Learning Discriminant Person-Specific Facial Models using Expandable Graphs

Stefanos Zafeiriou[†], Anastasios Tefas[†,*] *Member, IEEE,* and Ioannis Pitas[†] *Senior Member, IEEE,*

[†]Aristotle University of Thessaloniki

Department of Informatics

Box 451

54124 Thessaloniki, Greece

[*]Technological Educational Institute of Kavala

Department of Information Management

65404 Kavala, Greece

Address for correspondence :

Professor Ioannis Pitas

Aristotle University of Thessaloniki

54124 Thessaloniki

GREECE

Tel. ++ 30 231 099 63 04

Fax ++ 30 231 099 63 04

email: pitas@zeus.csd.auth.gr

Abstract

In this paper, a novel algorithm for finding discriminant person-specific facial models is proposed and tested for frontal face verification. The most discriminant features of a person's face are found and a deformable model is placed in the spatial coordinates that correspond to these discriminant features. The discriminant deformable models, for verifying the person's identity, that are learned through this procedure are elastic graphs that are dense in the facial areas considered discriminant for a specific person and sparse in other less significant facial areas. The discriminant graphs are enhanced by a discriminant feature selection method for the graph nodes in order to find the most discriminant jet features. The proposed approach significantly enhances the performance of elastic graph matching in frontal face verification.

Index Terms

Elastic graph matching, expandable graphs, linear discriminant analysis, frontal face verification.

I. INTRODUCTION

A well known technique for face recognition and verification is the *elastic graph matching* (EGM) algorithm [1]. In EGM, a reference object graph is created by overlaying a rectangular elastic sparse graph on the object image and then calculating a Gabor wavelet bank response at each graph node. The graph matching process is implemented by a stochastic optimization of a cost function which takes into account both jet similarities and grid deformations. A two stage coarse-to-fine optimization procedure suffices for the minimization of such a cost function.

Since its invention, EGM for face verification and recognition has been a very active research field [2]-[11]. In [2], it has been shown that EGM outperforms eigenfaces and autoassociation classification neural networks for face recognition. In [3], the graph structure has been enhanced by introducing a stack like structure, the so-called *bunch graph*, and has been tested for face recognition. For every node in the bunch graph structure, a set of jets has been measured for different instances of a face (e.g., with opened or closed mouth, eyes opened or shut). This way, the bunch graph representation could cover a variety of possible changes in the appearance of a face. In [4], the bunch graph structure

has been used for determining facial characteristics, such as beard or presence of glasses, or a even a person's sex.

In [6], EGM has been proposed and tested for frontal face verification and different choices for the elasticity of the graph have been investigated. A variant of the standard EGM, the so-called *morphological elastic graph matching* (MEGM), has been proposed for frontal face verification and tested for various recording conditions [7]-[9]. In MEGM, the Gabor features have been replaced by multiscale morphological features obtained through dilation-erosion of the facial image using a structuring function [12]. In [7], [9], the standard coarse to fine approach [6] for EGM has been replaced by a simulated annealing method that optimizes a cost function of the jet similarity measures subject to node deformation constraints. The multiscale morphological analysis has proven to be suitable for facial image analysis and MEGM has given better verification results than the standard EGM approach, without having to compute the computationally expensive Gabor filter bank output. Another variant of EGM has been presented in [10], where morphological signal decomposition has been used instead of the standard Gabor analysis [6].

Discriminant techniques have been employed in order to enhance the recognition and verification performance of the EGM. The use of linear discriminating techniques at the feature vectors for selecting the most discriminating features has been proposed in [6], [7], [9]. Several schemes that aim at weighting the graph nodes according to their discriminatory power have also been proposed in [7], [9], [11], [13].

Little or no research has been conducted concerning the type of graphs that is more appropriate for face recognition/verification. The sparse graph that has been used for face representation in the literature is:

- either an evenly distributed graph placed over a rectangular image region [6], [7], [9], [10], [13]
- or a graph that is placed on preselected nodes that correspond to some fiducial facial points (e.g., nose, eyes, etc.) [3], [4].

Intuitively, one may think that graphs with nodes placed at specified fiducial points may perform better than the rectangular graphs. However, such graphs are more difficult to be applied automatically, since they require a detection module to find the precise coordinates of the facial features in the reference

images or, in many cases, manual feature selection [3] is applied. On the contrary, an evenly distributed rectangular graph is easier to handle automatically, since only a face detection algorithm is needed in order to find an initial approximation of the rectangular facial region [6], [7], [9], [10], [13]. Figure 1 shows the two different types of facial graphs used in elastic graph matching algorithms.

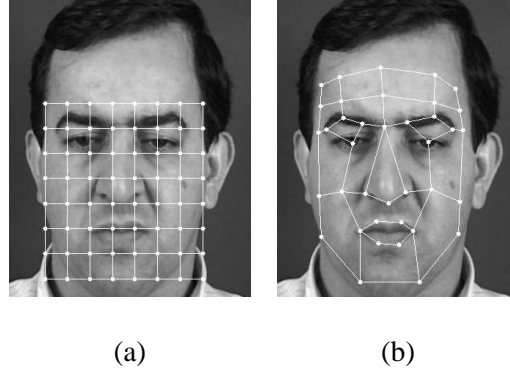(a)   (b)

Fig. 1.   Different types of facial graphs : a) rectangular graph c) graph with nodes at fiducial points.

In [4], [6], [7], [9], [11], [13] the discriminant analysis has been used for finding either linear discriminant transforms for feature selection at graph nodes or for discriminant weighting of the local node similarity measures. In this paper, we extend the use of discriminant analysis for finding the most discriminant facial features of a person's face. To do so, we introduce a discriminant analysis that produces a graph whose nodes correspond to discriminant facial points of a person.

In order to find such graphs, we introduce a heuristic cost optimization algorithm, which has as outcome the graph that optimizes a preselected discriminant cost. The cost is formed by calculating the significance of each node using discriminant values like the ones proposed in [6]-[8]. We assume that nodes with high discriminant values correspond to facial points with high discriminant capability. Then, we try to represent, in a better way, the corresponding neighborhood by adding more nodes around the original one. This practically means that we expand the nodes that are considered to be discriminant. This way, graphs that are person specific and have nodes placed at discriminant facial features, are obtained. Moreover, in order to improve the performance of the new discriminant graphs we apply a discriminant feature extraction method at each graph node which is a variant of the method proposed in [7], [8]. The proposed algorithm has been applied to frontal face verification on the XM2VTS database [14] and the Color FERET database [15], [16], [17], using a modified

morphological multiscale analysis proposed in [18].

Summarizing, the contributions of this paper are:

- The introduction of the use of person specific graphs placed on his/her discriminant facial features. These graphs substitute the rectangular evenly distributed graphs [7], [8] and the graphs having their nodes manually located at fiducial points [3].

- A novel heuristic cost optimization algorithm that automatically finds such graphs.

- The application of a two class (genuine versus impostor samples) linear discriminant transform at the nodes of the new graphs for discriminant feature selection.

Moreover, we illustrate the efficiency of our approach in the frontal face verification problem.

The rest of the paper is organized as follows. In Section II, the elastic graph matching algorithm is revisited and the problem is stated. In Section III, the algorithm that is used for learning person specific facial models along with the discriminant analysis of feature vectors are presented. Experimental results are depicted in Section IV. Finally, conclusions are drawn in Section V.

II. ELASTIC GRAPH MATCHING REVISITED

In the first step of the EGM algorithm, a sparse graph that is suitable for face representation is selected [3], [6], [7]. The facial image region is analyzed and a set of local descriptors is extracted at each graph node. Analysis is usually performed by building an information pyramid using scale-space techniques. In the standard EGM, a 2D Gabor based filter bank has been used for image analysis [1]. The output of multiscale morphological dilation-erosion operations or the morphological signal decomposition at several scales are nonlinear alternatives of the Gabor filters for multiscale analysis and both have been successfully used for facial image analysis [7], [9], [10], [18]. At each graph node that is located at image coordinates $\mathbf{x}$, a jet (feature vector) $\mathbf{j}(\mathbf{x})$ is formed:

$$\mathbf{j}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_M(\mathbf{x})]^T, \tag{1}$$

where $f_i(\mathbf{x})$ denotes the output of a local operator applied to the image f at the i-th scale or at the i-th pair (scale, orientation) and M is the jet dimensionality. The next step of EGM is to translate and deform the reference graph on the test image in order to find the correspondences of the reference graph nodes on the test image. This is accomplished by minimizing a cost function that employs

node jet similarities and, in the same time preserves, the node neighborhood relationships. Let the subscripts t and r denote a test and a reference facial image (or graph), respectively. The L_2 norm between the feature vectors at the l-th graph node of the reference and the test graph is used as a similarity measure between jets, i.e.:

$$C_f(\mathbf{j}(\mathbf{x}_t^l), \mathbf{j}(\mathbf{x}_r^l)) = ||\mathbf{j}(\mathbf{x}_r^l) - \mathbf{j}(\mathbf{x}_t^l)||. \tag{2}$$

Let $\mathcal{V}$ be the set of all graph vertices of a certain facial image. For the rectangular graphs, all nodes, except from the boundary nodes, have exactly four connected nodes. Let $\mathcal{H}(l)$ be the four-connected neighborhood of node l. In order to quantify the node neighborhood relationships using a metric, the local node deformation is used:

$$C_d(\mathbf{x}_t^l, \mathbf{x}_r^l) = \sum_{\xi \in \mathcal{H}(l)} ||(\mathbf{x}_t^l - \mathbf{x}_r^l) - (\mathbf{x}_t^\xi - \mathbf{x}_r^\xi)||. \tag{3}$$

The objective is to find a set of vertices $\{\mathbf{x}_t^l(r), l \in \mathcal{V}\}$ in the test image that minimizes the cost function:

$$C(\{\mathbf{x}_t^l\}) = \sum_{l \in \mathcal{V}} \{C_f(\mathbf{j}(\mathbf{x}_t^l), \mathbf{j}(\mathbf{x}_r^l)) + \lambda C_d(\mathbf{x}_t^l, \mathbf{x}_r^l)\}. \tag{4}$$

The jet of the l-th node that has been produced after the matching procedure of the graph of the reference person r in the image of the test person t, is denoted as $\mathbf{j}(\mathbf{x}_t^l(r))$. This notation is used due to the fact that different reference graphs r result to different test jets $\mathbf{j}(\mathbf{x}_t^l(r))$. Thus, the jet of the l-th node of the test graph t is a function of the reference graph r. The notation $\mathbf{j}(\mathbf{x}_r^l)$ is used only when the l-th node is in a preselected position of a facial image.

