Frontal face authentication using discriminating grids with morphological feature vectors

C. Kotropoulos A. Tefas I. Pitas

Department of Informatics, Aristotle University of Thessaloniki

Box 451, Thessaloniki 540 06, GREECE

E-mail :{costas,tefas,pitas}@zeus.csd.auth.gr

EDICS numbers: 4-KNOW Content Recognition and Understanding

3-MODA Multimodal and Multimedia Environments

Abstract

A novel elastic graph matching procedure based on multiscale morphological operations, the so called morphological dynamic link architecture, is developed for frontal face authentication. Fast algorithms for implementing mathematical morphology operations are presented. Feature selection by employing linear projection algorithms is proposed. Discriminatory power coefficients that weigh the matching error at each grid node are derived. The performance of morphological dynamic link architecture in frontal face authentication is evaluated in terms of the receiver operating characteristic on the M2VTS face image database. Preliminary results for face recognition using the proposed technique are also presented.

Corresponding author: I. Pitas

Face recognition from images is of particular interest in a wide range of applications, e.g., nonintrusive identification and verification for credit cards and automatic teller machine transactions, nonintrusive access control to buildings, identification for law enforcement, etc. Although other biometrics (e.g., fingerprints, iris, etc.) could also be used for person authentication, face recognition is still attractive for the following reasons: universality, collectability and acceptability [1]. In the following, a brief overview of related previous work is given, and the objectives of our work are outlined.

A. Previous work

A comprehensive survey of human and machine recognition techniques can be found in [2, 3]. There are several approaches in developing face recognition systems. For example, one approach employs linear projections of face images (treated as 1-D vectors) using either principal component analysis (PCA) [4] or linear discriminant analysis (LDA) [5, 6, 7]. PCA and LDA are parametric techniques closely related to matching pursuit filters that have also been applied for face identification [8]. There are also techniques stemming from neural network community like the *dynamic link architecture* (DLA), a general object recognition technique that represents an object by projecting its image onto a rectangular elastic grid where a Gabor wavelet bank response is measured at each node [9]. The aforementioned approach can be considered as a special form of the attributed graph matching [10]. For a more detailed description of the DLA, the interested reader may refer to [9, 15, 17].

The elastic graph matching and its applications have been an active research topic since its invention [11, 12, 13, 14, 15]. A particular problem in elastic graph matching that received much attention is the weighting of graph nodes according to their discriminatory power. Several methods have been proposed. For example, a Bayesian approach yields the more reliable nodes for gender identification, beard and glass detection in bunch graphs [16]. An automatic weighting of nodes according to their significance by employing local discriminants is proposed in [17]. A weighted average of feature vector similarities by a set of coefficients that take

into account the importance of each feature in assigning a test person to a specific class is investigated in [18].

B. Paper Outline

A novel elastic graph matching procedure based on multiscale morphological dilation-erosion, the so called *morphological dynamic link architecture* (MDLA), is developed for frontal face authentication in this paper. During the training phase, a sparse rigid grid is placed over the facial region of the image of each person in the reference set. Instead of using either the magnitude [9, 17] or phase [16] responses of a bank of Gabor filters tuned to different orientations and scales, as proposed in the original dynamic link architecture, we employ the *multiscale dilation-erosion* of the image by a structuring function [19] to derive a feature vector at each grid node. There are strong theoretical arguments that support this substitution. Signal extrema are of the fundamental signal features according to Witkin's scale-space theory. Since any linear convolution kernel introduces extrema with increasing scale in 2-D, the so-called *monotonic property* does not hold for linear filters and signal extrema. On the contrary, the multiscale dilation-erosion guarantees that, if a signal extremum appears at some scale σ_0, it also appears at zero scale and all finer scales in between. That is, the number of signal features does not increase with increasing scale [19]. In the test phase, the multiscale morphological operations are applied over the entire image of a candidate before elastic graph matching. Thus, the computation of the time consuming Gabor-based feature vectors that rely on floating point arithmetic operations (i.e., FFTs or convolutions) is avoided. Moreover, no filter design is needed. The novelty of the paper is not limited to the substitution of Gabor filters by the multiscale morphological ones, but in the detailed step by step design of a pattern matching scheme, i.e.: (i) the efficient feature extraction by studying the influence of the structuring function on verification performance and computational speed; (ii) the proper feature selection based on linear projections when there are not enough data to apply LDA from the beginning as opposed to the so called local discriminants approach proposed in [17]; (iii) the feature matching by revisiting the coarse-to-fine approach proposed in [9]; (iv) the automatic node

weighting that is not based on eigenanalysis which is a more demanding task with respect to data availability; and (v) the properly designed training and test procedure for evaluating the performance of any face authentication algorithm.

Efficient feature extraction is studied in Section IV-A, because feature computation is the second most time consuming task after graph matching. By reducing its computational time the method gains user's tolerance. We propose the combined use of MDLA with linear projection algorithms (Section IV-B) in order to enhance its verification capability. Principal component analysis precedes linear discriminant analysis for two reasons: (i) to reduce the dimensionality of feature vectors at grid nodes, and (ii) to yield uncorrelated feature vectors. It is interesting to note that the same approach has been used in *Fisherfaces* [6]. Additional explanation on this point can be found in Section V. The motivation behind using linear discriminant analysis is its property to provide the optimal linear discriminant among classes when the distribution of each class is Gaussian. A probabilistic hill climbing algorithm replaces the two-stage coarse-to-fine optimization procedure used in elastic graph matching (i.e., a rigid matching and a deformable matching) [9, 15], because such a procedure cannot be easily trapped in the local minima of the objective function to be minimized and, thus, yields better results in terms of the verification efficiency (Section IV-C). Moreover, we propose expressing the total matching error as a sum of matching errors weighted by discriminatory power coefficients (Section IV-D) in order to quantify the discriminating capability of each node and compensate for possible errors occurred during the probabilistic hill climbing algorithm due to the limited number of iterations allowed.

The basic block diagram of the MDLA in the training phase is depicted in Figure 1(a). A more advanced training procedure is sketched in Figure 1(b). According to the experimental protocol used (e.g., the one described in Section V), the test person during the training phase can be either a training client or a training impostor so that person-specific thresholds for all training clients are determined. It is seen that the system consists of four modules, namely: (i) the face detection module; (ii) the feature extraction module; (iii) the feature selection module shown in Figure 1(b), or alternatively, the more simple node weighting module shown

in Figure 1(a), and, (iv) the feature matching module.

The active modules during the test phase are shown in solid lines in Figure 1(c). It can be seen that only the feature extraction and feature matching modules are active. Discriminating information in the form either of PCA/LDA projection matrices or discriminatory power coefficients is recalled from the training database as well. It should be noted that face detection works in a semi-automatic mode during the training procedure to control the placement of the grid over the facial region of each reference person. It can also be used to initialize the elastic graph matching during the test phase.

The performance of the proposed morphological dynamic link architecture (MDLA) is evaluated in terms of the receiver operating characteristic for several threshold selections on the matching error on the database of the European Union ACTS project "Multimodal Verification for Teleservices and Security Applications" (abbreviated as M2VTS database throughout the paper) [28]. A comparative study of the verification capability of the proposed methods against other frontal face authentication algorithms is undertaken in Section V. It is demonstrated that the combined use of local discriminatory power coefficients with MDLA achieves an equal error rate of 3.7 % and is ranked as the best frontal face authentication algorithm with respect to the published equal error rates achieved on the M2VTS database using the experimental protocol described in Section V. However, we present preliminary results on face recognition by using MDLA to demonstrate its capability in this closely related problem as well.

Motivated by the analyses in [20, 21], we discuss on the number of claims that guarantees statistically significant results in a frontal face authentication task as a function of false acceptance and false rejection error rates specified a priori as *desired* system requirements (Section VI). We also test whether the performance achieved by morphological dynamic link architecture meets the aforementioned system specifications.

II. Face authentication versus face recognition

Before proceeding to the detailed description of the proposed system, let us briefly address the problem of face authentication. The majority of the approaches outlined in Section I-A

has been tested for face recognition. Face recognition algorithms are usually tested on the Face Recognition Technology (FERET) program database following the evaluation procedures developed as part of this program that is sponsored by the U.S. Department of Defense [8]. Therefore, FERET is a *de facto* standard in face recognition. In this paper we are interested in *face authentication*. Their differences are described subsequently. Although the algorithms employed in both face recognition and authentication systems are of common origin (for example, the dynamic link architecture [9]), there is no a priori guarantee that the same algorithm would enjoy the same performance level in both cases. Most published works in the area of face authentication are research outcomes using the databases collected and the experimental protocols defined in the M2VTS project [17, 32, 33, 34, 35]. We shall employ the M2VTS databases and the experimental protocols aiming at providing experimental evidence that is fully comparable to previously reported results.

