accuracy; 4) it demonstrates a self-learning ability, which suggests that instead of initializing the CAVIAR system with many training samples, we can trust the system's self-learning ability.

1. Introduction

There is a growing consensus among experts to the effect that it will be a long winter before automated classifiers can yield acceptable accuracy on some pattern recognition and image retrieval problems [3][4][5]. Nevertheless, research on interactive methods is still quite limited even in applications where there is enough time for limited interaction. The human role is usually confined to either preprocessing or postprocessing. The former is exemplified by recent work on marking the pupil-to-pupil baseline for face recognition [8], or finding text bounding-boxes for mobile sign recognition [2][9]. An example of postprocessing is relevance feedback in content-based image retrieval [7]. Here, however, interaction is limited to acceptable and not-acceptable responses, because there is no effective way to interact with arbitrary images.

In contrast, we have been investigating an approach where the action passes back and forth between the human and the machine, taking advantage of the relative strengths of both. Specifically, we exploit superior human gestalt perception for segmentation and for the judgment of the significance of pairwise differences, and superior computer storage and computation for extracting statistical features and for locating objects in a high-dimensional feature space. We find that in the narrow domain of flower recognition, our combined system outperforms both the unaided human and the unaided machine. In addition, we believe that our research points to effective methods for model-based interaction with images in other domains like faces and skin diseases, and in applications that require a mobile, hand-held interface.

In our interactive flower recognition system, a domain-specific model plays the central role in facilitating the communication (interaction) between human and computer. The key to efficient interaction is the display of the automatically fitted model that allows the human to retain the initiative throughout the classification process. We believe that such interaction must be based on a *visible model*, because 1) a high-dimensional feature space is incomprehensible to the human, and 2) not being familiar with the properties of the various classes, the human cannot judge the adequacy of the current decision boundaries, and therefore cannot interact efficiently with the feature-based classifier itself.

In the following sections, we briefly describe our flower recognition system, and report the results and conclusions of our evaluation of Computer Assisted Visual InterActive Recognition.

2. System description

We collected a database of 1078 flowers from 113 species with a digital camera. The flower recognition system discussed here was developed on a subset of 216 flowers with 29 classes [6] and evaluated on a subset of 102 classes with 6 samples per class. All pictures are 320 by 240 pixels. The pictures were taken under highly variable illumination. The majority of the flowers are yellow, white, red, or blue. Some of the flower pictures are quite out of focus, and several pictures contain multiple, tiny, overlapping flowers. The background is the real scene, which may include complex vegetation and sharp shadows. Humans have a remarkable ability to recognize such flowers, but it

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

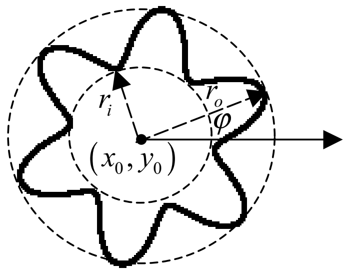


Figure 1. An example of the rose curve.

is difficult to conceive that a machine can recognize them reliably.

The *rhodonea* (*rose curve*) was defined by the Italian mathematician Guido Grandi between 1723 and 1728. We use a slightly modified rose curve to model the flowers:

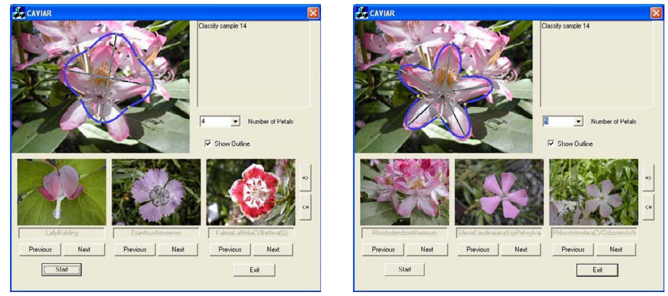
$$\rho = \frac{r_o + r_i}{2} + \frac{r_o - r_i}{2} cos(n\theta + n\varphi) \qquad (1)$$

A particular rose curve (Figure 1) is completely determined by 6 parameters: the center (x_0, y_0), the outer radius r_o, the inner radius r_i, the number of petals n, and the phase φ. We restrict the possible number of petals n to the range $[3, 8]$, and use a circle ($n = 0$) for the rest.

Eight features are derived from the rose curve model for classification. The petal number n and the ratio $\eta = r_o/r_i$ are two global shape features. The first three moments of the hue and saturation histograms of the pixels within the rose curve are the six color features.

At least one training sample per class is required to initialize the flower recognition system. The training process, described in [10], consists of the following steps: 1) the training pictures are interactively segmented; 2) rose curves are automatically fitted to the silhouettes; 3) eight classification features are extracted from each training picture based on the fitted rose curve.

