THEORETICAL ADVANCES

Gian Luca Marcialis · Fabio Roli

Fusion of appearance-based face recognition algorithms

Abstract Although many algorithms have been proposed, face recognition and verification systems can guarantee a good level of performances only for controlled environments. In order to improve the performance and robustness of face recognition and verification systems, multi-modal and mono-modal systems based on the fusion of multiple recognisers using different or similar biometrics have been proposed, especially for verification purposes. In this paper, a recognition and verification system based on the combination of two well-known appearance-based representations of the face, namely, principal component analysis (PCA) and linear discriminant analysis (LDA), is proposed. Both PCA and LDA are used as feature extractors from frontal view images. The benefits of such a fusion are shown for different environmental conditions, namely, *ideal* conditions, characterised by a very limited variability of environmental parameters, and *real* conditions with a large variability of lighting, scale and facial expression.

Keywords Biometrics · Face recognition and verification · Fusion of multiple classifiers · Multi-modal and mono-modal biometrics · Principal component analysis · Linear discriminant analysis

Introduction

The increase of terrorism and other kinds of criminal actions, such as fraud in e-commerce, increased the interest for more powerful and reliable ways to recognise the identity of a person [1, 2]. To this end, the use of behavioural or physiological characteristics, called *biometrics*, is proposed. Biometrics are unique from person-to-person, cannot be forgotten and it is very difficult to steal and reproduce them. Commonly used biometrics include voice, fingerprints and the face. For example, the automatic fingerprint identification systems (AFIS) are widely used for criminal investigations [1].

The face is another well-known biometric. First of all, it is the most natural physiological characteristic with which to recognise each other. Hence, people consider the face as a ''good'' biometric. Moreover, the acquisition process is basically non-intrusive and can be performed without the cooperation of the subject, if the person is not ''hostile'' (i.e. he/she does not forge his/her appearance, nor does not hide himself/herself from the camera). For these reasons, it is acknowledged that face recognition could play an important role in advanced video-based surveillance systems and personal authentication. Unfortunately, face recognition algorithms suffer a lot from the high variability of environmental conditions. As an example, the effectiveness of face recognition strongly depends on lighting conditions and on variations in the subject's pose and expression in front of the camera.

For the purposes of this paper, face recognition applications can be subdivided in two types: applications in controlled and uncontrolled environments. One of the main applications of the first type is the so called *identity authentication*. Identity authentication (or verification) is necessary for allowing or denying the access to restricted areas or resources. A person submits to the automatic identity verification system their face (frontal and/or profile view) and declares her/his identity. The system matches the acquired face with the *template* stored in its database, and classifies the person as a *genuine* (i.e. the claimed identity is accepted) or an *impostor*. Recognition is performed by evaluating the degree of similarity, or *score*, between the acquired face and the template face. If the score is higher than a certain value, named *accep-*

G. L. Marcialis (✉) · F. Roli (✉)
Department of Electrical and Electronic Engineering,
University of Cagliari, Piazza d'Armi,
09123 Cagliari, Italy
E-mail: marcialis@diee.unica.it
E-mail: roli@diee.unica.it
Tel.: +39-070-6755779
Fax: +39-070-6755782

tance threshold, the person is classified as *genuine*, otherwise, he/she is classified as *impostor* and the access to the required resource is denied. Automatic identity verification based on face recognition is usually performed in controlled environments, and requires co-operation with the person. This means that the face image is acquired in a certain pose or a certain number of fixed poses (e.g. frontal and profile views). Lighting, scale and other environmental conditions exhibit a very limited variability range.

Applications of the second type refer to the problem of recognition of an identity in a scene, and they are typical of video-surveillance tasks. The recognition system first detects the face [3] in the video sequence and normalises it with respect to pose, lighting and scale conditions. Then, it tries to associate the face with one or more faces stored in its database, and provides, as outputs, the set of faces that are considered "nearest" to the detected face. This problem is much more complex than the previous verification problem. It is computationally expensive and needs robust algorithms for detection, normalisation and recognition. In both verification and recognition applications, other problems can arise. For example, some kind of forgery in the subject appearance (e.g. beard).

