DISCRIMINANT GRAPH STRUCTURES FOR FACE VERIFICATION

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

[†] Aristotle University of Thessaloniki
Department of Informatics, Box 451, 54124 Thessaloniki, Greece

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 an algorithm for finding the most discriminant features 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 applying the proposed method in frontal face verification using the XM2VTS database.

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

1. INTRODUCTION

A popular class of techniques used for frontal face verification is elastic graph matching. In EGM the reference object graph is created by projecting the object's image onto a rectangular elastic sparse graph where a Gabor wavelet bank response is measured at each node. The graph matching process is implemented by a coarse-to-fine stochastic optimization of a cost function which takes into account both jet similarities and node deformation [1, 2].

A variant of the standard EGM, the so-called *morphological elastic graph matching* (MEGM), has been proposed for frontal face verification [2]. In MEGM the Gabor analysis has been superseded by multiscale morphological dilation-erosion using a scaled structuring function [2]. 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 [1, 2]. Several schemes that aim at weighting the graph nodes according to their discriminatory power have been proposed [2, 3].

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

Intuitively, one may think that graphs with nodes placed at specified fiducial points 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 [4] 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 [1, 2, 3].

In this paper we advance the research in EGM for frontal face verification by proposing the use of person-specific graphs placed at the person's 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 [1, 2]. We assume that nodes with high discriminant measures correspond to facial landmarks 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 as discriminant. 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 this Section we will briefly outline the EGM algorithm. In the first step of the EGM algorithm, a sparse graph that is suitable for face representation is selected [1, 2, 4]. 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. The output of multiscale morphological dilation-erosion operations at several scales is a nonlinear alternative of the Gabor filters for multiscale analysis and has been successfully used for facial image analysis [2]. At each graph node that is located at image coordinates $\mathbf{x}$, a jet (feature vector) $\mathbf{j}(\mathbf{x})$ is formed:

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

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.:

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

Let $\mathcal{V}$ be the set of all graph vertices of a certain facial image. For the rectangular graphs, all nodes, expect 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:

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

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:

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

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 [2], 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:

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

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 [1], [2].

3. FINDING DISCRIMINANT PERSON-SPECIFIC GRAPHS

3.1. 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. Let also, $\mathbf{m}(\mathcal{X})$ be the mean vector of a set of vectors $\mathcal{X}$ and $N(\mathcal{X})$ be its cardinality. 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 [2]:

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

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:

$$p^l(r) = \frac{\frac{1}{N(\mathcal{Y}_I^l(r))} \sum_{s_t^l(r) \in \mathcal{Y}_I^l(r)} s_t^l(r)}{\frac{1}{N(\mathcal{Y}_C^l(r))} \sum_{s_t^l(r) \in \mathcal{Y}_C^l(r)} s_t^l(r)}. \qquad (7)$$

The measure (7) 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:

$$E_g(r) = \frac{1}{L} \sum_{l=1}^{L} p^l(r) \qquad (8)$$

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 (8) 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.
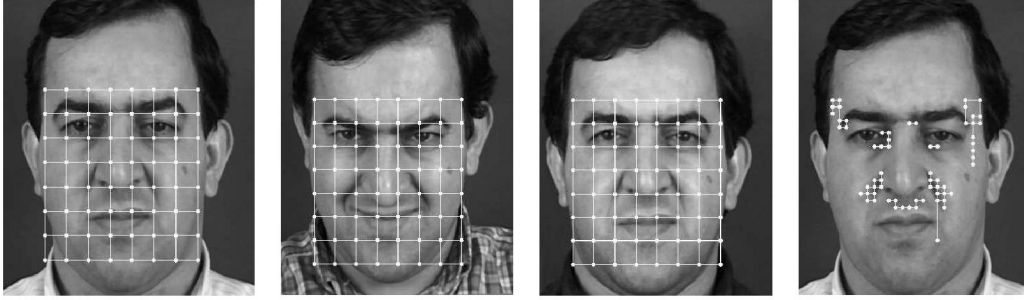
Fig. 1. The reference facial images with the reference graphs and the corresponding graph with nodes placed at his discriminant facial landmarks.