Face recognition and authentication are conceptually different problems. On the one hand, a face recognition system usually assists a human expert to determine the identity of a test face by computing all similarity scores between a test face and each human face stored in the system database and by ranking them. On the other hand, a face authentication system should decide *itself* if a test face is assigned to a *client* (i.e., one who claims his/her own identity) or to an *impostor* (i.e., one who pretends to be someone else).

The evaluation criteria for face recognition and face authentication systems are different. The performance of face recognition systems is quantified in terms of the percentage of correctly identified faces within the N best matches. By varying the rank N of the match, the curve of cumulative match score versus rank is obtained [8]. The objectives in face authentication is to design algorithms that: (i) provide a low *false rejection rate* (FRR) preventing the client inconvenience, and at the same time, (ii) guarantee that the impostors cannot be falsely interpreted as clients by offering the lowest possible *false acceptance rate* (FAR). In practice, there is a trade-off between these two objectives that is depicted by the so-called *receiver operating characteristic* (ROC) curve. The latter curve is obtained by varying the FAR. Accordingly, the performance of face authentication systems is measured in terms of the FRR achieved at

authentication algorithm, then we usually choose the operating point having $FAR = FRR$, the so called *equal error rate* (EER).

A third difference is in the requirements needed when face recognition/authentication systems are trained. Face recognition systems are usually trained on sets having one frontal image per person. Face authentication systems usually need more training frontal images per person in order to capture the intra-class variability (i.e., to model the variations of face images corresponding to the same person). The requirements increase dramatically when linear discriminant analysis is employed to accomplish feature selection, as is explained later on.

The selection of a face authentication technique is guided by the EER that is attained, its computational speed as well as its performance under real conditions that include lighting changes, changes in face position and scale, varying facial expressions, etc. In this paper, we address the selection of a face authentication system by employing only the EER under well-controlled conditions. This decision is grounded on the following arguments: (a) It is always possible to implement an algorithm in a more efficient way by speeding up certain steps. By studying a more efficient way to perform the multiscale dilation-erosion in Section IV-A, we succeeded to drop the computational time considerably. (b) We avoid using "tricks" (e.g., to reduce the resolution of the original image in order to claim a faster algorithm execution or to report computational time on more powerful platforms, etc.) (c) Simple compensation procedures (e.g., those proposed in [35]) can be applied prior to any face authentication algorithm aiming at compensating for the variable recording conditions. Having stated the face authentication problem, we proceed to the description of each module of the proposed system.

III. Frontal face detection

A very attractive approach for face detection based on multiresolution images (also known as *mosaic images*) has been proposed in [23]. Motivated by the simplicity of this face detection approach, we briefly describe a variant of this method that has the following features [24]: (a) It uses rectangular cells in contrast to the square cells used in [23]. (b) It is equipped with a

preprocessing step that determines an estimate of the cell dimensions and the offsets so that the mosaic model fits the face image of each person. (c) It has very low computational demands compared to the original algorithm, because the iterative nature of the latter is avoided due to the preprocessing step that has been used. (d) It employs more general rules that are close to our intuition for a human face.

The method attempts to find a resolution level where the main part of the face occupies an area of about 4×4 cells. The mosaic image that is created at this resolution level is the so called *quartet* image. The grey level of each cell equals the average value of the grey levels of all pixels included in the cell. An abstract face model at the resolution level of the quartet image is depicted in Figure 2. In this model, the main part of the face corresponds to the region of 4×4 cells having as origin the cell marked by "X". By subdividing each quartet image cell to 2×2 cells of half dimensions the *octet* image results, where the main facial features, such as eyebrows/eyes, nostrils/nose and mouth, are detected. Due to lack of space, the detailed description of the face detection algorithm employed is omitted and the interested reader is referred to [24].

Figure 3 depicts two frontal face images. Their quartet and octet images are shown in the same figure. The images in the last column are the results of the face detection algorithm. The octets for the facial features are shown overlaid in these images. The octets for eyebrows/eyes and nostrils/nose are shown as white overlaid rectangles. Mouth candidates are shown as black overlaid rectangles. The white cross indicates the characteristic bright octet between the eyes.

IV. Morphological dynamic link architecture

In this section, the variant of DLA under study is described. Fast algorithms for feature extraction are derived in Section IV-A. The combined use of linear projections and morphological dynamic link architecture is analyzed in Section IV-B. The proposed elastic graph matching procedure is presented in Section IV-C. Node weighting coefficients that depend on the first and second-order statistics of the matching errors are derived in Section IV-D.

An alternative to linear techniques used for generating an information pyramid is the scale-space morphological techniques. In this paper, we propose the substitution of Gabor-based feature vectors used in dynamic link architecture by the *multiscale morphological dilation-erosion* [19] for the following reasons: (1) Scale-space morphological techniques are able to find the true size of the object in an image smoothed to a particular size. (2) The scale parameter has a straightforward interpretation, since it is associated with the area of the domain of the structuring function. (3) Dilations and erosions can be computed very fast either by running min/max algorithms [25] or by recursive separable computations or by scale recursive computations, as is discussed in this Section. (4) Dilations and erosions deal with the local extrema in an image. Therefore, they are well-suited for facial feature representation, because key facial features are associated either to local minima (e.g., eyebrows/eyes, nostrils, endpoints of lips, etc.) or to local maxima (e.g., the nose tip).

The multiscale morphological dilation-erosion is based on the two fundamental operations of grayscale morphology, namely, the *dilation* and the *erosion*. Let $\mathcal{R}$ and $\mathcal{Z}$ denote the set of real and integer numbers, respectively. Given an image $f(\mathbf{x}) : \mathcal{D} \subseteq \mathcal{Z}^2 \rightarrow \mathcal{R}$ and a structuring function $g(\mathbf{x}) : \mathcal{G} \subseteq \mathcal{Z}^2 \rightarrow \mathcal{R}$, the dilation of the image $f(\mathbf{x})$ $g(\mathbf{x})$ is denoted by $(f \oplus g)(\mathbf{x})$ whereas the erosion is denoted by $(f \ominus g)(\mathbf{x})$ [25]. If the structuring function is chosen to be scale-dependent, that is $g_\sigma(\mathbf{z}) = |\sigma| g(|\sigma|^{-1}\,\mathbf{z})$ $\forall \mathbf{z} \in \mathcal{G}$: $\|\mathbf{z}\| \leq |\sigma|$, the morphological operations become scale-dependent as well. Suitable structuring functions are: (i) the scaled hemisphere, i.e., $g_\sigma(\mathbf{z}) = |\sigma| \left(\sqrt{1 - \left(|\sigma|^{-1}\|\mathbf{z}\|\right)^2} - 1 \right)$ [19], (ii) the flat structuring function, i.e., $g_\sigma(\mathbf{z}) = 0$ [25], and, (iii) the circular paraboloid, i.e., $g_\sigma(\mathbf{z}) = -|\sigma| \frac{\|\mathbf{z}\|^2}{\sigma^2}$ [22], where $\|\mathbf{z}\| \leq |\sigma|$. The choice of the structuring function affects the verification capability of the proposed technique to a margin of $\pm 0.5\%$ with respect to the equal error rate, but it does affect the time required to compute the dilation and erosion. The efficient computation of dilations and erosions is a factor of vital importance in practical applications. For this reason, we devote some space to developing efficient algorithms for dilation and erosion by a flat or a paraboloid structuring function.

