CAMERA ASSISTED VISUAL INTERACTIVE RECOGNITION (*CAVIAR*)

George Nagy (nagy@ecse.rpi.edu)

We have developed algorithms and software for an interactive recognition system. CAVIAR draws on sequential pattern recognition, image database, expert systems, pen computing, and digital camera technology. It recognizes wild flowers, cultivars, or Han characters more accurately than machine vision and faster than most laypersons. The novelty of the approach is that human perceptual ability is exploited through interaction with the image of the unknown object. The computer remembers the characteristics of all previously seen classes, *suggests* possible operator actions, and displays confidence scores based on already detected features.

Interactive feature extraction

The interaction is based on the few primitive actions that can be executed easily with a stylus and a small, touch-sensitive display. The richness of the interaction results from its *interpretation*. The system is aware that the operator is pointing at a petal, a stamen, a blemish, or the tip of a leaf. (If the currently proposed interaction has already been executed on the reference image, that result is displayed to help the operator. An operator who is unsure of what a *stamen* is can glance at the expected target highlighted on the currently displayed reference images.) A bounding box is interpreted, as appropriate, as that of the whole flower, of a leaf, or of a distinctive secondary color. When automated segmentation fails, the operator need only point to the incorrect part of the boundary. Standard color, shape and texture features are instantly computed on the designated part of the image. The new top candidates, based on the new confidence values, are displayed. The operator action leading to the potentially most discriminating feature is suggested.

Classification

The recognition engine is a sequential k-nearest-neighbors classifier. The system always displays exemplars of the three currently most-probable classes. With a single click, the operator can request other reference pictures of the same classes, or of a different class. When she finds the reference picture most similar to that of the unknown object, she clicks on the reference picture to assign its class to the unknown. If no acceptable match is found, she marks the image as *reject*. The purpose of the automated image processing and classifier functions is to save time by steering the operator to the most likely candidates.

Learning

The features extracted from newly classified objects are added to the reference database to improve the estimates of the class-conditional feature probabilities. At the same time, the operator gains familiarity with the most discriminating features for each type of object. The objective of the system is to minimize a weighted combination of classification error and operator time. We expect that a layperson will, after sufficient experience with the system, classify objects as accurately and as fast as a domain expert. At that time, the system will no longer be needed for that family of objects. This opens up whole families of educational applications.

System architecture

In developing CAVIAR, we observe the following guidelines:

- It should take very little time and expertise to adapt the system to a new family of objects.
- Interaction should be intuitive, fast, simple, and *consistent*.
- Only a few reference images of each class (currently *2*) should be necessary to prime the system.
- The human is always in charge: the computed suggestions can be heeded or discarded. The operator's decision is final.
- Every interaction and calculation is logged and time-stamped to allow experimentation to track improvement in the performance of the system and of the operator, and to reveal weak points.

Current status

An MS-Windows style prototype, using a public domain Intel computer vision library, is under development as a PhD project. Porting the system to a digital camera and a pocket computer is also a current MS. project Possible modes of deployment include web cameras with server-mediated classification, camera-back interaction, PDA-camera combinations, and self-contained stationary systems for industrial or luggage inspection. Our principal research objective is to establish a sound basis for partitioning the necessary tasks between the operator and the machine. We seek help to obtain suitable cameras and pocket computers, software development tools, and student support. We also hope to find partners to apply CAVIAR to industrial classification and *training*, and to K1-12 and university-level education.