Face Authentication with Gabor Information on Deformable Graphs

Benoît Duc, Stefan Fischer, *Associate Member, IEEE,* and Josef Bigün, *Senior Member, IEEE*

Abstract—Elastic graph matching has been proposed as a practical implementation of dynamic link matching, which is a neural network with dynamically evolving links between a reference model and an input image. Each node of the graph contains features that characterize the neighborhood of its location in the image. The elastic graph matching usually consists of two consecutive steps, namely a matching with a rigid grid, followed by a deformation of the grid, which is actually the elastic part. The deformation step is introduced in order to allow for some deformation, rotation, and scaling of the object to be matched. This method is applied here to the authentication of human faces where candidates claim an identity that is to be checked. The matching error as originally suggested is not powerful enough to provide satisfying results in this case. We introduce an automatic weighting of the nodes according to their significance. We also explore the significance of the elastic deformation for an application of face-based person authentication. We compare performance results obtained with and without the second matching step. Results show that the deformation step slightly increases the performance, but has lower influence than the weighting of the nodes. The best results are obtained with the combination of both aspects. The results provided by the proposed method compare favorably with two methods that require a prior geometric face normalization, namely the synergetic and eigenface approaches.

Index Terms— Biometrics, face authentication, face recognition, face verification, Fisher discriminants, Gabor filters, Gabor information.

I. INTRODUCTION

HUMAN identification systems based on biometrics other than the face have already led to commercial products with very high identification rates: the iris [1] and fingerprints [2] can be cited as examples. However, these systems are not always appreciated by users, as they require some close interaction with the machine often perceived as invasive. Moreover, they require the user to stop at the device and be cooperative, which is acceptable for access control to restricted areas, but not for other applications like surveillance. Face recognition may overcome some of these limitations.

B. Duc is with the Signal Processing Laboratory, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland.

S. Fischer is with FASTCOM Technology, S.A., Lausanne, Switzerland.

J. Bigün was with the Microprocessor and Interface Laboratory, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland. He is now with Halmstad University, S-30118 Halmstad, Sweden (e-mail: josef.bigun@cist.hh.se).

Recognition performed by the human being can be simultaneously seen as a holistic and a feature analysis approach [3]. Automatic face recognition often favors only one of these aspects. Features used for description of faces are either biometric features of the face, like distances between parts of the face like nose and mouth, or more abstract features, like filter responses on a grid [4]. Template-based methods that attempt to match well-defined portions of the face (eye, mouth) belong to the analysis category [5], [6]. The eigenface approach [7] describes images in terms of linear combinations of basis images, and thus represents a global holistic approach. Methods that constrain local features by adding geometrical constraints can be considered as a mixture of both aspects. One can cite here the *dynamic link architecture* (DLA) [4] and related graph-based feature matching approaches [8], as well as methods based on neural networks, and feature-based approaches where features are geometrical measures [5].

The mechanism for assessing connections between the image and model domain turns out to be complex and time-consuming. Therefore, a simplified implementation called *elastic graph matching* (EGM) is often preferred for finding objects in the scene with a known reference [9]. The elasticity of the matching provides some robustness to possible distortions of the object that may be due to a variation in pose, a scaling, or a deformation of the object. However, as the attributed graph is a two-dimensional (2-D) representation of three-dimensional (3-D) objects, this tolerance is limited. Other methods should be developed for large variations. Extensions have been proposed for rotations in facial depth [10].

One should note that these algorithms require a face segmentation preprocessing step, in order to extract faces in a complex environment, and to scale them coarsely. Yang and Huang [11] propose the use of mosaic images for detecting human faces in a complex background. A generalization to the low-resolution face template approach has been provided by Sung and Poggio [12], [13]. Finally, another face detection approach related to the eigenface approach is the distance-from-feature-space [14].

Recently, a relatively general method for attributing weights to features for a classification task has been proposed [15]. It is based on the assumption that a similarity measure should be high in case the test pattern is similar to the reference pattern, and low if they are different. The weighting is defined by a nonlinear function J that depends on a small set of parameters and on the training set. This function is the same for all nodes. These parameters are determined on a training

set by maximizing an evaluation function using the simplex method. Although quite general in the theory, the application of the method requires some fine tuning and some *a priori* choices. Furthermore, the optimal settings and the particular choices seem to have been obtained by trials and testing.

Here we are interested in face authentication, rather than face recognition. In authentication applications, the system is provided with a supposed identity of the candidate, which is known to the system. The goal is to state whether the identity is correct or not. Authentication is seldom encountered in the literature, as most of the time recognition problems are addressed. In this article, we propose to specialize a recognition method, the EGM approach, for authentication applications. For this reason, we propose local discriminant measures for face images, which lead to significantly improved performance for face authentication applications. In our contribution, subsets of the data are considered separately. While the method we tested is oriented to the authentication of people, it could also be used for recognition. The method leads to a faster training than in [15], as the solution to the optimization problem is known analytically.

In this presentation we also study the contribution of the deformation capability of the grid by comparing the performance of rigid and nonrigid graph matching in case of face authentication. To the best of our knowledge, such a study has not been undertaken before. We will compare the influence of grid deformability with respect to local discrimination. Both methods are independent and may be combined. Finally, the EGM combined with *local discriminants* (EGM-LD) is compared with two other authentication methods that use a prior geometric normalization of faces: principal component analysis and synergetic computers [16].

The paper is organized as follows. The elastic graph matching is described in Section II. The local discriminant measure is introduced in Section III. Alternative methods are described in Section IV. Experimental results are given in Section V. Finally, conclusions are drawn and future developments are discussed in Section VI.

II. Elastic Graph Matching for Face Authentication

Attributed graphs describe objects on sparse locations, by attaching to each node a feature vector that contains information on the local neighborhood of the node location. Here, we use the modulus of complex Gabor responses as features from filters with six orientations and three resolutions.

A. Gabor Filters

Each face is represented by a set of feature vectors positioned on nodes of a coarse, rectangular grid placed on the image. Comparing two face images is accomplished by matching and adapting a grid taken from one image to the features of the other image [4]. We use the modulus of complex Gabor responses as feature vectors.

Gabor decompositions have been proposed as an analysis tool for textures [17] motivated by psychophysics of the human visual system [18]. The Gabor decomposition we use can be
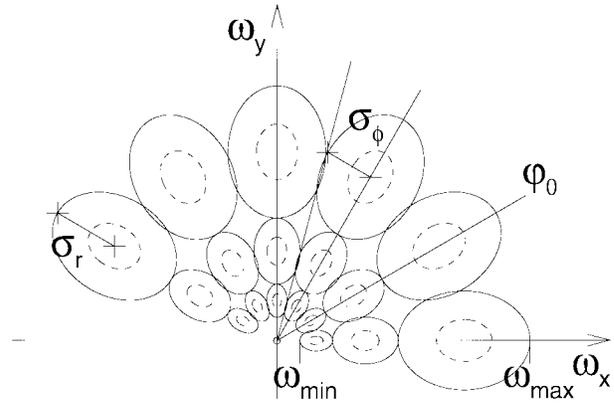
Fig. 1. Gabor decomposition of the Fourier domain. The lines show the inflexion points of the 2-D Gaussian-shaped filters. The dashed lines show an alternative decomposition scheme with a minimal amount of overlapping of the frequency bands.

considered as a *directional microscope*, with an orientation and scaling sensitivity.

The spectral plane is partitioned into m frequency and n orientation bands:

$$\hat{g}_{ij}(\boldsymbol{\omega}) = \exp(-\tfrac{1}{2}(\boldsymbol{\omega} - \boldsymbol{\omega}_{ij})^t R_j \Sigma_{ij}^{-1} R_j^t (\boldsymbol{\omega} - \boldsymbol{\omega}_{ij})) \quad (1)$$

where $1 \leq i \leq m$ and $1 \leq j \leq n$, R_j is a rotation matrix, and Σ_{ij} is a diagonal matrix. $A_{ij} = R\Sigma_{ij}^{-1}R^t$ is a positive-definite matrix whose principal axes define an ellipse oriented according to the orientation defined by j, see Fig. 1. More precisely

$$R_j = \begin{pmatrix} \cos(\phi_j) & \sin(\phi_j) \\ -\sin(\phi_j) & \cos(\phi_j) \end{pmatrix}$$

$$\Sigma_{ij} = \begin{pmatrix} \sigma_{r_i}^2 & 0 \\ 0 & \sigma_{\phi_j}^2 \end{pmatrix}$$

$$\boldsymbol{\omega}_{ij} = R_j \begin{pmatrix} \omega_{r_i} \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\phi_j)\omega_{r_i} \\ \sin(\phi_j)\omega_{r_i} \end{pmatrix} \quad (2)$$

where $\boldsymbol{\omega}_{ij}$ is the central frequency of the filter. The use of an asymmetric real transfer function implies that the corresponding point spread functions are complex.