Although many algorithms for both face recognition and verification tasks have been proposed, none of them can completely handle the above weaknesses (effective handling of lighting, pose and scale variations is a matter of on-going research). In general, face recognition and verification systems can achieve good performances only for controlled environments. In the last few years, the fusion of multiple recognisers using different biometrics has been proposed, especially for verification purposes (multi-modal biometrics systems). For example, the fusion of face, hand geometry and voice [4] has been proposed in order to overcome some limits of the individual biometrics. The main drawback of such systems is the difficult implementation, as they require different sensors, algorithms and hardware for the biometrics processing, which causes a big increase in their cost.

Another possibility is the fusion of multiple recognisers using the same biometric (mono-modal biometric systems). For example, the fusion of multiple face recognition algorithms. Mono-modal systems are obviously less expensive than multi-modal ones. However, "complementary" recognisers are necessary, based on different representations of the given biometric, in order to improve the performance of the individual recogniser. To this end, the fusion of recognisers using different views of the face was used [5, 6]. Few works have dealt with mono-modal systems using the same view of the face. In particular, the fusion of PCA and LDA has not received much attention [24, 25, 36]. We believe that the apparent strong correlation of LDA and PCA, especially when frontal views are used and PCA is applied before LDA, discouraged the fusion of such algorithms. However, our experiments show the potential of such a combination. In this paper, a recognition and verifica-

tion system based on the combination of PCA and LDA is presented. Both PCA and LDA are used as feature extractors from frontal view images. When an unknown face is submitted to the system, it is projected into the PCA and the LDA spaces. Then, the distances from the templates stored in the database are computed. The distance vectors in the PCA and LDA spaces are combined and the final recognition is performed by applying a simple decision scheme. We investigated the effectiveness of our system for both recognition and verification. Experimental results were carried out with three data sets that represent different environmental conditions, namely, "ideal" conditions, characterised by a very limited variability of environmental parameters, and "real" conditions with large variability of lighting, face scale and expression

Section 2 gives a short overview of previous works on the fusion of face recognition and verification algorithms. Section 3 briefly describes the two appearance-based approaches to face representation and recognition used, namely, PCA and LDA. Section 4 describes the algorithms for fusing PCA and LDA, both for face recognition and verification purposes. Section 5 reports experimental results and conclusions are drawn in Sect. 6.

Previous works on the fusion of face recognition algorithms

A good survey of face recognition systems can be found in [7]. In the following, we briefly review the main works on the fusion of multiple face recognisers.

Many face recognition algorithms have been proposed so far [7, 8, 9, 10, 11, 12, 13]. However, none of them is able to effectively handle the large variability of critical parameters, like pose, lighting, scale, facial expression or some kind of forgery in the subject appearance (e.g. beard).

In order to improve the performance and robustness of individual recognisers, the use of multiple classifier systems (MCSs) has been recently proposed. MCSs are currently a very active research field [14]. Multiple classifier systems cover a wide spectrum of applications: hand-written character recognition, fingerprint classification and matching, remote-sensing images classification, etc. The effectiveness of this approach is proved by many experimental results [14].

A few approaches for improving the performance and the robustness of face recognition using MCSs have also been proposed. In the following, we briefly review them.

Achermann and Bunke [6] proposed the fusion of two recognisers based on frontal and profile face views. The individual recognisers are based on hidden Markov models, eigenfaces (frontal views) and shape classifiers (profile view) [15]. The outcome of each expert, represented by a score, i.e. a level of confidence about the decision, is combined with the following fusion rules: consensus voting, majority voting, weighted majority

voting, rank summation, score summation and Bayes combination rule. The best results were reported with the voting methods; about 97.7% recognition accuracy for a home-made data set made up of 600 frontal face images and 150 profile face images. The number of classes (identities) was 30. Achermann and Bunke results show a sharp improvement with respect to the best individual classifier, that is, the eigenface classifier, which achieved 94.7% recognition accuracy.

Lucas [16] used a *n*-tuple classifier for combining the decisions of experts based on sub-sampled images. Testing was performed with the AT&T data set, made up of 400 face images and 40 classes (see Sect. 5 for more details about the AT&T data set). Lucas pointed out the advantages of his approach, that is, the simplicity and the effectiveness. Results were compared with others obtained by using more sophisticated approaches, such as the probabilistic decision-based neural net of Lin and Kung [17] and the convolutional neural net by Lawrence et al. [18]. Lucas showed that the *n*-tuple classifier performs much better than the other approaches, reporting 97.3% recognition accuracy with respect to 96.4% for the convolutional approach.