In [7], the optimization of (4) has been interpreted as a simulated annealing with additional penalties imposed by the graph deformations. Accordingly, (4) can be simplified to the minimization of:

$$\begin{aligned} D_t(r) &= \sum_{l \in \mathcal{V}} \{C_f(\mathbf{j}(\mathbf{x}_t^l), \mathbf{j}(\mathbf{x}_r^l))\} \text{ subject to} \\ \mathbf{x}_t^l &= \mathbf{x}_r^l + \mathbf{s} + \boldsymbol{\delta}_l, \ \ ||\boldsymbol{\delta}_l|| \le \delta_{\max}, \end{aligned} \tag{5}$$

where $\mathbf{s}$ is a global translation of the graph and $\boldsymbol{\delta}_l$ denotes a local perturbation of the graph nodes. The choices of λ in (4) and $\delta_{\max}$ in (5) control the rigidity/plasticity of the graph [6], [7]. After the matching procedure, the distance $D_t(r)$ is used as a quantitative measure for the similarity of two faces [6], [7], [8].

It is obvious that the standard EGM algorithm does not use any information about the facial classes used for verification. In order to treat this problem of EGM in [7], [8] linear discriminant techniques have been used for node feature selection or for node similarity measure weighting.

In this paper we propose a method that answers to the question "which facial points should be taken into consideration in order to create a graph that is discriminant for a particular reference person r?". Then, we create a graph using these fiducial points and use it as a model when an identity claim occurs. Moreover, we answer to the question "which features of each jet are discriminant?" by solving the optimization problem in [7], [8] using methods inspired by [19], [20], [21] for optimizing the Fisher's discriminant ratio.

III. FINDING DISCRIMINANT PERSON-SPECIFIC MODELS USING EXPANDABLE GRAPHS

In Section III-A, we formally define the optimization problem which produces the discriminant graphs. In Section III-B, we describe a heuristic algorithm that finds the person specific discriminant graphs. In Section III-C, the discriminant analysis used for feature selection at the graph nodes is presented. In the following, $\mathbf{m}(\mathcal{X})$ denotes the mean vector of a set of vectors $\mathcal{X}$ and $N(\mathcal{X})$ denotes its cardinality. When $\mathcal{X}$ is a set of scalar values their mean will be denoted as $m(\mathcal{X})$ and their variance as $\sigma^2(\mathcal{X})$.

A. Setting the Optimization Problem

Let $\mathcal{F}_C^l(r)$ and $\mathcal{F}_I^l(r)$ be the sets of the jets of the l-th node that correspond to genuine and impostor claims related to person r, respectively. In order to define the similarity of a test jet $\mathbf{j}(\mathbf{x}_t^l(r))$ with the class of reference jets for the same node, we use the following norm [7]:

$$c_t^l(r) = ||\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r))||^2. \tag{6}$$

Let $\mathcal{L}_C^l(r)$ and $\mathcal{L}_I^l(r)$ be the sets of local similarity values $c_t^l(r)$ that correspond to genuine and impostor claims, respectively. A possible measure for the discriminant power of the l-th node is the Fisher's discriminant ratio [22]:

$$p_1^l(r) = \frac{(m(\mathcal{L}_C^l(r)) - m(\mathcal{L}_I^l(r)))^2}{\sigma^2(\mathcal{L}_C^l(r)) + \sigma^2(\mathcal{L}_I^l(r))}. \tag{7}$$

In [7], [9] it has been proposed to weight the graph nodes after the elastic graph matching using the coefficients $p_1^l(r)$ in order to form a similarity measure between graphs. Another possible measure of the discriminant power of a graph node is the following:

$$p_2^l(r) = \frac{\frac{1}{N(\mathcal{L}_I^l(r))} \sum_{c_t^l(r) \in \mathcal{L}_I^l(r)} c_t^l(r)}{\frac{1}{N(\mathcal{L}_C^l(r))} \sum_{c_t^l(r) \in \mathcal{L}_C^l(r)} c_t^l(r)}. \qquad (8)$$

The measure (8) increases when the impostor local similarity measures for the graph node are high and/or the local similarity measures for the genuine class are small.

By summing the discriminant coefficients for a certain graph setup g we have:

$$E_g(r) = \frac{1}{L} \sum_{l=1}^{L} p^l(r) \qquad (9)$$

where L is the total number of nodes. This is the mean of all the discriminant ratios and is a characteristic measure for a particular graph setup of some reference person r. The scalar $p^l(r)$ could be any discriminant measure like the ones defined in (7) and (8) and is estimated using a training set of impostor and genuine claims. The measure defined in (9) creates an ordering relationship between graphs. That is, for two graphs g_1 and g_2 and for some reference person r if $E_{g_1}(r) < E_{g_2}(r)$ the graph g_2 is considered more discriminant than the graph g_1. Practically, the nodes of the graph g_2 are placed in more discriminant facial points than the nodes of g_1. Figure 2 shows two different graph setups g_1 and g_2 with different $E_g(r)$ values. Both graphs have 64 nodes. The graph depicted in Figure 2b is found experimentally to be more discriminant than the rectangular graph depicted in Figure 2a since $E_{g_1}(r) < E_{g_2}(r)$.
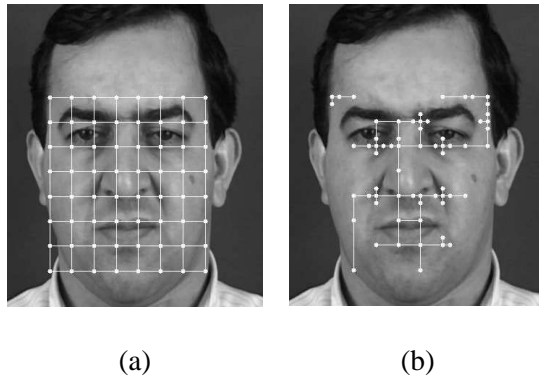

(a)   (b)

Fig. 2. a) A rectangular sparse graph; b) a graph that is more discriminant than the rectangular graph.

The previous analysis leads to an optimization procedure in order to find the graph g that has maximum $E_g(r)$. The desired properties (constraints) of the graph g apart from having maximum $E_g(r)$ are:

- The graph should have a relatively small number of nodes so that the elastic graph matching procedure has low computational cost.

- The nodes should not be very close to each other in order to avoid redundant use of the same discriminant information.

Formally, the above optimization problem can be written as:

$$\acute{g} = \arg\max_g E_g(r) \ \text{subject to}$$

$$||\mathbf{x}_r^l - \mathbf{x}_r^j|| \geq \Delta, \forall \ l, j \text{ nodes with } l \neq j \tag{10}$$

$$L = \text{constant}$$

where Δ is a preselected threshold that controls the density of the graph.

B. A Heuristic Optimization Approach

In order to solve the maximization problem (10), someone has to follow a heuristic optimization approach, since exhaustive search is not feasible. The ideal solution is a subgraph of a rectangular $N \times M$ graph that obeys the constraints (10), where N is the width and M is the height of the facial area (in pixels), respectively. The overall maximum of (10) can be found by measuring the discriminant power of each node (one node is placed at each pixel) and keeping a subset of nodes of the rectangular $N \times M$ graph that maximizes the constraint optimization problem (10). The previously described procedure guarantees that it will find one of the optimal solutions of (10) (since there is no guarantee that there is a unique solution) but has a prohibitive computational cost in the training phase, since the number of the nodes of the graphs that should be matched to impostor and to genuine images in order to calculate the discriminant power of every node is extremely high (the size of the graphs for a facial area of 100×150 pixels is 15.000 nodes).

In order to avoid such a computationally infeasible procedure, a sampling of the solution space should be performed prior to optimization. Thus, we should search for a sub-optimal solution of the constraint optimization problem (10) by assuming that the desired graph is a subgraph of a more

sparse graph than the rectangular $N \times M$ one. One possible solution is to assume that the solution is a subgraph of the graph of Δ-equilateral triangles with inter-node distance equal to Δ pixels (Δ is the threshold that controls the sparsity of the graph in (10)). Another option is the most dense Δ-rectangular graph (i.e., an evenly distributed graph with nodes placed every Δ pixels). Figure 3a shows the graph with Δ-equilateral triangles, having a total of 768 nodes. Figure 3b shows the most dense Δ-rectangular graph with nodes placed every 9 pixels apart, having a total of 700 nodes. Figure 3c shows a discriminant subset of the Δ-rectangular dense graph that has only 64 nodes.
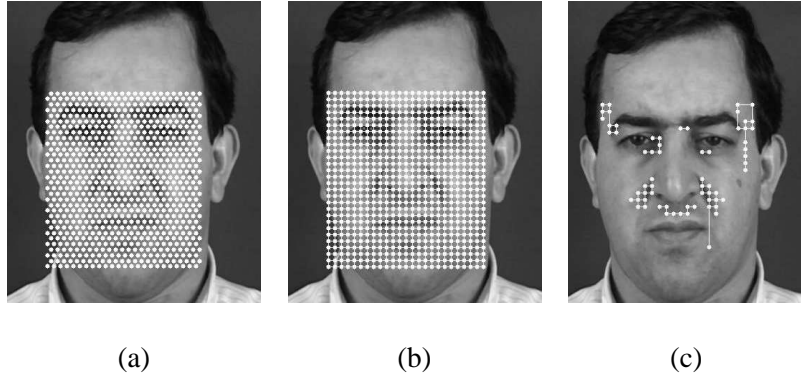

(a) (b) (c)

Fig. 3.   b) The dense graph of 9-equilateral triangles; b) a 9-rectangular dense graph; c) a discriminant subset of the rectangular dense graph.