After the training process is completed, each test flower is recognized as follows. The computer estimates an initial rose curve, then extracts features and classifies the test sample. It superimposes the estimated rose curve on the unknown flower picture, and displays the pictures of the top 3 candidates. The user then chooses to 1) click on one of the candidates, thereby assigning the label of the clicked picture to the unknown picture; 2) browse other candidates; or 3) adjust the rose curve. All six rose curve parameters can be adjusted with mouse operations. If the rose curve is adjusted, the computer accepts the user-adjusted parameters, re-estimates the remaining parameters, displays the new rose curve, extracts features, re-rank the classes, and displays the new top 3 candidates. This interactive process continues until the user concludes the recognition task. Figure 2 shows a difficult example with overlapping flowers of the same species in the background.



(a) (b)

Figure 2. The rose curve, superimposed on the unknown picture, plays a central role in the communication (interaction) between human and computer. (a) The initial automatic rose curve is poor, therefore so is the classification; (b) after some rose curve parameters have been adjusted, the computer displays the correct candidate.

3. Evaluation

There are two critical aspects of the performance: time and accuracy. The machine time depends on the hardware configuration and on the degree of software optimization. Since it is always much less than the human time, we assess only the human time, but report both human and machine accuracy.

3.1. Experiment protocol

We asked 30 naïve 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. Each subject first viewed an instructional PowerPoint presentation, and then classified the flower images. The order of the 102 flower images was randomized for each subject to avoid confusing class-specific effects with human learning. None of these test images was used in training the CAVIAR system used by that subject. Each session, consisting of instruction and classification, lasted about one hour. The subjects were compensated only by a framed and personalized color enlargement of a flower photo of their choice.

The 30 sessions addressed five different tasks (T1-T5). To reduce the variance of the observed variables, each task was replicated by 6 subjects, with images of different instances of the same 102 species. The tasks differed in

Table 1. Experiments

Task	Purpose	Classification	Training set composition	
			# of labeled samples	# of pseudo-labeled samples
T1	Unaided classification	Browsing only	None	None
T2	Interactive classification	Rose curve adjustment + browsing	510	None
T3	Same as T2	Same as T2	102	None
T4	Same as T2	Same as T2	102	204
T5	Same as T2	Same as T2	102	406

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 before labeling	8.5	52	79	4.0
T2 (CAVIAR)	10.7	93	N/A	N/A

whether the subjects would outline the flower for computer-aided classification, and in the size and composition of the training sets used for the algorithmic components of CAVIAR. Table 1 shows the experimental design, based on 612 distinct flower images.

We were interested primarily in: 1) comparing computer-aided classification (T2) with human-alone classification (T1) and machine-alone classification (initial automatic recognition of T2); 2) comparing supervised training of the automated components (T2) with semi-supervised training with an increasing number of samples labeled by the subjects (pseudo-labeled samples) in the course of the classification (T3-T5); and 3) human recognition strategy and learning as a function of experience with the system (throughout all tasks).

The times and locations of the mouse clicks of the subjects, and the responses of the system, were logged. This record was transferred to pre-formatted Excel worksheets after completion of the experiments, and subsequently aggregated and tabulated as reported below.

3.2. CAVIAR compared to human-alone and machine-alone

Table 2 shows the median performance of six subjects for human-alone (T1), machine-alone (T2 initial auto), and CAVIAR (T2). We observe that: 1) there is no obvious difference between CAVIAR and human-alone in accuracy. However, with the machine's help, the median time spent on each test sample is reduced to less than half of the human-alone time; 2) the accuracy of the machine alone is low (39%). With a little human help (10 seconds per flower), the median accuracy increases to 93%; 3) after some rose curve adjustments, the initial automatic Top-3 accuracy increases from 55% to 79%.

From the above observations, we conclude that: 1) combining human and machine can significantly reduce the recognition time compared to the unaided human, and significantly increase the accuracy compared to the unaided machine; 2) the visible rose curve model mediates human-computer communication effectively.

Table 3. Machine learning

	Initial Top-1 accuracy (%)	Initial Top-3 accuracy (%)	Initial rank order	Human time	Final Accuracy (%)
T3	27	44	12.7	16.4	90
T4	32	48	10.6	12.7	95
T5	37	55	8.6	10.7	92
T2	39	55	6.6	10.7	93

3.3. Machine learning

Table 3 shows the median values of the Top-1 accuracy, the Top-3 accuracy, the rank order after the initial automatic recognition, and the human time and the accuracy of the complete interactive recognition for T2, T3, T4, and T5. We observe that 1) the median Top-1 accuracy of the initial automatic recognition increases from T3 (27%) to T5 (37%), and approaches the median accuracy of T2 (39%); 2) the median Top-3 accuracy of the initial automatic recognition increases from T3 (44%) to T5 (55%), which is the same as T2 (55%); 3) the median rank order after the initial automatic recognition decreases from T3 (12.7) to T5 (8.6), approaching the median rank order of T2 (6.6); 4) the median final accuracy of T3, which has only one training sample, is still very high (90%); 5) there is not much difference in accuracy among these four tasks: the median accuracies are all above 90%; 6) the median time spent on each interactive recognition task decreases from T3 (16.4 seconds) to T5 (10.7 seconds), which is the same as the median time of T2.

