Elastic Graph Matching versus Linear Subspace Methods for Frontal Face Verification

Stefanos Zafeiriou, Anastasios Tefas and Ioannis Pitas
Dept. of Informatics
Aristotle University of Thessaloniki, Box 451, 54124 Thessaloniki, Greece
Tel: +30-2310-996361, Fax: +30-2310-998453
E-mail: pitas@zeus.csd.auth.gr

***Abstract*— In this paper, a comparative study between standard linear subspace techniques such as eigenfaces and fisherfaces and a novel morphological elastic graph matching for frontal face verification is presented. A set of experiments has been conducted in the M2VTS database in order to investigate the performance of each algorithm in different image alignment conditions. The experimental results indicate the superiority of the novel morphological elastic graph matching against all the other presented techniques.**

I. Introduction

Biometrics refer to the automatic identification of a person based on his/her physiological or behavioral characteristics. Research on biometrics has shown significant increase over the past few years due to the increasing demands on security applications. Biometrics can be used either for person recognition or person verification. The two problems are conceptually different. On the one hand, a person recognition system assists a human expert in determining the identity of a test face. On the other hand, person verification systems should decide whether an identity claim is valid or invalid. One of the most popular biometric modalities used by the scientific community is face verification. Face recognition/verification has attracted the attention of researchers for more than two decades and is among the most popular research areas in the field of computer vision and pattern recognition.

The most popular among the techniques used for frontal face recognition/verification are the subspace methods. Subspace methods project the original high dimensional image space into a low dimensional one. The classification is usually performed according to a simple distance measure in the final multidimensional space. Two of the most well studied subspace methods for face recognition are the eigenfaces [1] and the fisherfaces [2]. The main limitation of subspace methods is that they require perfect alignment of the face images in order to be functional.

Another popular class of techniques used for frontal face recognition/verification is elastic graph matching. Elastic graph matching is a simplified implementation of the Dynamic Link Architecture (DLA) [3]. DLA is a general object recognition technique that represents an object by projecting its image onto a rectangular elastic grid where a Gabor wavelet bank response is measured at each node [4]. A variant of elastic graph matching based on multiscale dilation-erosion, the so-called morphological elastic graph matching (MEGM) was proposed and tested for frontal face verification [5].

In this paper a comparative study is done between some standard linear subspace methods, like eigenfaces and fisherfaces, and morphological elastic graph matching using a novel multiscale analysis that is robust against illumination changes. This study was performed in order to investigate the sensitivity of these face verification systems under different image alignment conditions. All the experiments were conducted in the M2VTS database.

II. Subspace Methods

Let M be the number of samples in the image database $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, .., \mathbf{u}_M\}$ where $\mathbf{u}_i \in \Re^n$ is a database's image. A linear transformation of the original n-dimensional space onto a subspace with m-dimensions ($m << n$) is a matrix $\mathbf{W}^T \in \Re^{m\times n}$. The new feature vectors $\mathbf{y}_k \in \Re^m$ are given by:

$$\mathbf{y}_k = \mathbf{W}^T(\mathbf{u}_k - \bar{\mathbf{u}}), \quad k \in \{1, 2, \dots, M\} \tag{1}$$

where $\bar{\mathbf{u}} \in \Re^n$ is the mean image of all samples.

One of the oldest and well studied methods for low dimension representation of faces is the eigenface approach [6]. This representation was used in [1] for face recognition. The idea behind the eigenface representation is to choose a dimensionality reduction linear transformation that maximizes the scatter of all projected samples. The matrix that is connected to the scatter of multidimensional data is the total scatter matrix $\mathbf{S}_T \in \Re^{n\times n}$ defined as:

$$\mathbf{S}_T = \sum_{k=1}^{M}(\mathbf{u}_k - \bar{\mathbf{u}})(\mathbf{u}_k - \bar{\mathbf{u}})^T \tag{2}$$

The transformation matrix, $\mathbf{W}_e^T$, is chosen to be the one that maximizes the determinant of the total scatter matrix $\mathbf{S}_T$ of the projected samples, i.e.,

$$\mathbf{W}_e = \arg\max_{\mathbf{W}} |\mathbf{W}^T\mathbf{S}_T\mathbf{W}| = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_m] \tag{3}$$

where $\mathbf{w}_i \in \Re^n$ is the eigenvector that corresponds to the ith largest eigenvalue of $\mathbf{S}_T$. The matrix $\mathbf{S}_T$ is obvious a very high dimensional matrix. Thus, straightforward calculation of eigenvectors of $\mathbf{S}_T$ is not feasible. Fortunately, due to the fact that its rank is less or equal to $M-1$ there are some computational inexpensive ways to compute it [1].