This choice of Gabor decomposition makes the computation of the corresponding point spread function possible analytically. Alternatively, other decompositions using log-polar mappings could be used [19]. The impulse response g_{ij} corresponding to $\hat{g}_{ij}$ is given by:

$$g_{ij}(\boldsymbol{x}) = \exp(-\tfrac{1}{2}\boldsymbol{x}^t R_j \Sigma_{ij} R_j^t \boldsymbol{x}) \exp(i2\pi\boldsymbol{\omega}_{ij}^t \boldsymbol{x}). \quad (3)$$

The parameters are not completely determined yet. For a decomposition consisting of n orientations and m resolutions, the transfer functions are chosen here so that neighboring filters intersect at equal values $1/\sqrt{e}$ along their principal axes. In this we follow the choice in [19]. The n orientations are chosen equidistant, which results in a constant value σ_ϕ for all σ_{ϕ_j}:

$$\sigma_\phi = \sigma_{\phi_j} = \frac{\pi}{2n}. \quad (4)$$
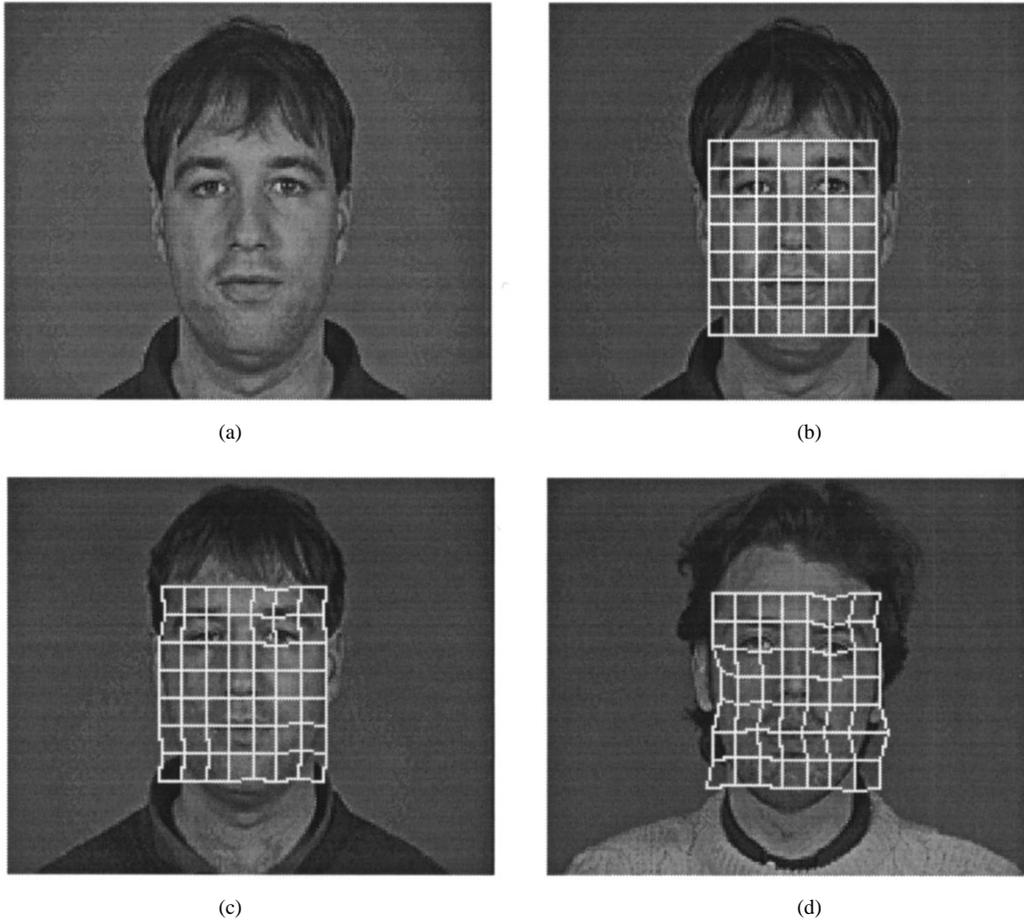
Fig. 2. Example of a grid matching. (a) Reference image. (b) Reference grid. (c) Matched grid on another image of the same person (residual matching error: 740). (d) Matched grid on another person (residual matching error: 1406).

The angular bandwidth of the filters is π/n, and the orientation centers are given by

$$\omega_{\phi_j} = 2\sigma_\phi(j - 1). \tag{5}$$

The radial frequency bands are distributed in octave steps, with a frequency bandwidth which doubles at each step, and cover a range from $\omega_{r_{\min}} > 0$ to $\omega_{r_{\max}} < 1/2$, in normalized frequencies where one represents the Nyquist frequency. By defining an intermediate variable

$$\sigma_0 = \frac{\omega_{r_{\max}} - \omega_{r_{\min}}}{2(2^m - 1)} \tag{6}$$

the radial centers and bandwidths are given by

$$\begin{aligned} \omega_{r_i} &= \omega_{r_{\min}} + \sigma_0(1 + 3(2^{i-1} - 1)) \\ \sigma_{r_i} &= \sigma_0 2^{i-1} \end{aligned} \tag{7}$$

In authentication applications a typical parameter set is the following: $\omega_{r_{\min}} = \frac{1}{16}, \omega_{r_{\max}} = \frac{1}{4}$, $m = 3$ and $n = 6$. Fig. 1 shows a typical decomposition.

Finally, one should note that Gabor functions are in many respects similar to 2-D Morlet wavelets [20].

B. Elastic Graph Matching

Elastic graph matching (EGM) consists in locating an attributed graph on the image that is as close as possible to the reference graph. The distance between two graphs is evaluated by a dissimilarity function, that considers both the feature vectors of each node and the deformation information attached to the edges. We consider dissimilarity measures where the contribution from nodes and edges are independent, more precisely

$$d(G, R) = \sum_{i=1}^{N_n} d_n(G_{n_i}, R_{n_i}) + \lambda \sum_{j=1}^{N_e} d_e(G_{e_j}, R_{e_j}) \tag{8}$$

where G_{n_i} represents the ith node of grid G, R_{e_j} is the jth node of grid R; N_n, N_e are the number of nodes and edges, respectively, and λ is a weighting factor that characterizes the stiffness of the graph. A *plastic* graph which opposes no reaction to deformation corresponds to $\lambda = 0$, while a totally rigid graph is obtained with very large values of λ.

The matching procedure consists of two consecutive steps [4]. The first step is used for obtaining a first match, by moving a rigid grid over a search area in the image. Starting from this initial guess, the grid is deformed in order to minimize (8).

The deformation is achieved by displacing each node around its current location, and by placing it where the minimum value of $d(G, R)$ is obtained. This operation is applied on each node successively, and the whole process is repeated until no further decrease is obtained. An example of grid matching is shown in Fig. 2.
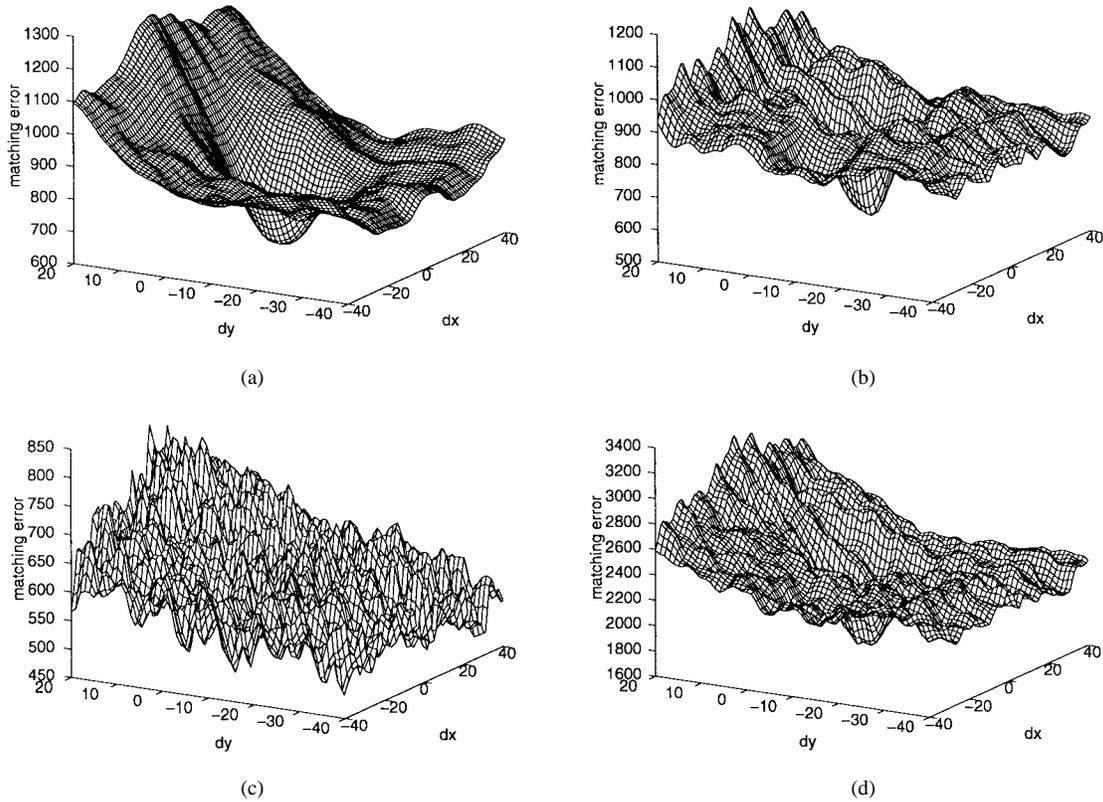
Fig. 3. Objective function for the rigid translation of the graph over the search window. Fig. 2(b) is used as reference grid, and Fig. 2(d) as test image. Three resolutions of Gabor responses are used. Here, the contribution of each resolution is shown separately. (a) Objective function with only the lowest resolution. (b) Objective function with only the medium resolution. (c) Objective function with only the highest resolution. (d) Total objective function. While low resolutions provide smooth, convex objective functions, high resolution responses provide sharper minima, and are used for refinement.