In this work we have used rectangular graphs, even though they are not the densest possible graphs (e.g. for $\Delta = 9$ the graphs with 9-equilateral triangles are more dense that the most dense 9-rectangular graph having about 10% more nodes). We have followed this solution in order to have compliance with the majority of the proposed approaches for elastic graph matching that use rectangular graphs as well [1], [6], [7], [9], [10]. Also, the rectangular graphs have less computational complexity than other graph structures that may require more nodes to be matched. We have experimented using graphs of various types and we have observed that the type of the dense graph, for which the sub-optimal solution will be a subgraph, is not significant. Nevertheless, the proposed algorithm can be easily extended in order to support other graphs types (like the graphs of Δ-equilateral triangles).

Moreover, it is also very computationally expensive to compute the discriminant measures of the nodes of the most dense Δ-rectangular graph. Thus, a second sampling step is required. In the second sampling step, a sparse subgraph of the most dense Δ-rectangular graph is selected as

the starting point of the optimization procedure. A possible solution for the initial sampling of the most dense rectangular graph is the evenly distributed rectangular graph with a total of L nodes. A solution for selecting the initial position of the sparse graph could be a selection based on a face detection/localization algorithm. These sampling steps are inevitable in order to have a computationally feasible solution of the constrained optimization problem (10).

After sampling the solution space, an iterative algorithm that uses expandable graphs is proposed in order to find the discriminant graph. We assume that the nodes that have high discriminant values should be placed in facial regions that are indeed discriminant for the specific person. These facial regions should be better represented. This can be achieved by expanding certain nodes that possess the highest discriminant power.

In the following, the steps of the proposed algorithm are described in more detail. This procedure should be repeated for every reference person r in the database. Before starting the optimization procedure, the reference graphs for the person r should be created. The reference graphs are created by overlaying a rectangular sparse graph on the facial image region in the positions indicated by a face localization algorithm. Figure 4 shows the reference facial images with the corresponding graphs for a person in the XM2VTS database. The procedure is person specific and is executed once for every person. That is, there is no need to rerun the procedure for the other already trained persons when a new person arrives.
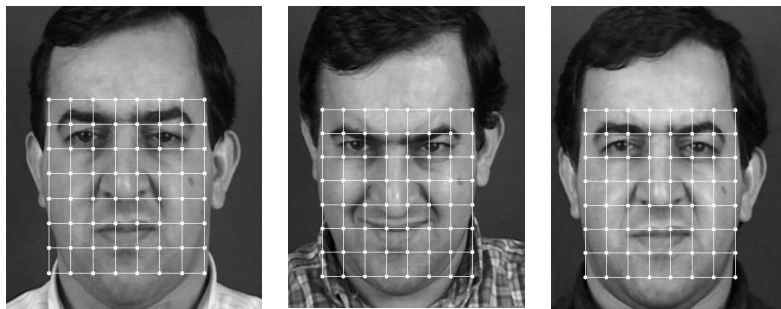

Fig. 4.   The reference facial images with the reference graphs.

Let the initial graph contain L vertices at the first iteration $i \leftarrow 1$. Let $\mathcal{B}_i$ be the set of graph vertices at the i-th iteration. The algorithm has the following steps:

Step 1 . Take the reference graphs and match them in all genuine and impostor images.

Step 2 . For each node l, calculate the measure $p^l(r)$.

Step 3 . Select a subset of the nodes with the higher discriminant value that have not been already expanded and expand them. The nodes that lie in the perimeter of the graph can be expanded only inside the facial region. Figure 5 describes pictorially this step for the rectangular graphs. For other types of graphs other expanding steps should be followed, i.e. for the graph of Δ-equilateral triangles, six nodes should be inserted in the graph at this step.
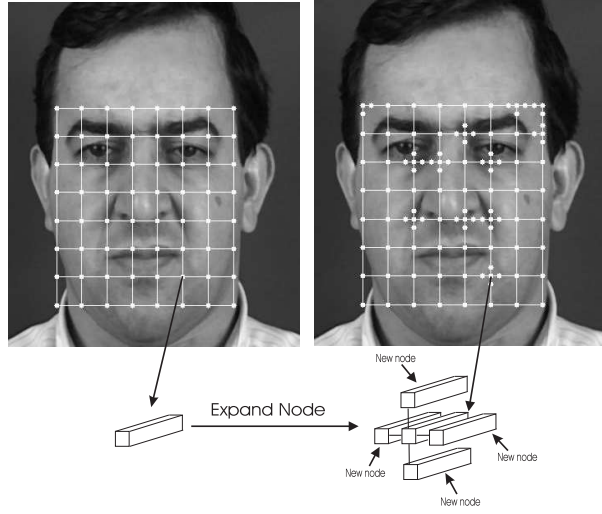

Fig. 5.   Expanding the rectangular graph.

Step 4 . Verify that the inserted nodes do not violate the graph sparseness criterion. That is, erase the new nodes that violate the criterion $||\mathbf{x}_r^l - \mathbf{x}_r^j|| < \Delta$, $\forall\ l, j$ (for the rectangular graphs used in this work, this is equivalent with checking if some of the inserted nodes have already been examined). The set of the final inserted nodes in the i-th iteration is denoted as $\mathcal{A}_i$.

Step 5 . Match locally the nodes of $\mathcal{A}_i$ in all the genuine and impostor facial images. Let $k \in \mathcal{A}_i$ be an inserted node and $\tilde{\mathbf{x}}_t^k$ be the initial coordinate vector for the node k in a test image t. The local matching procedure is the outcome of the local search:

$$\dot{\mathbf{x}}_t^k(r) = \arg\min_{\mathbf{x}_t^k} C_f(\mathbf{j}(\mathbf{x}_t^k), \mathbf{j}(\mathbf{x}_r^k)) \text{ subject to}$$
$$||\mathbf{x}_t^k - \tilde{\mathbf{x}}_t^k|| \leq \delta_{\max} \tag{11}$$

$\dot{\mathbf{x}}_t^k(r)$ is the final coordinate vector that gives the jet $\mathbf{j}(\dot{\mathbf{x}}_t^k(r))$.

Step 6 . For each node $k \in \mathcal{A}_i$, calculate its discriminant value $p^k(r)$.

Step 7 . Let $\mathcal{C}_i = \mathcal{A}_i \cup \mathcal{B}_i$. Order the nodes in $\mathcal{C}_i$ according to their discriminant power and obtain a graph g_{i+1} by keeping only the L nodes with the highest discriminant power. The set $\mathcal{B}_{i+1}$ contains the nodes of g_{i+1}.

Step 8 . If $|E_{g_{i+1}}(r) - E_{g_i}(r)| > \tau$ then $i \leftarrow i+1$ and goto Step 4 else stop.

The procedure described is a "greedy", hill climbing algorithm for finding the graph g with maximum $E_g(r)$. It always follows the direction of the best solution and, thus, it may get stuck at local maxima. An alternative way is to interpret this optimization procedure as a simulated annealing procedure [23] having $E_g(r)$ as the energy function. Moreover, in Step 4 we could expand not only some nodes that have high discriminant value but also some randomly selected ones. In this way, we may prevent the loss of some discriminant regions that have been captured neither by the initial graph nor by the expanding procedure.

Figure 6 shows the steps of the above algorithm for the first person in the training set of the XM2VTS database [14]. The algorithm starts from the rectangular 8×8 graph which is shown in the top left image of Figure 6. The second image from the top raw shows the first expanded instance of the rectangular graph. The third image shows the graph after the first node erasing step. The images that follow in a row-wise manner show a sequence of expanding/erasing steps. The algorithm converges after a total of seven expanding steps. The connections between nodes in all images have been created for visualization purposes. As can be seen in Figure 6 (lower right hand image), the discriminant graph nodes are concentrated in the areas between his cheeks and nose. This region is indeed characteristic for this particular person of the XM2VTS database.

Using as reference the facial images and graphs depicted in Figure 7, we demonstrate, the discriminant graph that is derived from the proposed procedure. As can be seen, the discriminant graph depicted at Figure 7d, is comprised of nodes around the mouth and in the lower forehead. This happens due to the fact that in some reference facial images (shown in Figure 7a,b,c), the facial region is partially occluded by hair while, in some other ones the hair does not overlap with the rectangular graph. This is of particular importance, since the proposed representation can handle the cases of people who tend to change hairstyle (or facial style in general e.g. eyeglasses, beard) frequently. In such cases, the graph nodes tend to concentrate in the interior facial region that is not affected at
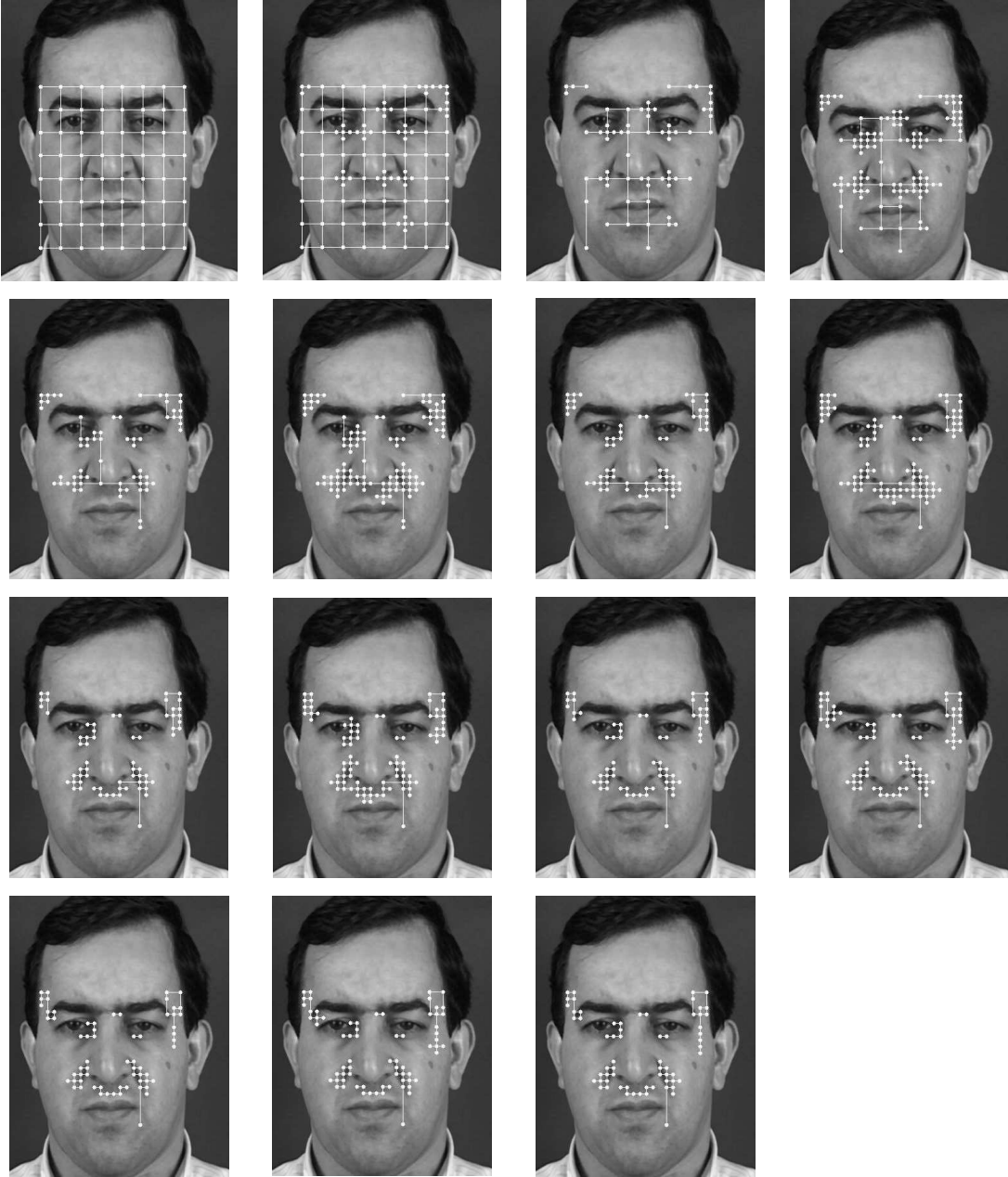
Fig. 6. The sequence of the steps for learning the graph that is comprised by expanding and erasing steps.