Formally, the above optimization problem can be written as:

$$\acute{g} = \arg\max_g E_g(r) \text{ subject to}$$
$$||\mathbf{x}_r^l - \mathbf{x}_r^j|| \geq \Delta, \forall \; l, j \text{ nodes with } l \neq j \qquad (9)$$
$$L = \text{constant}$$

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

3.2. A Heuristic Optimization Approach

In order to solve the maximization problem (9), someone has to follow a heuristic optimization approach since, exhaustive search is not feasible. Thus, we should search for a sub-optimal solution of the constraint optimization problem (9) by assuming that the desired graph is a subgraph of the most dense Δ-rectangular graph (i.e., an evenly distributed graph with nodes placed every Δ pixels).

Moreover, it is also very computational expensive to compute the discriminant measures of the nodes of the most dense Δ-rectangular graph (e.g., the most dense 9-rectangular graph for the facial images in Figure 1 has 700 nodes). Thus, a second sampling step is required. In the second sampling step, a sparse subgraph of the dense 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 computational feasible solution of the constrained optimization problem (9).

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 refer-
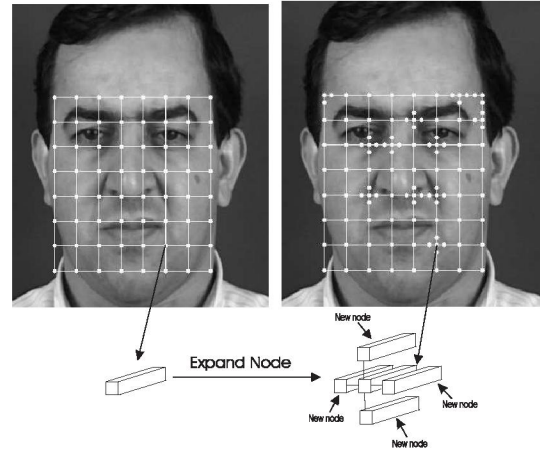
Fig. 2. Expanding the rectangular graph.

ence 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 $\mathcal{A}_i$.

Algorithm	Configuration I											
	Evaluation set			Test set								
				FAE=FRE		FRE=0		FAE=0		Total Error Rate(TER)		
	FAE=FRE	FAE(FRE=0)	FRE(FAE=0)	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-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
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

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 searching:

$$\dot{\mathbf{x}}_t^k(r) = \arg\min_{\mathbf{x}_t^k} C_f(\mathbf{j}(\mathbf{x}_t^k), \mathbf{j}(\mathbf{x}_r^k)) \text{ subject to}$$
$$||\mathbf{x}_t^k - \tilde{\mathbf{x}}_t^k|| \le \delta_{\max}$$
$$(10)$$

and $\dot{\mathbf{x}}_t^k(r)$ 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.

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 this 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).

4. EXPERIMENTAL RESULTS

The experiments were conducted in the XM2VTS database using the Configuration I protocol described in [5]. A 8×8 graph and a modified morphological analysis was used. The scores achieved during the testing procedure are summarized in Table 1 (the abbreviations will be explained below).

The EGM using rectangular graphs has given an TER equal to 12.9% in the test set of Configuration I. The best TER achieved for the rectangular graphs, using the feature vector discriminant analysis in [2], was 5.7%. The step of the discriminant feature selection in the rectangular graphs is denoted as EGM-FD (*feature discriminant*), in Table 1.

The procedure for finding discriminant person specific graphs, described in Section 3.2, without feature vector discriminant analysis, is abbreviated as *expandable*-EGM (E-EGM) in Table 1. 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%. When using feature vector discriminant analysis at the nodes of the discriminant graph structures an TER= 2.9% has been achieved. This approach is abbreviated as E-EGM-FD, in Table 1. That is, an increase in performance more that 50% in terms of TER is achieved when using the proposed graphs in comparison to the EGM approach that uses rectangular graphs.

5. CONCLUSION

A novel method that finds discriminant graph structures have been proposed. The discriminant graph structures are used along with morphological elastic graph matching and tested for frontal face verification where the verification performance of EGM is greatly enhanced.

6. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Messer, K. et al., "XM2VTSDB: The extended M2VTS database," in *AVBPA '99*, Washington, DC, USA, pp., 72–77.