C. Coarse-to-Fine Rigid Graph Matching

In our implementation, the rigid matching is obtained through full search matching on a predefined rectangular search window. If responses are needed at all points in the image, the optimal way of computing the responses consists in computing the 2-D fast Fourier transform (FFT) of the input image, and multiplying it with the frequency response of the filter. However, for a matching problem, the Gabor responses may be needed only on a small fraction of points in the image. If the number of points is small enough, direct convolution is less expensive.

By adopting a coarse-to-fine strategy, the number of required Gabor responses is significantly reduced. Multigrid relaxation methods have been proposed in applied numerical analysis [21]. They are particularly well-suited to the image analysis field [22].

We consider a multiresolution description of the image. First, the lowest resolution image is considered for matching. As a consequence, the objective function is smoothed, and the matching may be undertaken on a subsampled lattice. This property is intimately related to the fundamental sampling theorem: as the objective function has been lowpass filtered, it may be sampled at a coarser step without loss of information. Fig. 3 shows that the low-resolution image provides a smooth objective function, while the high frequency information generates a forest of local minima. However, the minima are more precisely localized when the high frequencies are incorporated.

Consecutive refinements are obtained by incorporating higher resolution information and by searching on a finer grid around the current estimate.

Coarse-to-fine strategies may get trapped in local minima [23]. A remedy to this weakness consists in the elaboration of mixed fine-to-coarse and coarse-to-fine strategies. However, we noticed that if only one head is present and occupies a significant part of the image, this problem does not occur. This assumption is satisfied when a prior face detection algorithm that extracts regions of interest is applied. In our experiments, the images may be considered as the output of such a segmentation, therefore a coarse-to-fine strategy has been retained.

In practice, a Gaussian pyramid is built [24]. The highest resolution level of the pyramid contains the original image, and the next levels contain lowpass versions of the original image, with a cut-off frequency decreasing as $\left(\frac{1}{2}\right)^n$, where n is the level of the pyramid. The lowest level is zero and has a normalized maximal frequency of $\frac{1}{2}$. In the pyramidal implementation adopted here, the image size depends on the resolution. The pyramid is built recursively, by building a new, lower-resolution level from the previous one by lowpass filtering and subsampling with a factor two, so that the size of the image is divided by two at each iteration. The lowpass filtering was achieved with separable Gaussian filters.

The definition of filters is simplified by using the same set of filters for a single resolution and a complete set of

orientations. These filters are applied to each level of the pyramid to obtain a complete set of resolutions. A significant reduction in the amount of computations is obtained for low frequency responses, compared to filtering the original image, as bandpass filters selecting low frequencies have a large support.

Finally, the grid matching serves two purposes: firstly, it aims at normalizing the input, in order to make the subsequent comparison invariant with respect to translation and a reasonable amount of deformation. Secondly, the residual error accounts for the difference between the normalized input and the reference pattern. Intuitively, the higher the error, the higher the probability of having an impostor.

III. FEATURE EXTRACTION

The first step of the authentication process consists in matching the image with the prototype grid of the claimed class (in the following, each person in the database is considered as a *class* of the classification problem). This prototype is taken as the mean of the feature vectors provided by all images of the considered person in the training set. It is expected that if the claimed identity is correct, the feature vector will be close to the class prototype; in case of an impostor, the matching will be poor. Unfortunately, early experiments showed that the *residual matching error* (RME), i.e., $d(G, R)$ after matching with $\lambda = 0$, is not sufficient to discriminate between an impostor and the authentic person, see Section V-B. This is partly due to the presence of noise in the measurement, but also due to the fact that not all nodes are discriminative. Indeed, the feature space considered here is very large: for an 8×8 grid comprising 18 Gabor responses at each node, a total of $N_G = 1152$ features is obtained.

Reducing the dimensionality is an efficient way to reduce the influence of noise [25], [26]. From a training set consisting of several frontal views of each person, one establishes subspaces which maximize the dispersion of all classes while minimizing the dispersion within the classes.

However, the number of training samples is small compared to the number of features. Also, the features on two graph nodes may be considered as independent. Therefore, it is reasonable to address dimensionality reduction independently at each node of the graph. If features are considered locally, the number of training samples is larger than the dimension of the feature space, which allows us to apply standard feature reduction methods.

A. Local Discriminants

Suppose that the dimensionality of the considered feature space is small compared to the number of training elements in each of the c considered classes. One would like to establish a decision criterion for the acceptance or rejection of the candidate. This criterion should be "small" if the candidate is the right person, and "large" in case of an impostor. Obviously, this decision has to be made on the difference between the prototype of the claimed class and the measured feature vector. The components of this difference do not have the same significance, as some may be more relevant than others for the

given class. Therefore, we propose the following discriminant criterion:

$$d_k(\boldsymbol{r}) = \left(\sum_{i=1}^{N_g} v_{k_i}(r_i - \mu_{k_i}) \right)^2 = (\boldsymbol{v}_k^t(\boldsymbol{r} - \boldsymbol{\mu}_k))^2 \quad (9)$$

for class k, $k = 1 \cdots c$, where r_i are the components of the measurement vector $\boldsymbol{r}$, N_g is the dimension of the local feature space, and $\boldsymbol{\mu}_k$ is the mean of vectors $\boldsymbol{r}$. The unknown coefficient vector $\boldsymbol{v}_k$'s are determined on the training set by minimizing the ratio:

$$
\begin{aligned}
D_k &= \frac{\displaystyle\sum_{\boldsymbol{r} \in S_k} d_k(\boldsymbol{r})}{\displaystyle\sum_{\boldsymbol{r} \in (S - S_k)} d_k(\boldsymbol{r})} \\
&= \frac{\displaystyle\sum_{\boldsymbol{r} \in S_k} \boldsymbol{v}_k^t(\boldsymbol{r} - \boldsymbol{\mu}_k)(\boldsymbol{r} - \boldsymbol{\mu}_k)^t \boldsymbol{v}_k}{\displaystyle\sum_{\boldsymbol{r} \in (S - S_k)} \boldsymbol{v}_k^t(\boldsymbol{r} - \boldsymbol{\mu}_k)(\boldsymbol{r} - \boldsymbol{\mu}_k)^t \boldsymbol{v}_k} \\
&= \frac{\boldsymbol{v}_k^t W_k \boldsymbol{v}_k}{\boldsymbol{v}_k^t B_k \boldsymbol{v}_k}, \quad (10)
\end{aligned}
$$

where S_k is the set of training vectors belonging to class k, and S is the whole training set, so that $(S - S_k)$ is the set of all impostors for class k. Here, $\boldsymbol{\mu}_k$ is the mean on S_k. By this, we are back to a two-class classification problem, where the classes are S_k and $(S - S_k)$. This formulation leads to a generalized eigenvalue problem: $W_k \boldsymbol{v}_k = \lambda B_k \boldsymbol{v}_k$, and $\boldsymbol{v}_k$ is given by the eigenvector corresponding to the smallest generalized eigenvalue. This is known as *Fisher's discriminant ratio* [25].

All local responses have to be combined in order to provide a unique, global dissimilarity measure for the considered face. This is a problem related to sensor or decision fusion [27]. Here, we build the global response by simply adding the contributions from the local discriminants. This discriminant measure will be abbreviated "LD."

B. Separation Parameters

It is necessary to choose a threshold for defining acceptance/rejection intervals in the domain of possible responses from training data. Here we assume that the system will provide a soft decision between $[0, \infty[$, therefore a mapping between the original response interval and the $[0,1]$ interval is needed.