all by the hair style. Similarly, the proposed method will discard nodes lying in the chin area, if the

person sometimes has beard and some times not. On the contrary, if a person always has a beard

and, particularly, if he is the only one having a beard in the entire database, then the beard nodes

will be expanded and will be considered as highly significant. Unfortunately, this procedure some

times may cause overfitting (i.e., the discriminant graph may catch some facial points that are not

truly discriminant but have been accidentally upon the facial area during the training procedure). This issue will be discussed in more detail in Section IV-B.1.
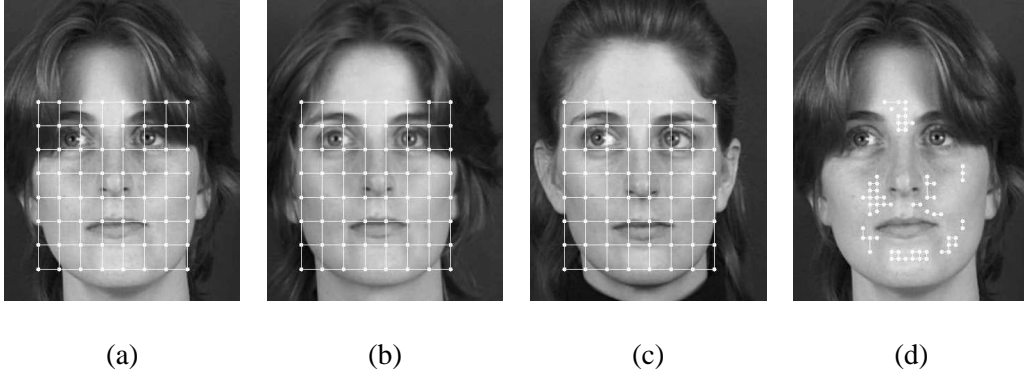

| (a) | (b) | (c) | (d) |

Fig. 7. The reference facial images with the reference graphs and the corresponding discriminant graph.

The plots of the measure $E_g(r)$ versus the number of iterations until convergence are illustrated in Figure 8a,b for the two persons in Figures 6 and 7, respectively. As can be seen for both persons, the algorithm converges after a total of seven iterations. In the eighth iteration no new node has been inserted and the algorithm converges. The elastic graph matching procedure of the
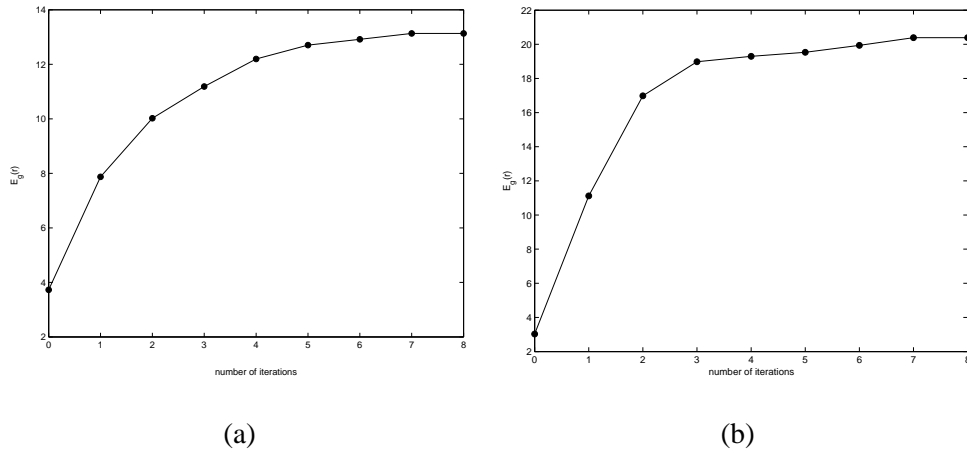

| (a) | (b) |

Fig. 8. The plot of $E_g(r)$ versus the number of iterations (a) for the person in Figure 6; (b) for the person in Figure 7

new graphs is performed using the minimization procedure indicated in the optimization problem (5). The optimization problem (5) uses global translation of the graph. That is, the components cannot be translated independently but only as part of the entire graph. In the second step, every node can be locally matched (deformed) independently as it imposed by the optimization problem (5).

C. Discriminant Feature Selection

We can improve the performance of the new graph by imposing discriminant analysis in the feature vectors of the graphs as in [7], [9]. Therefore, we should try to find a discriminant dimensionality reduction transform matrix $\mathbf{\Psi}^l(r)$ for every node l of the graph of the reference person r. In the reduced space we use the following metric in order to define the similarity of test matched jets as:

$$\hat{c}_t^l(r) = ||\mathbf{\Psi}^l(r)^T (\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))||^2. \tag{12}$$

The discriminant criterion used is similar to the one given in (8) and can be expanded as:

$$J(\mathbf{\Psi}^l(r)) = \frac{\frac{1}{N(\mathcal{L}_I^l(r))} \sum_{c_t^l(r) \in \mathcal{L}_I^l(r)} \hat{c}_t^l(r)}{\frac{1}{N(\mathcal{L}_C^l(r))} \sum_{c_t^l(r) \in \mathcal{L}_C^l(r)} \hat{c}_t^l(r)} = \frac{\mathrm{tr}[\mathbf{\Psi}^l(r)^T \mathbf{W}^l(r) \mathbf{\Psi}^l(r)]}{\mathrm{tr}[\mathbf{\Psi}^l(r)^T \mathbf{B}^l(r) \mathbf{\Psi}^l(r)]} \tag{13}$$

where the matrices $\mathbf{W}^l(r)$ and $\mathbf{B}^l(r)$ are given by:

$$\mathbf{W}^l(r) = \frac{1}{N(\mathcal{L}_I^l(r))} \sum_{\widehat{\mathbf{j}}(\mathbf{x}_t^l(r)) \in \mathcal{F}_I^l(r)} (\widehat{\mathbf{j}}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))(\widehat{\mathbf{j}}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))^T \tag{14}$$

and

$$\mathbf{B}^l(r) = \frac{1}{N(\mathcal{L}_C^l(r))} \sum_{\widehat{\mathbf{j}}(\mathbf{x}_t^l(r)) \in \mathcal{F}_C^l(r)} (\widehat{\mathbf{j}}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))(\widehat{\mathbf{j}}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))^T. \tag{15}$$

The maximum of J, in cases that $\mathbf{B}^l(r)$ is invertible, is obtained when $\mathbf{\Psi}$ is a matrix having columns the generalized eigenvectors of $\mathbf{B}^l(r)^{-1}\mathbf{W}^l(r)$. In order to satisfy the invertibility of $\mathbf{B}^l(r)$, the training set should contain at least as many samples from the genuine class as the feature dimensionality. Thus, in many cases, it is difficult to satisfy the invertibility of $\mathbf{B}^l(r)$ (refer to Section IV for more details). In [7], [9] *Principal Component Analysis* (PCA) has been initially used in order to satisfy the invertibility of $\mathbf{B}^l(r)$.

Recently, *Direct Linear Discriminant Analysis* (D-LDA) algorithms for discriminant feature extraction have been proposed [24]-[26] in order to prevent the loss of discriminatory information that occurs when a PCA step is applied prior to LDA [27]. In D-LDA the search of discriminant directions is applied in the null space (i.e., formed by the eigenvectors that correspond to null eigenvalues) of the between scatter matrix. Moreover, in [19], [20] it has been shown that two different discriminant transforms can be derived from the Fisher's LDA optimization problem, the so-called regular and irregular discriminant transforms. The novelty of the proposed approach is that it exploits all the

available information of the discriminant criterion (13), in contrast to the methods proposed in [7], [9], that discard useful information using an initial PCA step.

