

## ADVANCES IN ELASTIC GRAPH MATCHING FOR FRONTAL FACE VERIFICATION

Stefanos Zafeiriou , Anastasios Tefas and Ioannis Pitas

Dept. of Informatics, Aristotle University of Thessaloniki, Box 451, 54124 Thessaloniki, Greece  
e-mail: {dralbert,tefas,pitas}@zeus.csd.auth.gr

### ABSTRACT

*Elastic graph matching is one of the most well known techniques for frontal face recognition/verification and one of the few techniques that can be combined successfully with fully automatic face localization and alignment methods. In this paper we propose a series of techniques that enhance the performance of elastic graph matching in frontal face verification by exploiting the individuality of human facial features in many ways. First the use of discriminant analysis in the feature vectors of the graph nodes is explored. The use of the node deformation for discrimination is also proposed. Moreover, the local similarity values at the nodes of the elastic graph, are weighted by coefficients that are also derived from some discriminant analysis in order to form a total similarity measure between faces. We present an algorithm that combines all the above discriminant steps. Moreover, we propose an algorithm for finding the most discriminant landmarks upon a person's face and a person-specific graph is placed in the spatial coordinates that correspond to these discriminant features. We illustrate the improvements in performance by the proposed advances in frontal face verification using the XM2VTS database.*

### 1. INTRODUCTION

A popular class of techniques used for frontal face recognition/verification is elastic graph matching [1], [2], [3]. In [2] the Gabor based EGM has been proposed and tested for frontal face verification. A variant of the Gabor-EGM, the so-called *morphological elastic graph matching* (MEGM), has been proposed for frontal face verification [3]. In MEGM the typical Gabor analysis [2] has been superseded by multiscale morphological dilation-erosion using a scaled structuring function [3]. The multiscale morphological analysis has been proven to be suitable for facial image analysis and MEGM has given superior verification results compared to the Gabor-EGM approach without having to create a computational expensive filter bank setup.

Discriminant techniques have been employed in order to enhance the recognition and verification performance of the EGM algorithm. The use of linear discriminant techniques at the feature vectors for selecting the most discriminant features has been proposed in [2], [3]. Several schemes

that aim at weighting the graph nodes according to their discriminant power have been proposed [3], [4].

Little or no research has been conducted concerning what type of graphs are 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 [2], [3], [4] or a graph that is placed on preselected nodes that correspond to some fiducial facial landmarks (e.g., nose, eyes, etc.) [1].

Intuitively, one may think that graphs with nodes placed at specified fiducial landmarks may perform better. 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 [1] is applied. On the contrary, an evenly distributed rectangular graph is easier to be handled automatically, since only a face detection algorithm is needed in order to find an initial approximation of the rectangular facial region [2], [3], [4].

In this paper we advance the research in EGM for frontal face verification in many ways by exploiting the individuality of human facial features. Firstly, we illustrate where and how discriminant techniques can be employed in the EGM and propose a combined discriminant architecture. Moreover, we propose the use of person-specific graphs placed at his discriminant facial landmarks. To do so, 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 measures like the ones proposed in [2], [3]. This way, graphs that are person specific and have nodes placed at discriminant facial features, are obtained. The proposed methodologies can be applied to all EGM algorithms.

### 2. ELASTIC GRAPH MATCHING

In the first step of the EGM algorithm, a sparse graph that is suitable for face representation is selected [1], [2], [3]. The facial image region is analyzed and a set of local descriptors is extracted at each graph node. That is, at each graph node that is located at image coordinates  $x$ , a jet

(feature vector)  $\mathbf{j}(\mathbf{x})$  is formed:

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

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.:

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

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:

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

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:

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

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 [3], 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:

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

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  $\lambda$  in (4) and  $\delta_{\max}$  in (5) control the rigidity/plasticity of the graph [2], [3]. After the matching procedure, the distance  $D_t(r)$  is used as a quantitative measure for the similarity of two faces [2], [3].

### 3. EMPLOYING DISCRIMINANT ANALYSIS IN EVERY PHASE OF EGM

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  $\sigma^2(\mathcal{X})$ .