For a flat structuring function, dilations can be efficiently computed by applying running max calculations (e.g., the MAXLINE algorithm [25]) in which the computation of $(f \oplus g_\sigma)(x_1, x_2)$ exploits the previous outcome $(f \oplus g_\sigma)(x_1 - 1, x_2)$. Similar running min calculations can be exploited in the efficient computations of erosions. In this paper, we propose scale-recursive computations that speed up feature calculation considerably. For a flat structuring function, scale-recursive max computations are based on the observation that:

$$\begin{aligned}(f \oplus g_{\sigma+1})(x_1, x_2) &= \max\{(f \oplus g_\sigma)(x_1, x_2), \max_{(z_1,z_2)\in\Delta G(\sigma+1)}\{f(x_1 + z_1, x_2 + z_2)\}, \\ &\qquad \max\{f(x_1 \pm (\sigma+1), x_2 \pm (\sigma+1))\}\} \end{aligned} \tag{1}$$

where the set $\Delta G(\sigma+1) = \{(z_1, z_2) \in \mathcal{Z}^2 : (z_1^2 + z_2^2) > \sigma^2, \quad (z_1^2 + z_2^2) \leq (\sigma+1)^2, \quad |z_1| \leq \sigma, \ |z_2| \leq \sigma\}$ possesses a symmetry and can easily be computed prior to the computation of dilations. A similar recursive computation can be applied for minima as well.

Let us consider next the case of a circular paraboloid structuring function in the 2-D case. It can be decomposed as $g_\sigma(z_1, z_2) = g_\sigma^{(1)}(z_1) + g_\sigma^{(1)}(z_2)$, where $g_\sigma^{(1)}(\cdot)$ denotes the one-dimensional structuring function [22]. Therefore, the computations are separable. It can easily be seen that [22]:

$$(f \oplus g_\sigma)(x_1, x_2) = \max_{z_1 \in \mathcal{G}_\sigma^{(1)}} \left(\gamma(x_1 - z_1, x_2) + g_\sigma^{(1)}(z_1)\right) \tag{2}$$

$$\gamma(x_1, x_2) = \max_{z_2 \in \mathcal{G}_\sigma^{(2)}} \left(f(x_1, x_2 - z_2) + g_\sigma^{(2)}(z_2)\right) \tag{3}$$

where $\mathcal{G}_\sigma^{(i)}, i = 1, 2$ are the projections of $\mathcal{G}$ on the axes. Following similar lines as in [22], let us suppose that the maximum occurs in (3) for $z_2 = \xi$. The recursive separable implementation of grayscale dilation is:

$$\gamma(x_1, x_2 + 1) = \max_{z_2 = -\sigma}^{\xi+1} \left(f(x_1, x_2 + 1 - z_2) + g_\sigma^{(1)}(z_2)\right). \tag{4}$$

A similar recursive separable implementation of grayscale erosion can also be derived.

The multiscale dilation-erosion of the image $f(\mathbf{x})$ by $g_\sigma(\mathbf{x})$ is defined by [19]:

$$(f \star g_\sigma)(\mathbf{x}) = \begin{cases} (f \oplus g_\sigma)(\mathbf{x}) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ (f \ominus g_\sigma)(\mathbf{x}) & \text{if } \sigma < 0. \end{cases} \tag{5}$$

The outputs of multiscale dilation-erosion for $\sigma = -\sigma_m, \ldots, \sigma_m$ form the feature vector located at grid node $\mathbf{x}$:

$$\mathbf{j}(\mathbf{x}) = ((f \star g_{\sigma_m})(\mathbf{x}), \ldots, (f \star g_1)(\mathbf{x}), f(\mathbf{x}), (f \star g_{-1})(\mathbf{x}), \ldots, (f \star g_{-\sigma_m})(\mathbf{x})) \,. \tag{6}$$

The parameter σ_m is upper bounded by the half of the minimal distance between two nodes of the sparse grid that is to be created. The value $\sigma_m = 9$ has been used in all experiments reported in this paper. An 8×8 sparse grid has been created by measuring the feature vectors $\mathbf{j}(\mathbf{x})$ at equally spaced nodes over the output of the face detection algorithm described in Section III. Fig. 4 depicts the output of multiscale dilation-erosion for the scales that have been used. The first nine pictures starting from the upper left picture are dilated images and the remaining nine are eroded images. It is seen that multiscale dilation-erosion captures important information for key facial features such as eyebrows, eyes, nose tip, nostrils, lips, face contour etc.

It can be seen that the method in its present form does not contain any directional information. The main reason for this decision is that the rotational symmetry satisfies the spatial isotropy which constitutes a basic principle of scale-space theory [26]. Moreover, the circular paraboloid is shown to be the morphological Green's function for the partial differential equation $\frac{\partial}{\partial \sigma}(f \oplus g_\sigma)(\mathbf{z}) = \|\nabla_{\mathbf{z}}(f \oplus g_\sigma)(\mathbf{z})\|^2$. However, departing from scale-space theory arguments, it is possible to employ directional morphological operations in the feature extraction module. Then, a major constraint is that of the integer geometry that has to be taken into consideration in the definition of the structuring functions at different angles.

B. Feature selection

The most popular linear projection algorithms that can be used to reduce the dimensionality of the feature vectors are the *Karhunen-Loeve or principal component analysis* (PCA) and the *linear discriminant analysis* (LDA) [30].

Representations based on PCA are useful to image reconstruction/ compression tasks [5, 7, 31]. In addition to dimensionality reduction, PCA decorrelates the feature vectors and facilitates the LDA applied subsequently in both eigenvalue/eigenvector computations and

matrix inversion. Let $\mathbf{j}'(\mathbf{x}_l) = \mathbf{j}(\mathbf{x}_l) - \mathbf{m}(\mathbf{x}_l)$ be the normalized feature vector at node $\mathbf{x}_l$, where $\mathbf{j}(\mathbf{x}_l)$ is given by (6) and $\mathbf{m}(\mathbf{x}_l)$ be the mean feature vector. The eigenvectors $\mathbf{e}_i(\mathbf{x}_l)$ that correspond to the p largest eigenvalues of the covariance matrix of feature vectors $\mathbf{j}'(\mathbf{x}_l)$ are computed and the PCA projected feature vector is given by:

$$\tilde{\mathbf{j}}(\mathbf{x}_l) = \begin{bmatrix} \mathbf{e}_1^T(\mathbf{x}_l) \\ \vdots \\ \mathbf{e}_p^T(\mathbf{x}_l) \end{bmatrix} \mathbf{j}'(\mathbf{x}_l) = \mathbf{P}(\mathbf{x}_l)\mathbf{j}'(\mathbf{x}_l) \tag{7}$$

where T denotes the transposition operator. The resulting vector $\tilde{\mathbf{j}}(\mathbf{x}_l)$ is comprised by the so-called *most expressive features* [5]. It is of dimensions $p \times 1$ with $p \leq (2\sigma_m + 1)$.

Most expressive features preserve the shape of the class distributions and they are appropriate for signal representation/reconstruction tasks, because the subspace spanned by the p top eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_p$ minimizes the mean square error between the raw feature vectors and their projections onto this subspace. Therefore, there is no guarantee that the most expressive features alone are necessarily good candidates for discriminating among classes defined by a set of samples [5, 7]. Subsequently, LDA is applied to the most expressive feature vectors. LDA searches for the best linear subspace which maximizes class separability. Class separability is not related to Bayes error directly. Accordingly, LDA is not always optimal in Bayes sense [27]. The feature vectors produced after LDA projection are called *most discriminating features* (MDFs) [5]. We are interested in applying LDA locally at each grid node. Let $\mathcal{S}$ be the entire set of feature vectors at a grid node and $\mathcal{S}_k$ be the subset of features vectors at this node extracted from the frontal face images of the k-th person in the database. The local LDA scheme determines a weighting matrix $(d \times p)$, $\mathbf{V}_k$, such that the ratio:

$$\mathcal{M}_k = \frac{\mathrm{tr}\left[\mathbf{V}_k \left\{\sum_{\tilde{\mathbf{j}} \in \mathcal{S}_k} (\tilde{\mathbf{j}} - \tilde{\mathbf{m}}_k)(\tilde{\mathbf{j}} - \tilde{\mathbf{m}}_k)^T\right\} \mathbf{V}_k^T\right]}{\mathrm{tr}\left[\mathbf{V}_k \left\{\sum_{\tilde{\mathbf{j}} \in (\mathcal{S} - \mathcal{S}_k)} (\tilde{\mathbf{j}} - \tilde{\mathbf{m}}_k)(\tilde{\mathbf{j}} - \tilde{\mathbf{m}}_k)^T\right\} \mathbf{V}_k^T\right]} = \frac{\mathrm{tr}\left[\mathbf{V}_k \mathbf{W}_k \mathbf{V}_k^T\right]}{\mathrm{tr}\left[\mathbf{V}_k \mathbf{B}_k \mathbf{V}_k^T\right]} \tag{8}$$

is minimized, where $\tilde{\mathbf{m}}_k$ is the class-dependent mean vector of the most expressive feature vectors. The solution of the generalized eigenvalue problem (8), i.e., the row vectors of $\mathbf{V}_k$ ($\mathbf{v}_{ik}$, $i = 1, \ldots, d$), is given by the eigenvectors that correspond to the d smallest in magnitude eigenvalues of $\mathbf{B}_k^{-1}\mathbf{W}_k$ or equivalently by the eigenvectors that correspond to the d largest in

magnitude eigenvalues of $\mathbf{W}_k^{-1}\mathbf{B}_k$ provided that both $\mathbf{W}_k$ and $\mathbf{B}_k$ are invertible. It is worth noting that the eigenvalue problem could be computationally unstable, because the matrix $\mathbf{W}_k^{-1}\mathbf{B}_k$ is not symmetric in general. The elegant method proposed in [5] that diagonalizes the symmetric matrices $\mathbf{W}_k$ and $\mathbf{B}_k$ has been used to solve the generalized eigenvalue problem (8).

