# Why am I here?

George Nagy, Rensselaer Polytechnic Institute, Troy, NY (nagy@ecse.rpi.edu)

My ticket to this workshop may be my forty years' in *OCR*. Over the years I have taught computers to recognize typescript, printed legal documents, patents, business letters, children's books, Moby Dick, technical journals, newspaper articles, hand-printed zip-codes, cursive handwriting, tables, topographic maps, engineering drawings, serial numbers on revolver butts, telephone-exchange minute counters, and Han.

My PhD dissertation was on neural networks for continuous speech recognition. My dominant interest ever since has been adaptation or learning, i.e., systems *that improve with use.* This includes the application of *linguistic context* at the morphological, lexical and syntactic levels, and the application of substitution-cipher decryption algorithms to OCR (with Dick Casey, Sharad Seth, and Tin Ho). More recently, my students and I have been exploiting spatial context, or *style*, which is nothing but the statistical dependence between features of patterns from the same source (printer, scanner, writer, speaker).

On a break from OCR, Haralambopoulos and my analysis of the log files of an entire year's throughput of a university computing center (when most universities had only one computer) won a prize at a computer performance evaluation conference. Later David Embley and I were captivated by *human-computer interaction* (which at that time meant mainly text editors), tested Ss, and wrote a *Computing Survey*s review. Beginning with simulated data for the forthcoming ERTS (later Landsat) from the Michigan Aircraft, my colleagues and I analyzed *multispectral remote sensing* imagery. On a recent sabbatical leave in Sweden, I was pleased to be able to pick up this thread after a twenty-year lapse. Now we are studying 4-D phenomena at the other end of the size scale: X-ray microtomographic observations of crack propagation in concrete. Other interests include *computational geometry* (we proved a theorem that Euclid had missed), *digital terrain models, visibility*, and *geometric modeling* of mechanical parts: Wagle and I wrote the first *Computing Surveys* review of GIS.

Another current project where some of my experience is proving handy, and which may be of interest here, is *CAMERA ASSISTED VISUAL INTERACTIVE RECOGNITION* (CAVIAR). We have developed algorithms and software for a camera-based 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.

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. 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.

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 partners to apply CAVIAR to industrial classification and *training*, and to K1-12 and university-level education.