We apply the theory developed in [19], [20] in order to find the regular and the irregular discriminant information for the discriminant criterion (13). First, we have to deal with the null space of $\mathbf{B}^l(r)$. The vectors contained in the null space of the matrix $\mathbf{B}^l(r)$ maximize the criterion (13). That is, all the projections $\boldsymbol{\psi}_i^l(r)$ that satisfy both $\mathbf{B}^l(r)\boldsymbol{\psi}_i^l(r) = \mathbf{0}$ and $\mathbf{W}^l(r)\boldsymbol{\psi}_i^l(r) \neq \mathbf{0}$ (or equivalently $\boldsymbol{\psi}_i^l(r)^T \mathbf{B}^l(r)\boldsymbol{\psi}_i^l(r) = 0$ and $\boldsymbol{\psi}_i^l(r)^T \mathbf{W}^l(r)\boldsymbol{\psi}_i^l(r) > 0$) maximize the criterion (13). For these vectors we define the following criterion:

$$J_b(\boldsymbol{\Psi}^l(r)) = \mathrm{tr}[\boldsymbol{\Psi}^l(r)^T \mathbf{W}^l(r)\boldsymbol{\Psi}^l(r)], \boldsymbol{\Psi}^l(r) = [\boldsymbol{\psi}_1^l(r) \ldots \boldsymbol{\psi}_n^l(r)] , \|\boldsymbol{\psi}_i^l(r)\| = 1. \tag{16}$$

Using the criteria (13) and (16), two different discriminant feature extraction transforms can be derived that correspond to regular and irregular discriminant information, respectively. The first step of the discriminant dimensionality reduction is to find the orthonormal eigenvectors that correspond to non-null and null eigenvalues of $\mathbf{B}^l(r)$. Let $\boldsymbol{\Xi}_1^l(r)$ and $\boldsymbol{\Xi}_2^l(r)$ be the matrices having as columns the orthonormal eigenvectors of $\mathbf{B}^l(r)$ that correspond to non-null and to null eigenvectors, respectively. The eigenanalysis of $\mathbf{B}^l(r)$ is an easy task, since, in most cases, the graph jets do not have more than 20 features [6], [7], [9], [10].

1) Regular Discriminant Transform: Let $\hat{\mathbf{B}}^l(r) = \boldsymbol{\Xi}_1^l(r)^T \mathbf{B}^l(r)\boldsymbol{\Xi}_1^l(r)$ and $\hat{\mathbf{W}}^l(r) = \boldsymbol{\Xi}_1^l(r)^T \mathbf{W}^l(r)\boldsymbol{\Xi}_1^l(r)$ be the $\mathbf{B}^l(r)$ and $\mathbf{W}^l(r)$ matrices projected in the space spanned by the vectors in $\boldsymbol{\Xi}_1^l(r)$. In this space, $\hat{\mathbf{B}}^l(r)$ is invertible. Thus, the discriminant transform in this space can be given by the matrix $\boldsymbol{\Theta}_1^l(r)$ that has as columns the generalized eigenvectors of $\hat{\mathbf{B}}^l(r)^{-1}\hat{\mathbf{W}}^l(r)$. The regular discriminant features are given by:

$$\boldsymbol{\Gamma}_1^l(r) = \boldsymbol{\Xi}_1^l(r)\boldsymbol{\Theta}_1^l(r). \tag{17}$$

This step gives the regular discriminant feature vectors $\mathbf{j}_1(\mathbf{x}_t^l(r)) = \boldsymbol{\Gamma}_1^l(r)^T \mathbf{j}(\mathbf{x}_t^l(r))$, $\mathbf{j}_1(\mathbf{x}_t^l(r)) \in \Re^m$ with $m \leq \min(M, N(\mathcal{L}_C^l(r)) - 1)$.

2) Irregular Discriminant Transform: The irregular discriminant transform can be found by projecting $\mathbf{W}^l(r)$ in the space spanned by $\boldsymbol{\Xi}_2^l(r)$, only if $\mathbf{B}^l(r)$ is singular. This happens when the number of training genuine feature vectors is smaller than the dimensionality of the feature vector.

Let $\tilde{\mathbf{W}}^l(r) = \mathbf{\Xi}_2^l(r)^T \mathbf{W}^l(r) \mathbf{\Xi}_2^l(r)$ be the projected matrix $\mathbf{W}^l(r)$ in the null space of $\mathbf{B}^l(r)$. The discriminant transform in this space is given by $\mathbf{\Theta}_2$ that has as columns the orthonormal eigenvectors of $\tilde{\mathbf{W}}^l(r)$. The irregular transform matrix is given by:

$$\mathbf{\Gamma}_2^l(r) = \mathbf{\Xi}_2^l(r)\mathbf{\Theta}_2^l(r). \tag{18}$$

The irregular discriminant transform gives the vectors $\mathbf{j}_2(\mathbf{x}_t^l(r)) = \mathbf{\Gamma}_2^l(r)^T \mathbf{j}(\mathbf{x}_t^l(r))$ and $\mathbf{j}_2(\mathbf{x}_t^l(r)) \in \Re^n$ with $n \leq M - N(\mathcal{L}_C^l(r)) + 1$. It is easy to prove that the irregular discriminant transform performs perfect classification in the training set. That is, every genuine training jet falls in the same point in the projected space [28].

For a test jet $\mathbf{j}(\mathbf{x}_t^l(r))$, two kind of discriminant feature vectors are available using the regular and the irregular discriminant transform, i.e., $\mathbf{j}_1(\mathbf{x}_t^l(r))$ and $\mathbf{j}_2(\mathbf{x}_t^l(r))$, respectively. The similarity of these two feature vectors with the reference facial class r can be estimated using two distances:

$$\dot{c}_t^l(r) = ||\mathbf{\Gamma}_1^l(r)^T(\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))|| \text{ and } \ddot{c}_t^l(r) = ||\mathbf{\Gamma}_2^l(r)^T(\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))||.$$

The final distance is a fusion of the two measures $\dot{c}_t^l(r)$ and $\ddot{c}_t^l(r)$. The total normalized distance at the l-th node between the reference r and the test person t is:

$$\tilde{c}_t^l(r) = \frac{\dot{c}_t^l(r)}{\sum_{i \in \mathcal{U}} \dot{c}_i^l(r)} + \frac{\ddot{c}_t^l(r)}{\sum_{i \in \mathcal{U}} \ddot{c}_i^l(r)} \tag{19}$$

where $\mathcal{U}$ is the training facial image database. The new distance between faces, after discriminant analysis, is given by:

$$\tilde{D}_t(r) = \sum_{l \in V} \tilde{c}_t^l(r). \tag{20}$$

IV. EXPERIMENTAL RESULTS

A. Databases and Evaluation Protocols

The experiments were conducted in the XM2VTS database using the protocols described in [14] and in the Color FERET database using the face verification evaluation methodology presented in [15]. The description of the XM2VTS database is available at [29]. There have been many versions of FERET database released from time to time, the current version being released is the Color FERET database [30].

1) The XM2VTS database and Testing Protocol: The XM2VTS database contains 295 subjects, 4 recording sessions and two shots (repetitions) per recording session. The XM2VTS database provides two experimental setups namely, Configuration I and Configuration II [14]. Each configuration is divided in three different sets: the training set, the evaluation set and the test set. The training set is used to create genuine and impostor models for each person. The evaluation set is used for learning the verification decision thresholds. In case of multimodal systems, the evaluation set can be also used for training the fusion manager. For both configurations the training set has 200 clients, 25 evaluation impostors and 70 test impostors. The two configurations differ in the distribution of the client training and the client evaluation data. For additional details concerning XM2VTS database the interested reader can refer to [14]. Recently, frontal face verification competitions using the XM2VTS [31], [32] have been conducted. The interested reader can refer to [31], [32] and to the references therein for the tested face verification algorithms.

The performance of face verification systems is measured in terms of the *False Rejection Rate* (FRR) achieved at a fixed *False Acceptance Rate* (FAR). There is a trade-off between FAR and FRR, producing the so-called *Receiver Operating Characteristic* (ROC) curve [8], [13]. The performance of a verification system is often quoted by a particular operating point of the ROC curve where FAR=FRR. This operating point is called *Equal Error Rate* (EER).

When a verification technique is to be evaluated for a real application then the decision thresholds should be set a priori. The evaluation set is used for setting the thresholds. The same thresholds will then be used on the test set. Let FAE and FRE be the corresponding *False Acceptance Rate* (FAR) and *False Rejection Rate* (FRR) obtained from the evaluation set. Since application requirements might constrain the FAR or FRR to be within certain limits, the system is evaluated for three different vectors of decision thresholds that correspond to the operating points where FAE=0, FRE=0 or FAE=FRE. For each given threshold, the *Total Error Rate* (TER) can be obtained as the sum of FAR and FRR.

2) The FERET database and Testing Protocol: The Color FERET that has been used in our experiments contains 994 persons [30]. The evaluation methodology of the FERET database is not as strict as the XM2VTS protocol (i.e., it does not define who would be the clients or the impostors). It only requires that the training should be performed using the FA set (also referred to as gallery set)

that contains one frontal view per person. The testing sets (also referred to as probe sets) include the [15]:

- FB images: the second frontal image taken from the same set as the gallery image (FA set). Usually, the person has a different expression from the one posed in FA. The FB set contains 992 individuals.

- Duplicate images 1 (Dup1): This probe set is used for analyzing the effect of aging on the verification performance. The images are taken in a different session (different date) or taken under special circumstances (such as the subject was wearing glasses, different hair length, etc). The Dup1 probe set contains 736 images of 249 individuals.

- Duplicate images 2 (Dup2): The Dup2 probe set is a subset of the Dup1 probe set consisting of frontal images captured at least 540 days after the capture of the subject gallery image (FA set). The Dup2 probe set size is 228 images of 75 individuals.

B. Experiments in the XM2VTS database

For initializing the algorithm, described in Section III, a rectangular 8×8 graph setup has been used [6], [7]. The multiscale analysis used for filling the jets is the one proposed in [18] (i.e., *Normalized Morphological Elastic Graph Matching* (NMEGM)) and the jet dimension has been set to 19. The structuring element used in all experiments was cylindrical [7], preferred for computational reasons. Only the luminance information at a resolution of 720×576 has been considered in our experiments. We have used the value $\Delta = 9$ for the parameter that controls the sparsity of the graph.

Unlike most of the subspace techniques [33]-[35] that require a perfect manual alignment in order to perform well, the proposed algorithm has been combined with a fully automatic alignment method according to the eyes position of each facial image using the eye coordinates that have been derived from the method reported in [36]. The impact of inaccurate facial feature localization upon face recognition is pointed in [37], [38].