We shall confine the discussion to $d = 2$ for facilitating the description. Let the superscripts t and r denote a test and a reference person (or grid), respectively. Having found the weighting matrix $\mathbf{V}_k(\mathbf{x}_l)$, for the l-th node of the k-th person, we project the reference most expressive feature vector at this node onto the plane defined by $\mathbf{v}_{1k}(\mathbf{x}_l)$ and $\mathbf{v}_{2k}(\mathbf{x}_l)$, i.e., $\breve{\mathbf{j}}(\mathbf{x}_l^r) = \mathbf{V}_k\left[\mathbf{P}(\mathbf{x}_l)\left(\mathbf{j}(\mathbf{x}_l^r) - \mathbf{m}_l\right) - \tilde{\mathbf{m}}_{kl}\right]$ to yield the reference most discriminating feature vector at this node. Let us suppose that a test person claims the identity of the k-th person. Then, the test most discriminating feature vector at the l-th node is given by $\breve{\mathbf{j}}(\mathbf{x}_l^t) = \mathbf{V}_k\left[\mathbf{P}(\mathbf{x}_l)\left(\mathbf{j}(\mathbf{x}_l^t) - \mathbf{m}_l\right) - \tilde{\mathbf{m}}_{kl}\right]$.

The row vectors $\mathbf{v}_{ik}$, $i = 1, 2, \ldots, d$, of $\mathbf{V}_k$ weigh the most expressive features $\tilde{\mathbf{j}} - \tilde{\mathbf{m}}_{kl}$ automatically according to their discriminatory power [3]. If a component corresponds to pure random noise, its contribution in the subspace defined by $\mathbf{v}_{ik}$, $i = 1, 2, \ldots, d$ will be approximately zero. This is not the case with the subspace defined by $\mathbf{e}_1, \ldots, \mathbf{e}_p$, where the contribution of such a noisy component will be roughly proportional to the noise variance. More MEFs do not necessarily give a better recognition/authentication rate. On the contrary, the more most discriminating features are employed, the better recognition/authentication rate is obtained.

C. Elastic graph matching

The L_2 norm of the difference between the raw feature vectors at the i-th grid node has been used as a (signal) similarity measure, i.e.: $C_v(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r)) = \|\mathbf{j}(\mathbf{x}_l^t) - \mathbf{j}(\mathbf{x}_l^r)\|$. If linear projections of feature vectors are employed in MDLA, then the signal similarity measure is rewritten as $C_v(\breve{\mathbf{j}}(\mathbf{x}_l^t), \breve{\mathbf{j}}(\mathbf{x}_l^r)) = \|\breve{\mathbf{j}}(\mathbf{x}_l^t) - \breve{\mathbf{j}}(\mathbf{x}_l^r)\|$. In the following, we omit the distinction between raw feature vectors and most discriminating ones for notation simplicity. Let us define by $\mathcal{V}$ the set of grid nodes. Grid nodes are considered as graph vertices. Let also $\mathcal{N}(l)$ denote the

four-connected neighborhood of vertex l. The objective in DEA [9] is to find the set of test grid node coordinates $\{\mathbf{x}_l^t,\ l \in \mathcal{V}\}$ that minimizes the cost function:

$$C(\{\mathbf{x}_l^t\}) = \sum_{l \in \mathcal{V}} \left\{ C_v(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r)) + \lambda \sum_{\xi \in \mathcal{N}(l)} C_e(l, \xi) \right\}. \tag{9}$$

where $C_e(l, \xi)$ penalizes the grid deformations, i.e., $C_e(l, \xi) = \|(\mathbf{x}_l^t - \mathbf{x}_l^r) - (\mathbf{x}_\xi^t - \mathbf{x}_\xi^r)\|$, $\xi \in \mathcal{N}(l)$. The cost function (9) defines a distance measure $D(t, r)$ between the test person t and the reference person r. In [9] the authors argue that a two stage coarse-to-fine optimization procedure suffices for the minimization of (9). Clearly, the aforementioned two stage procedure is a heuristic algorithm that aims at solving an optimization problem within reasonable time [16]. The coarse step is essentially a "block-matching" step. That is, the authors assume that it is possible to place a grid on the face region of the test person by translating an undistorted replica of the reference grid, that contains feature vectors from the test image at its nodes, with a predefined step. Such an assumption is valid, when the objective function (9) is a smooth, convex function with a prominent global minimum. If the objective function contains multiple local minima, the coarse step can easily be trapped in local minima. By employing a simulated annealing algorithm, we avoid such a possibility. Accordingly, we propose: (i) to exploit the face detection results that are provided by the hierarchical rule-based system described in Section III for initializing the minimization of the cost function, and (ii) to replace the two stage optimization procedure by a probabilistic hill climbing algorithm (i.e., a simulated annealing algorithm) which is reminiscent of Algorithm 1.4 [29, p. 12] that does not make distinction between coarse and fine matching. Indeed, one may interpret (9) as a simulated annealing with an additional penalty (i.e., a constraint on the objective function). Since $C_e(l, \xi)$ does not penalise translations of the whole graph the random configuration $\mathbf{x}_l$ can be of the form of a random translation $\mathbf{s}$ of the (undeformed) reference grid and a bounded local perturbation $\boldsymbol{\delta}_l$, i.e., $\mathbf{x}_l^t = \mathbf{x}_l^r + \mathbf{s} + \boldsymbol{\delta}_l$, $\|\boldsymbol{\delta}_l\| \leq \delta_{\max}$. The choice of $q_{\max}$ controls the rigidity/plasticity of the graph. Accordingly, only the first sum contributes to the matching error, i.e., $C(\{\mathbf{x}_l^t\}) = \sum_{l \in \mathcal{V}} C_v(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r))$.

Figure 5 depicts the grids formed in the matching procedure of a test person with himself

and with another person for a pair of test persons.

D. Weighting the elastic graph matching error

In this section, we propose an alternative scheme aiming at weighting the grid nodes after elastic graph matching by coefficients that depend on the first and second-order statistics of matching errors. That is, we want to weigh the signal similarity measure $C_v(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r))$ using a class-dependent coefficient, the so-called *Discriminatory Power Coefficient* (DPC), $DP_l(\mathcal{S}_r)$, so that when person t claims the identity of person r the distance measure between them is computed by:

$$D(t,r) = \sum_{l \in \mathcal{V}} \frac{DP_l(\mathcal{S}_r)\, C_v(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r))}{\sum_{i \in \mathcal{V}} DP_i(\mathcal{S}_r)} \tag{10}$$

where $\mathcal{S}_r$ denotes the class of the reference person r. The DPC of the l-th node for the class $\mathcal{S}_r$, $DP_l(\mathcal{S}_r)$, is a factor that shows how well the intra-class matching errors are separated from the inter-class matching errors at this node. Let $m_{\text{intra}}(\mathcal{S}_r, l)$ be the mean intra-class matching error for the class $\mathcal{S}_r$ and $m_{\text{inter}}(\mathcal{S}_r, l)$ be the mean inter-class matching error between the class $\mathcal{S}_r$ and the class $(\mathcal{S} - \mathcal{S}_r)$ at grid node l. Let $\text{var}_{\text{intra}}(\mathcal{S}_r, l)$ and $\text{var}_{\text{inter}}(\mathcal{S}_r, l)$ be the corresponding variances. Obviously, grid nodes that do not possess any discriminatory power should be discarded, e.g., the nodes for which $m_{\text{inter}}(\mathcal{S}_r, l) < m_{\text{intra}}(\mathcal{S}_r, l)$. A plausible measure of the discriminatory power of grid node l for the class $\mathcal{S}_r$ is the *Fisher's Linear Discriminant function* [30]:

$$DP_l(\mathcal{S}_r) = \frac{(m_{\text{inter}}(\mathcal{S}_r, l) - m_{\text{intra}}(\mathcal{S}_r, l))^2}{\text{var}_{\text{inter}}(\mathcal{S}_r, l) + \text{var}_{\text{intra}}(\mathcal{S}_r, l)}. \tag{11}$$

The DPCs of grid nodes for several persons in the database are shown pictorially in Figure 6. The nodes that correspond to facial features with higher discriminatory power (e.g., eyes, eyebrows) are weighted with coefficients having a larger value. The discriminatory power coefficients proposed in this section can easily be computed during the application of the authentication algorithm to any database following any experimental protocol. They can easily be modified when persons are added or deleted from the database, because the computation of the mean matching errors and their variances can be made incrementally. This is not the case with linear projections when they are applied to select the most discriminating features,

because, so far, only the asymptotic convergence of recursively computed eigenvectors to the true ones has been theoretically established. In our case, a limited number of feature vectors is only available.

