EXPLOITING DISCRIMINANT INFORMATION IN ELASTIC GRAPH MATCHING

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

In this paper, we investigate the use of discriminant techniques in the elastic graph matching (EGM) algorithm. First we use discriminant analysis in the feature vectors of the nodes in order to find the most discriminant features. The similarity measure for discriminant feature vectors and the node deformation are combined in a discriminant manner in order to form a local similarity measure between nodes. Moreover, the local similarity values at the nodes of the elastic graph, are weighted by coefficients that are also derived by some discriminant analysis in order to form a total similarity measure between faces. We illustrate the improvements in performance in frontal face verification using a modified multiscale morphological analysis.

1. INTRODUCTION

A popular class of techniques used for frontal face recognition/verification is EGM [1]. 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 procedure is implemented by a coarse-to-fine stochastic optimization of a cost function which takes into account both jet similarities and node deformation [1].

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 In MEGM the Gabor analysis has been superseded by multiscale morphological dilation-erosion by a scaled structuring function [2].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 discriminant techniques at the feature vectors for selecting the most discriminant 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]. In [3] it has been shown that the verification performance of the EGM can be highly improved by proper node weighting strategies.

In this paper we illustrate where and how discriminant techniques can be employed in the EGM. More precisely, each node is considered as a local expert and discriminant feature selection techniques are employed for enhancing its recognition/verification performance. The deformation of each node is considered as a second local similarity metric that can quantify the relationships with its neighboring nodes. The new local similarity value at each node is produced by discriminant weighting of both the feature vector similarity measure and the node deformation. As a final discriminant step the local similarity measures at grid nodes are weighted by coefficients according to their discriminant power. The problem of frontal face verification is used in the following of the paper in order to describe in detail the different discriminant steps.

2. ELASTIC GRAPH MATCHING

In this Section we will briefly outline the problem of frontal face verification and the framework under which EGM performs face verification. Let $\mathcal{U}$ be a facial image database and each facial image $\mathbf{u} \in \mathcal{U}$ belongs to one of the C person classes $\{\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_C\}$ with $\mathcal{U} = \bigcup_{i=1}^{C} \mathcal{U}_i$. For a face verification system that uses the database $\mathcal{U}$ a genuine (or client) claim is performed when a person t provides its facial image, $\mathbf{u}$, claiming that $\mathbf{u} \in \mathcal{U}_r$ and $t = r$. When a person t provides its facial image $\mathbf{u}$ while claiming that $\mathbf{u} \in \mathcal{U}_r$, with $t \neq r$, an impostor claim occurs. The scope of a face verification system is to handle properly these claims by accepting the genuine claims and rejecting the impostor ones.

The first step of EGM is to analyze the facial image region of the image $\mathbf{u}$. Then, a set of local descriptors is extracted at each graph node. In the standard EGM a 2D Gabor based filter bank has been used for image analysis. The output of multiscale morphological dilation-erosion operations 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

This work is funded by the integrated project BioSec IST-2002-001766 (Biometric Security, http://www.biosec.org), under Information Society Technologies (IST) priority of the 6th Framework Programme of the European Community.

coordinates $\mathbf{x}$ a jet $\mathbf{j}(\mathbf{x})$ is formed as:

$$\mathbf{j}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_S(\mathbf{x})), \quad (1)$$

where $f_i(\mathbf{x})$ denotes the output of a local operator applied to the image f at the ith scale or at the ith pair (scale, orientation) and S is the dimensionality of the jet.

The next step of the 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 relationships. Let the superscripts t and r denote a test and a reference person (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.:

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

Let $\mathcal{V}$ be the set of graph vertices. Let also $H(l)$ 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:

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

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

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

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 $\mathbf{j}(\mathbf{x}_t^l(r))$. The optimization of (4) has been interpreted in [2] as a simulated annealing with additional penalties imposed by the graph deformations. Accordingly, (4) can be simplified to:

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

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 $\boldsymbol{\delta}_{\max}$ in (5) and of λ in (4) control the rigidity/plasticity of the graph [1],[2]. Obviously, both functions (4) and (5) define a similarity measure between two faces.