Tolba and Abu-Rezq [19] presented four combination rules tailored to the individual classifiers used: a radial basis function network (RBF) and a linear vector quantisation classifier (LVQ). Both classifiers were trained and tested on the AT&T data set. The combination rules are basically "expert" rules (e.g. "if the pattern is rejected by the LVQ classifier, then reject it"). The authors reported 99.5% recognition accuracy for a rejection rate of 0.5%. The individual LVQ network reported 99.0% recognition accuracy, while the RBF network reported 98.0%.

The above works focused on the recognition task. In the field of face verification, most of the literature used multi-modal approaches. For example, the face is coupled with one or more biometrics, such as fingerprints and voice.

Roli et al. [20] described the combination of two speech classifiers with six face classifiers using frontal views. Speech recognisers were based on static and dynamic statistical models, while the frontal face recognisers were based on appearance-based (LDA) and fractal approaches [21]. The investigated combination rules were: majority voting, sum rule, behaviour-knowledge space (BKS) and decision templates (DT). The XM2VTS data set [22] was used for experiments according to the protocol described in [23]. The XM2VTS data set is made up of 2,360 images (295 classes). Experiments were aimed to investigate the conditions under which the combination rules needing a training phase (such as BKS and DT rules) could perform better than the simpler ones.

Verlinde et al. [5] investigated a mono-modal system based on different face views and a multi-modal system. The multi-modal system combined the face-based mono-modal system with a speech-based system. The authors focused on the performance improvement achievable with the multi-modal system. The rules for classifier combinations were: majority voting, linear and quadratic classifiers, K-NN, decision trees, logistic regression and Bayes combination rule. Experiments were performed with the M2VTS data set (the initial version of the XM2VTS data set) [22].

Ross and Jain [4] described the fusion of face, speech and hand geometry for personal authentication. A home-made data set was used for the experiments; such a data set was made up of 250 images per biometric (50 classes). Experiments were performed by simple combination rules such as sum, decision trees and linear discriminant functions. Reported results showed that such a fusion allows improving performance with respect to the best individual classifier.

Finally, the authors [24, 25] reported preliminary experiments on the fusion of two appearance-based approaches, PCA and LDA, for face verification and recognition.

PCA and LDA for face representation

Let X be a d-dimensional feature vector. In our case, d is equal to the number of pixels of each face image. The high dimensionality of X is a well-known problem for the design of a good face recognition algorithm. Therefore, methods for reducing the dimensionality of this image space are required. To this end, principal component analysis (PCA) and linear discriminant analysis (LDA) are widely used.

Principal component analysis [8, 26] is defined by the transformation:

$$y_i = W_{\text{PCA}}^t x_i \tag{1}$$

where $x_i \in X \subseteq \Re^d$, $i = 1,..., n$ (n samples). W_{PCA} is a d-dimensional transformation matrix whose columns are the eigenvectors related to the eigenvalues computed according to the formula:

$$\lambda e_i = S e_i \tag{2}$$

where S is the scatter matrix (i.e. the covariance matrix):

$$S = \sum_{i=1}^{n} (x_i - m) \cdot (x_i - m)^t, \ m = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3}$$

This transformation is called Karuhnen-Loeve transform. It defines the d-dimensional space in which the covariance among the components is zero, because the covariance matrix is diagonal. The eigenvalues correspond to the variances of each component in the transformed space. After ordering the eigenvalues by increasing order, it is possible to consider a small number of *principal* components exhibiting the highest variance. The principal components of the transformed space are also called the most *expressive* features, and the eigenvectors related to the most expressive features are called *eigenfaces*.

The linear discriminant analysis (LDA, also called Fisher discriminant analysis) [9, 26] is defined by the transformation:

$$y_i = W_{LDA}^t x_i \tag{4}$$

The columns of W_{LDA} are the eigenvectors of $S_w^{-1} S_b$, where S_w is the *within-class scatter matrix*, and S_b is the *between-class scatter matrix*. It is possible to show that this choice maximises the ratio $\det(S_b)/\det(S_w)$.