V. Experimental results

The morphological dynamic link architecture has been tested on the M2VTS database [28]. The database contains 37 persons' video data, which include speech consisting of uttering digits and image sequences of rotated heads. Four recordings (i.e., shots) of the 37 persons have been collected. In our experiments, the sequences of rotated heads have been considered by using only the luminance information at a resolution of 286×350 pixels. From each image sequence, one frontal image has been chosen based on symmetry considerations. Four experimental sessions have been implemented by employing a combination of the "leave-one-out" principle and the rotation estimates, i.e., a variant of the jack-knife method. Each experimental session consists of a training and a test procedure that is applied to its training set and its test set, respectively. Let BP, BS, CC, ..., XM be the identity codes of the persons included in the database. Figure 7 depicts a configuration of the experimental protocol when the person BP is excluded. Additional configurations are obtained by permutations of the "left-out" persons and the shots that form the training and the test sets.

First, let us describe the training procedure. The training set is built of 3 (out of 4) shots of 36 (out of 37) persons. This amounts to 108 frontal face images. For each of the 36 trained classes (i.e., clients), 6 intra-class distance measures and 210 inter-class distance measures (i.e., 35 trained impostors $\times$ 6 inter-class distances per impostor) are computed by using the morphological dynamic link architecture as authentication mechanism. The objective of the training procedure is to determine a client-specific threshold on the distance measures. Let $D_{(l)}(BS; 4, BP)$ denote the l-th order statistic in the set of impostor distances for the trained client BS, when the frontal face image of person BP from shot 4 is excluded. A threshold for person BS can by chosen as:

$$T_{BS}(4, BP) = D_{(1+\varrho)}(BS; BP, 4) \qquad \varrho = 0, 1, 2, \ldots \tag{12}$$

In the test procedure, three shots create the training set, while the fourth one is used as the test set. Each person is considered as an impostor in turn, while the remaining 36 persons are used as clients. Each client claims his/her own identity, while each impostor pretends the identity of each client in turn. By repeating the procedure four times, $4 \times 37 \times 36 = 5328$ authentic tests and the same number of impostor claims is realized in total. In each claim, the reference grids derived for each client during the training procedure are matched and adapted to the feature vectors computed at every pixel of a test frontal face image using MDLA. The resulting distance measure is compared to the threshold then. By varying the parameter ϱ, a pair of FAR and FRR can be computed. Accordingly, we may create a plot of FRR versus FAR with ϱ being an implicit parameter. This plot is the receiver operating characteristic (ROC) curve of the authentication technique. The ROC curve of the MDLA is plotted in Figure 8. It has been found that the selection of the structuring function affects marginally the verification performance of MDLA, because the area under the ROC does not change dramatically. The equal error rate of MDLA is 9.35%, when the scaled hemisphere is used. However, the choice of the structuring function affects seriously the computational time for performing multiscale dilation-erosion. The dilation-erosion by a scaled hemisphere cannot be computed in a recursive separable fashion and is computationally more demanding. The dilation-erosion by a flat structuring function can easily be computed using a running max/min algorithm, e.g. MAXLINE [25]. Moreover, it is amenable to a scale-recursive computation, as is proposed in Section IV-A. Fast algorithms for dilation-erosion by a paraboloid structuring function also exist, as has already been described in Section IV-A. The time needed to compute the multiscale dilation-erosion for $\sigma = -9, \ldots, 9$ is tabulated in Table I. All times refer to a typical facial image of dimensions 286×350 pixels. The computations have been performed in a SUN Sparc Ultra 1 Enterprise 3000 workstation with 64 MB RAM. It is seen that the scale-recursive implementation by a flat structuring function is the fastest method.

To enhance the verification capability of the proposed method, linear projection algorithms are employed for feature selection. In addition to the reasons explained in Section IV-B, another practical reason is the following: Let $\mathcal{J}$ and N denote the dimensionality of the feature vector

(i.e., $\mathcal{J} = 19$ for raw feature vectors) and the number of feature vectors available at each grid node. LDA will break down if the inequalities $N \geq \mathcal{J} + 2$ and $\mathcal{J} \geq 2$ are not satisfied. The number of feature vectors for impostors is quite large and does not pose any difficulty. The only way to anticipate the lack of feature vectors for clients is to apply principal component analysis first for raw feature vector dimensionality reduction and to apply discriminant analysis afterwards. It has been found that six principal components can approximate the feature vectors with a mean squared error less than 5%. Moreover, an EER $\approx 9.56\%$ has been achieved by a scaled hemisphere structuring function. Thus, we have achieved feature dimensionality reduction without sacrificing the verification efficiency of the MDLA. Having solved the problem of dimensionality by applying PCA, we proceed to the performance evaluation of the combined scheme of MDLA with both PCA and LDA projections. Two cases are considered, namely, the derivation of one and two most discriminating features at each grid node. The corresponding ROCs are plotted in Figure 8. The EER of MDLA with one and two MDFs is 6.8 % and 5.4 %, respectively. It is seen that the incorporation of linear projections improves the EER by 2.44–4%. Further improvements are obtained when the weighting procedure of Section IV-D is applied to the raw MDLA. The ROC of MDLA with node weighting coefficients given by (11) is plotted in Figure 8. The EER of MDLA with node weighting coefficients given by (11) is found to be 3.7%.

Next, we compare the EER achieved by several frontal face authentication algorithms. We report results for the implementations of these algorithms tested on the same database following the above-described experimental protocol in order to offer the reader the possibility to compare their performance. To demonstrate the power of each concept, we present also the EER achieved by the raw algorithm and that achieved by fine tuning each algorithm. The EERs and the corresponding citations are tabulated in Table II. For completeness, we mention that an EER of 25.2% has been reported when 20 eigenfaces are retained [17]. It can be seen that the proposed MDLA with DPCs is ranked as the first method with respect to EER.

Recently, we have tested the performance of the raw MDLA on the extended M2VTS database that contains 295 persons' video data including other faces besides European ones.

The false rejection rates at approximately the same false acceptance rate have been measured a posteriori in the test procedure for morphological elastic graph matching, standard elastic graph matching and optimized robust correlation and are tabulated in Table III [35]. From an inspection of Table III, it is seen that MDLA outperforms the elastic graph matching and the optimized robust correlation by 4.0% and 1.75%, respectively. MDLA has also been tested under real conditions that include changes in illumination, face size differences, varying face position as well as varying facial expressions [35]. Preliminary results indicate that MDLA is less sensitive to face size variations. However, it is sensitive to lighting variations.

We conclude this section by studying the performance of MDLA with one or two most discriminating features in face recognition. Figure 9 shows the percentage of correctly identified faces within the N best matches for $N = 1, 2, \ldots, 37$, i.e., the so called cumulative match score versus rank curve. The cumulative match score has been averaged on 12 face recognition trials on the M2VTS database using all possible permutations of the available shots. It is seen that the proposed technique can also address the question "Who is present in a picture?" (i.e., a face recognition problem) besides confirming if "person X is present in a picture" (i.e., a face authentication problem).

VI. Discussion

The experimental results presented in Section V have been collected according to the common experimental protocol used in M2VTS project. In this section, we briefly discuss on the accuracy of the frontal face authentication algorithm that is based on morphological dynamic link architecture and we evaluate the performance of the algorithm following the methods presented in [20, 21].