3. 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 $\mathbf{j}(\mathbf{x}_t^l(r))$:

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

We will use linear techniques for finding the transform $\mathbf{g}_r^l$ but non-linear techniques can be also used. Before calculating the linear projections we normalize all the jets that have been produced during the match of the graphs of the reference person r to all other facial images in the training set in order to have zero mean and unit magnitude. Let $\widehat{\mathbf{j}}(\mathbf{x}_t^l(r))$ be the normalized jet at l-th node. Let $\mathcal{F}_C^l(r)$ and $\mathcal{F}_I^l(r)$ be the sets of the normalized jets of the l-th node that correspond to genuine claims and impostor claims related to person r, respectively.

We use the same criterion as [1],[2] that can give more than one discriminant directions. Let $\mathbf{W}^l(r)$ and $\mathbf{B}^l(r)$ be the matrices:

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

and

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

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

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

where 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, $\acute{\mathbf{\Psi}}^l(r)$, 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 $\acute{\mathbf{\Psi}}^l(r)$ 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:

$$\acute{\mathbf{j}}(\mathbf{x}_t^l(r)) = \mathbf{g}_l^r(\widehat{\mathbf{j}}(\mathbf{x}_t^l(r)) = \acute{\mathbf{\Psi}}^l(r)^T \widehat{\mathbf{j}}(\mathbf{x}_t^l(r)), \quad (10)$$

The similarity measure of the new feature vectors can be given by a simple distance metric. We have used the L_2 norm for forming the new feature vector similarity measure in the final multidimensional space:

$$C_f(\acute{\mathbf{j}}(\mathbf{x}_t^l(r)), \acute{\mathbf{j}}(\mathbf{x}_r^l)) = ||\acute{\mathbf{j}}(\mathbf{x}_t^l(r)) - \acute{\mathbf{j}}(\mathbf{x}_r^l)||. \quad (11)$$

4. LOCAL SIMILARITY MEASURE DISCRIMINANT WEIGHTING

In [1, 2] 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 distance 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 $f_t^l(r) = C_f(\acute{\mathbf{j}}(\mathbf{x}_t^l(r)), \acute{\mathbf{j}}(\mathbf{x}_r^l))$ and the node deformation is $d_t^l(r) = C_d(\mathbf{x}_t^l(r), \mathbf{x}_r^l)$. Let $\mathbf{d}_t^l(r) \in \Re^2$ be a column vector that is comprised by the two similarity measures for the node l between the test person t and the reference person r, i.e.:

$$\mathbf{d}_t^l(r) = \begin{bmatrix} f_t^l(r) \\ d_t^l(r) \end{bmatrix} \tag{12}$$

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

$$c_t^l(r) = f_t^l(r) + \lambda d_t^l(r) = \begin{bmatrix} 1 & \lambda \end{bmatrix} \mathbf{d}_t^l(r) = \mathbf{e}^T \mathbf{d}_t^l(r) \tag{13}$$

where λ is the constant that controls the rigidity/plasticity of the graph [1]. In general $\mathbf{e}^T$ does not contain any discriminant information. Thus, when forming the local similarity measure the vector $\mathbf{e}^T$ should be superseded by a discriminant function μ_l^r that is person and node specific. The new local similarity measure is:

$$c_t^l(r) = \mu_r^l(\mathbf{d}_t^l(r)). \tag{14}$$

The discriminant transforms can be constructed by using linear or non-linear methods for building discriminant function. We have used LDA in order to find the discriminant transform μ_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 $\acute{\mathbf{q}}^l(r)$:

$$J(\mathbf{q}^l(r)) = \frac{\mathbf{q}^l(r)^T \mathbf{D}_S^l(r) \mathbf{q}^l(r)}{\mathbf{q}^l(r)^T \mathbf{D}_W^l(r) \mathbf{q}^l(r)}. \tag{15}$$

