Internal Report 96–08

Face Recognition by Elastic Bunch Graph Matching

by

Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger,
and Christoph von der Malsburg

Ruhr-Universität Bochum
Institut für Neuroinformatik
44780 Bochum

IR-INI 96–08
April 1996
ISSN 0943-2752

Face Recognition by Elastic Bunch Graph Matching*

Laurenz Wiskott[1†], Jean-Marc Fellous[2‡],
Norbert Krüger[1], and Christoph von der Malsburg[1,2]

[1] Institut für Neuroinformatik
Ruhr-Universität Bochum
D-44780 Bochum, Germany
http://www.neuroinformatik.ruhr-uni-bochum.de

[2] Computer Science Department
University of Southern California
Los Angeles, CA 90089, USA

April 18, 1996

Abstract

We present a system for recognizing human faces from single images out of a large database with one image per person. The task is difficult because of image variance in terms of position, size, expression and pose. The system collapses most of this variance by extracting concise face descriptions in the form of *image graphs*. In these, fiducial points on the face (eyes, mouth etc.) are described by sets of wavelet components (*jets*). Image graph extraction is based on a novel approach, the *bunch graph*, which is constructed from a small set of sample image graphs. Recognition is based on a straightforward comparison of image graphs. We report here on recognition experiments with galleries of 250 images from the FERET database, also across different poses.

Keywords: face recognition, different poses, Gabor wavelets, elastic graph matching, bunch graph, ARPA/ARL FERET database.

*Supported by grants from the German Federal Ministry for Science and Technology (413-5839-01 IN 101 B9) and from ARPA and the U.S. Army Research Lab (01/93/K-109).

†Current address: Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, San Diego, CA 92186-5800, http://www.cnl.salk.edu/CNL, wiskott@salk.edu.

‡Current address: Volen Center for Complex Systems, Brandeis University, Waltham, MA 02254-9110.

1 Introduction

We set ourselves the task of recognizing persons from single images by reference to a gallery, which also contained only one image per person. Our problem was to address image variation due to differences in head pose, position, and size and due to changing facial expression (to name only the most important). Our task is thus a typical discrimination-in-the-presence-of-variance problem, where one has to try to collapse the variance and to emphasize discriminating features. This is in general only possible with the help of information about the structure of variations to be expected.

Classification systems differ vastly in terms of nature and origin of their knowledge about image variations. Systems in Artificial Intelligence and Computer Vision often stress specific designer-provided structures, for instance explicit models of three-dimensional objects or of the image-generation process, whereas Neural Network models tend to stress absorption of structure from examples with the help of statistical estimation techniques. Both of these extremes are expensive in their own way and fall painfully short of the ease with which natural systems pick up essential information from just a few examples. Part of the success of natural systems must be due to general properties and laws as to how object images transform under natural conditions.

Our system has an important core of structure which refers to images in general and to the fact that the images of coherent objects tend to translate, scale, rotate and deform elastically in the image plane. Our basic object representation is the labeled graph; edges are labeled with distance information and nodes are labeled with wavelet responses locally bundled in *jets*. Stored *model graphs* can be matched to new images to generate *image graphs*, which can then be incorporated into a gallery and become model graphs. Wavelets as we use them are robust to moderate lighting changes and small shifts and deformations. Model graphs can easily be translated, scaled, oriented or deformed during the matching process, thus compensating for a large part of the variance of the images. Unfortunately, having only one image for each person in the galleries does not provide sufficient information to handle rotation in depth analogously. Though, we present results on recognition across different poses.

This general structure is useful for handling any kind of coherent object and may be sufficient for discriminating between structurally different object types. However, for in-class discrimination of objects, of which face recognition is an example, it is necessary to have information specific to the structure common to all objects in the class. This is crucial for the extraction of those structural traits from the image which are important for discrimination ("to know where to look and what to pay attention to"). In our system, class-specific information has the form of *bunch graphs*, one for each pose, which are stacks of a moderate number (70 in our experiments) of different faces, jet-sampled in an appropriate set of fiducial points (placed over eyes, mouth, contour etc.). Bunch graphs are treated as combinatorial entities in that for different fiducial points jets from different sample faces can be selected, thus creating a highly adaptable model. This model is matched to new facial images in order to reliably find the fiducial points in the image. Jets at these points and their relative positions are extracted and are combined into an image graph, a representation of the face which has no remaining variation due to size, position (or in-plane orientation, not implemented here).

A bunch graph is created in two stages. Its qualitative structure as a graph (a set of nodes plus edges) as well as the assignment of corresponding labels (jets and distances) for one initial image is designer-provided, whereas the bulk of the bunch graph is extracted semi-automatically from sample images by matching the embryonic bunch graph to them, less and less often intervening to correct incorrectly identified fiducial points. Image graphs are rather robust to small in-depth rotations of the head. Larger rotation angles, i.e. different poses, are handled with the help of bunch graphs with different graph structure and designer-provided correspondences between nodes in different poses.

After these preparations our system can extract from single images concise invariant face descriptions in the form of image graphs (called model graphs when in a gallery). They contain all information relevant for the face discrimination task. For the purpose of recognition, image graphs can be compared with model graphs at small computing cost by evaluating the mean jet similarity.

In summary, our system is based to a maximum on a general data structure — graphs labeled with wavelet responses — and general transformation properties. These are designer-provided, but due to their generality and simplicity the necessary effort is minimal. At the present stage of development our system makes use of hand-crafted object-specific graph structures and a moderately labor-intensive procedure to generate bunch-graphs. We plan to eliminate this need for human intervention and guess-work with the help of statistical estimation methods. Our system comes close to the natural model by needing only a small number of examples in order to handle the complex task of face recognition.

We will compare our system to others and to our own previous work in the discussion.

2 The System

2.1 Preprocessing with Gabor Wavelets

The representation of local features is based on the Gabor wavelet transform (see Figure 1). Gabor wavelets are biologically motivated convolution kernels in the shape of plane waves restricted by a Gaussian envelope function (Daugman 1988). The set of convolution coefficients for kernels of different orientations and frequencies at one image pixel is called a jet. In this section we define jets, different similarity functions between jets, and our procedure for precise localization of jets in an image.

2.1.1 Jets

A *jet* describes a small patch of grey values in an image $\mathcal{I}(\vec{x})$ around a given pixel $\vec{x} = (x, y)$. It is based on a wavelet transform, defined as a convolution

$$\mathcal{J}_j(\vec{x}) = \int \mathcal{I}(\vec{x}')\psi_j(\vec{x} - \vec{x}')d^2\vec{x}' \tag{1}$$

with a family of *Gabor kernels*

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} \exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right)\left[\exp(i\vec{k}_j\vec{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \tag{2}$$

in the shape of plane waves with wave vector $\vec{k}_j$, restricted by a Gaussian envelope function. We employ a discrete set of 5 different frequencies, index $\nu = 0, ..., 4$, and 8 orientations, index $\mu = 0, ..., 7$,

$$\vec{k}_j = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_\nu \cos\varphi_\mu \\ k_\nu \sin\varphi_\mu \end{pmatrix}, \; k_\nu = 2^{-\frac{\nu+2}{2}}\pi, \; \varphi_\mu = \mu\frac{\pi}{8}, \tag{3}$$

with index $j = \mu + 8\nu$. This sampling evenly covers a band in frequency space. The width σ/k of the Gaussian is controlled by the parameter $\sigma = 2\pi$. The second term in the bracket of (2) makes the kernels *DC-free*, i.e. the integral $\int \psi_j(\vec{x})d^2\vec{x}$ vanishes. One speaks of a wavelet transform since the family of kernels is self-similar, all kernels being generated from one *mother wavelet* by dilation and rotation.