These matrices are computed as follows:

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{n_j} \left(x_i^j - m_j \right) \cdot \left(x_i^j - m_j \right)^t, \quad m_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i^j \tag{5}$$

where x_i^j is the *i*-th pattern of the *j*-th class, and n_j is the number of patterns for the *j*-th class.

$$S_b = \sum_{j=1}^{c} \left(m_j - m \right) \cdot \left(m_j - m \right)^t, \quad m = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{6}$$

The eigenvectors of the LDA are called *fisherfaces* and the components of the transformed space are the most discriminant features. LDA transformation is strongly dependent on the number of classes, *c*, the number of samples, *n*, and the original space dimensionality, *d*. It is possible to show that there are almost $c-1$ non-zero eigenvectors, $c-1$ being the upper bound of the discriminant space dimensionality. We need $d+c$ samples at least to have a non-singular S_w. It is impossible to guarantee this condition in real applications. Consequently, an appropriate "regularisation" technique is necessary [9, 26, 27, 28. 29, 30, 31, 32] to make the S_w matrix invertible. A possible solution is to "perturb" S_w by a regularisation coefficient, σ, and to perform LDA using the "regularised" S_w, indicated in the following as S_w^{reg} (regularised LDA [30]):

$$S_w^{reg} = S_w + \sigma \cdot I \tag{7}$$

where I is the identity matrix.

Another solution is to perform a transformation to an intermediate feature space with reduced dimensionality $d_{intermediate} << d$. As an example, the PCA transformation has been proposed to generate such an intermediate space by Belhumeur et al. [9] and Zhao et al. [27]. We also used the LDA after the application of

the PCA transform, thereby, using the S_w^{reg} defined as follows:

$$S_w^{reg} = W_{PCA}^t S_w W_{PCA} \tag{8}$$

where W_{PCA} is the PCA transformation matrix according to Eq. 1.

It is worth noting that other regularisation techniques have been proposed for solving the small sample size problem of the LDA [28, 29, 30, 31, 32].

## Fusion of PCA and LDA for face verification and recognition

In this section, we present our methodology for fusing the two appearance-based approaches previously described. Figure 1 gives an overview of the proposed method. Recognition and verification are implemented by the following steps:

- Representation of the input face image by PCA and LDA features
- The Euclidean distance vectors, d^{PCA} and d^{LDA}, of the input image from all of the *N* face templates stored in the database are computed
- For face recognition, we fuse the two distance vectors by a transformation rule, and apply two types of classification algorithms, namely, *nearest mean* and *nearest neighbour*
- For face verification, we fuse the two distance vectors and make the final decision by applying a weighted acceptance threshold

In this paper, we use the Euclidean distance because it is the simplest way to compute a distance among feature vectors. A more sophisticated distance could be used [33], but here, we preferred to simplify the distance computation in order to focus on the benefits of fusion.

It is worth noting that many works have analysed the differences between PCA and LDA (see, in particular, [9, 34]), but, to the best of our knowledge, no work has investigated the possibility of fusing them. In our opinion, the apparent strong correlation of LDA and PCA, especially when frontal views are used and PCA is applied before LDA, discouraged the fusion of such algorithms. However, it should be noted that LDA and

Fig. 1 Scheme of the proposed face recognition and verification system. The mugshot has been taken from the AT&T data set
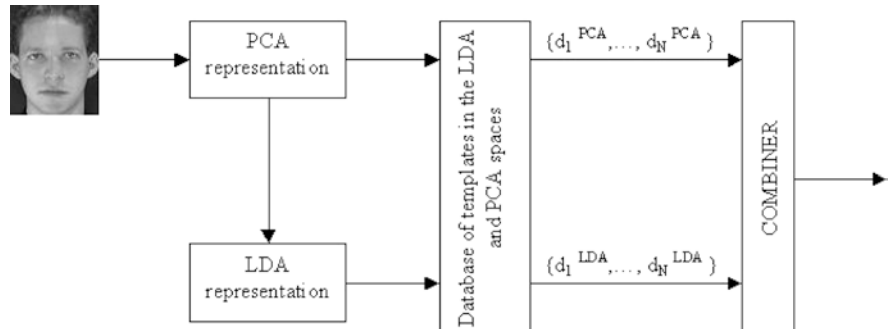
PCA are not as correlated as one might think, since the LDA transformation applied to the principal components can generate a feature space significantly different from the PCA one. Results reported in the literature appear to suggest that PCA and LDA could have a certain degree of complementarity [34]. Results reported in Sect. 5 confirmed this claim.