1) Learning the Discriminant Person Specific Graphs: The training set of the Configuration I contains 200 persons with 3 images per person. The evaluation set contains 3 images per client for genuine claims and 25 evaluation impostors with 8 images per impostor. Thus, evaluation set gives

a total of $3 \times 200 = 600$ genuine claims and $25 \times 8 \times 200 = 40.000$ impostor claims. The test set has 2 images per client and 70 impostors with 8 images per impostor and gives $2 \times 200 = 400$ genuine claims and $70 \times 8 \times 200 = 112.000$ impostor claims. The training set is used for learning the discriminant graph for each reference person, according to the algorithm presented in Section III. For initializing the algorithm, the face localization procedure proposed in [8] has been used. We have used the discriminant measure defined in (8) for learning the discriminant graphs. We have seen that similar discriminant graphs have been acquired using the measure defined in (7). We have also seen that the proposed procedure for finding discriminant graphs is quite fast, since the inserted nodes are only matched locally and in all cases the algorithm has converged using no more than seven expanding steps for the training persons in the XM2VTS database. Figure 9 shows some of the graphs that have been learned during the training procedure.
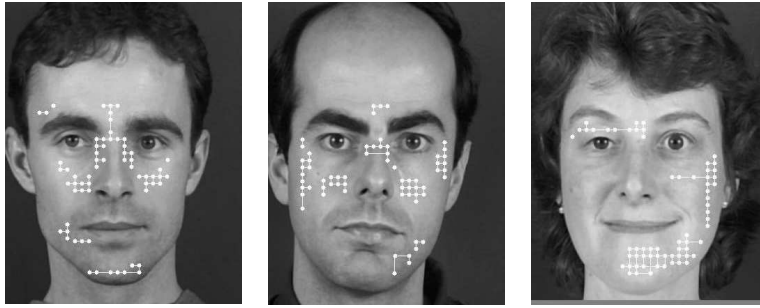

Fig. 9. Some examples of discriminant graphs learning during the training procedure in the XM2VTS database.

One possible drawback of the proposed approach is overfitting, which may occur in any discriminant technique. In our case, overfitting can occur when the discriminant graph is placed in facial points that are not truly discriminant but have accidentally appeared in the facial area during the training procedure. For example, when all the training facial images of a person contain glasses, cosmetics or tattoos, it is highly possible that the nodes of the discriminant graph are placed in the spatial coordinates that contain a visual instance of these artifacts. This fact is exemplified in Figure 10. In the training phase, only facial images with eye glasses with black thick skeleton have been present. Thus, many nodes of the discriminant graph have been placed in the spatial coordinates that correspond to eye glasses. If the test facial images contain no eyeglasses, we may have false rejections. In order to solve these problems eyeglass removing algorithms, like the ones proposed in [39], [40] and/or the

enhancement of the training set with virtual facial image samples, can be applied. There are many methods for enriching the training set with virtual samples [41]-[44].
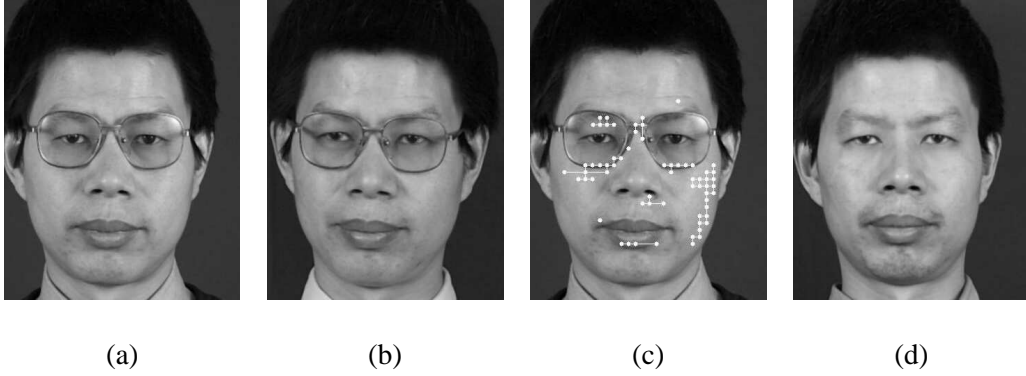

Fig. 10. (a)-(b) Some of the training samples; (c) the corresponding discriminant graph; (d) test image.

2) Learning the Discriminant Node Transforms: After learning a discriminant graph for each person, the training set is used for calculating the discriminant transforms for feature selection for each reference person r and for each node l. In the training set, 3 reference graphs per person are created. The $3 \times 2 = 6$ graphs that comprise the genuine class are created by applying elastic graph matching having one image as reference (i.e., in order to create the graph) and the other 2 images are used as test images. The impostor class contains $3 \times 3 \times 199 = 1797$ graphs. The jet dimension is 19. Thus, for a reference person r and a node l the matrix $\mathbf{B}^l(r)$ has 5 non-null dimensions and 14 null dimensions. For every node jet, two discriminant transforms can be found as described in Section III-C. The regular discriminant transform can give discriminant feature vectors with up to 5 dimensions. The feature vectors derived through the irregular discriminant transform can give feature vectors with up to 14 feature dimensions.

3) Thresholds on Similarity Measure: With or without feature vector discriminant transforms, a similarity measure between faces either by using $D_t(r)$ or $\tilde{D}_t(r)$ can be found. In order to reject or accept an identity claim, a threshold should be used on this similarity measure. For choosing the thresholds, the method proposed in [7] has been used. In detail, the similarity measures for every person are calculated in the training set and form the distance vector $\mathbf{o}(r)$. The elements of the vector $\mathbf{o}(r)$ are sorted in ascending order and are used for the person specific thresholds on the distance measure. Let $T_Q(r)$ denote the Q-th order statistic of the vector of distances, $\mathbf{o}(r)$. The threshold of

the person r is chosen to be equal to $T_Q(r)$. Let r_1, r_2 and r_3 be the 3 instances of the person r in the training set. A claim of a person t is considered valid if $\min_j\{\hat{D}_t(r_j)\} < T_Q(r)$ where $\hat{D}_t(r_j)$ is the distance between the graph of test person t and the reference graph r_j (the similarity measure $\hat{D}_t(r_j)$ could be either $D_t(r)$ or $\tilde{D}_t(r)$). Obviously when varying Q, different pairs of FAR and FRR can be created to produce the ROC curve.

4) Experimental results in Configuration I: In order to illustrate the contribution of finding discriminant person specific graphs, we have conducted the following experiments:

- NMEGM using rectangular 8×8 evenly distributed graphs without any discriminant analysis.

- NMEGM using the procedure for finding discriminant person specific graphs, described in Section III-B, without feature vector discriminant analysis (abbreviated as E-NMEGM).

- NMEGM using rectangular 8×8 evenly distributed graphs and using the discriminant analysis described in Section III-C (abbreviated as NMEGM-FD).

- NMEGM using the procedure for finding discriminant person specific graphs and using the proposed feature vector discriminant analysis (abbreviated as E-NMEGM-FD).

Since all the tested approaches are fully automatic, we have not conducted experiments using graphs placed at fiducial facial points [3] as they require manually localization of these points in all reference facial images. The error rates according to the XM2VTS Configuration I protocol are illustrated in Table I. The NMEGM without any discriminant step has given a TER=12.9% at FAE=FRE in the test set of Configuration I. When replacing the rectangular evenly distributed graphs with the person specific discriminant graphs proposed in this paper, the TER has been reduced to 6.05%. That is, an increase in performance more that 50% in terms of TER is achieved when using the proposed graphs.

When using the proposed discriminant analysis in the rectangular graphs, a TER=5.1% at FAE=FRE has been achieved (NMEGM-FD). We have kept the number of dimensions that have given the best EER in the evaluation set. It has been experimentally found that the best EER has been achieved when keeping the first 3 dimensions of the regular discriminant transform and the first 3 of the irregular transform. The best results have been achieved when using feature vector discriminant analysis in the proposed discriminant graphs keeping, as well, 3 regular and 3 irregular dimensions. This setup has given a TER= 2.9% at FAE=FRE (E-NMEGM-FD). Thus, an increase in performance of about

50% has been verified in the case that we use feature vector discriminant analysis in the discriminant graphs. For comparisons we have implemented a Fisherfaces (PCA+LDA) approach [27]. For the Fisherfaces a TER 10% has been measured at FAE=FRE using the same sets of images.

The ROC curves are depicted in Figure 11 where the EER and the operating point using the thresholds of the operating point that corresponds to EER in the evaluation set are also shown (abbreviated as EER-E in Figure 11). The operating points that corresponds to the thresholds of the EER measured in the evaluation set are shown in columns 5 and 6 of Table I.
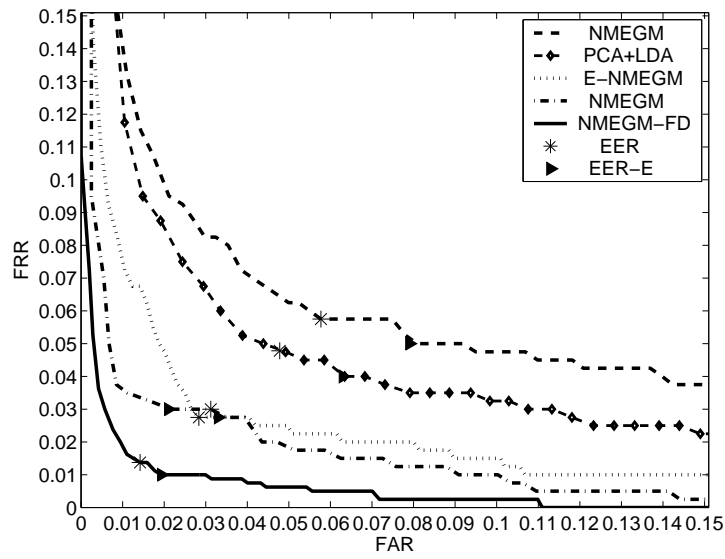

Fig. 11. ROC curves for the different discriminant variants of NMEGM in test set of the Configuration I experimental protocol of the XM2VTS database.