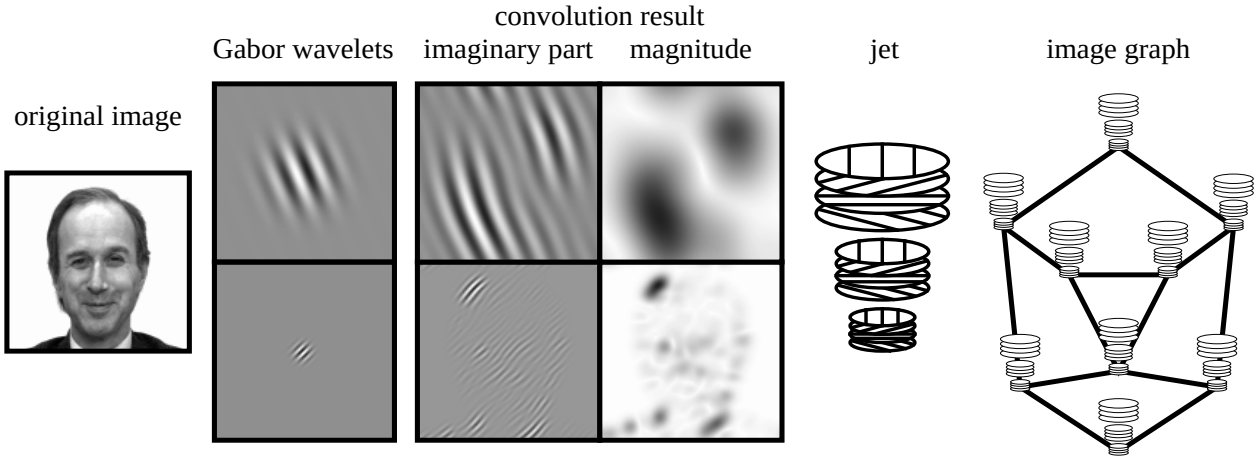

Figure 1: The graph representation of a face is based on the Gabor wavelet transform, a set of convolutions with kernels in the shape of wavelets. These have the shape of plane waves restricted by a Gaussian envelope function. We compute 40 coefficients (5 frequencies × 8 orientations). Coefficient phase varies with wavelet frequency (see imaginary part), magnitude varies slowly. The set of 40 coefficients obtained for one image point is referred to as a jet. A sparse collection of such jets together with some information about their relative location constitutes an image graph, used to represent an object, such as a face.

A jet $\mathcal{J}$ is defined as the set $\{\mathcal{J}_j\}$ of 40 complex coefficients obtained for one image point. It can be written as

$$\mathcal{J}_j = a_j \exp(i\phi_j) \tag{4}$$

with amplitudes $a_j(\vec{x})$, which slowly vary with position, and phases $\phi_j(\vec{x})$, which rotate with a rate set by the spatial frequency or wave vector $\vec{k}_j$ of the kernels (see Figure 1).

Gabor wavelets were chosen for their robustness as a data format and for their biological relevance. Since they are DC-free, they provide robustness against varying brightness in the image. Robustness against varying contrast can be obtained by normalizing the jets. The limited localization in space and frequency yields a certain amount of robustness against translation, distortion, rotation, and scaling. Only the phase changes drastically with translation. This phase variation can be either ignored, or it can be used for estimating

displacement, as will be shown later. A disadvantage of the large kernels is their sensitivity to background variations. It was shown, however, that if the object contour is known the influence of the background can be suppressed (PÖTZSCH 1994). Finally, the Gabor wavelets have similar shape as the receptive fields of simple cells found in the visual cortex of vertebrate animals (POLLEN & RONNER 1981; JONES & PALMER 1987; DEVALOIS & DEVALOIS 1988).

2.1.2 Comparing Jets

Due to phase rotation, jets taken from image points only a few pixels from each other have very different coefficients, although representing almost the same local feature. This can cause severe problems for matching. We therefore either ignore the phase or compensate for its variation explicitly. The similarity function

$$\mathcal{S}_a(\mathcal{J},\mathcal{J}') = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}, \tag{5}$$

already used by BUHMANN et al. (1992) and LADES et al. (1993), ignores the phase completely. With a jet $\mathcal{J}$ taken at a fixed image position and jets $\mathcal{J}' = \mathcal{J}'(\vec{x})$ taken at variable position $\vec{x}$, $\mathcal{S}_a(\mathcal{J},\mathcal{J}'(\vec{x}))$ is a smooth function with local optima forming large attractor basins (see Figure 2a), leading to rapid and reliable convergence with simple search methods such as gradient descent or diffusion.

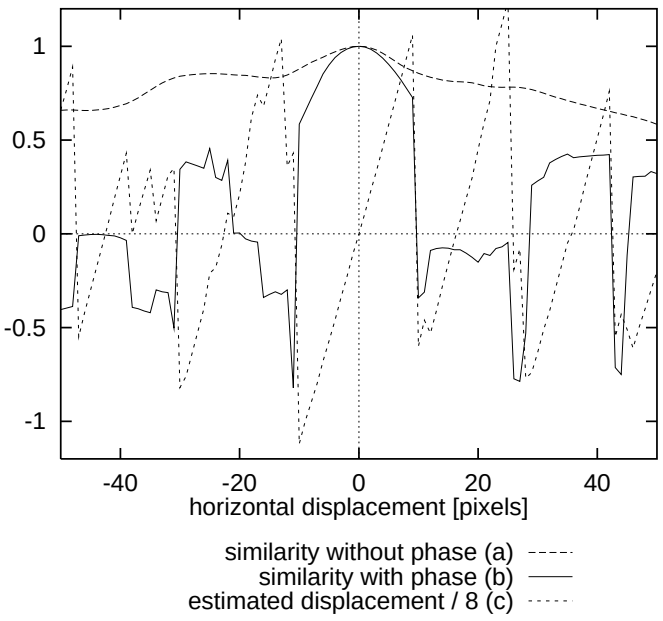

Figure 2: a) Similarity $\mathcal{S}_a(\mathcal{J}(\vec{x}_1),\mathcal{J}'(\vec{x}_0))$ with jet $\mathcal{J}'$ taken from the left eye of the face shown in Figure 1, and jet $\mathcal{J}$ taken from pixel positions of the same horizontal line, $\vec{x}_1 = \vec{x}_0 + (d_x, 0), d_x = -50, ..., 50$ (The image in Figure 1 has a width of 128 pixels). The similarity potential is smooth and has a large attractor basin. **b)** Similarity $\mathcal{S}_\phi(\mathcal{J}(\vec{x}_1),\mathcal{J}'(\vec{x}_0))$ and **c)** estimated displacement $\vec{d}(\mathcal{J}(\vec{x}_1),\mathcal{J}'(\vec{x}_0))$ for the same jets as in a (using focus 1). The similarity potential has many more local optima. The right eye is 24 pixels away from the left eye, generating a local maximum for both similarity functions close to $d_x = -24$. The estimated displacement is precise around the 0-position and rougher at other local optima, especially at the other eye. (The displacement values are divided by 8 to fit the ordinate range.)