An invertible mapping from $[0, \infty[$ to $[0, 1]$ is provided by the $\tanh$ function. For our purpose, the soft score $S \in [0, 1]$ should be one for an identity claim acceptance, and zero for an identity claim rejection, whereas the global discriminant value tends to zero for a perfect match and to infinity for a perfect mismatch. We suggest the mapping

$$S(x) = \tanh\left(\frac{\log(3)}{2x} t \right) \quad (11)$$

where t is an empirically chosen constant. By definition t is called the *separation parameter* (SP), as it acts like a

decision point on x between acceptance and rejection intervals ($S(t) = 0.5$). In the case of a soft decision, SP acts as a parameter selecting the mapping function. In the case of a hard decision, SP can be viewed as a possible threshold. The threshold itself may be chosen in a number of ways.

The first possibility of separation parameter selection consists in choosing the minimal dissimilarity measure among the training impostors for t. This quite natural choice leads to more false rejections than false acceptances. Another possibility consists in choosing the maximal dissimilarity measure among training data of the correct person. Statistical tests provide a third possibility. Under the assumption that all Gabor responses g_i after matching are independent Gaussian random variables centred at the corresponding references r_i, it is possible to design a statistical test [28]. Indeed, if the Gabor responses are considered independently, the quantity $\Sigma_{i=1}^{N_G}\left((g_i - r_i/\sigma_i)^2\right.$ follows a $\chi_{N_G}^2$ distribution. Therefore, the u percentile $\chi_u^2(N_G)$ provides the SP for the confidence level u. For high values of N_G ($N_G > 50$), the following formula may be applied for estimating $\chi_u^2(N_G)$ [28]:

$$\chi_u^2(N_G) \simeq \tfrac{1}{2}(z_u + \sqrt{2N_G - 1}) \qquad (12)$$

where z_u is the u percentile for the normal distribution. For 1152 Gabor responses, the 0.95 percentile amounts to $\chi_{0.95}^2(1152) = 1232$.

The situation is somewhat different when the local discriminants are applied, because we do not sum the squares of the residuals ($g_i - r_i$), but the squares of their linear combinations

$$d_{\mathrm{disc}_k}^2(l) = \sum_{i=1}^{N_d}\left(\sum_{j=1}^{N_g} v_{k_{ij}}(g_{ij} - r_{ij})\right)^2$$

where N_d is the number of local discriminants and $N_g = N_G/N_d$ is the dimension of local feature spaces. If ($g_{ij} - r_{ij}) \sim N(0, \sigma_{ij})$, one can take into account the fact that a linear combination of Gaussian random variables is also Gaussian. More precisely, if $x_i \sim N(\mu_i, \sigma_i)$, then a linear combination given by $y = \Sigma_{i=1}^n a_i x_i$ is such that $y \sim N(\Sigma_{i=1}^n a_i\mu_i, \sqrt{\Sigma_{i=1}^n a_i^2\sigma_i^2})$. Taking into account this result, one obtains that each local discriminant i is $N(\Sigma_{j=1}^{N_g} v_{k_{ij}}r_{ij}, \sqrt{\Sigma_{j=1}^{N_g} v_{k_{ij}}^2\sigma_{ij}^2})$.

In the case of the residual matching error as well as for local discriminants, the mean and variances of the Gaussian variables are estimated on the training set.

IV. ALTERNATIVE APPROACHES

The proposed algorithm has been compared with two other approaches that require a geometric normalization of images as a preprocessing step: the eigenface and a simplified synergetic computer. These methods require a prior image normalization because the image content is scanned and put into a 1-D structure, which is seen as an element of a vector space. The alignment of data is necessary if one wants to keep some spatial coherence between vectors after the 1-D "projection" of the image. Both approaches describe subspaces generated by the training images by means of an appropriate basis.

A. Principal Component Analysis

The eigenface approach [7], [29] is based on the *principal component analysis* (PCA), or Karhunen-Loève transform (KLT). It takes into account the statistical dependencies between the pixel values of images in a training set to design an orthogonal basis optimal for representing images. For this purpose, the eigenvectors of the covariance matrix of the training set of face images are computed; they are called *eigenfaces*. Only the most significant eigenfaces, i.e., the ones corresponding to the largest eigenvalues of the covariance matrix, are retained for describing the images. Each training and test face is then characterized by its projection on the eigenfaces, and the comparison of two faces is achieved by comparing two sets of projections.

B. Adjoint Vectors and Synergetic Computers

The eigenface approach is optimal for a compact description of face images in the least square sense: it finds an orthonormal basis that defines a subspace for which the projection error is as small as possible, on average and in a least square sense.

Another type of approach also considers the prototype faces as basis vectors generating a subspace. Here, the basis vectors $\boldsymbol{v}_u$ are not orthogonal. Approximating new input images as a linear combination of prototype faces requires the computation of a set of *adjoint vectors* $\boldsymbol{v}_{u'}^+$, which satisfy an orthogonality relationship with the $\boldsymbol{v}_u$'s:

$$\boldsymbol{v}_{u'}^+ \cdot \boldsymbol{v}_u = \delta_{uu'}. \qquad (13)$$

By constraining the feature vectors to be in the subspace defined by the prototypes $\{\boldsymbol{v}_u, u = 1 \cdots c\}$, the problem becomes well-posed, provided that all $\boldsymbol{v}_u$ are linearly independent, which is assumed to be the case since they are prototypes representing distinct persons. By adopting the matrix notation, we denote by V the matrix containing vector $\boldsymbol{v}_u$ in the uth column, and by V^+ the matrix of adjoint vectors, written in rows. By definition of adjoint vectors, one has

$$V^+V = I. \qquad (14)$$

Expressing adjoint vectors as linear combinations of prototype vectors gives

$$V^+ = AV^t. \qquad (15)$$

Introducing (15) in (14) yields

$$V^+ = (V^tV)^{-1}V^t. \qquad (16)$$

Relationship (14) yields an optimal separation of classes (persons), instead of a compact description, as obtained with the PCA. An input vector $\boldsymbol{q}$ may be approximated by a linear combination of the reference vectors

$$\boldsymbol{q} \approx \sum_u \xi_u\boldsymbol{v}_u. \qquad (17)$$

If the input image belongs to person u, then it is expected that ξ_u is close to one while the remaining $\xi_{u'}, u' \neq u$ are close to 0.

This approach may be viewed as a particular case of synergetic computers [16], [30]. Synergetic computers model the
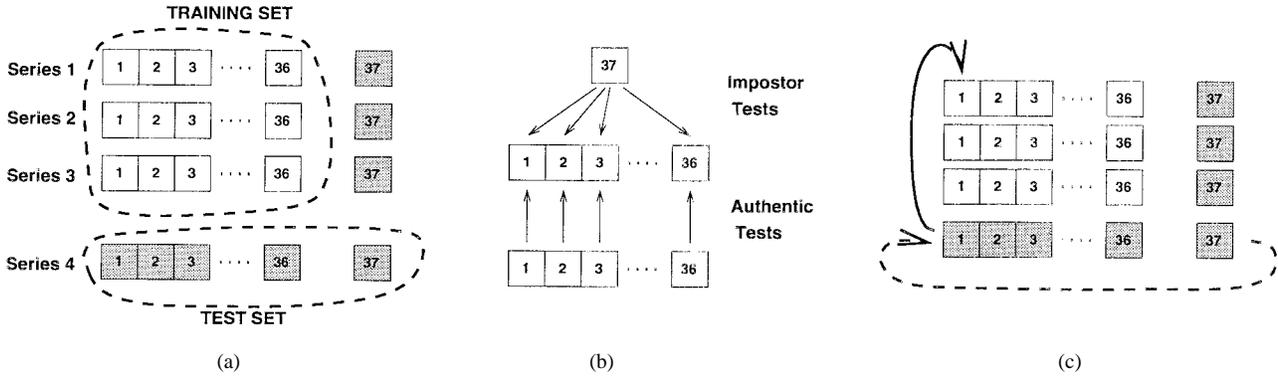
Fig. 4. Experimental protocol. (a) The database is divided into a training set of three shots of 36 persons and a test set consisting of all 37 persons. This defines a configuration. (b) For authentic tests, the 36 persons in the training set access under their identity; for impostor tests, the left-out person (here person 37) attempts to access under the 36 other identities. Each configuration brings a total of 36 authentic accesses and 36 impostor accesses. (c) Permutation of shots and left-out persons for defining a new configuration.

process of pattern recognition by a dynamic process expressed by a differential equation. The most general model allows us to deform and move patterns. In case no such transformation is allowed, the dynamic process can be dropped, as the order parameter which is closer to one in the initial computation always converges to one.

The adjoint vectors are used here for authentication. The training consists in building V^+ from a reference view of each person by using (16). A separation parameter is computed from the training set as explained in Section III-B.

V. EXPERIMENTS

A. Face Database and Experimental Setup

Large image databases for authentication purposes are much harder to collect than those for recognition purposes. In order to account for intraperson appearance variability due to time, many images of the same person must be recorded. We are aware of many recognition results in the literature based on databases consisting of hundreds of persons having at most two or three frontal *images*. Evidently, such databases are of less value for establishing *authentication* results on clients than a smaller database that offers larger intraperson variability. Therefore, a database with a small number of people but much information for each person was prefered to databases with many people and few data available per person.