From the above observations, we conclude that: 1) the CAVIAR system can be initialized with a single training sample per class, but still achieve high accuracy; 2) the CAVIAR system shows self-learning: adding pseudo-labeled training samples improves automatic recognition, which in turn helps the subjects to identify the flowers faster; 3) both automatic performance (initial rank order) and interactive performance (human time) for T5 are near the corresponding values of T2, which is expected to achieve the best performance. This suggests that instead of initializing the CAVIAR system with many training samples, we can trust the system's self-learning ability (although, of course, the first users would need more time).

adjustments; 3) subjects remember the flowers to become "connoisseurs" of the flower database. With CAVIAR, lay persons need little practice to become faster than unaided "connoisseurs".

4. Conclusions

We introduced and evaluated the CAVIAR concept. Collecting photographs of flowers and conducting experiments on recognizing them quickly and accurately has been pleasant. We are now porting CAVIAR to a hand-held device [1] to investigate the applicability of our approach to flowers in the field, and to more mundane tasks anywhere.

Acknowledgments

We thank Borjan Gagoski, Adam Callahan, and Greenie Cheng for collecting the flower database, and Borjan Gagoski for supervising the experiments.

References

[1] A. Evans, J. Sikorski, P. Thomas, J. Zou, G. Nagy, S.-H. Cha, and C.C. Tappert, "Interactive Visual System," *Pace CSIS Tech. Rep. no. 196*, 2003.

[2] I. Haritaoglu, "Scene Text Extraction and Translation for Handheld Devices," *Proc. CVPR01*, vol. 2, pp. 408-413, 2001.

[3] R. Jain, *US NSF Workshop on Visual Information Management Systems*, 1992.

[4] A.C. Kak and G.N. Desouza, "Robotic Vision: What Happened to the Visions of Yesterday," *Proc. ICPR02*, vol. 2, pp. 839-847, 2002.

[5] R. J. Mericsko, "Introduction of 27th AIPR Workshop - Advances in Computer-Assisted Recognition," *Proc. of SPIE*, vol. 3584, October 1998.

[6] G. Nagy and J. Zou, "Interactive Visual Pattern Recognition," *Proc. ICPR02*, vol. 2, pp. 478-481, 2002.

[7] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, No. 5, pp. 644-655, 1998.

[8] J. Yang, X. Chen, and W. Kunz, "A PDA-based Face Recognition System," *Proc. of the 6th IEEE Workshop on Applications of Computer Vision*, pp. 19-23, December, 2002.

[9] J. Zhang, X. Chen, J. Yang, and A. Waibel, "A PDA-based Sign Translator," *Proc. of the 4th IEEE Int. Conf. on Multimodal Interfaces*, pp. 217-222, 2002.

[10] J. Zou, *Computer Assisted Visual InterActive Recognition*, Ph.D. thesis, ECSE Department, RPI, May, 2004.

Table 4. Average percentage of successive rose curve adjustments (%)

	0	1	2	3	4	5	6	7	8	9	10	>10
T2	58.5	16.5	8.2	5.6	5.1	1.6	1.3	1.6	0.5	0.3	0.5	0.3
T3	45.9	19.3	10.5	9.5	4.4	3.8	1.5	1.8	1.1	0.7	0.2	1.5
T4	46.1	21.1	12.6	8.2	5.1	2.5	1.0	1.0	0.8	0.5	0.5	0.8
T5	57.4	16.8	8.5	7.2	4.6	1.6	1.1	0.3	0.5	0.5	0.3	1.1
Mean	52.0	18.4	9.9	7.6	4.8	2.4	1.2	1.2	0.7	0.5	0.4	0.9

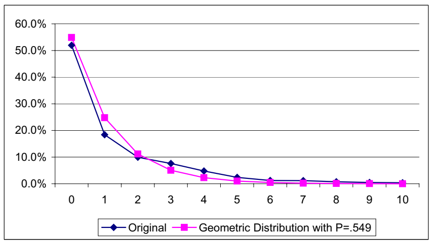


Figure 3. The percentage of successive rose curve adjustments.

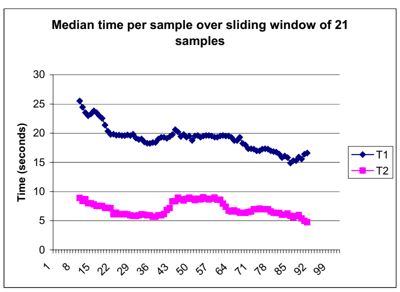


Figure 4. Human learning.

3.4. Human recognition strategy and learning

Table 4 shows the average percentage of successive rose curve adjustments for T2-T5. We observe that 1) more than 90% of the samples are identified by 3 adjustments; 2) there is little difference among T2, T3, T4, and T5. We therefore average them in Figure 3, and show that a geometric distribution with $p = 0.549$ fits the curve well, i.e., the probability of success on each adjustment is just over one half.

Figure 4 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. We observe that 1) in T1, the human time decreases from 26 to 17 seconds as the subjects become more familiar with the database; 2) in T2, the human time decreases from 9 to 5 seconds.

We conclude that: 1) on average, 52% of the samples are immediately confirmed; 2) subjects do adjust the rose curve when necessary and, on average, each sample requires 1.3