Using phase has two potential advantages. Firstly, phase information is required to discriminate between patterns with similar amplitudes, should they occur, and secondly, since phase varies so quickly with location, it provides a means for accurate jet localization in an image. Assuming that two jets $\mathcal{J}$ and $\mathcal{J}'$ refer to object locations with small relative displacement $\vec{d}$, the phase shifts can approximately be compensated for by the terms $\vec{d}\vec{k}_j$, leading to a phase-sensitive similarity function

$$\mathcal{S}_\phi(\mathcal{J},\mathcal{J}') = \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j - \vec{d}\,\vec{k}_j)}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}. \tag{6}$$

In order to compute it, the displacement $\vec{d}$ has to be estimated. This can be done by maximizing $\mathcal{S}_\phi$ in its Taylor expansion, as explained in the following section. It is actually a great advantage of this second similarity function that it yields this displacement information. Profiles of similarities and estimated displacements are shown in Figure 2.

2.1.3 Displacement Estimation

In order to estimate the displacement vector $\vec{d} = (d_x, d_y)$, we have adopted a method used for disparity estimation (Theimer & Mallot 1994) based on (Fleet & Jepson 1990). The idea is to maximize the similarity $\mathcal{S}_\phi$ in its Taylor expansion:

$$\mathcal{S}_\phi(\mathcal{J},\mathcal{J}') \approx \frac{\sum_j a_j a'_j [1 - 0.5(\phi_j - \phi'_j - \vec{d}\,\vec{k}_j)^2]}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}\ . \tag{7}$$

Setting $\frac{\partial}{\partial d_x}\mathcal{S}_\phi = \frac{\partial}{\partial d_y}\mathcal{S}_\phi = 0$ and solving for $\vec{d}$ leads to

$$\vec{d}(\mathcal{J},\mathcal{J}') = \begin{pmatrix} d_x \\ d_y \end{pmatrix} = \frac{1}{\Gamma_{xx}\Gamma_{yy} - \Gamma_{xy}\Gamma_{yx}} \times \begin{pmatrix} \Gamma_{yy} & -\Gamma_{yx} \\ -\Gamma_{xy} & \Gamma_{xx} \end{pmatrix} \begin{pmatrix} \Phi_x \\ \Phi_y \end{pmatrix}, \tag{8}$$

if $\Gamma_{xx}\Gamma_{yy} - \Gamma_{xy}\Gamma_{yx} \neq 0$, with

$$\begin{aligned} \Phi_x &= \textstyle\sum_j a_j a'_j k_{jx}(\phi_j - \phi'_j), \\ \Gamma_{xy} &= \textstyle\sum_j a_j a'_j k_{jx} k_{jy}, \end{aligned}$$

and $\Phi_y, \Gamma_{xx}, \Gamma_{yx}, \Gamma_{yy}$ defined correspondingly.

This equation yields a straightforward method for estimating the displacement or disparity between two jets taken from object locations close enough that their Gabor kernels are highly overlapping. Without further modifications, this equation can determine displacements up to half the wavelength of the highest frequency kernel, which would be two pixels for $k_0 = \pi/2$. The range can be increased by using low frequency kernels only. For the largest kernels the estimated displacement may be 8 pixels. One can then proceed with the next higher frequency level and refine the result, possibly by correcting the phases of the higher frequency coefficients by multiples of 2π according to the displacement estimated

on the lower frequency level. We refer to the number of frequency levels used for the first displacement estimation as *focus*. A focus of 1 indicates that only the lowest frequency level is used and that the estimated displacement may be up to 8 pixels. A focus of 5 indicates that all five levels are used, and the disparity may only be up to 2 pixels. In any case, all five levels are eventually used in the iterative refinement process described above.

If one has access to the whole image of jets, one can also work iteratively. Assume a jet $\mathcal{J}$ is to be accurately positioned in the neighborhood of point $\vec{x}_0$ in an image. Comparing $\mathcal{J}$ with the jet $\mathcal{J}_0 = \mathcal{J}(\vec{x}_0)$ yields an estimated displacement of $\vec{d_0} = \vec{d}(\mathcal{J}, \mathcal{J}(\vec{x}_0))$. Then a jet $\mathcal{J}_1$ is taken from position $\vec{x}_1 = \vec{x}_0 + \vec{d_0}$ and the displacement is estimated again. But since the new location is closer to the correct position, the new displacement $\vec{d_1}$ will be smaller and can be estimated more accurately with a higher focus, converging eventually to subpixel accuracy. We have used this iterative scheme in the matching process described in Section 2.3.

2.2 Face Representation

2.2.1 Individual Faces

For faces, we have defined a set of *fiducial points*, e.g. the pupils, the corners of the mouth, the tip of the nose, the top and bottom of the ears, etc. A *labeled graph* $\mathcal{G}$ representing a face consists of N nodes on these fiducial points at positions $\vec{x}_n, n = 1, ..., N$ and E edges between them. The nodes are labeled with jets $\mathcal{J}_n$. The edges are labeled with distances $\Delta\vec{x}_e = \vec{x}_n - \vec{x}_{n'}, e = 1, ..., E$, where edge e connects node n' with n. Hence the edge labels are two-dimensional vectors. (When wanting to refer to the geometrical structure of a graph, unlabeled by jets, we call it a *grid*.) This face graph is *object-adapted*, since the nodes are selected from face-specific points (fiducial points, see Figure 4).

Graphs for different head pose differ in geometry and local features. Although the fiducial points refer to corresponding object locations, some may be occluded, and jets as well as distances vary due to rotation in depth. To be able to compare graphs for different poses, we manually defined pointers to associate corresponding nodes in the different graphs.

2.2.2 Face Bunch Graphs

In order to find fiducial points in new faces, one needs a general representation rather than models of individual faces. This representation should cover a wide range of possible variations in the appearance of faces, such as differently shaped eyes, mouths, or noses, different types of beards, variations due to gender, age and race, etc. It is obvious that it would be too expensive to cover each feature combination by a separate graph. We instead combine a representative set of individual model graphs into a stack-like structure, called a *face bunch graph* (FBG) (see Figure 3). Each model has the same grid structure and the nodes refer to identical fiducial points. A set of jets referring to one fiducial point is called a *bunch*. An eye bunch, for instance, may include jets from closed, open, female, and male eyes etc. to cover these local variations. During the location of fiducial points in a face not seen before, the procedure described in the next section selects the best fitting jet, called the *local expert*, from the bunch dedicated to each fiducial point. Thus, the full combinatorics of jets in the

bunch graph is available, covering a much larger range of facial variation than represented in the constituting model graphs themselves.