#### 3-A. Feature Vector Discriminant analysis

It is obvious that the standard EGM treats uniformly all the different features that form the jets. Thus, it sounds reasonable to use discriminant techniques in order to find the most discriminant features. In other words, we should learn a person and node specific discriminant function  $\mathbf{g}_r^l$ , for the  $l$ -th node of the reference person  $r$ , that transforms the jets  $\mathbf{j}(\mathbf{x}_t^l(r))$ :

$$\hat{\mathbf{j}}(\mathbf{x}_t^l(r)) = \mathbf{g}_r^l(\mathbf{j}(\mathbf{x}_t^l(r))). \quad (6)$$

We will use linear techniques for finding the transform  $\mathbf{g}_r^l$ . 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.

We use the same criterion as [2],[3] that can give more than one discriminant projections. Let  $\mathbf{W}^l(r)$  and  $\mathbf{B}^l(r)$  be the matrices:

$$\mathbf{W}^l(r) = \sum_{\mathcal{F}_I^l(r)} (\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))(\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))^T \quad (7)$$

and

$$\mathbf{B}^l(r) = \sum_{\mathcal{F}_C^l(r)} (\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))(\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r)))^T. \quad (8)$$

The optimal discriminative directions  $\boldsymbol{\Psi}^l(r)$  are given by maximizing the criterion:

$$J(\boldsymbol{\Psi}^l(r)) = \frac{\text{tr}[\boldsymbol{\Psi}^l(r)^T \mathbf{W}^l(r) \boldsymbol{\Psi}^l(r)]}{\text{tr}[\boldsymbol{\Psi}^l(r)^T \mathbf{B}^l(r) \boldsymbol{\Psi}^l(r)]} \quad (9)$$

where  $\text{tr}[\mathbf{R}]$  is the trace of the matrix  $\mathbf{R}$ . This criterion is well suited for the face verification problem due to the fact that it tries to find the feature projections that maximize the distance of impostor jets from the genuine class center while minimizing the distance of genuine jets from genuine class center. If  $\mathbf{B}^l(r)$  is not singular then (9) is maximized when the column vectors of the projection matrix,  $\boldsymbol{\Psi}^l(r)$ , are the eigenvectors of  $\mathbf{B}^l(r)^{-1} \mathbf{W}^l(r)$ .

In order to proceed to feature dimensionality reduction in  $M < S$  dimensions the matrix  $\boldsymbol{\Psi}^l(r)$  should be comprised by the eigenvectors of  $\mathbf{B}^l(r)^{-1} \mathbf{W}^l(r)$  that correspond to the  $M$  greatest eigenvalues. The feature vector after discriminant dimensionality reduction is:

$$\hat{\mathbf{j}}(\mathbf{x}_t^l(r)) = \mathbf{g}_r^l(\mathbf{j}(\mathbf{x}_t^l(r))) = \boldsymbol{\Psi}^l(r)^T \mathbf{j}(\mathbf{x}_t^l(r)). \quad (10)$$

Now, the  $L_2$  norm is used for forming the new feature vector similarity measure in the final multidimensional space.

### 3-B. Local Similarity Measure Discriminant Weighting

In [2], [3] only the jet similarity measure has been considered when forming the total similarity measure between two graph nodes. The node deformation was only employed implicitly in the matching stage by imposing additional rigidity/plasticity penalties. We propose to combine the feature vector similarity measure and the node deformation in a discriminant manner in order to form the new local similarity measure. The node feature similarity measure between the reference person  $r$  and the test person  $t$  for the  $l$ -th node is  $f_t^l(r) = C_f(\mathbf{j}(\mathbf{x}_t^l(r)), \mathbf{j}(\mathbf{x}_r^l))$  and the node deformation is  $d_t^l(r) = C_d(\mathbf{x}_t^l(r), \mathbf{x}_r^l)$ . Let  $\mathbf{d}_t^l(r) \in \mathbb{R}^2$  be a column vector that is comprised by the two local similarity measures, for the node  $l$ , between the test person  $t$  and the reference person  $r$ , i.e.:

$$\mathbf{d}_t^l(r) = [ f_t^l(r) \ d_t^l(r) ]^T \quad (11)$$

According to the standard EGM [1] the node similarity value after the matching procedure is given by:

$$c_t^l(r) = f_t^l(r) + \lambda d_t^l(r) = [ 1 \ \lambda ] \mathbf{d}_t^l(r) = \mathbf{e}^T \mathbf{d}_t^l(r) \quad (12)$$

where  $\lambda$  is the constant that controls the rigidity/plasticity of the graph [2]. In general  $\mathbf{e}$  does not contain any discriminant information. Thus, when forming the local similarity measure the vector  $\mathbf{e}$  should be superseded by a discriminant function  $\mu_r^l$  that is person and node specific. The new local similarity measure is:

$$c_t^l(r) = \mu_r^l(\mathbf{d}_t^l(r)). \quad (13)$$

LDA is used in order to find the discriminant transform  $\mu_r^l$ .

Let  $\mathcal{L}_C^l(r)$  and  $\mathcal{L}_I^l(r)$  be the sets of local similarity vectors  $\mathbf{d}_t^l(r)$  that correspond to genuine and impostor claims, respectively. In order to form the optimization criterion, the between class scatter matrix,  $\mathbf{D}_S^l(r)$ , and the within class scatter matrix,  $\mathbf{D}_W^l(r)$ , of the local similarity vectors  $\mathbf{d}_t^l(r)$  are employed. The optimization criterion used for finding the discriminant weighting vector  $\mathbf{q}^l(r)$  :

$$J(\mathbf{q}^l(r)) = \frac{\mathbf{q}^l(r)^T \mathbf{D}_S^l(r) \mathbf{q}^l(r)}{\mathbf{q}^l(r)^T \mathbf{D}_W^l(r) \mathbf{q}^l(r)}. \quad (14)$$

The optimal weighting coefficients are given by [5]:

$$\mathbf{q}^l(r) = \frac{\mathbf{D}_W^l(r)^{-1}(\mathbf{m}(\mathcal{L}_I^l(r)) - \mathbf{m}(\mathcal{L}_C^l(r)))}{\|\mathbf{D}_W^l(r)^{-1}(\mathbf{m}(\mathcal{L}_I^l(r)) - \mathbf{m}(\mathcal{L}_C^l(r)))\|}. \quad (15)$$

The new similarity value between the  $l$ -th node of the reference graph and the same node of the test graph is now:

$$c_t^l(r) = \mu_r^l(\mathbf{d}_t^l(r)) = \mathbf{q}^l(r)^T \mathbf{d}_t^l(r). \quad (16)$$

### 3-C. Discriminant Node weighting

In the standard EGM all nodes are treated uniformly when forming the final similarity measure between faces. Thus, it sounds reasonable to weight the similarity measures of nodes that correspond to different facial landmarks with weights that correspond to their discriminant power. The weights should be person specific due to the fact that different persons have different discriminant facial landmarks. Let  $\mathbf{c}_t(r) \in \mathbb{R}^L$  be a column vector comprised by the new local similarity values at every node:

$$\mathbf{c}_t(r) = [c_t^1(r) \ c_t^2(r) \ \dots \ c_t^L(r)]^T \quad (17)$$

where  $L$  is the number of graph nodes. The vector  $\mathbf{c}_t(r)$  is the total similarity vector between the reference face  $r$  and a test face  $t$ . The standard EGM algorithm approach [2] treats uniformly all the similarity values  $c_t^l(r)$ . That is, the total similarity measure between a reference person  $r$  and a test person  $t$  is simply the sum of all node similarity measures:

$$D_t(r) = \sum_{i=1}^L c_t^i(r) = \mathbf{1}^T \mathbf{c}_t(r), \quad (18)$$

where  $\mathbf{1}$  is an  $L \times 1$  vector of ones. The algorithm should learn a discriminant function  $\beta_r$  that is person specific and form the total similarity measure between faces:

$$\hat{D}_t(r) = \beta_r(\mathbf{c}_t(r)). \quad (19)$$

The transform  $\beta_r$  could be just a weighting vector or a more complicated nonlinear support vector machine [4]. We will use LDA to create a total similarity measure between the reference person  $r$  and a test person  $t$ .