It has been also verified that the proposed discriminant graphs using feature vector discriminant analysis have very good performance compared with other state-of-the-art face verification algorithms. Table II shows a comparison of E-NMEGM-FD (which gave the best results) with other methods that use fully automatic alignment. The results have been acquired by the competition in XM2VTS database [32]. Obviously, E-NMEGM-FD outperforms all the approaches tested in [32] using fully automatic alignment. In [45] there has been another competition using the XM2VTS database in order to explore the effects of severe illumination changes in face verification algorithms. Only one algorithm has been tested using fully automatic alignment and has achieved a TER= 2.14%.

TABLE I

ERROR RATES ACCORDING TO XM2VTS PROTOCOL FOR CONFIGURATION I

Algorithm	Configuration I											
	Evaluation set			Test set								
	FAE=FRE	FAE(FRE=0)	FRE(FAE=0)	FAE=FRE		FRE=0		FAE=0		Total Error Rate(TER)		
				FA	FR	FA	FR	FA	FR	FAE=FRE	FRE=0	FAE=0
NMEGM	9.2	98.2	65.0	7.9	5.0	98.8	0.0	0.0	61.0	12.9	98.8	61.0
E-NMEGM	3.4	34.5	55.5	3.3	2.75	22.7	0.0	0.0	44.7	6.05	22.7	44.7
NMEGM-FD	2.2	26.5	52.3	2.1	3.0	11.0	0.5	0.2	13.2	5.1	11.5	13.4
E-NMEGM-FD	**1.2**	**14.2**	**24.5**	**1.93**	**1.0**	**11.0**	0.0	0.0	**10.7**	**2.93**	**11.0**	**10.7**

TABLE II

A COMPARISON OF TER FOR CONFIGURATION I USING FULLY AUTOMATIC REGISTRATION.

Algorithm	TER
IDIAP-Cardinaux[32]	4.7
UPV[32]	3.98
UNIS-NC[32]	3.86
E-NMEGM-FD	2.93
CAS[45]	2.14

C. Experiments in the FERET database

In many applications and experimental protocols, like the experimental protocol of the FERET database, there is only one genuine facial image available in the training set. In this case, the measures defined in (7) and (8) cannot be used for measuring the discriminant power of a graph node. Hence, in order to apply the proposed algorithm to the FERET database using the FERET evaluation protocol [15], [16], [17] we have modified the algorithm so that it can be functional having one sample available for training. For that case, we use an alternative measure based only on the numerator of (8):

$$p_3^l(r) = \frac{1}{N(\mathcal{L}_I^l(r))} \sum_{c_t^l(r) \in \mathcal{L}_I^l(r)} ||\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{j}(\mathbf{x}_r^l)||^2, \qquad (21)$$

where $\mathbf{j}(\mathbf{x}_r^l)$ is the jet of the l-th node in the only reference image of the person r. This measure is based on the impostor class and on the one genuine sample and emphasizes the client facial features that are far from the facial features of the impostors in terms of (21). Since we have only one sample available for training, it is not possible to use the methods in Section III-C for discriminant feature selection. Thus, we have used only PCA for feature selection in the nodes of the elastic graphs. For PCA feature selection the number of dimensions has been chosen to be 5. Some discriminant graphs from the FERET database are shown in Figure 12. In the FERET database, NMEGM algorithm has been combined with a fully automatic alignment method according to the eyes position of each facial image using the eye coordinates that have been derived from the method reported in [46]. For comparisons we have also implemented a PCA approach [47], the so-called Eigenfaces, and we have applied it to the same set of images. Finally, we have trained an LDA approach using the XM2VTS database and tested it in the FERET database but, unfortunately the results have been very poor even in comparison with PCA. This is partially attributed to the different environmental conditions between the two databases.
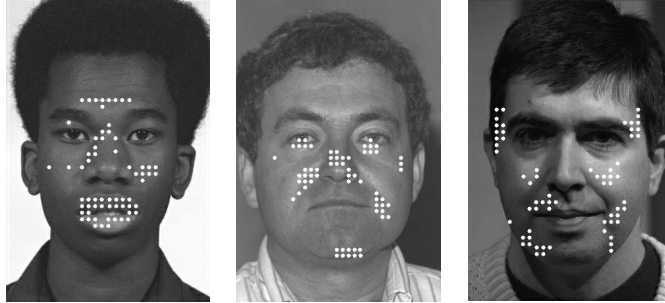

Fig. 12. Some examples of discriminant graphs learning during the training procedure in the FERET database.

The FERET evaluation methodology requires testing to three sets of images namely FB, Dup1 and Dup2. For performing the experiments in the FB set, the FA set has been randomly partitioned to clients and impostors having 50% (i.e., 497 individuals) as clients and 50% as impostors. In other words, the gallery consists of 497 individuals. The FB contains one image per person. Thus we have a total of 497 genuine claims. The remaining 495 images of the FB set have been used for impostor claims leading to a total of $495 \times 497 = 246015$ impostor claims. We have repeated the experiments four times, using different partitions for clients and impostors, and the average ROC curve has been

estimated. In the FERET evaluation protocol the ROC curves, the Probability of correct verification (PCV) versus the FA is used [15]. PCV is equal to 1-FR. The tested approaches have been the raw NMEGM (without any discriminant analysis), the E-NMEGM and two approaches NMEGM-PCA and E-NMEGM-PCA using PCA feature selection in the nodes of the rectangular graphs, in NMEGM, and in the nodes of the discriminant graphs of the E-NMEGM approach. The ROC curves for the FB set are shown in Figure 13a. As can be seen, the E-NMEGM-PCA approach outperforms all other approaches.

The Dup1 set consists of 736 images of 249 individuals. Thus, a subset of FA has been used for experiments consisting of 249 individuals (the same ones that participate in the Dup1 set). The subset of FA set has been randomly partitioned to clients and impostors having 50% (i.e., 125 individuals) as clients and the remaining 124 individuals as impostors. We have repeated the experiments four times and the average ROC curves has been estimated for all the tested approaches. In Dup1 the number of images per individual are not the same for each person. The mean number of genuine claims have been 367 and the mean number of impostor claims has been $367 \times 125 = 45875$. The final test set is the Dup2 set that consists of 228 images of 75 individuals. As in the experiments in Dup1, a subset of FA has been used for experiments in Dup2 consisting of 75 individuals. Accordingly, the subset of FA has been randomly partitioned to clients and impostors having 50% (i.e., 38 individuals) as clients and remaining 37 individuals as impostors. We have repeated the experiments four times and the average ROC curve has been estimated. In Dup2 the number of images per individuals are not the same for each person. The mean number of genuine claims has been 112 and the mean number of impostor claims has been $36 \times 116 = 4408$. The ROC curves for the Dup1 and Dup2 set are shown in Figures 13b and c, respectively. The tests in Dup1 and Dup2 have been shown that the proposed discriminant graphs outperform all other tested approaches, even in very different conditions to the conditions that the training has been performed.

The experiments in the Color FERET database verified the results from the XM2VTS database, i.e. that the selection of nodes that consist the facial graph is important and can significantly increase the performance of elastic graph matching for frontal face verification even when only one sample is available in the training set.
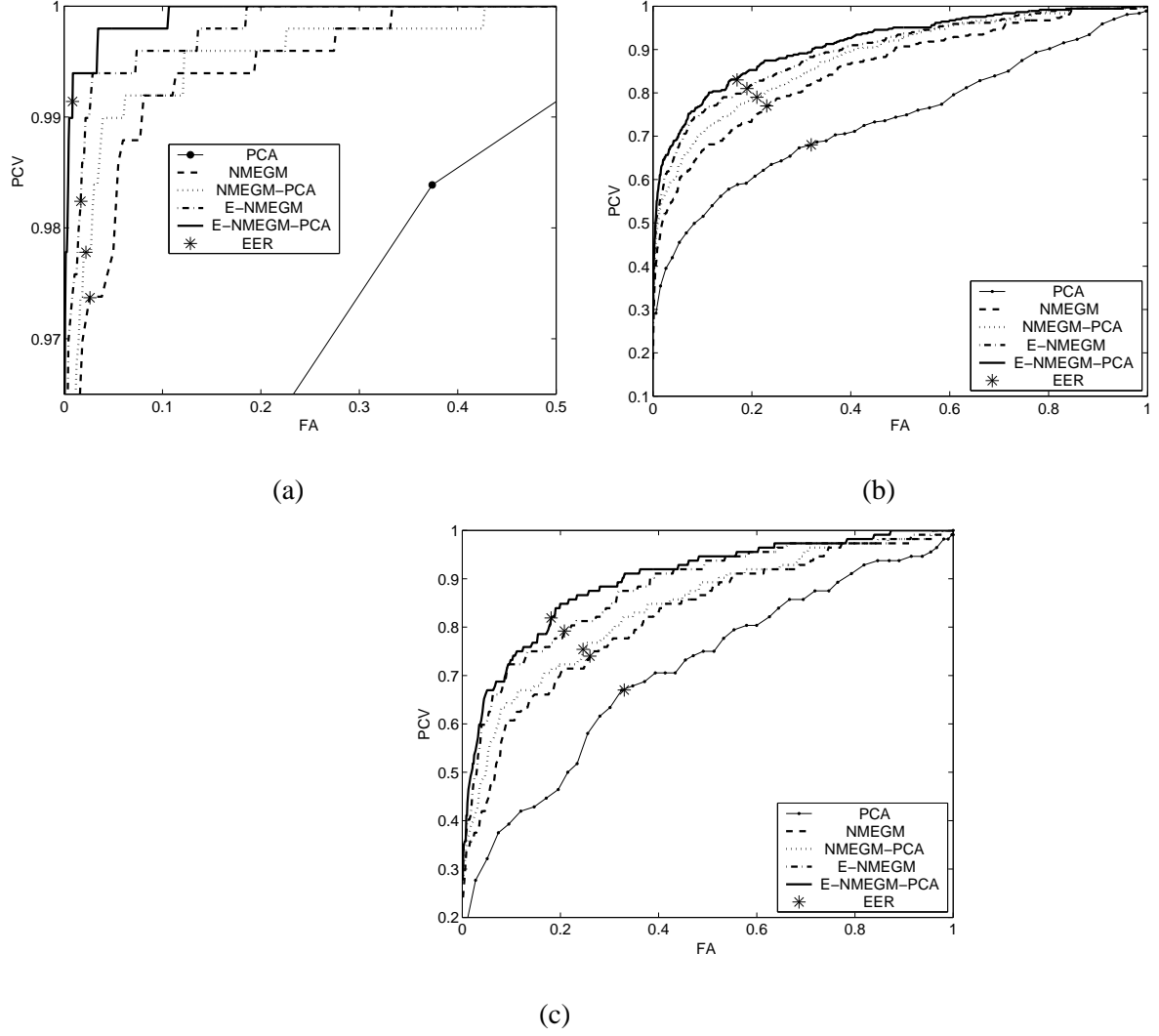
(a)

(b)

(c)

Fig. 13. (a) ROC curves for the FB probe set; (b) ROC curves for the Dup1 probe set; (c) ROC curves for the Dup2 probe set.