Assume for a particular pose there are M model graphs $\mathcal{G}^{\mathcal{B}m}$ $(m = 1, ..., M)$ of identical structure, taken from different model faces. The corresponding FBG $\mathcal{B}$ then is given the same structure, its nodes are labeled with bunches of jets $\mathcal{J}_n^{\mathcal{B}m}$ and its edges are labeled with the averaged distances $\Delta\vec{x}_e^{\mathcal{B}} = \sum_m \Delta\vec{x}_e^{\mathcal{B}m}/M$.

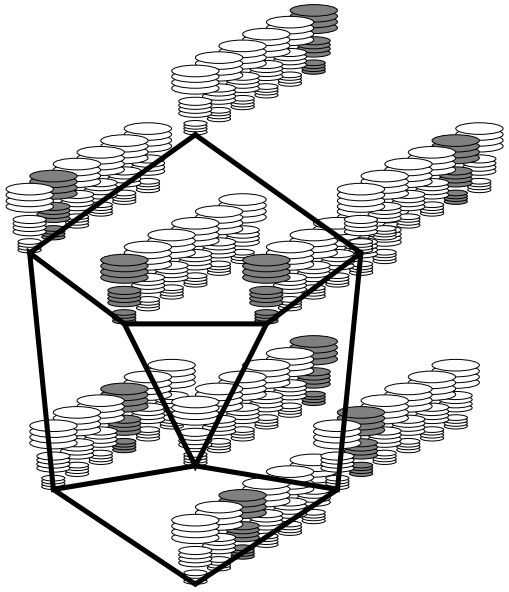

Figure 3: The Face Bunch Graph (FBG) serves as a representation of faces in general. It is designed to cover all possible variations in the appearance of faces. The FBG combines information from a number of face graphs. Its nodes are labeled with sets of jets, called bunches, and its edges are labeled with averages of distance vectors. During comparison to an image the best fitting jet in each bunch is selected independently, indicated by grey shading.

2.3 Generating Face Representations by Elastic Bunch Graph Matching

So far we have only described how individual faces and general knowledge about faces are represented by labeled graphs and the FBG, respectively. We are now going to explain how these graphs are generated. The simplest method is to do so manually, by marking a set of fiducial points for a given image and letting edge labels be computed as the differences between node positions. Finally the Gabor wavelet transform provides the jets for the nodes. We have used this manual method to generate initial graphs for the system, one for each pose, together with pointers to indicate which pairs of nodes in graphs for different poses correspond to each other.

If the system has a FBG (possibly consisting of one model only), graphs for new images can automatically be generated by Elastic Bunch Graph Matching. Initially, when the FBG contains only few faces, it is necessary to review and correct the resulting matches, but once the FBG is rich enough (approximately 70 graphs) one can rely on the matching and generate large galleries of model graphs automatically.

2.3.1 The Graph Similarity Function

A key role in Elastic Bunch Graph Matching (EBGM) is played by a function evaluating the *graph similarity* between an image graph and the FBG of identical pose. It depends on the jet similarities and the distortion of the image grid relative to the FBG grid. For an image graph $\mathcal{G}^{\mathcal{I}}$ with nodes $n = 1, ..., N$ and edges $e = 1, ..., E$ and a FBG $\mathcal{B}$ with model graphs $m = 1, ..., M$ the similarity is defined as

$$\mathcal{S}_{\mathcal{B}}(\mathcal{G}^{\mathcal{I}}, \mathcal{B}) = \frac{1}{N} \sum_{n} \max_{m} \left(\mathcal{S}_{\phi}(\mathcal{J}_n^{\mathcal{I}}, \mathcal{J}_n^{\mathcal{B}m}) \right) - \frac{\lambda}{E} \sum_{e} \frac{(\Delta \vec{x}_e^{\mathcal{I}} - \Delta \vec{x}_e^{\mathcal{B}})^2}{(\Delta \vec{x}_e^{\mathcal{B}})^2}, \tag{9}$$

where λ determines the relative importance of jets and metric structure. $\mathcal{J}_n$ are the jets at node n and $\Delta \vec{x}_e$ are the distance vectors used as labels at edges e. Since the FBG provides several jets for each fiducial point, the best one is selected and used for comparison. These best fitting jets serve as *local experts* for the image face.

2.3.2 Matching Procedure

The goal of EBGM on a probe image is to find the fiducial points and thus to extract from the image a graph that maximizes the similarity with the FBG. In practice, one has to apply a heuristic algorithm to come close to the optimum within reasonable time. We use a coarse to fine approach. The matching schedule described here assumes normalized face images of known pose such that only one FBG is required. The more general case of varying size is sketched in the next section.

Stage 1 Find approximate face position: Condense the FGB into an *average graph* by taking the average amplitude of the jets in each bunch of the FBG (or, alternatively, select one arbitrary graph as a representative). Use this as a rigid model ($\lambda = \infty$) and evaluate its similarity at each location of a square lattice with a spacing of 4 pixels. At this stage the similarity function $\mathcal{S}_a$ without phase is used instead of $\mathcal{S}_{\phi}$. Repeat the scanning around the best fitting position with a spacing of 1 pixel. The best fitting position finally serves as starting point for the next stage.

Stage 2 Refine position and estimate size: Now the FBG is used without averaging, varying it in position and size. Check the four different positions $(\pm 3, \pm 3)$ pixels displaced from the position found in Stage 1, and at each position check two different sizes which have the same center position, a factor of 1.18 smaller or larger than the FBG average size. This is without effect on the metric similarity, since the vectors $\vec{x}_e^{\mathcal{B}}$ are transformed accordingly. We still keep $\lambda = \infty$. For each of these eight variations the best fitting jet for each node is selected and its displacement according to Equation 8 is computed. This is done with a focus of 1, i.e., the displacements may be of a magnitude up to eight pixels. The grids are then rescaled and repositioned in order to minimize the square sum over the displacements. Keep the best of the eight variations as starting point for the next stage.

Stage 3 Refine size and find aspect ratio: A similar relaxation process as described for Stage 2 is applied, relaxing, however, the x- and y-dimensions independently. In addition, the focus is increased successively from 1 to 5.

Stage 4 Local distortion: In a pseudo-random sequence the position of each individual image node is varied in order to further increase the similarity to the FBG. Now the metric similarity is taken into account by setting $\lambda = 2$ and using the vectors $\vec{x}_e^{\mathcal{B}}$ as obtained in Stage 3. In this stage only those positions are considered for which the estimated displacement vector is small ($d < 1$, see Equation 8). For this local distortion the focus again increases from 1 to 5.

The resulting graph is called the *image graph* and is stored as a representation for the individual face of the image.