The transformed feature vectors $\mathbf{y}_k$, produced by this dimensionality reduction method, are called most expressive features because they best express the population [6], [7]. The main drawback of the eigenfaces approach, as a subspace method, is that it does not deal directly with discrimination between classes. In order to use the information of how the data are separated to different classes, Fisher's Linear Discriminant (FLD) is used to produce the linear transformation. Let that each image $\mathbf{u}_i$, in the image database $\mathbf{U}$, belongs to one of the C person classes $\{U_1, U_2, \ldots, U_C\}$. Let the between-class scatter matrix be defined as:

$$\mathbf{S}_B = \sum_{i=1}^{C} N_i (\bar{\mathbf{u}}_i - \bar{\mathbf{u}})(\bar{\mathbf{u}}_i - \bar{\mathbf{u}})^T \tag{4}$$

and the within-class scatter matrix be defined as:

$$\mathbf{S}_W = \sum_{i=1}^{C} \sum_{\mathbf{u}_k \in U_i} (\mathbf{u}_k - \bar{\mathbf{u}}_i)(\mathbf{u}_k - \bar{\mathbf{u}}_i)^T, \tag{5}$$

where $\bar{\mathbf{u}}_i$ is the mean of class U_i, and N_i is the cardinality of class U_i. The goal of the linear transformation $\mathbf{W}_f^T$ is to maximize the between class scatter while minimizing the within class scatter, i.e.,

$$\mathbf{W}_f = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_W \mathbf{W}|} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \ldots \ \mathbf{w}_m]. \tag{6}$$

The advantage of using ratio (6) is that if $\mathbf{S}_W$ is not singular then (6) is maximized when the column vectors of the projection matrix, $\mathbf{W}_f$, are the eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$. For a face database with C classes and M total images, the rank of $\mathbf{S}_W$ is at most $M-C$ and the rank of $\mathbf{S}_B$ is at most $C-1$. Thus, there are at most $C-1$ eigenvectors that correspond to non zero eigenvalues of $\mathbf{S}_W^{-1}\mathbf{S}_B$. To cope with the fact that $\mathbf{S}_W$ has rank $(M-C) << n$, fisherfaces where proposed in [2]. Fisherfaces, is a two step dimensionality reduction method. First the feature dimensionality is reduced to $M-C$ dimensions using the eigenfaces approach in order for $\mathbf{S}_W$ to become non-singular. After that, the dimension of the new features is reduced further using the criterion (6). The total dimensionality reduction transformation to $l \leq C-1$ dimensions is:

$$\mathbf{W}_t^T = \mathbf{W}_f^T \mathbf{W}_e^T \in \Re^{l \times n} \tag{7}$$

where $\mathbf{W}_e^T$ and $\mathbf{W}_f^T$ are the first and the second dimensionality reduction transformations respectively. In [8] was shown that fisherfaces outperform eigenfaces only when large and representative training data sets are available. The main problem of subspace methods is that they require the facial images to be perfectly aligned [8]. That is, all the facial images should be aligned in order to have all the fiducial points (e.g. eyes, nose, mouth, e.t.c.) represented at the same position inside the feature vector. For this purpose the facial images are very often aligned manually and moreover they are anisotropically scaled. Perfect automatic alignment is in general a difficult task to be assessed.

III. Morphological Elastic Graph Matching

A technique for face verification/recognition that does not require perfect alignment in order to perform well is the elastic graph matching [4], [5], [9] algorithm. In all cases [4], [5], [9] no alignment preprocessing step was used. Recently it was shown that morphological elastic graph matching combined with support vector machines had very good performance for frontal face authentication [10]. A more detailed description of elastic graph matching follows.