Fusion of PCA and LDA for face verification

Any face verification system is characterised by the acceptance threshold computation. Our system uses a single threshold. By tuning only one real value, named threshold, we increase or decrease the ''security'' degree of our system. In the following section, further details about threshold computation are given.

Before proceeding, it is worth noting that in our experiments:

(a) All data sets were subdivided into two parts: one was used as training set and the other as test set
(b) We computed the transformation matrix of PCA and LDA, and the template vector for each class (''identity'') from the training set (we selected the mean vector)
(c) We used the test set in order to compute the detection-error trade-off curves (DET) according to [35]

Our approach to threshold computation

The acceptance threshold is usually computed by using some kind of metric. The Euclidean distance is the most commonly used metric in PCA- or LDA-based systems. A discussion about the opportunity of this choice can be found in [33]. In our work, we used the Euclidean distance as the metric. As remarked above, our decision scheme is based on a single threshold. Multiple thresholds related to different identities can be used as well. In order to overcome the limitation due to the use of a single threshold, we used a ''weighted'' threshold. For a given identity, we computed the Euclidean distance of all patterns from the template, thereby, generating a vector of distances. Then, for each identity, the related weight can be computed in various ways. We used the average of all of the distances. Other choices are possible as well (e.g. the minimum distance can be used as weight). The rationale behind this is that the identity's weight can be seen if we consider that multiplying the acceptance threshold for this value makes the final decision (acceptance/rejection). The obtained weighted threshold is a simple way to handle the differences among classes. We do not have to tune different thresholds for the different classes, as the single threshold is tuned by the class-related weights. Obviously, different weights for PCA and LDA are used.

The fusion algorithms

In the following, we describe:

(a) The methods for combining the two weights associated to PCA and LDA, called w_j^{PCA} and w_j^{LDA}
(b) The methods for combining the distances from a candidate pattern to the templates associated to PCA and LDA, called d^{PCA} and d^{LDA}

The term w_j defines the weight derived from the combination of the PCA and LDA weights. Item (b) refers to the final distance, d, that will be compared with the product of the threshold, γ, with w_j. We call such a product the *weighted threshold*. The claimed identity will be accepted if:

$$d < w_j \cdot \gamma \tag{9}$$

The normalisation of all weights and distances for each class is required. We used the well-known formula:

$$\mathrm{valueNorm} = \frac{\mathrm{value} - \mathrm{valueMin}}{\mathrm{valueMax} - \mathrm{valueMin}} \tag{10}$$

According to this formula, distances take values in the range [0,1], therefore, the same meaning of a matching score can be attributed to them. The minimum distance value means that the compared faces are the same face, while the maximum distance value means that the compared faces are definitely different. The opposite holds for a matching score.

In the following, the fusion algorithms used in our experiments are described in terms of the computations of the above weight, w_j, and distance, d:

''Mean'' algorithm:

$$d = \frac{d^{\mathrm{PCA}} + d^{\mathrm{LDA}}}{2}; \ w_j = \frac{w_j^{\mathrm{PCA}} + w_j^{\mathrm{LDA}}}{2} \tag{11}$$

''Max'' algorithm:

$$d = \frac{d^{\mathrm{PCA}} + d^{\mathrm{LDA}}}{2}; \ w_j = \max\left\{ w_j^{\mathrm{PCA}}, w_j^{\mathrm{LDA}} \right\} \tag{12}$$

''MaxMin'' algorithm:

$$d = \min\left\{ d^{\mathrm{PCA}}, d^{\mathrm{LDA}} \right\}; \ w_j = \max\left\{ w_j^{\mathrm{PCA}}, w_j^{\mathrm{LDA}} \right\} \tag{13}$$