This work is part of a project developing authentication methods based on several modalities, such as speech, frontal, and profile views of the face. The use of several modalities required the acquisition of a multimodal database, which contains both sound and image information [31]. It includes four shots of 37 individuals, which were taken at one week intervals. For each shot, people were asked to rotate their head from $0°$ to $-90°$, again to $0°$, then to $+90°$ and back to $0°$ degrees. Currently this effort is being continued to increase the size to 300 persons.

For image-based authentication, the rotation sequences were considered by using the luminance information of images in QCIF format (144×176). Frontal images are selected automatically using a symmetry measure. The number of selected images varies for each person. We select images

that have a symmetric gray-level distribution in a rectangular region. This region has the shape of a horizontal bar that covers the face region and the background on the two sides of the face. We compute the horizontal center of gravity and the symmetry with respect to this center and select images with an extremal symmetry measure. However, the use of this method has been semiautomatic as it fails sometimes. As the frontal view extraction is not the main focus of this contribution, we examined the output of the detection algorithm and discarded the false detections. In total, 551 frontal face images are used, yielding approximately 15 frontal views per person taken at four different weeks.

The experiments were conducted following a combination of the left-one-out and the rotation estimates, which is a variant of the jackknife method [25], [32]. This protocol, which also supports multimodal authentication, has been adopted by all partners of the project for multimodal authentication, so that results from all modalities can be compared. Alternatively, each person is labeled as an *impostor*, while the other 36 are considered as *clients*; see Fig. 4. For each combination, three shots of the 36 clients build the training set while the fourth shot is used as an evaluation set in the following way: each client attempts to access under its own identity, and the impostor attempts to access under the identity of the 36 clients. This sums up to 36 authentic tests and 36 impostor tests. This procedure is repeated four times, by successively considering each shot for evaluation. In total, the client and impostor verifications amount each $37 \times 4 \times (37 - 1) = 5328$.

The local discriminants are estimated from the training set consisting of 36 people. For this purpose, a reference view is selected for each person among all available images at random. Authentic accesses are obtained by simulating accesses with all available images of a person to his reference. Accesses imposting a given person are simulated by considering the reference views of each of the remaining 35 persons in the training set.

The separation parameter is chosen in several ways as described in Section III-B, and the discriminant measures are mapped to the [0,1] interval. For obtaining a hard decision, a threshold is necessary. While its default value is 0.5, varying the threshold between zero and one allows us to vary the *false*
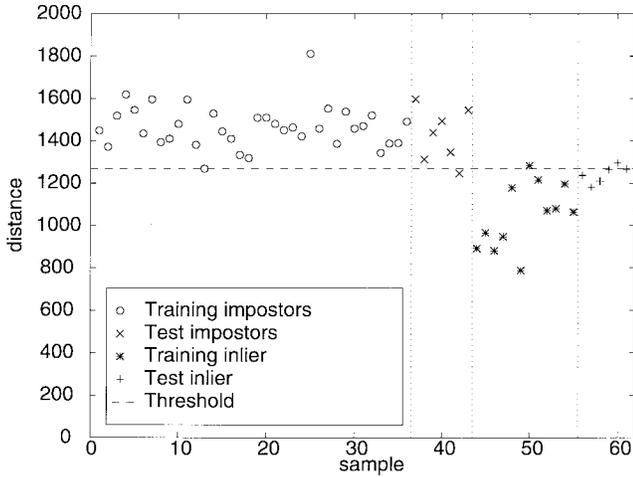
Fig. 5. Plot of distances for person (or class) 15. The distance of the grids of different kinds of images, namely impostors in the training and the test set, members of the class in the training and test set, are shown. If one uses the minimal distance on the training impostors as a threshold for the decision, some members of the class in the training and test set are misclassified. This figure is given as an illustration and does not follow the experimental protocol described in Section V-A.

Fig. 6. Plot of discriminant values for person (or class) 15. The image indices correspond to the ones of Fig. 5. If one uses the minimal discrimination measure over the training impostors as a threshold for the decision, then all points are correctly classified. Members of the class and impostors are better separated than in Fig. 5.

acceptance (FA) and *false rejection* (FR) rates, the decrease of one of them being balanced by the increase of the other. By varying the threshold continuously, one may draw curves of FA as a function of FR, which are denoted as receiver operating characteristic (ROC) curves in the following.

B. Elastic Graph Matching and Feature Reduction

In order to justify the process of feature extraction, we first want to show that the Euclidean distance between features, i.e., the residual matching error $d(G, R)$ with $\lambda = 0$, is not sufficient for a reliable decision. Fig. 5 shows distances of training and test samples with person 15 used as reference. It turns out that the distance to the reference view is clearly not sufficient to detect impostors.

A representation of discriminant values for the same person is shown in Fig. 6. Now the discrimination of impostors is much more powerful. One can notice that there seems to be some overtraining, as the discrimination measure is almost zero for all members of the considered class in the training set, and significantly larger for images of the same class in the test set, while remaining smaller than the threshold. This is due to the small number of training samples for each person in the database.

At that point, the discriminant values in the $[0, \infty[$ interval are normalized to the $[0,1]$ interval, so that they can be combined with or compared to other verification modalities like speech [33]. Nevertheless, using a hard threshold is useful for comparing performance of different alternatives. As an illustration of the usefulness of the discriminant measure over all classes, we show the ROC for the RME and the local discriminants (LD) in Fig. 7. Such curves reflect the performance of a given solution averaged over *all* classes. The points on the ROC were obtained by scaling the minimum threshold displayed in Figs. 5 and 6 with a varying factor. For safety critical applications, a low value of FA is desired. The
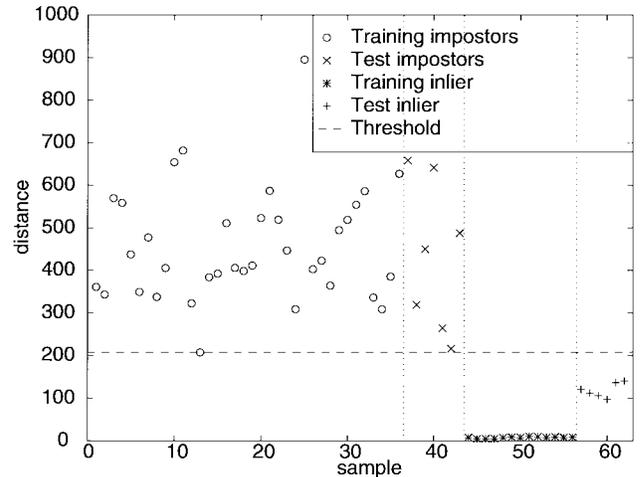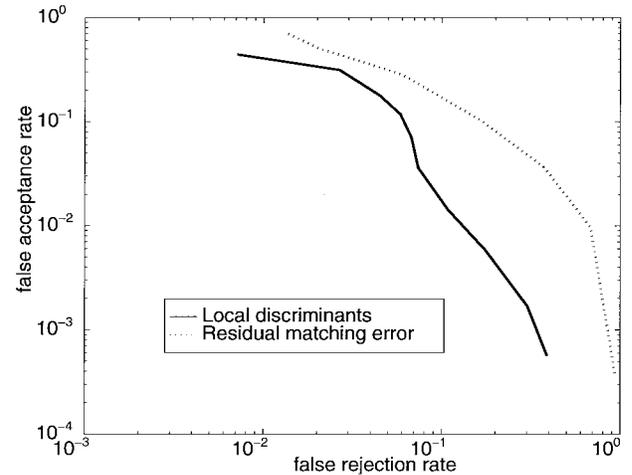
Fig. 7. Experimental ROC curve for the residual matching error and the local discriminants in a log-log scale. Results were obtained with $\lambda = 2$.

best method is the one which provides the lowest FR for the imposed value of FA. For other applications, one might want the smallest FA for a given FR. Fig. 7 shows that the LD outperforms the RME everywhere.