Let  $\mathcal{T}_C(r)$  and  $\mathcal{T}_I(r)$  be the sets of the total similarity vectors for the genuine and impostor claims of the reference person  $r$ , respectively. Let the within-class scatter matrix and the between-class scatter for the total similarity vectors  $\mathbf{c}_t(r)$  be  $\mathbf{V}_W(r)$  and  $\mathbf{V}_B(r)$ , respectively. The optimal weighting coefficients that are derived from the maximization of:

$$J(\mathbf{w}(r)) = \frac{\mathbf{w}(r)^T \mathbf{V}_B(r) \mathbf{w}(r)}{\mathbf{w}(r)^T \mathbf{V}_W(r) \mathbf{w}(r)} \quad (20)$$

are the elements of the vector  $\hat{\mathbf{w}}(r)$  [5]:

$$\hat{\mathbf{w}}(r) = \frac{\mathbf{V}_W(r)^{-1}(\mathbf{m}(\mathcal{T}_I(r)) - \mathbf{m}(\mathcal{T}_C(r)))}{\|\mathbf{V}_W(r)^{-1}(\mathbf{m}(\mathcal{T}_I(r)) - \mathbf{m}(\mathcal{T}_C(r)))\|}. \quad (21)$$

The similarity measure between the reference person  $r$  and the test person  $t$ , after all the successive discriminant steps, is given by:

$$\hat{D}_t(r) = \beta_r(\mathbf{c}_t(r)) = \hat{\mathbf{w}}(r)^T \mathbf{c}_t(r). \quad (22)$$

#### 4. FINDING DISCRIMINANT PERSON-SPECIFIC GRAPHS

After denoting the different steps of EGM where discriminant analysis can be employed and having proposed a combined discriminant architecture we are interesting on how the use of discriminant analysis can be extended in the structure of the graph itself. To do so, we will formulate and solve an optimization problem that finds discriminant person-specific graphs placed at his discriminant facial feature.

##### 4-A. Setting the Optimization Problem

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 [3]:

$$s_t^l(r) = \|\mathbf{j}(\mathbf{x}_t^l(r)) - \mathbf{m}(\mathcal{F}_C^l(r))\|^2. \quad (23)$$

Let  $\mathcal{Y}_C^l(r)$  and  $\mathcal{Y}_I^l(r)$  be the sets of local similarity values  $s_t^l(r)$  that correspond to genuine and impostor claims, respectively. A possible measure for the discriminant power of the  $l$ -th node is the following:

$$p^n(r) = \frac{\frac{1}{N(\mathcal{Y}_I^l(r))} \sum_{s_t^l(r) \in \mathcal{Y}_I^l(r)} s_t^l(r)}{\frac{1}{N(\mathcal{Y}_C^l(r))} \sum_{s_t^l(r) \in \mathcal{Y}_C^l(r)} s_t^l(r)}. \quad (24)$$

The measure (24) 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 we have:

$$E_g(r) = \frac{1}{L} \sum_{l=1}^L p^l(r) \quad (25)$$

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 measure defined in (25) 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 landmarks than the nodes of  $g_1$ . Figure 1 shows two different graph setups  $g_1$  and  $g_2$  with different values for the measure  $E_g(r)$ . Both graphs have 64 nodes. The graph depicted in the right hand side of Figure 1 is found experimentally to be more discriminant than the rectangular graph depicted in left hand side of the Figure 1 since  $E_{g_1}(r) < E_{g_2}(r)$ .

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 in order to have low computational cost for the elastic graph matching procedure.
- 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:

$$\begin{aligned} \hat{g} = \arg \max_g E_g(r) \quad \text{subject to} \\ \|\mathbf{x}_r^l - \mathbf{x}_r^j\| \geq \Delta, \forall l, j \text{ nodes with } l \neq j \\ L = \text{constant} \end{aligned} \quad (26)$$

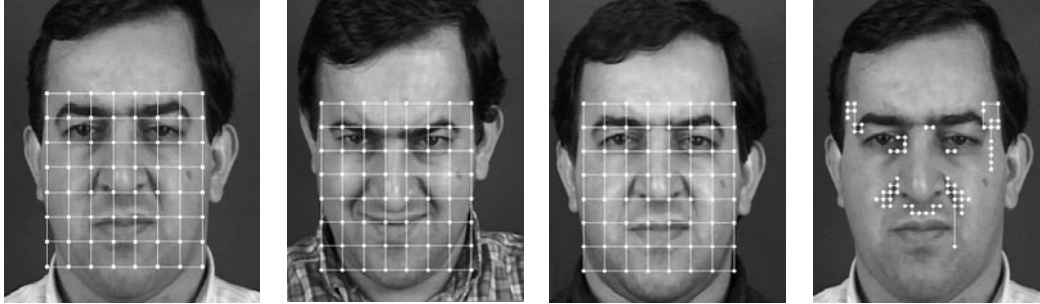
where  $\Delta$  is a preselected threshold that controls the density of the graph.