V. Conclusion

In this paper, a novel method for learning person specific facial models has been proposed. The method is comprised of two steps. In the first step, a graph having its nodes at discriminant facial points is found. In the second step, a novel discriminant feature extraction method is used in order to select the most discriminant features for the graph nodes. The discriminant models are used along with morphological elastic graph matching and tested for frontal face verification where state of the art verification performance has been achieved.

Acknowledgment

REFERENCES

[1] 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 Transactions on Computers*, vol. 42, no. 3, pp. 300–311, Mar. 1993.

[2] Jun Zhang, Yong Yan, and Martin Lades, "Face recognition: Eigenface, elastic matching, and neural nets," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1423–1435, Sept. 1997.

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

[4] L. Wiskott, "Phantom faces for face analysis," *Pattern Recognition*, vol. 30, no. 6, pp. 837–846, 1997.

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

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

[7] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using discriminating grids with morphological feature vectors," *IEEE Transactions on Multimedia*, vol. 2, no. 1, pp. 14–26, Mar. 2000.

[8] C. Kotropoulos, A. Tefas, and I. Pitas, "Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions," *Pattern Recognition*, vol. 33, no. 12, pp. 31–43, Oct. 2000.

[9] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using morphological elastic graph matching," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 555–560, Apr. 2000.

[10] A. Tefas, C. Kotropoulos, and I. Pitas, "Face verification using elastic graph matching based on morphological signal decomposition," *Signal Processing*, vol. 82, no. 6, pp. 833–851, 2002.

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

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

[13] A. Tefas, C. Kotropoulos, and I. Pitas, "Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 735–746, 2001.

[14] K. Messer, J. Matas, J.V. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *AVBPA'99*, Washington, DC, USA, 22-23 March 1999, pp. 72–77.

[15] S.A. Rizvi, P.J. Phillips, and H. Moon, "The FERET verification testing protocol for face recognition algorithms," in *Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 14-16 April 1998, pp. 48–53.

[16] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.

[17] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[18] S. Zafeiriou, A. Tefas, and I. Pitas, "Elastic graph matching versus linear subspace methods for frontal face verification," in *International Workshop on Nonlinear Signal and Image Processing*, 2005.

[19] J. Yang, A.F. Frangi, J. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.

[20] J. Yang and J.-Y. Yang, "Why can LDA be performed in PCA transformed space?," *Pattern Recognition*, vol. 36, no. 2, pp. 563–566, 2003.

[21] J. Yang, Z. Jin, J. Yang, A.F. Frangi, and D. Zhang, "Essence of kernel Fisher discriminant: KPCA plus LDA," *Pattern Recognition*, vol. 37, no. 10, pp. 2097–2100, 2004.

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

[23] R. H. J. M. Otten and L. P. P. P. van Ginneken, *The annealing algorithm*, Kluwer, B.V., Deventer, The Netherlands, The Netherlands, 1989.

[24] L.F. Chen, H.Y.M. Liao, M.T. Ko, J.C. Lin, and G.J. Yu, "A new LDA-based face recognitions system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, pp. 1713–1726, 2000.

[25] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognition*, vol. 34, pp. 2067–2070, 2001.

[26] L. Juwei, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 195–200, 2003.

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

[28] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 4–13, 2005.

[29] "XM2VTS database," Available information in http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb.

[30] "Color FERET database," Available information in http://www.itl.nist.gov/iad/humanid/colorferet/home.html.

[31] J. Matas, M. Hamou, K. Jonsson, J. Kittler, Y. Li, C. Kotropoulos, A. Tefas, I. Pitas, T. Tan, H. Yan, F. Smeraldi, J. Bigun, N. Capdevielle, W. Gerstner, S. Ben-Yacouba, Y. Abdelaoued, and E. Mayoraz, "Comparison of face verification results on the XM2VTS database," in *ICPR*, Barcelona, Spain, 3-8 September 2000, pp. 858–863.

[32] K. Messer, J.V. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz,

L. Vandendorpe, S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, R. Paredes, B. Kepenekci, F.B. Tek, G.B. Akar, F. Deravi, and N. Mavity, "Face verification competition on the XM2VTS database," in *AVBPA03*, Guildford, United Kingdom, 9-11 June 2003, pp. 964–974.

[33] L. Chengjun and H. Wechsler, "Independent component analysis of gabor features for face recognition," *IEEE Transactions on Neural Networks*, vol. 14, no. 4, pp. 919–928, July 2003.

[34] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 683 – 695, 2006.

[35] L. Chengjun, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725–737, 2006.

[36] K. Jonsson, J. Matas, and Kittler, "Learning salient features for real-time face verification," in *Proc. Second International Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA'99)*, 1999, pp. 60–65.

[37] W. Zhao, R. Chellappa, P.-J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Survey*, vol. 35, pp. 399–458, 2003.

[38] M.-H. Yang, D.J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[39] C. Wu, C. Liu, H.-Y. Shum, Y.-Q. Xy, and Z. Zhang, "Automatic eyeglasses removal from face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 3, pp. 322 – 336, 2004.

[40] J.-S. Park, Y.H. Oh, S.C. Ahn, and S.W. Lee, "Glasses removal from facial image using recursive error compensation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 805 – 811, 2005.

[41] D. Beymer and T. Poggio, "Image representations for visual learning," *Science*, vol. 272, no. 5250, 1996.

[42] D. Beymer and T. Poggio, "Face recognition from one example view," in *Proc. Int. Conf. on Computer Vision*, Boston, MA, June 1995, pp. 500–507.

[43] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of the Optical Society of America A*, vol. 14, no. 8, pp. 1724–1733, 1997.

[44] P. Niyogi, F. Girosi, and T. Poggio, "Incorporating prior information in machine learning by creating virtual examples," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2196–2209, 1998.

[45] K. Messer, J. Kittler, J. Short, G. Heusch, F. Cardinaux, S. Marcel, Y. Rodriguez, S. Shan, Y. Su, W. Gao, and X. Chen, "Performance characterisation of face recognition algorithms and their sensitivity to severe illumination changes," in *International Conference on Biometrics, ICB 2006*, Hong Kong, China, January 5-7 2006.

[46] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas, "An eye detection algorithm using pixel to edge information," in *Proc. of 2nd IEEE-EURASIP International Symposium on Control Communications, and Signal Processing (ISCCSP 2006)*, Marrakech, Morocco, March 13-15 2006.

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

Stefanos Zafeiriou was born in Thessaloniki, Greece in 1981. He received the BS degree from the Department of Informatics of Aristotle University of Thessaloniki with honors in 2003. He is currently a researcher and teaching assistant and he is studying towards a PhD at the Department of Informatics at the University of Thessaloniki. His current research interests lie in the areas of signal and image processing, pattern recognition and computer vision as well as in the area of watermarking for copyright protection and authentication of digital media.

Anastasios Tefas received the B.Sc. in informatics in 1997 and the Ph.D. degree in informatics in 2002, both from the Aristotle University of Thessaloniki, Greece. Since 2006, he has been an Assistant Professor at the Department of Information Management, Technological Educational Institute of Kavala. From 1997 to 2002, he was a researcher and teaching assistant in the Department of Informatics, University of Thessaloniki. From 2003 to 2004, he was a temporary lecturer in the Department of Informatics, University of Thessaloniki where he is currently, a senior researcher. He has co-authored over 50 journal and conference papers. His current research interests include computational intelligence, pattern recognition, digital signal and image processing, detection and estimation theory, and computer vision.

Ioannis Pitas received the Diploma of Electrical Engineering in 1980 and the PhD degree in Electrical Engineering in 1985 both from the Aristotle University of Thessaloniki, Greece. Since 1994, he has been a Professor at the Department of Informatics, Aristotle University of Thessaloniki. From 1980 to 1993 he served as Scientific Assistant, Lecturer, Assistant Professor, and Associate Professor in the Department of Electrical and Computer Engineering at the same University. He served as a Visiting Research Associate at the University of Toronto, Canada, University of Erlangen- Nuernberg, Germany, Tampere University of Technology, Finland, as Visiting Assistant Professor at the University of Toronto and as Visiting Professor at the University of British Columbia, Vancouver, Canada. He was lecturer in short courses for continuing education. He has published over 150 journal papers, 440 conference papers and contributed in 22 books in his areas of interest. He is the co-author of the books Nonlinear Digital Filters: Principles and Applications (Kluwer, 1990), 3-D Image Processing Algorithms (J. Wiley, 2000), Nonlinear Model-Based Image/Video Processing and Analysis (J. Wiley, 2001) and author of Digital Image Processing Algorithms and Applications (J. Wiley, 2000). He is the editor of the book Parallel Algorithms and Architectures for Digital Image Processing, Computer Vision and Neural Networks (Wiley, 1993). He has also been an invited speaker and/or member of the program committee of several scientific conferences and workshops. In the past he served as Associate Editor of the IEEE Transactions on Circuits and Systems, IEEE Transactions on Neural Networks, IEEE Transactions on Image Processing, EURASIP Journal on Applied Signal Processing and co-editor of Multidimensional Systems and Signal Processing. He was general chair of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing (NSIP95), technical chair of the 1998 European Signal Processing Conference and general chair of IEEE ICIP 2001. His current interests are in the areas of digital image and video processing and analysis, multidimensional signal processing, watermarking and computer vision.