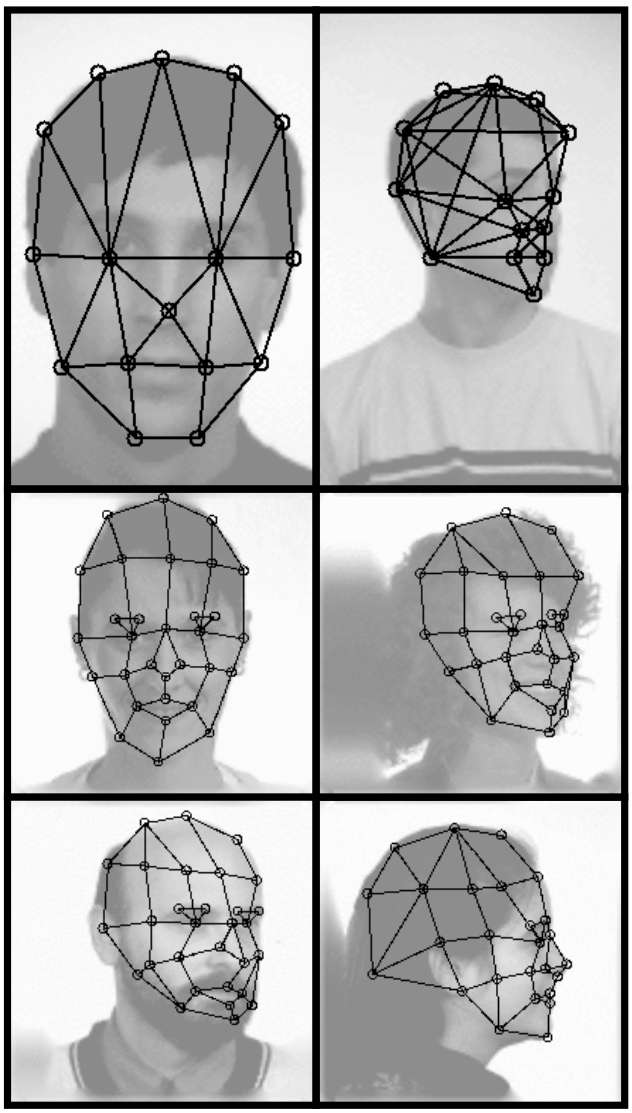

Figure 4: Object-adapted grids for different poses. The nodes are positioned automatically by elastic graph matching against the corresponding face bunch graphs. The two top images show originals with widely differing size and grids as used for the normalization stage with many nodes on the outline. The other images are already rescaled to normal size. Their grids have more nodes on the face, which is more appropriate for recognition (The grids used in Section 3.2 had about 14 additional nodes which are not shown here for simplicity). One can see that in general the matching finds the fiducial points quite accurately. But mismatches occurred for example for the left face in the bottom row. The chin was not found accurately; the leftmost node and the node below it should be at the top and the bottom of the ear respectively.

2.3.3 Schedule of Graph Extraction

To minimize computing effort and to optimize reliability we extract a face representation in two stages. The first stage, called *normalization stage* and described in greater detail in (Krüger et al. 1995), has the purpose of estimating the position and size of the face in

the original image, so that the image can be scaled and cut to standard size. The second stage takes this image as input and extracts a precise image graph appropriate for face recognition purposes. The two stages differ in emphasis. The first one has to deal with greater uncertainty as to size and position of the head and has to optimize the reliability with which it finds the face, but there is no need to find fiducial points with any precision or extract data important for face recognition. The second stage can start with little uncertainty as to position and size of the head but has to extract a detailed face graph with high precision.

In the experiments described below, original images had a format of 256×384 pixels, and the faces varied in size by a factor of three (see Figure 4). The poses were given and needed not to be determined. The normalization stage used three FBGs of appropriate pose which differ in face size. We somewhat arbitrarily picked approximately 30 images to form each FBG. More careful selection of images to cover a wider range of variations can only improve system performance. The grids used in the construction of the FBGs put little emphasis (i.e., few nodes) on the interior of the face and have fewer nodes than those used for the second stage, see Figure 4 for two examples. The smaller number of nodes speeds the process of face finding. Using an EBGM scheme similar to the one described in Section 2.3.2 we match each of the three FBGs to the input image. We select the graph that matches best, cut a frame of appropriate size around it from the image and resize it to 128×128 pixels. The poses could be determined analogously (Krüger et al. 1996), but here the poses are assumed to be known. In our experiments, normalization took approximately 20 seconds on a SPARCstation 10-512 with a 50 MHz processor and identified face position and scale correctly in approximately 99% of the images.

The second stage uses the matching procedure exactly as described in Section 2.3.2, starting the match at standard size and position. The face bunch graphs used in this stage have more nodes, which we have placed in positions which we believe are important for person identification, emphasizing the interior of the face. Each of the three principal poses (frontal, half-profile, and profile; left-facing poses are flipped to right-facing poses) is matched with a different grid structure and with a different FBG, formed by using 70 arbitrarily chosen images. This stage took approximately 10 seconds.

2.4 Recognition

After having extracted model graphs from the gallery images and image graphs from the probe images, recognition is possible with relatively little computational effort, by comparing an image graph to all model graphs and picking the one with the highest similarity value. The similarity function we use here for comparing graphs is an average over the similarities between pairs of corresponding jets. For image and model graphs referring to different pose, we compare jets according to the manually provided correspondences. If $\mathcal{G}^{\mathcal{I}}$ is the image graph, $\mathcal{G}^{\mathcal{M}}$ the model graph, and if node $n_{n'}$ in the model graph corresponds to node n' in the image graph, we define *graph similarity* as:

$$\mathcal{S}_{\mathcal{G}}(\mathcal{G}^{\mathcal{I}}, \mathcal{G}^{\mathcal{M}}) = \frac{1}{N'} \sum_{n'} \mathcal{S}_a(\mathcal{J}^{\mathcal{I}}_{n'}, \mathcal{J}^{\mathcal{M}}_{n_{n'}}), \tag{10}$$

where the sum runs only over the N' nodes in the image graph with a correspondent in the model graph. We use the jet similarity function without phase here. It turned out to

be more discriminative, possibly because it is more robust with respect to change in facial expression and other variations. We here ignore the jet distortions created by rotation in depth, but will take up the subject in a later paper.

This graph similarity induces a ranking of the model graphs relative to an image graph. A person is recognized correctly if the correct model yields the highest graph similarity, i.e., if it is of rank one. A confidence criterion on how reliably a person is recognized can easily be derived from the statistics of the ranking (see LADES et al. 1993). However, we have restricted our results to unqualified recognition rates, which already give a good impression of system performance.

3 Experiments

3.1 Database

The image galleries we used in our experiments are taken from the ARPA/ARL FERET database provided by the US Army Research Laboratory. That database explicitly distinguishes different poses, and images are labeled with pose identity. These poses are: frontal, half-profile right or left (rotated by about 40-70 degrees), and profile right or left (see Figure 5 for examples). For most faces there are two frontal views with different facial expression. Apart from a few exceptions there are no disguises, variations in hairstyle or in clothing. The background is always a homogeneous light or grey, except for smoothly varying shadows. The size of the faces varies by about a factor of three (but is constant for each individual, information which we could have used to improve recognition rates, but didn't). The format of the original images is 256×384 pixels. Inquiries about the database should be directed to Jonathon Phillips (e-mail:`jphillip@nvl.army.mil` or `jonathon@arl.mil`).