First, we determine the number of test client and impostor claims that gives statistically significant results. Such a consideration should be part of the database design and/or the experimental protocol and it is usually based on the error rates of the state-of-the art systems. The specifications of a desired frontal face authentication algorithm on M2VTS database are shown in Table IV. Let z_u denote the u-percentile of the standardized normal distribution

(with mean 0 and variance 1). If we consider the authentication errors as Bernoulli trials, we can assert then with risk α of being wrong that a number of client claims:

$$\nu_C = \left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2 p_{FR}\,(1 - p_{FR}) \tag{13}$$

is sufficient to guarantee that the expected value of the false rejection rate, p_{FR}, and the empirical value of the false rejection rate $\hat{p}_{FR}$, are related by $|p_{FR} - \hat{p}_{FR}| \leq \varepsilon$. By substituting the specifications given in Table IV into (13), we obtain $\nu_C = 2167$. If we are interested in one-sided bounds, then $\nu_C = (\frac{z_\alpha}{\varepsilon})^2\, p_{FR}\,(1 - p_{FR}) = 1536$ [21]. For completeness, we note that the more pessimistic Chernoff bound yields $\nu_C = 3580$ client claims [21]. The experimental protocol described in section V furnishes us with 5328 client claims, but they can hardly be considered as independent identically distributed. Subsequently, the number of test impostor claims is estimated. Let p_0 be the probability that a particular impostor is accepted by the authentication system when he/she pretends to be a particular client. The probability that a given person matched against κ clients produces at least one false acceptance is the probability of false acceptance [20], i.e.:

$$\text{Prob}\{1 \leq Y \leq \kappa\} = \sum_{y=1}^{\kappa} \binom{\kappa}{y} p_0^y\,(1 - p_0)^{\kappa - y} = 1 - (1 - p_0)^\kappa = p_{FA} \tag{14}$$

where Y is a binomial random variable that counts the number of false acceptances in κ claims. By replacing p_{FA} by p in (14) and solving for p_0, we find an estimate of p_0, $\hat{p}_0 = 1.717 \times 10^{-3}$. If we consider the false acceptances as Bernoulli trials, we can assert with risk α of being wrong that the number of impostor claims:

$$\nu_I \approx \left(\frac{z_{\alpha/2}}{\varepsilon'}\right)^2 p_0\,(1 - p_0) \qquad \varepsilon' \approx \frac{\varepsilon}{\kappa} \tag{15}$$

is sufficient to guarantee that the expected value of the false acceptance rate, p_{FA}, and the empirical value of the false acceptance rate $\hat{p}_{FA}$, are related by $|p_{FA} - \hat{p}_{FA}| \leq \varepsilon$. Letting $\alpha = 5\%$ ($z_{\alpha/2} = 1.96$), $\varepsilon = 1\%$ (i.e., $\varepsilon' = 2.777 \times 10^{-4}$) and by substituting $\hat{p}_0$ into (15), we find $\nu_I = 85339$ test impostor claims. If we relax the requirement of 1% false acceptance error margin and we set $\varepsilon' \propto p_0$, i.e., $\varepsilon' = 10^{-3}$, we need $\nu_I = 6585$ test impostor claims which is at the order of magnitude provided by the protocol.

Second, we test the statistical significance of the error rates measured during the test procedure of morphological dynamic link architecture with two most discriminating features. Our frontal face authentication algorithm yields 251 false acceptances and 318 false rejections. We are interested in finding the threshold τ that determines whether we should accept the null hypothesis that the probability of false rejection (or false acceptance) is less than 6% based on the observed value, $\hat{p}_{FR} = 5.96\%$ (or $\hat{p}_{FA} = 4.71\%$). Following standard arguments of hypothesis testing [36], it can be shown that the threshold τ is given by solving the expression:

$$0.05 = \sum_{y=\tau+1}^{5328} \binom{5328}{y} 0.06^y \, 0.94^{5328-y}. \tag{16}$$

Let $\mathsf{G}(\xi)$ denote the cumulative density function of the standardized normal distribution. By approximating the binomial distribution with the normal one, we obtain:

$$\mathsf{G}\left(\frac{5328 - 5328 \cdot 0.6}{\sqrt{5328 \cdot 0.06 \cdot 0.94}}\right) - \mathsf{G}\left(\frac{\tau + 1 - 5328 \cdot 0.6}{\sqrt{5328 \cdot 0.06 \cdot 0.94}}\right) = 0.05 \tag{17}$$

which yields $\tau = 347$. Since the number of both false rejections and false acceptances is less than the threshold, we accept the null hypothesis that EER $\leq 6\%$.

VII. Conclusions

A novel dynamic link architecture based on multiscale morphological dilation-erosion has been presented for frontal face authentication. Instead of a set of Gabor filters tuned to different orientations and scales, multiscale morphological operations have been employed. Fast algorithms for multiscale morphological operations have been derived. Linear projection algorithms have been used for feature selection. An automatic weighting of the nodes according to their discriminatory power has been proposed. The performance of the proposed morphological dynamic link architecture has been tested in terms of the receiver operating characteristic for several threshold selections on the matching error in the M2VTS database. The comparison with other frontal face authentication algorithms developed within M2VTS project indicates that morphological dynamic link architecture with discriminatory power coefficients is ranked as the best algorithm with respect to the equal error rate achieved.

[1] A. Jain, R. Bolle, and S. Pankanti, Eds., *Biometrics: Personal Identification in Networked Society*, Boston, MA: Kluwer Academic, 1999.

[2] R. Chellapa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995.

[3] J. Weng, and D.L. Swets, "Face Recognition," in *Biometrics: Personal Identification in Networked Society* (A. Jain, R. Bolle, and S. Pankanti, Eds.), pp. 67–86, Boston, MA: Kluwer Academic, 1999.

[4] M. Turk, and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[5] D.L. Swets, and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence* vol. 18, no. 8, pp. 831–837, August 1996.

[6] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence* vol. 19, no. 7, pp. 711–720, July 1997.

[7] K. Etemad, and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," in *Lecture Notes in Computer Science: Audio- and Video- based Biometric Person Authentication* (J. Bigün, G. Chollet, and G. Borgefors, Eds.), vol. 1206, pp. 127–142, 1997.

[8] P. Jonathon Phillips, "Matching pursuit filters applied to face identification," *IEEE Trans. on Image Processing*, vol. 7, no. 8, pp. 1150–1164, August 1998.

[9] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R.P. Würtz, and W. Konen, "Distortion invariant object recognition in the Dynamic Link Architecture," *IEEE Trans. on Computers*, vol. 42, no. 3, pp. 300–311, March 1993.

[10] D.J. Burr, "Elastic matching of line drawings," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 3, pp. 708–713, 1981.

[11] B.S. Manjunath, R. Chellapa, and C. v.d. Malsburg, "A feature based approach to face recognition,' in *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 373–378, 1992.

[12] W. Konen, E.S. Krüger, "ZN-Face: A system for access control using automated face recognition," in *Proc. of the First Int. Conf. on Automatic Face and Gesture Recognition*, pp. 18–23, Zürich, Switzerland, 1995.

[13] L. Wiskott, J.-M. Fellous, N. Krüger, and C. v.d. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, July 1997.

[14] R.P. Würtz, "Object recognition robust under translations, deformations, and changes in the background," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 769–775, July 1997.

[15] J. Zhang, Y. Yan, and M. Lades, "Face recognition: Eigenface, elastic matching and neural nets," *Proceed-*

ings of the IEEE, vol. 85, no. 9, pp. 1423-1435, September 1997.

[16] L. Wiskott, "Phantom faces for face analysis," in *Lecture Notes in Computer Science: Computer Analysis of Images and Patterns* (G. Sommer, K. Daniilidis, J. Pauli, Eds.), vol. 1296, pp. 480–487, 1997.

[17] B. Duc, S. Fischer, and J. Bigün, "Face authentication with Gabor information on deformable graphs," *IEEE Trans. on Image Processing*, vol. 8, no. 4, pp. 504–516, April 1999.

[18] N. Krüger, "An algorithm for the learning of weights in discrimination functions using a priori constraints," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 764–768, July 1997.

[19] P.T. Jackway, and M. Deriche, "Scale-space properties of the multiscale morphological dilation-erosion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 38–51, January 1996.

[20] W. Shen, M. Surette, and R. Khanna, "Evaluation of automated biometrics-based identification and verification systems," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1464-1478, September 1997.