##### 4-B. A Heuristic Optimization Approach

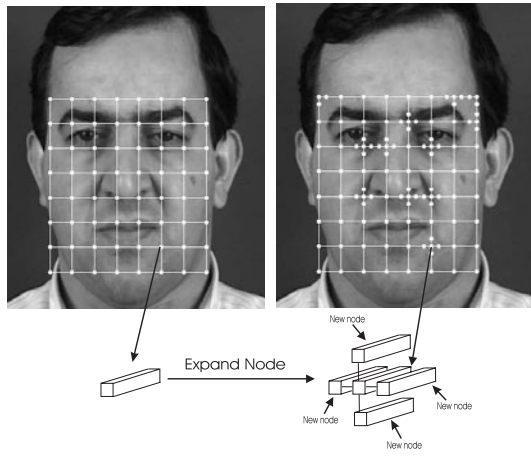
In order to solve the maximization problem (26), someone has to follow a heuristic optimization approach since, exhaustive search is not feasible. In the following, the steps of the proposed heuristic algorithm are described in more detail. This procedure should be repeated for every reference person  $r$  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. The first three images of Figure 1 show the reference facial images with the corresponding graphs for a person in the XM2VTS database.

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.

- Step 1. Take the reference graphs and match them in all genuine and impostor images.
- Step 2. For each node  $l$  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 2 describes pictorially this step.
- Step 4. Verify that the inserted nodes do not violate the graph sparseness criterion. That is, erase the new nodes that violate the criterion  $\|\mathbf{x}_r^l - \mathbf{x}_r^j\| < \Delta, \forall l, j$  neighboring nodes. 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



**Fig. 1.** The reference facial images with the reference graphs and the corresponding graph with nodes placed at his discriminant facial features.



**Fig. 2.** Expanding the rectangular graph.

search:

$$\hat{\mathbf{x}}_t^k(r) = \arg \min_{\mathbf{x}_t^k} C_f(\mathbf{j}(\mathbf{x}_t^k), \mathbf{j}(\mathbf{x}_r^k)) \text{ subject to } \|\mathbf{x}_t^k - \hat{\mathbf{x}}_t^k\| \leq \delta_{\max} \quad (27)$$

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

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.

Using as reference the facial images and graphs depicted in Figure 1, we demonstrate, in Figure 1, the discriminant graph that is derived from the proposed procedure. As it can be seen in Figure 1 (right hand image), the nodes of the discriminant graph are concentrated in the areas between his cheeks and nose. This region is indeed characteristic

for this particular person in the XM2VTS database. The elastic graph matching procedure of the new graphs is performed using the minimization procedure indicated in the optimization problem (5).

## 5. EXPERIMENTAL RESULTS

The experiments were conducted in the XM2VTS database using the Configuration I protocol described in [6]. The node features have been derived from the modified morphological analysis proposed in [7]. The images were aligned using an automatic alignment method. An  $8 \times 8$  graph and a modified morphological analysis was used. The training set is used for calculating for each reference person  $r$  and for each node  $l$  a matrix  $\hat{\Psi}^l(r)$  for feature selection. A PCA step is used prior to discriminant analysis in order to obtain the invertibility of  $\mathbf{B}^l(r)$ . The evaluation set is used for learning the discriminant vector  $\hat{\mathbf{q}}^l(r)$  for weighting the local similarity vector and the vector,  $\hat{\mathbf{w}}(r)$ , that weights the total similarity vector of the graph nodes. The evaluation set is also used for learning the thresholds.