3.2 Results

We used various model and probe galleries with faces of different pose. Each gallery contained 250 faces with just one image per person. We relied on the explicitly labeled pose identity instead of using our own pose recognition capability. Recognition results are shown in Table 1.

The very high correct recognition rate for frontal against frontal images (first row) in spite of the variation in facial expression and the slight variations in pose shows the great potential of our system. However, all images were taken in single sessions, such that lighting conditions are very similar from picture to picture and hairstyle and clothing are identical. Further work will have to show the degree of robustness against greater image variation. Initial experiments are encouraging

Before comparing left against right poses we flipped all right pose images over. Since human heads are bilaterally symmetric to some degree and since at least our present system performs poorly on rotation in depth (see below) we proceeded under the expectation that it would be easier to deal with differences due to facial asymmetry than with differences caused by substantial head rotation. This assumption is born out at least by the high recognition rate of 84% for right profile against left profile (third row). The sharply reduced

Figure 5: Sample faces from the ARPA/ARL FERET database: frontal views, half-profiles, and profiles. Pictures for left-facing poses are flipped around a vertical axis, and all images have been rescaled to standard size by our normalization stage (Section 2.3.3). Notice the large variation in the rotation angle of half-profiles and that some faces have no variation in facial expression.

---

recognition rate of 57% (second row) when comparing left and right half-profiles could be due to inherent facial asymmetry, but the more likely reason is the poor control in rotation angle in the database — inspection of images shows that right and left rotation angles differ by up to 30 degrees, cf. Figure 5.

When comparing half profiles with either frontal views or profiles another reduction in recognition rate is observed (although even a correct recognition rate of 10% out of a gallery of 250 is still high above chance level, which would be 0.4%!). The results are asymmetrical, performance being better if frontal or profile images serve as model gallery than if half-profiles are used. This is due to the fact that both frontal and profile poses are much more standardized than half-profiles, for which the angle varies between 40 and 70 degrees. We interpret this as being due to the fact that similarity is more sensitive to depth-rotation than to inter-individual face differences. Thus, when comparing frontal probe images to a half-profile gallery, a 40 degree half-profile gallery image of a wrong person is often favored over the correct gallery image if in the latter the head is rotated by a larger angle. A large number of such false positives degrades the correct-recognition rate considerably.

Model gallery	Probe images	First rank #	First rank %	First 10 ranks #	First 10 ranks %
250 fa	250 fb	245	98	248	99
250 hr	181 hl	103	57	147	81
250 pr	250 pl	210	84	236	94
249 fa + 1 fb	171 hl + 79 hr	44	18	111	44
171 hl + 79 hr	249 fa + 1 fb	42	17	95	38
170 hl + 80 hr	217 pl + 33 pr	22	9	67	27
217 pl + 33 pr	170 hl + 80 hr	31	12	80	32

Table 1: Recognition results for cross-runs between different galleries (f: frontal views; a, b: expression a and b; h: half profiles; p: profiles; l, r: left and right). Each gallery contained only one image per person; the different compositions in the four bottom lines are due to the fact that not for all persons all poses were available. Given are numbers on how often the correct model was identified as rank one and how often it was among the first 10 (4%).

4 Discussion

The system we describe is general and flexible. It is designed for an *in-class recognition* task, i.e. for recognizing members of a known class of objects. We have applied it to face recognition but the system is in no way specialized to faces and it can be directly applied to other in-class recognition tasks, such as recognizing individuals of a given animal species, given the same level of standardization of the images. In contrast to many neural network systems, no extensive training for new faces or new object classes is required. Only a moderate number of typical examples have to be inspected to build up a bunch graph, and individuals can then be recognized after storing a single image.

The performance is high on faces of same pose. Robustness against rotation in depth up to about 20 degrees has already been demonstrated with a previous version of the system (Lades et al. 1993). The constraint of working with single images deprives the system from information necessary for invariance to rotation in depth. Therefore the system performance in recognizing people in unfamiliar poses is significantly degraded. It is known from psychophysical experiments that also human subjects perform poorly on this task. Bruce et al. (1987) have shown that reliability on judging whether two unfamiliar faces are the same degrades significantly with depth rotation angle. In the absence of easily distinguished features such as hairstyle, beard or glasses a similar result was obtained by Kalocsai et al. (1994).

4.1 Comparison to Previous Work

In comparison to the system (Lades et al. 1993) on the basis of which we have developed the system presented here we have made several major modifications. We now utilize wavelet phase information for relatively precise node localizations, which could be used as an additional recognition cue (though topography is not used at all in the recognition step of the current system) — previously, node localization had been rather imprecise. We have intro-

duced the potential to specialize the system to specific object types and to handle different poses with the help of object-adapted grids. These improvements and the introduction of the Face Bunch Graph makes it possible to extract from images sparse and largely variance-free structural descriptions in the form of image graphs, which reliably refer to fiducial points on the object. This accelerates recognition from large databases considerably since for each probe image data variance needs to be searched only once instead of in each attempted match to a gallery image, as was previously necessary. The ability of the new system to refer to object-fixed fiducial points irrespective of pose represents an advantage in itself and is essential for some interesting graph operations (cf. Section 4.3).

There is a considerable literature on face recognition, and many different techniques have been applied to the task (see Samal & Iyengar 1992; Valentin et al. 1994 for reviews).

Many systems are based on user defined face-specific features. Yuille (1991), for example, represents eyes by a circle within an almond-shape and defines an energy function to optimize a total of 9 model parameters for matching it to an image. Brunelli & Poggio (1993) similarly employ specific models for eyebrows, nose, mouth, etc. and derive 35 geometrical features such as eyebrow thickness, nose width, mouth width, and eleven radii describing the chin shape. The drawback of these systems is that the features as well as the procedures to extract them must be defined and programmed by the user for each object class again, and the system has no means to adapt to samples for which the features fail. For example, the eye models mentioned above may fail for faces with sun glasses or have problems if the eyes are closed, or chin radii cannot be extracted if the face is bearded. In these cases the user has to design new features and new algorithms to extract them. With this type of approach, the system can never be weaned from designer intervention. Our system, in contrast, can be taught exceptional cases, such as sun glasses or beards, or entirely new object classes, by the presentation of examples and incorporation into bunch graphs. The still necessary designer intervention to craft appropriate grids will soon be replaced by automatic methods, see below.

