The Bochum/USC Face Recognition System and How it Fared in the FERET Phase III Test

Kazunori Okada[1], Johannes Steffens[1,2,3], Thomas Maurer[2], Hai Hong[1], Egor Elagin[1,3], Hartmut Neven[1,3], and Christoph von der Malsburg[1,2]

[1] Computer Science Department *and* Center for Neural Engineering
University of Southern California
Los Angeles, CA 90089-2520, USA
kazunori@selforg.usc.edu

[2] Institut für Neuroinformatik
Ruhr-Universität Bochum
D–44780 Bochum, Germany

[3] Now also at Eyematic Interfaces, Inc
827 20th Street, Santa Monica, CA 90403, USA

Summary. This paper summarizes the Bochum/USC face recognition system, our preparations for the FERET Phase III test, and test results as far as they have been made known to us. Our technology is based on Gabor wavelets and elastic bunch graph matching. We briefly discuss our technology in relation to biological and PCA based systems and indicate current activities in the lab and potential future applications.

1. Introduction

Vision is the most important of our senses by which we establish continuity between past and present. Vision is difficult for the simple fact that present scenes never repeat past examples in detail. Bridging that difference is the challenge. Vision has many aspects among which object recognition is but one. Object recognition requires the detection of similarity in spite of image variation in terms of translation, rotation scaling, pose (rotation in depth), deformation, illumination, occlusion, noise and background. Moreover, depending on the specific task, an object may have changing attributes, e.g., surface markings.

In principle, there are three types of information as a basis for generalization from past samples to present instances. One is the information in those samples themselves. It is of extreme biological importance to generalize from minimal sample bases. A second is the structural commonality of an individual object with others. This is a prominent aspect of face recognition, but plays an important role also for more variegated objects as far as they are composed of common shape primitives [Biederman, 1987]. A third type of information is based on first principles which can be built into a system and need not be derived from experience at all. An example of first principles are the transformation laws within the image plane — translation, rotation and scaling. In general, a vision system will exploit a mixture of all three information sources.

Face recognition is a rather particular example of object recognition in that all faces are qualitatively similar to each other and the distinctions to be made are of a gradual nature. For a discussion of face recognition in distinction to other object

H. Wechsler et al. (eds.), *Face Recognition*

recognition tasks see [Biederman and Kalocsai, 1997]. There is the common expectation that technical systems are potentially superior to human face recognition in being able to make precise metric measurements. Unfortunately, this is vitiated by even small variations in pose and facial expression.

We are dealing here with the problem of recognizing a person from a single photograph against a gallery of hundreds of persons, each represented again by a single photograph — the task set by the FERET program. The task as such is virtually impossible for humans to perform, due to the practical impossibility of memorizing (or repeatedly looking through) data bases of thousands of images as was required in the program's test. But even deciding whether two images presented in direct sequence do or do not refer to the same person is made difficult by variation in pose (or expression) [Kalocsai et al., 1994], [Biederman and Kalocsai, 1997]. The difficulty arises from the fact that a single photo doesn't contain enough information about a face's depth profile to predict images of different pose. The face recognition system we have developed is distinguished from others by a larger extent to which its generalization capabilities are based on general principles instead of on statistical learning. We will come back to this point at the end of the paper.

This report succinctly describes the basic system as developed previously [Lades et al., 1993], [Wiskott et al., 1997] and the particular improvements in preparation for the latest FERET test, as well as some details of our system implementation. We then discuss performance of our system resulting from in-house preparation tests and the FERET phase III test, which we have taken in March of 1997 and which has been partially reported [Phillips and Rauss, 1997]. We conclude by mentioning current activities in the lab and potential future applications of our technology, and by discussing our technology in relation to biological and PCA-based systems.

2. The System as Previously Developed

2.1 The Wavelet Transform

Previous versions of our system are described in [Lades et al., 1993],[Wiskott et al., 1997]. The basic data format of our system is the Gabor-based wavelet

$$\psi_{\mathbf{k}}(\mathbf{x}) = \frac{k^2}{\sigma^2} e^{-\frac{k^2}{2\sigma^2}x^2} \left\{ e^{i\,\mathbf{k}\,\mathbf{x}} - e^{-\frac{\sigma^2}{2}} \right\}. \tag{2.1}$$

The wavelet is a plane wave with wave vector $\mathbf{k}$, restricted by a Gaussian window, the size of which relative to the wavelength is parameterized by σ. The second term in the brace removes the DC component. A wavelet, centered at image position $\mathbf{x}$, is used to extract the wavelet component $J_{\mathbf{k}}$ from the image with gray level distribution $I(\mathbf{x})$,

$$J_{\mathbf{k}}(\mathbf{x}) = \int d\mathbf{x}'\ I(\mathbf{x}')\ \psi_{\mathbf{k}}(\mathbf{x} - \mathbf{x}'). \tag{2.2}$$

We typically sample the space of wave vectors $\mathbf{k}$ in a discrete hierarchy of 5 resolution levels (differing by half-octaves) and 8 orientations at each resolution level, thus giving 40 complex values for each sampled image point (the real and imaginary components referring to the cosine and sine phases of the plane wave). We designate the samples in $\mathbf{k}$-space by the index $j = 1, \ldots, 40$ and consider all wavelet components centered in a single image point as a vector which we call a *jet*. A jet describes the local features of the area surrounding $\mathbf{x}$. If sampled with sufficient density, the image can be reconstructed from jets within the bandpass covered by the sampled frequencies.