The facial image region is analyzed and a set of local descriptors extracted at the node of a sparse grid, is created. There are various types of grids proposed in the literature [4], [5], [9]. The simplest is an evenly distributed grid over a rectangular image region. This type of grid was used in the experiments presented in this paper. In all cases, the first step of the elastic graph matching algorithm is to build an information pyramid in the reference face image. In the morphological elastic graph matching this information pyramid is build using multiscale morphological dilation-erosions [11]. Given an image $f(\mathbf{x}) : D \subseteq Z^2 \rightarrow \Re$ and a structuring function $g(\mathbf{x}) : G \subseteq Z^2 \rightarrow \Re$, the dilation of the image $f(\mathbf{x})$ by $g(\mathbf{x})$ is denoted by $(f \oplus g)(\mathbf{x})$. Its complementary operation, the erosion, is denoted by $(f \ominus g)(\mathbf{x})$ [12]. The multiscale dilation-erosion pyramid of the image $f(\mathbf{x})$ by $g_\sigma(\mathbf{x})$ is defined by [11]:

$$(f * g_\sigma)(\mathbf{x}) = \begin{cases} (f \oplus g_\sigma)(\mathbf{x}) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ (f \ominus g_{|\sigma|})(\mathbf{x}) & \text{if } \sigma < 0 \end{cases} \tag{8}$$

where σ denotes the scale parameter of the structuring function.

Such morphological operation can highlight and capture important information for key facial features such as eyebrows, eyes, nose tip, nostrils, lips, face contour, etc. but can be affected by different illumination conditions and noise [5]. To compensate for these conditions, the normalized multiscale dilation-erosion is proposed for facial image analysis. It is well known that the different illumination conditions affect the facial region in a non uniform manner. However, it can safely be assumed that the illumination changes are locally uniform inside the area of the structuring element used for multiscale analysis. The proposed analysis is:

$$(f * g_\sigma)\mathrm{n}(\mathbf{x}) = \begin{cases} (f \oplus g_\sigma)(\mathbf{x}) - \mu_{\mathbf{z} \in G_\sigma}(f(\mathbf{x}-\mathbf{z})) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ (f \ominus g_{|\sigma|})(\mathbf{x}) - \mu_{\mathbf{z} \in G_{|\sigma|}}(f(\mathbf{x}+\mathbf{z})) & \text{if } \sigma < 0 \end{cases} \tag{9}$$

where $\mu_{\mathbf{z} \in G_\sigma}(f(\mathbf{x}-\mathbf{z}))$ and $\mu_{\mathbf{z} \in G_\sigma}(f(\mathbf{x}+\mathbf{z}))$ are the mean values of the image $f(\mathbf{x}-\mathbf{z})$, $\mathbf{x}-\mathbf{z} \in D$ and $f(\mathbf{x}+\mathbf{z})$, $\mathbf{x}+\mathbf{z} \in D$ inside the support area of the structuring element $G_\sigma = \{\mathbf{z} \in G : ||\mathbf{z}|| < \sigma\}$, respectively. The structuring element used in all experiments was cylindrical for computational complexity reasons [5], [12]. The output of these morphological operations form the feature vector $\mathbf{j}(\mathbf{x})$ at the grid node located at image coordinates $\mathbf{x}$. Figure 1 depicts the output of the normalized dilation erosion for various scales. The first nine pictures starting from the upper left corner are eroded images and the remaining nine are dilated images. The new dynamic link architecture will be denoted as normalized morphological elastic graph matching (NMEGM) in the rest of the paper.

Fig. 1. Output of normalized multi-scale dilation-erosion for nine scales.