An approach to face recognition also avoiding user defined features is based on Principal Component Analysis (PCA) (Sirovich & Kirby 1987; Kirby & Sirovich 1990; Turk & Pentland 1991; O'Toole et al. 1993). In this approach, faces are first aligned with each other and then treated as high-dimensional pixel vectors from which eigenvectors, so-called eigenfaces, are computed, together with the corresponding eigenvalues. A probe face is decomposed with respect to these eigenvectors and is efficiently represented by a small number, say 30, of expansion coefficients. (The necessary image alignment can be done automatically within the PCA framework, see Turk & Pentland 1991). PCA is optimal with respect to data compression and seems to be successful for recognition purposes. In its original, simple form, PCA treats an entire face as one vector. It is then sensitive to partial occlusion or to object deformation. This problem has been dealt with by treating small image regions centered on fiducial points (eyes, nose, mouth) as additional pixel vectors from which to extract more features by PCA (Pentland et al. 1994). These fiducial regions can easily be located by matching subspaces of a few PCs to the image. The PCA approach in this form and our approach begin to look very similar: A graph of fiducial points, labeled with example-derived feature vectors (small sets of eigenvalues in the PCA approach, bunches of jets in our approach) is matched to the image, and the optimally matching "grid" is used to extract a structural description from the image. Recognition is then based on this.

A remaining difference between the approaches lies in the nature of the underlying feature types: statistically derived principal components, or "prejudice-derived" wavelets. It remains to be seen which of the approaches has the greater potential for development.

4.2 Performance Comparison

To obtain a meaningful performance comparison between different face recognition systems, the Army Research Laboratory has established a database of face images (ARPA/ARL FERET database) and compared our and several other systems in a blind test. Official results will be released soon (Phillips et al. 1996). Here we summarize results which other groups have reported for their systems tested on the FERET database. The recognition rates are given in Table 2.

Reference	Method	Model gallery	Probe images	First rank %
Gordon (1995)	normalized cross correlation on different regions in a face			
	manually located normalization points	202 fa + pl	202 fb + pr	96
	fully automatic system	194 fa + pl	194 fb + pr	72
		194 fa	194 fb	62
Gutta et al. (1995)	radial basis function network			
	fully automatic system	100 fa	100 fb	83
Moghaddam & Pentland (1994)	principal component analysis on the whole face			
	fully automatic system	150 fa	150 fb	99
		150 hr	150 hl	38
		150 pr	150 pl	32
Phillips & Vardi (1995)	trained matching pursuit filters for different regions in a face			
	manually located feature points	172 fa	172 fb	98
	fully automatic system	172 fa	172 fb	97
Wiskott et al. (1996)	Gabor wavelets, labeled graphs and elastic bunch graph matching			
	fully automatic system	250 fa	250 fb	98
		250 hr	181 hl	57
		250 pr	250 pl	84

Table 2: Methods, and performances of the different systems discussed. Our results are repeated for comparison. For some systems it is not reported whether fa or fb has been used for the model gallery; we consistently indicate the frontal model gallery by fa. When comparing the results, notice that the first rank recognition rates depend on gallery size. For lower recognition rates the rates decrease approximately as 1/(gallery size). Only Moghaddam & Pentland (1994) have reported results on half profiles and on profiles; none of the groups has reported results across different poses, such as half profile probe images against profile gallery.

Gordon (1995) has developed a system which automatically selects regions around left eye, right eye, nose, and mouth for frontal views and a region covering the profile for profile

views. The faces are then normalized for scale and rotation. The recognition is based on normalized cross correlation of these five regions as compared to reference models. Results are given on the fully automatic system, also for frontal views only, and on a system where the normalization points, i.e. pupil centers, nose and chin tip, are selected by hand. For the combined gallery (fa + pl) there is a great difference between the performance of the fully automatic system and that with manually located normalization points. That indicates that the automatic location of the normalization points is the main weakness of this system.

Gutta et al. (1995) have collected the images for the FERET database. They have tested the performance of a standard RBF (radial basis function) network and a system based on geometrical relationships between facial features, such as eyes, nose, mouth etc. The performance of the latter is very low and not summarized in Table 2.

Moghaddam & Pentland (1994) present results based on the PCA approach discussed in the previous section. A front-end system normalizes the faces with respect to translation, scale, lighting, contrast, as well as slight rotations in the image plane. The face images are then decomposed with respect to the first eigenvectors and the corresponding coefficients are used for face representation and comparison. The performance on frontal views, which are highly standardized, is high and comparable to that of our system, but the performance on half profiles and profiles is relatively low. That indicates that the global PCA-approach is more sensitive to variations such as rotation in depth or varying hairstyle.

Phillips & Vardi (1995) have trained two sets of matching pursuit filters for the tasks of face location and identification. The filters focus on different regions: the interior of the face, the eyes, and the nose for location; tip of the nose, bridge of the nose, left eye, right eye, and interior of the face for identification. The performance is high and comparable to that of our system. The small performance difference between the fully automatic system and the identification module indicates that the location module works reliably.

In the following section we discuss some methods which can potentially or have been shown to further improve the performance of our system. A detailed analysis of our basic system and some of these extensions will be given by Lyons et al. (1996).

4.3 Further Developments

The newly introduced features of the system open various possibilities to improve the system further. Especially the ability to reliably find surface points on an object can be useful, for instance when the issue is learning about local object properties (Shams & Spoelstra 1996). One such local property is differential degree of robustness against disturbances. Krüger et al. have developed a system for learning weights emphasizing the more discriminative nodes (Krüger 1995; Krüger et al. 1996). On model galleries of size 130–150 and probe images of different pose an average improvement of the first rank recognition rates of 6% was achieved, from 25% without to 31% with weights on average. Another individual treatment of the nodes has been developed by Maurer & von der Malsburg (1995). They applied linear jet transformations to compensate for the effect of rotation in depth. On a frontal pose gallery of 90 faces and half profile probe images an average improvement of the first rank recognition rate of 15% was achieved, from 36% without rotation to 50% and 53% with rotation, depending on witch pose was rotated. However, linear transformation is obviously not sufficient, and one may have to train and apply more general transformations.

By using phase information and the FBGs, matching accuracy has improved significantly. However, many partial mismatches still occur. This is probably due to the primitive way graph topography is handled, distortions being controlled by independent elastic stretching of node-to-node vectors. It would be of great help if these many degrees of freedom were replaced by just the few parameters necessary to describe typical global distortion patterns, due, for instance, to rotation in depth, variation in facial expression or hairstyle, or to the regular and symmetrical face shape variations from person to person or with age. These would have to be learned from examples, of course. Better deformation models would yield information on the type of disturbance, which would be of value in itself, and improve recognition. Some research in this direction has, for example, been done by LANITIS et al. (1995). When determined with sufficient reliability, graph topography can itself be used as a basis for recognition (cf. BRUNELLI & POGGIO 1993).

In (WISKOTT 1996) the bunch graph technique has been used to fairly reliably determine facial attributes from single images, such as gender or the presence of glasses or a beard. If this technique was developed to extract independent and stable personal attributes, such as age, skin color, hair type or again gender, recognition from large databases could be improved and speeded considerably by preselecting corresponding sectors of the database.

If the restriction to single probe and gallery images were relaxed and several images or a brief video sequence could be inspected, the system could be enriched with information on depth profile and on different poses. The merging of different poses into a coherent representation and the recognition of objects on this basis in a pose-independent way has been demonstrated in (VON DER MALSBURG & REISER 1995). Depth-profile information could be added to individual nodes as discriminatory feature to improve recognition directly and it could be used to predict poses not previously seen.