[21] I. Guyon, J. Makhoul, R. Schwartz, and V. Vapnik, "What size test gives good error rates estimates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 52–64, January 1998.

[22] P.T. Jackway, "Morphological scale-space with application to three-dimensional object recognition," Ph.D. Dissertation, Queensland University of Technology, Australia, 1994.

[23] G. Yang, and T.S. Huang, "Human face detection in a complex background," *Pattern Recognition*, vol. 27, no. 1, pp. 53–63, 1994.

[24] C. Kotropoulos, and I. Pitas, "Rule-based face detection in frontal views," in *Proc. of the IEEE Int. Conf. on Acoust, Speech and Signal Proc.(ICASSP 97)*, pp. 2537–2540, Munich, Germany 1997.

[25] I. Pitas, and A.N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*. Norwell, MA: Kluwer Academic Publ., 1990.

[26] R. van den Boomgaard, and L. Dorst, "The morphological equivalent of Gaussian scale-space," in *Gaussian Scale-Space Theory* (J. Sporring, M. Nielsen, L. Florack, and P. Johansen, Eds.), pp. 203–220, Boston, MA: Kluwer Academic, 1997.

[27] K. Fukunaga, *Statistical Pattern Recognition*. San Diego, CA: Academic Press, 1990.

[28] S. Pigeon, and L. Vandendorpe, "The M2VTS multimodal face database," in *Lecture Notes in Computer Science: Audio- and Video- based Biometric Person Authentication* (J. Bigün, G. Chollet and G. Borgefors, Eds.), vol. 1206, pp. 403–409, 1997.

[29] R.H.J.M. Otten, and L.P.P.P. van Ginneken, *The Annealing Algorithm*. Norwell, MA: Kluwer Academic Publ., 1989.

[30] A.K. Jain, and R.C. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[31] D.L. Swets, and J. Weng, "Discriminant analysis and eigenspace partition tree for face and object recognition from views," in *Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp.192–197, Killington, Vermont, U.S.A., October 14-16, 1996.

[32] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, March 1998.

[33] J. Matas, K. Jonsson, and J. Kittler, "Fast face localisation and verification," *Image and Vision Computing*, vol. 17, pp. 575–581, 1999.

[34] S. Pigeon, and L. Vandendorpe, "Image-based multimodal face authentication," *Signal Processing*, vol. 69, pp. 59–79, August 1998.

[35] C. Kotropoulos, A. Tefas, and I. Pitas, "Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions," *Pattern Recognition*, accepted for publication, 1999.

[36] A. Papoulis, *Probabilities, Random Variables and Stochastic Processes*, 3/e. N.Y.:McGraw-Hill, 1991.

TABLE I

COMPUTATIONAL TIME FOR MULTISCALE DILATION-EROSION.

Structuring function	Algorithm	Time (sec)
hemisphere	straightforward	65.18
flat	straightforward	25
	running	6.98
	scale-recursive	2.00
paraboloid	separable	4.63
	recursive-separable	4.61

TABLE II

EERs achieved by several frontal face authentication algorithms

Method	EER (%)
Raw morphological dynamic link architecture	9.35
Morphological dynamic link architecture with two most discriminating features	5.4
Morphological dynamic link architecture with discriminatory power coefficients	3.7
Optimized robust correlation [32]	12.8
Optimized robust correlation [33]	5.4
Elastic graph matching based on Gabor wavelets [17]	11.8
Elastic graph matching based on Gabor wavelets with local discriminants [17]	5.4
Grey level frontal face matching [34]	8.5

TABLE III

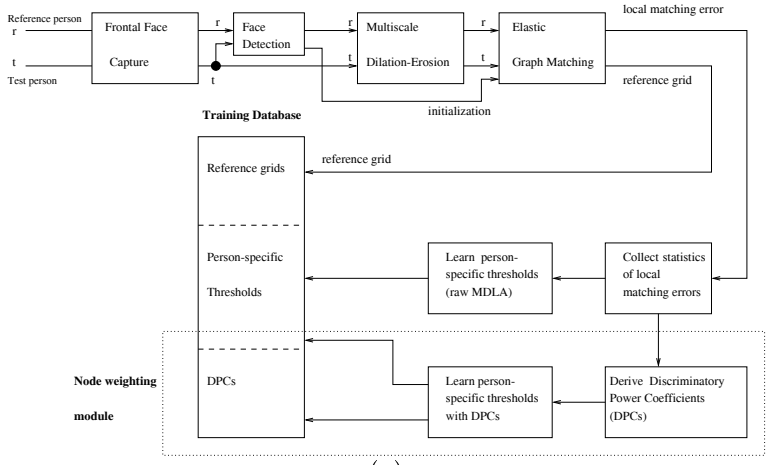
False rejection rates measured a posteriori at approximately the same false acceptance rates in the test procedure. The algorithms are tested on the extended M2VTS database.

Method	FAR (%)	FRR (%)
Elastic graph matching based on Gabor wavelets	7.67	7.25
Raw morphological dynamic link architecture	7.94	3.25
Optimized robust correlation	5.71	5.75
Raw morphological dynamic link architecture	5.52	4.00

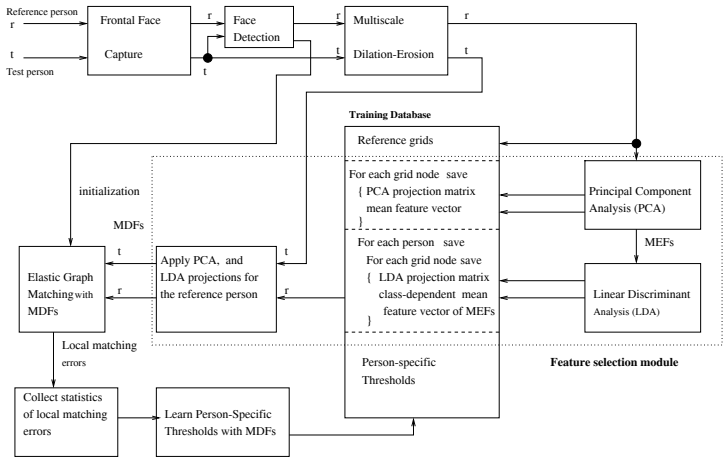
TABLE IV

DESIRED SPECIFICATIONS OF A FRONTAL FACE AUTHENTICATION SYSTEM.

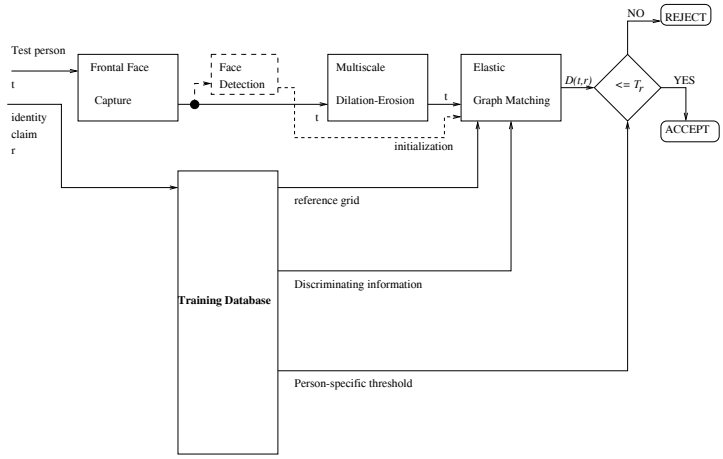
EER p	number of clients κ	confidence $1-\alpha$	margin of error ε
6 %	36	95%	1 %



(a)

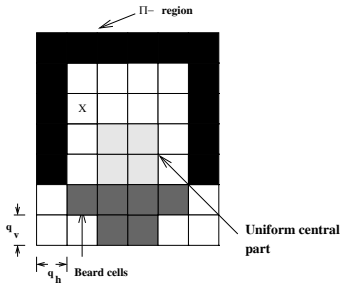


(b)



(c)

Fig. 1. (a) Basic block diagram of MDLA during the training phase. (b) A more advanced feature selection procedure that can be used during the training phase. (c) Block diagram of MDLA during the test phase.



(a)

Fig. 2. Abstract face model at the resolution level of the quartet image.