Another interesting point is that it is not necessary to take into account all local discriminants in order to obtain good performance. By retaining only a fraction of the discriminants that reach the largest values of the following criterion:

$$C_k = \sum_{\boldsymbol{r} \in (S - S_k)} d_k(\boldsymbol{r}) \qquad (18)$$

results are almost as good as with all discriminants taken into account. D_k, defined in (10), is not chosen to order LD's according to their relevance because it is not suitable for comparing information in different regions. Indeed, it has been observed that in homogeneous regions, much lower values of D_k may be reached than in textured ones, however, homogeneous regions are not significant for authentication.
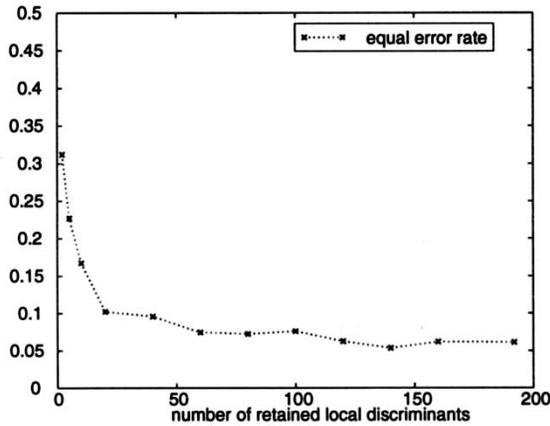
(a)
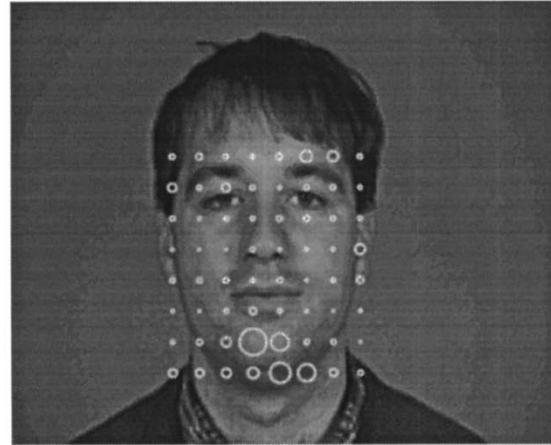
(b)
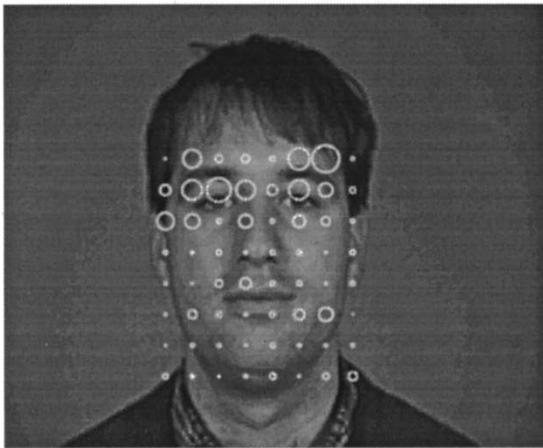
(c)
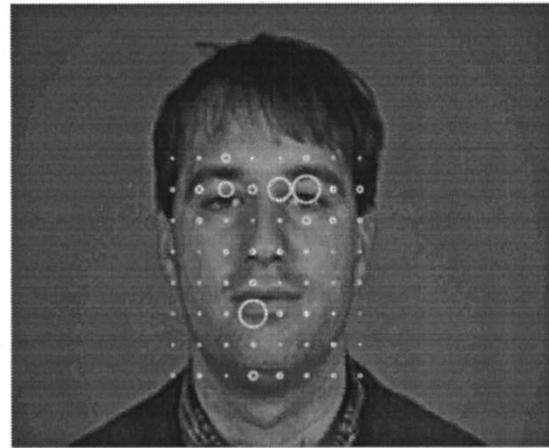
(d)

Fig. 8. (a) Equal error rate as a function of the number of most significant LD's retained. A local minimum is obtained when 140 LD's are kept. The significance of the nodes is displayed in (b)–(d), for the three different resolutions. The radiuses of the circles located at each node are proportional to the significance of the nodes, i.e., to C_k as defined in (18).

When considering the *equal error rate* (EER) as a criterion, a local minimum is obtained when 140 LD's are kept [see Fig. 8(a)]. Fig. 8(b)–(d) illustrate the significance of each node at each of the three resolutions. It is intuitively correct that the weights are higher around the eyes at high frequencies while they are larger in the chin and cheek areas at lower frequencies. For frontal face authentication, the tip of the nose region is not significant.

C. Separation Parameter Comparison

All results shown before were obtained by hard decision using a threshold, the minimum of training impostors; see Section III-B. Fig. 9 shows a comparison of results obtained with three types of SP's. It turns out that the difference is not significant when the RME is used. On the contrary, if the LD's are used as dissimilarity measure, the minimum of training impostors provides the best results, followed by the statistical test and the maximum of training authentic accesses. This can be explained by comparing Figs. 5 and 6. When the local discriminants are used (Fig. 6), and due to the relatively small number of authentic accesses in the training phase as

compared to the impostor accesses, the dissimilarity measure tends to zero for all authentic accesses: we are at the limit of overtraining. In this situation, it is difficult to estimate reliably the variance required for normalization in the statistical test, and the maximum of authentic accesses also loses part of its significance, as it always tends to zero. On the contrary, the minimum of training impostors keeps a finite value. Therefore, the minimum of training impostors is chosen as the standard separation parameter and thresholding in the following.

D. Evaluation of Elasticity Significance

In order to assess the effectiveness of grid elasticity, we compare an elastic and a nonelastic graph matching procedure. The nonelastic graph matching is obtained by dropping the second step of the matching procedure described in Section II-B, which is equivalent to choosing a very large λ in (8). A completely "plastic" grid is obtained with $\lambda = 0$: as the second term vanishes, each grid node is free to move everywhere in the image. By running the simulations according to the experimental protocol of Section V-A with several values of λ, it is possible to assess the usefulness of the elastic step, and
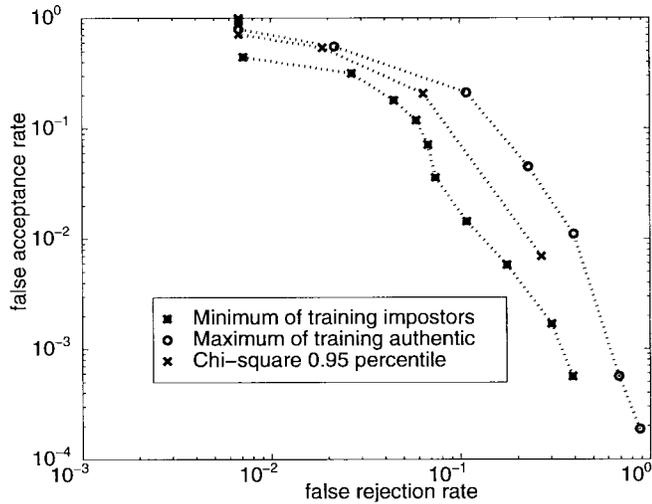
(a)
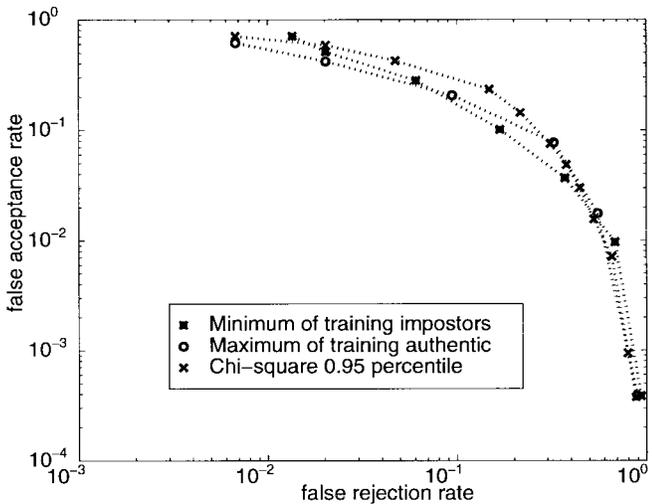
(b)

Fig. 9. Comparison of thresholds, in case of (a) sum of LD's and (b) RME. These plots were obtained with $\lambda = 2$.

Fig. 10. Total error rates according to the threshold for the rigid matching and the elastic graph matching, with $\lambda = 2$.

TABLE I
EQUAL ERROR RATES, WITH RESIDUAL MATCHING ERROR AND WITH LOCAL DISCRIMINANTS AS DISSIMILARITY MEASURES. THE RIGID CASE CORRESPONDS TO A VERY LARGE λ, DENOTED HERE BY $\lambda = \infty$. THE EQUAL ERROR RATES ARE OBTAINED BY INTERPOLATION. FOR PREVENTING ANY CONVERGENCE PROBLEMS AT LOW VALUES OF λ, THE NUMBER OF ITERATIONS ON THE ELASTIC MATCHING WAS LIMITED TO 100 FOR $\lambda = 0$

λ	RME	LD
∞	14.4	8.5
10.0	12.0	7.3
5.0	11.9	8.4
4.0	12.1	9.2
3.0	12.3	6.5
2.0	11.8	6.1
1.0	11.2	5.4
0.5	10.8	6.2
0.0	11.6	6.0

also to study the tolerance of the discriminant approach with respect to the rigidity of the grid.

Fig. 10 shows the total error rate defined by TE = FA + FR, for the rigid matching and the elastic graph matching, for both types of discriminant measures. Clearly, the presence of the local discrimination has a larger influence on the results than the elastic deformation.