The next step of the elastic graph matching is to translate and deform the reference graph on the test image so that a cost function is minimized. Let the superscripts t and r denote a test and a reference person (or grid), respectively. The L_2 norm, is used as a similarity measure, between the feature vectors at the lth grid node of the reference and the test graph, i.e. $C_u(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r)) = ||\mathbf{j}(\mathbf{x}_l^t) - \mathbf{j}(\mathbf{x}_l^r)||$. The objective is to find a set of vertices $\{\mathbf{x}_l^t, l \in V\}$ that minimize the cost function:

$$D(t,r) = \sum_{l \in V} \{C_u(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r))\} \text{ subject to} \\ \mathbf{x}_l^t = \mathbf{x}_l^r + \mathbf{s} + \mathbf{q}_l, \;\; ||\mathbf{q}_l|| \leq \mathbf{q}\text{max}, \tag{10}$$

where $\mathbf{s}$ is a global translation of the graph and $\mathbf{q}_l$ denotes a local perturbation of the grid nodes. The choice of $\mathbf{q}\text{max}$ controls the rigidity/plasticity of the graph. The cost function given by (10) defines the similarity measure between two persons, in the morphological elastic graph matching.

IV. Experimental Results

The linear subspace techniques and the morphological elastic graph matching presented in the previous sections have been tested on the M2VTS database [13]. The database contains 37 persons' video data. Four recordings (i.e., shots) of the 37 persons have been collected. Only the luminance information has been considered in all verification techniques. The experimental protocol is depicted in Figure 2 and is an implementation of the "leave one out" principal. When this protocol is applied to the M2VTS database, it gives a total of 5328 impostor and 5328 client claims. The objective in the training procedure is to determine a threshold per person on the distance measure. The linear transformations should also be learned in case of subspace methods. The strategy used for choosing the thresholds is the one described in [5]. An identity claim of a test person is considered as valid if the resulting similarity distance between the test and a reference graph is less than or equal to the predefined threshold. In case of subspace methods the claim is considered valid if the euclidian distance between the test feature vector and one of the reference feature vectors is less than or equal to the predefined threshold. For a given set of thresholds, the performance of the verification system can be expressed using the false acceptance rate (FAR) and the false rejection rate (FRR). Accordingly, the plot of FRR versus FAR can be created using as implicit parameter, the set of thresholds. This plot forms the Receiver Operating Characteristic (ROC) curve of the verification technique. The operating point used more frequently for performance evaluation is the one having FAR equal to FRR and is called Equal Error Rate (EER).

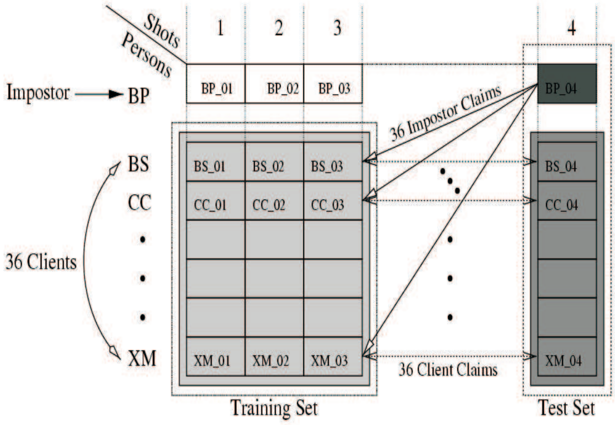

Fig. 2. The M2VTS protocol.

The experiments were conducted in order to illustrate how the performance of each verification system varies when a perfect image alignment system is not available. In the first set of experiments all images were aligned manually according to the eyes position of each facial image. For subspace techniques the background was eliminated in order to diminish its impact in the verification procedure. This step is crucial for subspace techniques since the remaining background around the facial region reduces the verification performance. The achieved EERs for the different verification methods described in this paper are summarized in Table I. The best EER, 6.05%, is achieved using the novel multiscale morphological analysis, presented in Section III.

TABLE I
COMPARISON OF EQUAL ERROR RATES FOR SUBSPACE AND ELASTIC GRAPH MATCHING VERIFICATION TECHNIQUES IN THE ALIGNED (A) AND ATTACKED (BY SCALE (S) AND ROTATION (R)) M2VTS DATABASE.