The manual definition of appropriate grid structures and semi-autonomous process of bunch graph acquisition will have to be replaced by an autonomous process. Automatic reduction of a FBG to its essential jets in the bunches has been demonstrated by KRÜGER et al. (1996). The creation of a new bunch graph is most easily based on image sequences, which contain many cues for grouping, segmentation, and detecting correspondences. The grouping of nodes for salient points on the basis of common motion has been demonstrated in (MANJUNATH et al. 1992). Monitoring a rotating object by continuously applying EBGM can then reveal which nodes refer to corresponding fiducial points in different poses (cf. VON DER MALSBURG & REISER 1995; MAURER & VON DER MALSBURG 1996). In a boot-strapping fashion, object-specific grids and object bunch graphs could be established by starting with a crude or general grid, such as utilized in (LADES et al. 1993), to first recognize object classes and establish approximate correspondences between images within a class, on the basis of which more object-specific grids and bunch graphs could be formed in stages, a little bit reminiscent of the way in which we find precise image graphs in two stages here.

Acknowledgements

We wish to thank Irving Biederman, Ladan Shams, Michael Lyons, and Thomas Maurer for very fruitful discussions and their help in evaluating the performance of the system. Thanks

goes to Thomas Maurer also for reviewing and optimizing the code. We acknowledge helpful comments on the manuscript by Jonathon Phillips. For the experiments we have used the FERET database of facial images collected under the ARPA/ARL FERET program.

References

Bruce, V., Valentine, T., and Baddeley, A. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, 1:109–120.

Brunelli, R. and Poggio, T. (1993). Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052.

Buhmann, J., Lange, J., von der Malsburg, C., Vorbrüggen, J. C., and Würtz, R. P. (1992). Object recognition with Gabor functions in the dynamic link architecture: Parallel implementation on a transputer network. In Kosko, B., editor, *Neural Networks for Signal Processing*, pages 121–159. Prentice Hall, Englewood Cliffs, NJ 07632.

Daugman, J. G. (1988). Complete discrete 2-D Gabor transform by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(7):1169–1179.

DeValois, R. L. and DeValois, K. K. (1988). *Spatial Vision*. Oxford Press.

Fleet, D. J. and Jepson, A. D. (1990). Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104.

Gordon, G. G. (1995). Face recognition from frontal and profile views. In Bichsel, M., editor, *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, IWAFGR*, pages 47–52, Zurich.

Gutta, S., Huang, J., Singh, D., Shah, I., Takacs, B., and Wechsler, H. (1995). Benchmark studies on face recognition. In Bichsel, M., editor, *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, IWAFGR*, pages 227–231, Zurich.

Jones, J. and Palmer, L. (1987). An evaluation of the two dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. of Neurophysiology*, 58:1233–1258.

Kalocsai, P., Biederman, I., and Cooper, E. E. (1994). To what extent can the recognition of unfamiliar faces be accounted for by a representation of the direct output of simple cells. In *Proceedings of the Association for Research in Vision and Ophtalmology, ARVO*, Sarasota, Florida.

Kirby, M. and Sirovich, L. (1990). Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108.

Krüger, N. (1995). Learning weights in discrimination functions using a priori constraints. In Sagerer, G., S.Posch, and Kummert, F., editors, *Mustererkennung 1995*, pages 110–117. Springer Verlag.

Krüger, N., Pötzsch, M., and von der Malsburg, C. (1995). Face finding with a learned representation based on labeled graphs. Submitted to *IEEE Transactions on Neural Networks*.

Krüger, N., Pötzsch, M., and von der Malsburg, C. (1996). Determination of face position and pose with a learned representation based on labeled graphs. Submitted to *IEEE Transactions on Neural Networks*.

Lades, M., Vorbrüggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. P., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311.

Lanitis, A., Taylor, C. J., and Cootes, T. F. (1995). An automatic face identification system using flexible appearance models. *Image and Vision Computing*, 13(5):393–401.

Lyons, M., Shams, L., Maurer, T., Wiskott, L., Fellous, J.-M., Krüger, N., and von der Malsburg, C. (1996). Performance of the eidos face recognition system in the FERET competition. In preparation.

Manjunath, B. S., Chellappa, R., and von der Malsburg, C. (1992). A feature based approach to face recognition. Technical Report CAR-TR-604 or CS-TR-2834, Computer Vision Laboratory, University of Maryland, Colledge Park, MD 20742–3411.

Maurer, T. and von der Malsburg, C. (1995). Linear feature transformations to recognize faces rotated in depth. In *Proceedings of the International Conference on Artificial Neural Networks, ICANN'95*, pages 353–358, Paris.

Maurer, T. and von der Malsburg, C. (1996). Tracking and learning graphs on image sequences of faces. Submitted to the *International Conference on Artificial Neural Networks ICANN'96*, Bochum.

Moghaddam, B. and Pentland, A. (1994). Face recognition using view-based and modular eigenspaces. In *Proc. SPIE Conference on Automatic Systems for the Identification and Inspection of Humans*, volume SPIE 2277, pages 12–21.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., and Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America A*, 10(3):405–411.

Pentland, A., Moghaddam, B., and Starner, T. (1994). View-based and modular eigenspaces for face recognition. In *IEEE Proc. Computer Vision and Pattern Recognition*, pages 84–91.

Phillips, P. J., Rauss, P., and Der, S. (1996). FERET (face recognition technology) recognition algorithm development and test report. Technical report, U.S. Army Research Laboratory.

Phillips, P. J. and Vardi, Y. (1995). Data driven methods in face recognition. In Bichsel, M., editor, *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, IWAFGR*, pages 65–70, Zurich.

Pollen, D. A. and Ronner, S. F. (1981). Phase relationship between adjacent simple cells in the visual cortex. *Science*, 212:1409–1411.

Pötzsch, M. (1994). Die Behandlung der Wavelet-Transformation von Bildern in der Nähe von Objektkanten. Technical Report IR-INI 94-04, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany. Diploma thesis.

Samal, A. and Iyengar, P. A. (1992). Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25(1):65–77.

Shams, L. and Spoelstra, J. (1996). Learning Gabor-based features for face detection. Submitted to *World Congress on Neural Networks '96*.

Sirovich, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519–524.

Theimer, W. M. and Mallot, H. A. (1994). Phase-based binocular vergence control and depth reconstruction using active vision. *CVGIP: Image Understanding*, 60(3):343–358.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

Valentin, D., Abdi, H., O'Toole, A. J., and Cottrell, G. W. (1994). Connectionist models of face processing: A survey. *Pattern Recognition*, 27(9):1209–1230.

von der Malsburg, C. and Reiser, K. (1995). Pose invariant object recognition in a neural system. In *Proceedings of the International Conference on Artificial Neural Networks ICANN'95*, pages 127–132, Paris. EC2 & Cie.

Wiskott, L. (1996). Phantom faces for face analysis. Technical Report IR-INI 96-06, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany. Submitted to *Pattern Recognition*.

Yuille, A. L. (1991). Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70.