The EER defined as the point where FA = FR is shown for several values of λ in Table I. There is a transition from elastic to rigid matching. The local discrimination is able to provide almost constant results for λ between 0.5 and 3.0. For larger values of λ, the performance degrades. The elastic graph matching improved the rigid graph matching, which can be observed by inspecting Fig. 10. Table I shows that the EER is improved from 14% down to 11%. However, combining the rigid graph matching with local discriminants is better than elastic graph matching without local discriminants. Not surprisingly, combining the elastic deformation with local discrimination yielded the best results.

In conclusion, it has been shown that a small degree of elasticity provides an improvement in the performance. The behavior remains constant over a certain range of λ, but from a certain rigidity on, the performance degrades.

D. Comparison with Other Approaches

Our method has been compared with two other approaches, namely the eigenface approach (see Section IV-A) and the synergetic computer approach (see Section IV-B) using the same database and test protocol.

The geometrical normalization method considered here consists of a face detection algorithm followed by an affine transformation that places the eyes and the center of the mouth at predefined locations in the normalized image. The face detection is modular and consists of a series of three consecutive steps. At each step, candidates in the previous step are examined. As each step imposes new conditions, this approach allows us to progressively get rid of false face detections. Of course, the first steps should not be too restrictive, so that the good candidates are almost certainly retained. The three modules are the following.

1) *Template matching on a multi-resolution pyramid.* Templates of the head with conditions on gray-level values that are to be fulfilled are applied on a Gaussian pyramid. Here, the subsampling factor of 1.7 is chosen in order to allow finer intermediate scales than with the standard subsampling factor of two. First, head templates are applied on a range of pyramid levels, then eye region templates are applied. This step is inspired from mosaic images proposed by Yang and Huang [11].

2) *Eye localization and verification.* Eyes are localized precisely and checked against grey-level and gradient templates selected on a training set. Finally, the size and exact position of the eyes are determined.

3) *Nose and mouth detection* is achieved by looking for grey-level minima along the median line between the eyes. Precisely, the bottom of the nose and the lower side of the upper lip are detected, which correspond to the first and second significant minima along this line, respectively.

For authentication, it is obvious that a reliable normalization is highly desirable. If this prior step fails, the candidate will certainly be rejected, as a badly normalized image is usually very different from its reference.

The normalization method described above was applied to the M2VTS database (Section V-A), where *exactly* one head is present in each image. From the 551 images that were presented to the head detection algorithm, 112 images were considered as containing no head, while in all others exactly one head was detected, i.e., there was no false detection. Therefore, a false rejection rate of 18.5% was observed. By visual inspection of the 439 normalized images, 41 of them were undoubtedly badly normalized, yielding a 9.3% bad normalization rate.

From a detailed analysis of the reports, it turns out that the method is very sensitive to the eye detection step. If the person wears glasses, or in case the eyes are not looking straight ahead or are closed, the detection usually fails. One remedy is to incorporate additional templates specifically for these situations into the model. For authentication tests according to the protocol of Section V-A, all rejected and badly normalized images were excluded, which means that 398 images were used, so that the authentication results reflect as much as possible the performance of the methods, and not that of the prior normalization.

Fig. 11 shows the ROC curves for various numbers of eigenvalues retained. The performance of the system increases with the number of eigenfaces retained, until a minimum is reached at about 20 eigenfaces.

Fig. 12 shows a comparison between ROC curves for the eigenface, the synergetic computer, and the proposed EGM-LD approach. Table II gives the equal error rates of the three methods. The principal component analysis provides the poorest results. This may be explained in the following way: theoretically, the PCA provides the best orthogonal description of normalized face images in the sense that the reconstruction from PCA coefficients yields on average the smallest sum of squared errors. This does not assure that PCA is the best discrimination approach. The synergetic computers provided
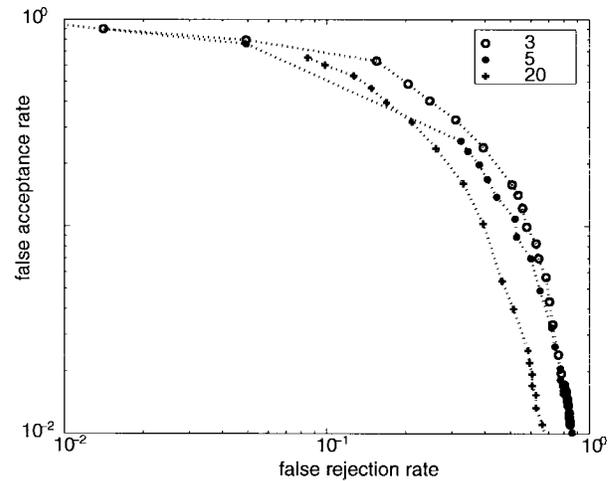
(a)
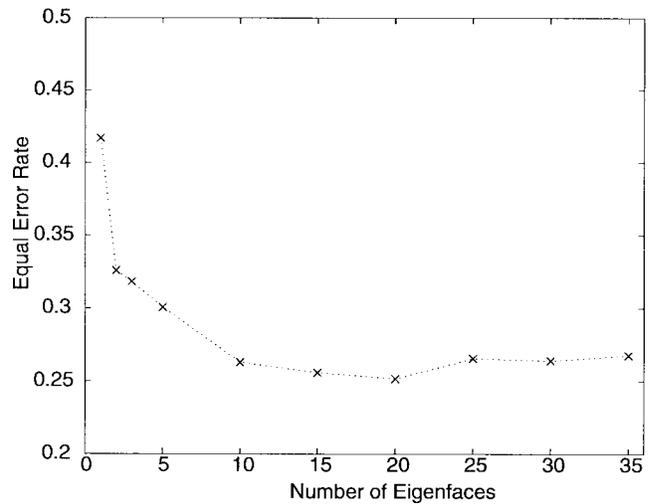
(b)

Fig. 11.　(a) ROC curves for various numbers n of eigenvalues retained. The n most significant vectors are always retained. (b) EER as a function of the number of retained eigenvalues. A minimum is reached for $n = 20$, approximately.

significantly better performance in our tests. Still, the best results were obtained with the elastic graph matching with Gabor features and local discriminants.

The poor performance of the two alternative methods may also be due to the normalization step. As mentioned above, the quality of the normalization was judged by visual inspection. It is possible that small errors decrease the performance of the adjoint vectors method and of the PCA approach drastically.

VI. CONCLUSION

In this contribution, we have shown how the matching of sparse local frequency information arranged on a regular grid may be used for face authentication. A local, linear discrimination approach has been presented, that weighs the contribution of each feature according to its significance for the considered class. This discrimination can be seen as a specialization of the elastic graph matching approach to authentication. It improves the performance significantly. This

(a)

(b)
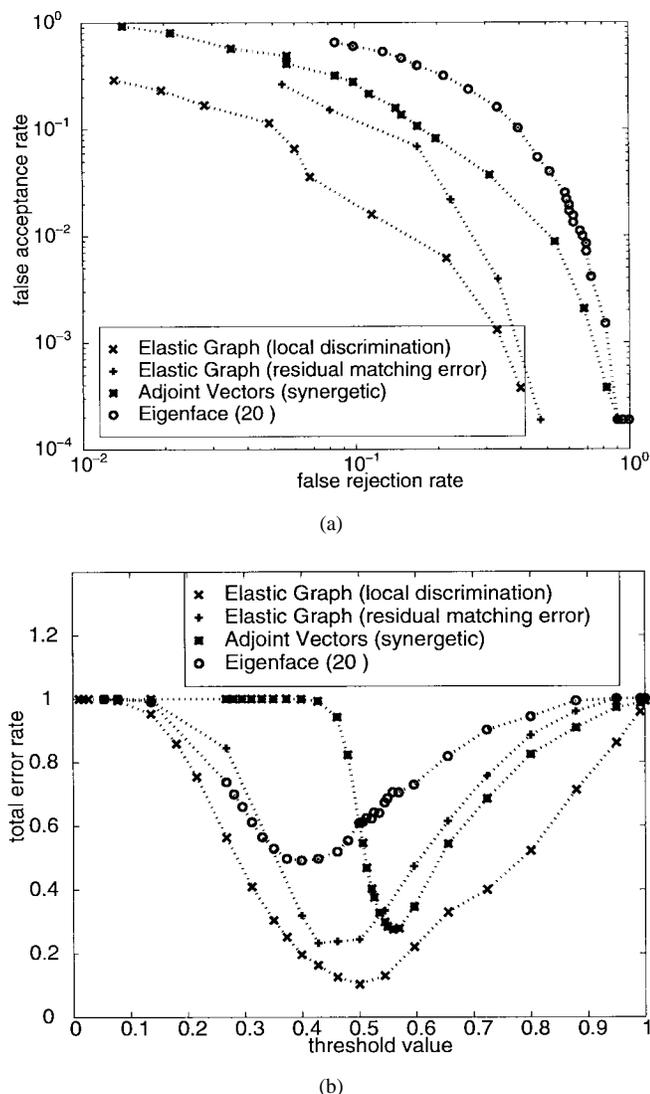
Fig. 12. (a) ROC curves for elastic graph matching (with and without local discrimination), synergetic computers, and eigenface approaches. For the eigenvalue method, 20 eigenvalues were retained. The elastic graph matching results were obtained with $\lambda = 2.0$. (b) Total error curves as a function of normalized threshold.