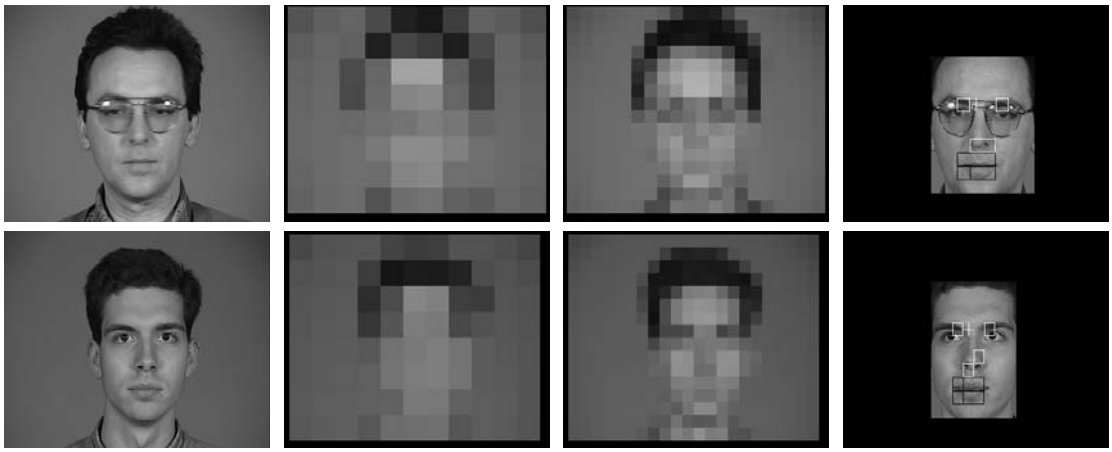


Fig. 3. Two frontal face images from M2VTS database. Their quartet and octet images are shown in the second and third column, respectively. The outcome of face detection algorithm is depicted in the last column.

Fig. 4. Dilated and eroded images for scales 1–9 by a scaled hemisphere.

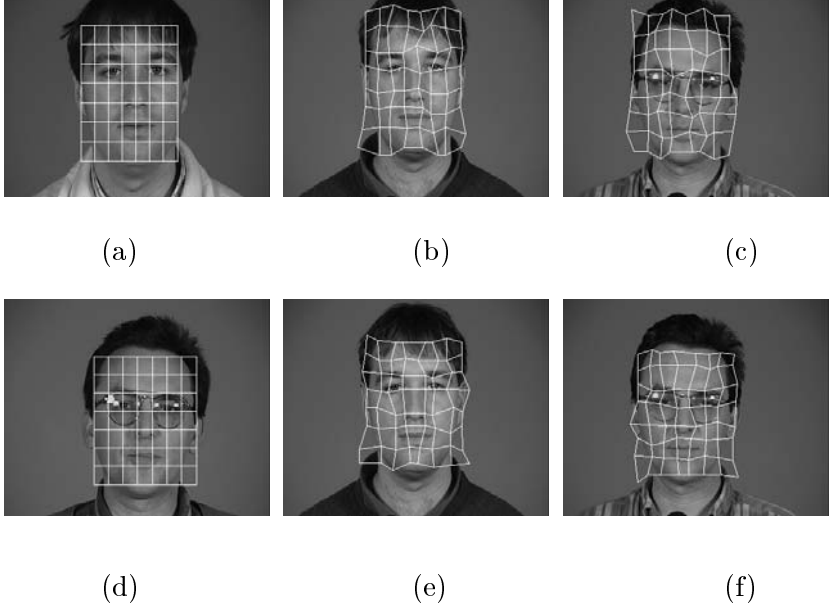


Fig. 5. Grid matching procedure in MDLA: (a) Model grid for person BS. (b) Best grid for test person BS after elastic graph matching with the model grid. (c) Best grid for test person LV after elastic graph matching with the model grid for person BS. (d) Model grid for person LV. (e) Best grid for test person BS after elastic graph matching with the model grid for LV. (f) Best grid for test person LV after elastic graph matching with the model grid.

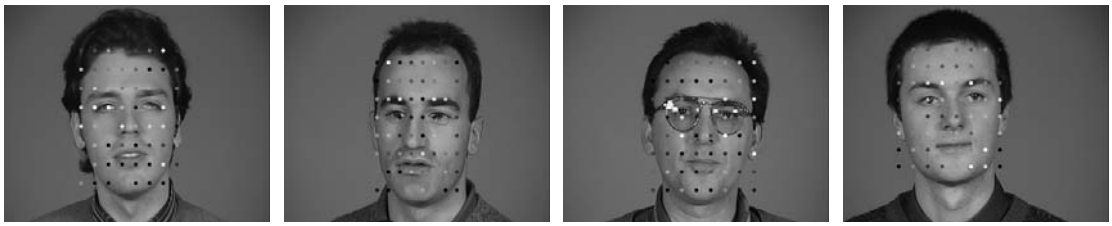


Fig. 6. Discriminatory Power Coefficients of the grid nodes in MDLA. The brighter a node is the bigger discriminatory power possesses. The intensity of the nodes is normalized for visualization purposes.

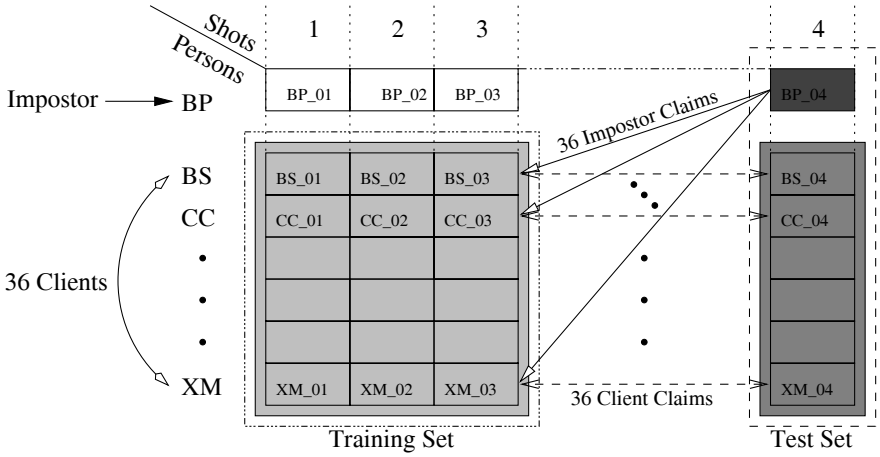


Fig. 7. Experimental Protocol.

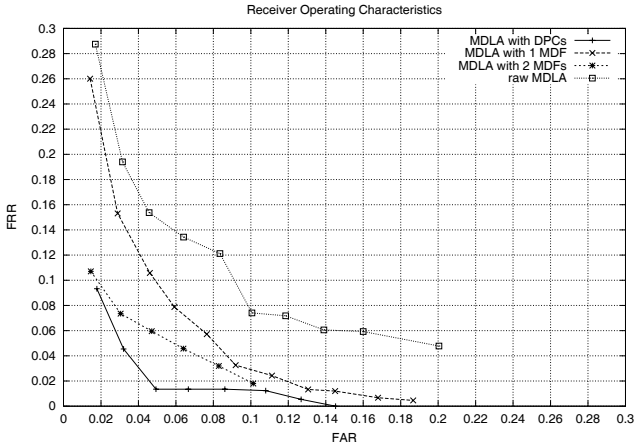


Fig. 8. Receiver Operating Characteristics of MDLA with one and two MDFs as well as when discriminatory power coefficients are derived at each grid node.

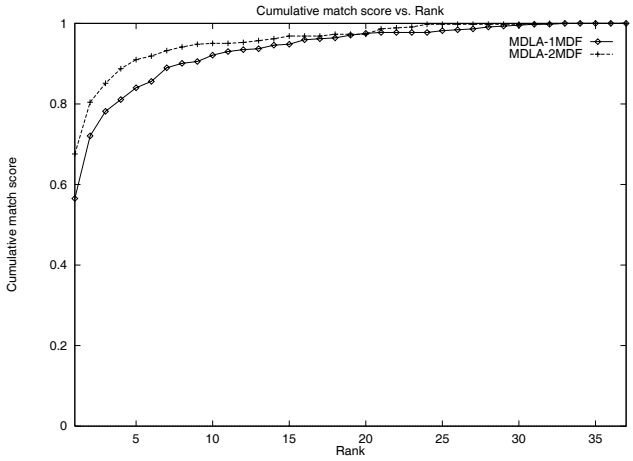


Fig. 9. Performance of MDLA with one or two MDFs in face recognition measured by the cumulative match score versus the top rank averaged on 12 trials on the M2VTS database.