The EGM using no discriminant step has given an TER equal to 12.9% in the test set of Configuration I. The best TER achieved, using only feature vector discriminant analysis, was 5.7% and was achieved when we kept the first 3 discriminant projections. The step of the discriminant feature selection using the EGM will be denoted as EGM-FD (*feature discriminant*).

We also investigated the contribution of the discriminant weighting of the local similarity vector. This was conducted by using no feature projections and by treating uniformly all the local similarity measures. That way we achieved an TER equal to 9.2%. When only discrimination between local similarity measures is considered the acronym EGM-LD (*local discriminant*) is used.

The contribution of weighting the local similarity measure with coefficients that are derived from LDA without other discriminant steps was also investigated. To do so, we applied only discriminant weighting in the graph level by calculating,  $\hat{\mathbf{w}}_r$ , without applying prior discriminant

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
EGM	9.2	98.2	65.0	7.9	5.0	98.8	0.0	0.0	61.0	12.9	98.8	61.0
EGM-ND	6.3	62.8	56.3	6.7	4.2	63.8	0.0	0.0	61.0	10.7	63.8	61.0
EGM-LD	5.2	45.5	20.0	5.2	4.0	45.0	0.5	0.0	17.0	9.2	45.5	17.0
EGM-FD	2.5	29.9	55.3	2.5	3.2	11.2	0.2	0.2	14.7	5.7	11.4	14.9
DEGM	0.2	0.7	6.5	1.6	1.2	10.2	0.0	0.0	13.1	2.8	10.2	13.1
E-EGM	3.4	34.5	55.5	3.3	2.75	22.7	0.0	0.0	44.7	6.05	22.7	44.7
E-EGM-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 1. Error Rates for XM2VTS Configuration I

analysis. The TER obtained was 10.7%. EGM-ND (*node discriminant*) will denote the EGM when only discriminant weighting of the total similarity vector is performed. The best TER achieved was 2.8% using successively all the discriminant steps. The acronym *discriminant* EGM (DEGM) will be used when all the discriminant steps were used. Table 1 shows the error rates according to the protocol described in [6].

Afterwards, we have applied the procedure for finding discriminant person specific graphs, described in Section 4-B, without feature vector discriminant analysis, abbreviated as *expandable*-EGM (E-EGM). When replacing the rectangular evenly distributed graphs with the person specific discriminant graphs proposed in this paper, the TER has been measured at 6.05%. That is, an increase in performance more than 50% in terms of TER is achieved when using the proposed graphs in comparison to the EGM approach that uses rectangular graphs. Similar to the best results of the DEGM, has been achieved when using feature vector discriminant analysis in the proposed discriminant graphs (i.e., using the discriminant feature extraction method at Section 3-A in the person-specific graphs). This set up has given a TER= 2.9% and is abbreviated as E-EGM-FD. The results of the DEGM and the E-EGM-FD approach are the best that have been reported using an automatic alignment method [8] in the Configuration I of the XM2VTS database.

## 6. CONCLUSION

In this paper a series of discriminant techniques that enhance the performance of the EGM algorithm in frontal face verification is proposed. A combined discriminant architecture is proposed that is independent to the structure of the graph is proposed. Accordingly, a new discriminant analysis that exploits the individuality of human facial features and finds a person specific graph placed at his discriminant facial features is also introduced. It is shown that the proposed advances can achieve state of the art face verification performance.

## 7. ACKNOWLEDGMENT

This work has been partially funded by the integrated project BioSec IST-2002-001766 (Biometric Security, <http://www.biosec.org>), and by the network of excellence BioSecure IST-2002-507634 (Biometrics for Secure Authentication, <http://www.biosecure.info>), both under Information Society Technologies (IST) priority of the 6th Framework Programme of the European Community.

## 8. REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] K. Fukunaga, *Statistical Pattern Recognition*, CA: Academic, San Diego, 1990.
- [6] 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.
- [7] S. Zafeiriou, A. Tefas, and I. Pitas, "Elastic graph matching versus linear subspace methods for frontal face verification," in *International Workshop on Non-linear Signal and Image Processing*, 2005.
- [8] K. Messer and et. al, "Face verification competition on the XM2VTS database," in *AVBPA03*, Guildford, United Kingdom, 9-11 June 2003, pp. 964–974.