The optimal weighting coefficients are given by [4]:

$$\acute{\mathbf{q}}^l(r) = \frac{\mathbf{D}_W^l(r)^{-1}(\mathbf{m}(\mathcal{L}_I^l(r)) - \mathbf{m}(\mathcal{L}_C^l(r)))}{||\mathbf{D}_W^l(r)^{-1}(\mathbf{m}(\mathcal{L}_I^l(r)) - \mathbf{m}(\mathcal{L}_C^l(r))||}. \tag{16}$$

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

$$c_t^l(r) = \mu_r^l(\mathbf{d}_t^l(r)) = \acute{\mathbf{q}}^l(r)^T \mathbf{d}_t^l(r). \tag{17}$$

5. 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 fiducial points with weights that correspond to their discriminant power. The weights should be person specific due to the fact that different persons have different discriminant fiducial points. Let $\mathbf{c}_t(r) \in \Re^L$ be a column vector comprised by the new local similarity values at every node:

$$\mathbf{c}_t(r) = \begin{bmatrix} c_t^1(r) \\ c_t^2(r) \\ \vdots \\ c_t^L(r) \end{bmatrix} \tag{18}$$

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 [1] 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 simple the sum of all node similarity measures:

$$D_t(r) = \sum_{i=1}^{L} c_t^i(r) = \mathbf{1}^T \mathbf{c}_t(r), \tag{19}$$

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

$$\acute{D}_t(r) = \beta_r(\mathbf{c}_t(r)). \tag{20}$$

The transform β_r could be just a weighting vector or a more complicated nonlinear support vector machine [3]. 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 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:

$$J(\mathbf{w}(r)) = \frac{\mathbf{w}(r)^T \mathbf{V}_B(r) \mathbf{w}(r)}{\mathbf{w}(r)^T \mathbf{V}_W(r) \mathbf{w}(r)} \tag{21}$$

are the elements of the vector $\acute{\mathbf{w}}(r)$ [4]:

$$\acute{\mathbf{w}}(r) = \frac{\mathbf{V}_W(r)^{-1}(\mathbf{m}(\mathcal{T}_I(r)) - \mathbf{m}(\mathcal{T}_C(r)))}{||\mathbf{V}_W(r)^{-1}(\mathbf{m}(\mathcal{T}_I(r)) - \mathbf{m}(\mathcal{T}_C(r))||}. \tag{22}$$

The similarity distance between the reference person r and the test person t, after all the successively discriminant steps, is given by:

$$\acute{D}_t(r) = \beta_r(\mathbf{c}_t(r)) = \acute{\mathbf{w}}(r)^T \mathbf{c}_t(r). \tag{23}$$

Table 1. Error Rates according to XM2VTS protocol for Configuration 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-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

6. EXPERIMENTAL RESULTS

The experiments were conducted in the XM2VTS database using the protocol described in [5]. The images were aligned using an automatic alignment method. A 8×8 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 $\acute{\mathbf{\Psi}}^l(r)$ 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 $\acute{\mathbf{q}}^l(r)$ for weighting the local similarity vector and the vector, $\acute{\mathbf{w}}(r)$, that weights the total similarity vector of the graph nodes. The evaluation set is also used for learning the thresholds. Table 1 shows the error rates according to the protocol described in [5].

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 denoted as EGM-FD.

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 distances is considered we will use the acronym EGM-LD.

The contribution of weighting the local similarity measure with coefficients that are derived by LDA without other discriminant steps was also investigated. To do so, we applied only discriminant weighting in the graph level by calculating, $\acute{\mathbf{w}}_r$, without applying prior discriminant analysis. The TER obtained was 10.7%. EGM-ND 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. These results are the best that have been reported using an automatic alignment method [6]. The acronym DEGM will be used when all the discriminant steps were used.

7. CONCLUSIONS

The use of discriminant techniques in the EGM framework is explored. The different phases of EGM that discriminant information can be used are indicated. The successively discriminant steps are applied in modified morphological EGM algorithm.

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

[5] K. Messer, J. Matas, J.V. Kittler, J. Luettin, and G. Maitre, "Xm2vtsdb: The extended m2vts database," in *AVBPA'99*, 1999, pp. 72–77.

[6] 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*, 2003, pp. 964–974.