Verification Technique	EER(A) (%)	EER(R) (%)	EER(S) (%)
NMEGM	6.05	6.65	7.4
MEGM	9.4	10.2	11.1
Eigenfaces	10-40	13.1-38	13-39
Fisherfaces	8.3-26	9.5-26	11-31

Assuming that a perfect alignment method is not available, every image of the aligned M2VTS database is attacked with a random rotation, uniformly distributed in $(-3,3)$ degrees. The minimum EER, for the subspace methods, is increased about $1.2-3.1\%$. Another attack considered, is scaling. Every image of the aligned M2VTS database is attacked with a random scaling, uniformly distributed in $(-5,5)$ pixels. That is, the aligned face images are randomly scaled in such a way that the distance between the eyes to be increased or decreased by $a \in (-5,5)$ pixels. The minimum EER, for the linear subspace, methods is increased about $2.7-3.0\%$.

In Figure 2 it is shown that the training set, at each circle of the "leave-one-out" protocol, is comprised of 36 different client classes and $36 \times 3 = 108$ images. Thus, 107 is the maximum number of eigenfaces and 35 is the maximum number for fisherfaces that can be produced in the training phase. Figure 3 shows the performance of the eigenfaces for various feature dimensions, in the aligned and in the attacked M2VTS database. Figure 4 describes the performance of fisherfaces verification system for various feature dimensions. It is obvious that the performance of the subspace methods greatly depends on the alignment method used by the verification system. Another interesting observation is that in the attacked M2VTS database, the fisherfaces subspace method has very unstable behavior and the performance deteriorates as more discriminant dimensions are kept. On the contrary, elastic graph matching is not so sensitive to geometric distortions. Thus, elastic graph matching is more reliable for real face verification systems where perfect alignment is not available.

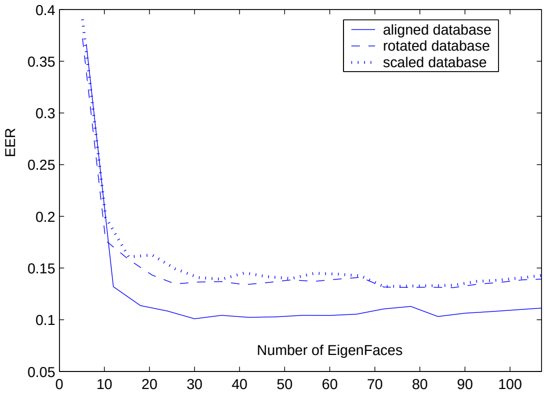

Fig. 3. EER for various number of kept eigenfaces.

V. Conclusions

In this paper a comparison between linear subspace methods and elastic graph matching that use a novel morphological multiscale analysis for frontal face verification was given. Different alignment conditions were considered. The experimental results confirmed the fact that the performance of subspace methods greatly depends on the alignment system used. On the contrary, morphological elastic graph matching is not so sensitive to geometric distortions.

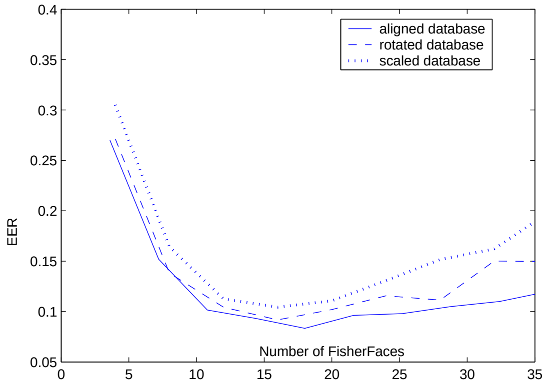

Fig. 4. EER for various number of kept fisherfaces.

Acknowledgment

This work is funded by the network of excellence BioSecure IST-2002-507634 (Biometrics for Secure Authentication, http://www.biosecure.info), under Information Society Technologies (IST) priority of the 6th Framework Programme of the European Community.

References

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

[2] 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.

[3] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der 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.

[4] 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.

[5] 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.

[6] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces.," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

[7] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.

[8] A. Martinez and A. Kak, "Pca versus lda," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[9] L. Wiskott, J. Fellous, N. Krüger, and C. von der 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.

[10] 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.

[11] 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.

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

[13] S. Pigeon and L. Vandendorpe, "The m2vts multimodal face database," *in Lecture Notes in Computer Science: Audioand Video- based Biometric Person Authentication (J. Bigun, G. Chollet and G.Borgefors, Eds.)*, vol. 1206, pp. 403–409, 1997.