TABLE II
EQUAL ERROR RATES OF THE ELASTIC GRAPH
MATCHING, COMPARED TO OTHER METHODS

METHOD	EER
EGM (Local discriminants)	6.1
EGM (Residual distance)	11.8
Synergetic Computers	15.8
Eigenface (20)	25.2

approach compares favorably with methods that require an explicit geometric normalization of images.

We could also show that the elastic step of the graph matching leads to some performance improvements. However, the elasticity does not play a decisive role for the success of the matching, as established by authentication performance: the EER increases marginally from 11% to 14%, when going from elastic to rigid matching (EGM). On the other hand, the

increase in computational complexity brought by the elastic deformation is acceptable. Therefore, for best performance, elastic matching together with local discrimination should be used.

We would also like to point out that the requirements in terms of computational complexity are different in the training and in the operational phase. As the training is usually done off-line, computationally expensive methods may be used. On the contrary, the authentication of a candidate in the operational phase should be achieved with minimum delay. The LD's add a significant increase in the computational complexity of the training. However, in the operational phase, the cost of adding local discrimination is insignificant compared to the graph matching.

The different aspect of frequency information provided by the Gabor decomposition, namely the complex response, its phase and its modulus, need to be compared. In particular, phase information shows interesting properties, like robustness with respect to illumination, and may be used for the refinement of the matching.

Finally, this verification application is embedded into a multimodal approach for person authentication, which promises significantly lower error rates than monomodal methods [33], [34]. Here, local discriminants are simply added to provide a global response. However, it is possible to improve this strategy by training [35].

REFERENCES

[1] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 1148–1161, Nov. 1993.
[2] R. R. Rogers, "Biometric identification systems: The science of transaction facilitation," in *Automatic Systems for the Identification and Inspection of Humans*, R. J. Mammone and J. D. Murley, Eds., vol. 2277 SPIE, 1994, pp. 194–199.
[3] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, pp. 705–740, 1995.
[4] M. Lades *et al.*, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, pp. 300–311, Mar. 1993.
[5] R. Brunelli and T. Poggio, "Face recognition; Features versus templates," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 1042–1043, Oct. 1993.
[6] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," in *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*. New York: IEEE Comput. Soc. Press, 1989, pp. 104–109.
[7] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, pp. 71–86, 1991.
[8] B. S. Manjunath, R. Chellappa, and C. v. d. Malsburg, "A feature based approach to face recognition," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, June 1992, pp. 373–378.
[9] L. Wiskott, "Labeled graphs and dynamic link matching for face recognition and scene analysis," Reihe Physik 53, 1995.
[10] L. Wiskott, J.-M. Fellous, N. Krüger, and C. v. d. Malsburg, "Face recognition and gender determination," in *Proc. Int. Workshop on Automatic Face- and Gesture Recognition*, M. Bichsel, Ed., Zürich, Switzerland: MultiMedia Lab., Dept. Comput. Sci., Univ. Zürich, June 1995, pp. 92–97.
[11] G. Yang and T. S. Huang, "Human face detection in a complex background," *Pattern Recognit.*, vol. 27, pp. 53–63, 1994.

[12] D. Sung and T. Poggio, "Learning human face detection in cluttered scenes," in *Proc. 6th Int. Conf. Computer Analysis of Images and Patterns*, Prague, Czech Republic, Sept. 6–8, 1995, pp. 432–439.

[13] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," Tech. Rep. A.I. Memo 1521, C.B.C.L. Paper no. 112, Artif. Intell. Lab., Mass. Inst. Technol., Cambridge, Dec. 1994.

[14] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. 1994 IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition*, Seattle, WA, June 1994, pp. 84–90.

[15] N. Krüger, "An algorithm for the learning of weights in discrimination functions using *a priori* constraints," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 764–768, July 1997.

[16] H. Haken, *Synergetic Computers and Cognition: A Top-Down Approach to Neural Nets*. Berlin, Germany: Springer-Verlag, 1991.

[17] J. Bigün and J. M. H. du Buf, "N-folded symmetries by complex moments in Gabor space," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 80–87, Jan. 1994.

[18] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," *J. Opt. Soc. Amer.*, vol. 7, pp. 923–932, May 1990.

[19] J. Bigün and J. M. H. du Buf, "N-folded symmetries by complex moments in Gabor space and their application to unsupervised texture segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 80–87, Jan. 1994.

[20] J.-P. Antoine, R. Murenzi, B. Piette, and M. Duval-Destin, "Image analysis with 2D continuous wavelet transform: Detection of position, orientation and visual contrast of simple objects," in *Proc. Int. Conf. Wavelets and Applications*, May 1989, pp. 144–159.

[21] W. Hackbusch and U. Trottenberg, Eds., *Lecture Notes in Mathematics: Multigrid Methods*, vol. 1228. Berlin, Germany: Springer, 1986.

[22] D. Terzopoulos, "Image analysis using multigrid relaxation methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 129–139, Mar. 1986.

[23] F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: A review and a new contribution," *Proc. IEEE*, vol. 83, pp. 858–876, June 1995.

[24] P. J. Burt, "Fast filter transforms for image processing," *Comput. Graph. Image Process.*, vol. 16, pp. 20–51, 1981.

[25] P. a. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. London, U.K.: Prentice-Hall Int., 1982.

[26] J. Bigün, "Unsupervised feature reduction in image segmentation by local transforms," *Pattern Recognit. Lett.*, vol. 14, pp. 573–583, 1993.

[27] B. V. Dasarathy, *Decision Fusion*. New York: IEEE Computer Soc. Press, 1994.

[28] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.

[29] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Amer., A*, vol. 4, pp. 519–524, 1987.

[30] T. Wagner and F. G. Boebel, "Testing synergetic algorithms with industrial classification problems," *Neural Networks*, vol. 7, pp. 1313–1321, 1994.

[31] S. Pigeon and L. Vandendorpe, "The M2VTS multimodal face database," in *Proc. 1st Int. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA'97)*, Crans-Montanta, Switzerland, Mar. 12–14, 1996, pp. 403–409 (http://www.tele.ucl.ac.be/M2VTS).

[32] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. London, U.K.: Chapman & Hall, 1993.

[33] B. Duc, G. Maître, S. Fischer, and J. Bigün, "Person authentication by fusing face and speech information," in *1st Int. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA'97)*, J. Bigün, G. Chollet, and G. Borgefors, Eds., Crans-Montana, Switzerland, Mar. 12–14, 1997.

[34] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 955–966, Oct. 1995.

[35] E. S. Bigün, B. Duc, and S. Fischer, "Expert conciliation for multi modal person authentication systems by Bayesian statistics," in *1st Int. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA'97)*, J. Bigün, G. Chollet, and G. Borgefors, Eds., Crans-Montana, Switzerland, Mar. 12–14, 1997.

Benoît Duc graduated in physics in 1992, and received the Ph.D. degree in image processing from the Communications Systems Division, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

From 1992 to 1997, he was a Research Assistant at the Signal Processing Laboratory, EPFL. Since 1998, he has been with Motorola, Inc., Geneva, Switzerland, as a research engineer. His research interests include pattern recognition, feature extraction, segmentation, motion estimation, and image compression.

Stefan Fischer (A'96) graduated from the University of Stuttgart, Germany, in 1992 as an engineer in technical cybernetics, and received the Ph.D. in communications systems from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1997.

He is currently with FASTCOM Technology, S.A., Lausanne, a company he founded. FASTCOM is active in the domain of network embedded systems. His research interests include pattern recognition, multiscale image analysis and biometric person identification.

Josef Bigün (M'88–SM'98) received the M.Sc. degree in applied mathematics and the Ph.D. degree in computer vision from Linköping University, Linköping, Sweden, in 1983 and 1988, respectively.

Between 1988 and 1998, he was a Senior Researcher with the Swiss Federal Institute of Technology, Lausanne, Switzerland. In 1999, he became Professor of Signal Analysis at Halmstad University and Chalmers University of Technology, Sweden. His main interest has been in feature extraction modeling, including multifrequency and orientation filtering, local symmetry theories and deformation transformations, for pattern recogniton and image analysis applications. He is also interested in understanding the biological recognition mechanisms and their signal processing. He has been the principal investigator of several national as well as international research projects.

Dr. Bigün has served several professional conferences as reviewer, program committee member, or chairman. He has also served a a reviewer and editorial member for professional journals. He is an executive board member of the International Association for Pattern Recognition and the Swedish Society for Automated Image Analysis.