It is easy to see that, for the same threshold value, γ, an impostor pattern can be accepted using Eq. 13 and rejected using Eq. 11. Let us suppose that $\bar{d}$ and $d_{\min}$ are the combined distance values of Eqs. 11 and 12 and Eq. 13 for a certain pattern, respectively, and $\bar{w}_j$ and $w_j^{\max}$ are the combined weights using Eq. 11 and Eqs. 12 and 13, respectively. Let us suppose that $\gamma = \frac{\bar{d}}{w_j^{\max}} + \varepsilon$, with ϵ being a small positive or negative value. Under this assumption, the pattern is rejected using Eq. 11 and accepted using Eq. 13. Nevertheless, the pattern is accepted using Eq. 12 if $\epsilon > 0$. So, for the same threshold value, the rejection rate decreases from Eq. 11 to Eq. 13. Conse-

quently, the above fusion strategies have an intrinsically decreasing degree of "security".

Fusion of PCA and LDA for face recognition

We propose two kinds of approaches to fuse PCA and LDA face representations: the nearest neighbour approach (NN) and the nearest mean approach (NM) [26].

First of all, we normalise the distance vectors, d^{PCA} and d^{LDA}, in order to map the range of these distances to the interval [0,1] according to Eq. 10.

Then, a *combined distance vector*, d, that must contain both PCA and LDA information, is computed. To this end, the following two techniques have been used:

– The combined distance vector is computed as the mean vector:

$$d = \left\{ \frac{d_1^{PCA} + d_1^{LDA}}{2}, ..., \frac{d_N^{PCA} + d_N^{LDA}}{2} \right\} \qquad (14)$$

– The combined distance vector is computed by appending the d^{PCA} and d^{LDA} vectors:

$$d = \left\{ d_1^{PCA}, ..., d_N^{PCA}, d_1^{LDA}, ..., d_N^{LDA} \right\} \qquad (15)$$

where N is the number of images in the database. If C is the number of the identities, also called *classes*, an identity, c, is associated to each couple $\left(d_j^{LDA}, d_j^{PCA} \right)$, $j = 1, ..., N$.

After computing and ordering the combined distance vector, d, we followed the NN strategy: the class associated to the minimum distance of d is selected. If the combined distance vector follows Eq. 14, we call our algorithm "Mean-NN" (M-NN); if it follows Eq. 15, we call our algorithm "Append-NN" (A-NN).

In the case of the NM approach, we first compute a template for each identity in the database. We select the average image for both the PCA and the LDA representations. Consequently, our distance vectors, d^{PCA} and d^{LDA}, are composed of C components instead of N. These vectors are combined according to Eqs. 14 or 15. The identity associated to the smallest combined distance is selected. The related algorithms are called "Mean-NM" (M-NM), and "Append-NM" (A-NM), respectively.

Experimental results

In this section, we report our experiments for two well-known face databases: the AT&T and the Yale data sets.

For the case of the face recognition task, we also report the results for the MIT data set.

The data sets

The AT&T data set is made up of ten different images of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All of the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The data set was subdivided into a training set, made up of five images per class/identity (200 images), and a test set, made up of five images per class (200 images). In order to assess verification performances, we used all possible combinations of five images out of ten to generate the training and the test sets (252 cases). Reported results refer to the average detection-error trade-off curves (DET) for the 252 cases. In order to assess recognition performances, we repeated our experiment for ten random partitions of the data set. Reported results refer to the average rank order statistics of the ten runs. Figure 2 shows an example of the face images from the AT&T data set. The AT&T data set is publicly available at the http://www.cam-orl.co.uk/facedatabase.html. This data set was used to simulate an "ideal" environment for face recognition ad verification because the environmental variations are very small.

The Yale data set is made up of 11 images per 15 classes/identities (165 total images). Each face is characterised by different facial expressions or configurations: centre-light, with/without glasses, happy, left-light, normal, right-light, sad, sleepy, surprised and wink. The data set was subdivided into a training set, made up of five images per class (75 images), and a test set, made up of six images per class (90 images). To assess verification performances, we used all possible combinations of five out of eleven images to generate the training and the test sets (330 cases). Reported results refer to the average DET curves for such 330 cases. To assess recognition performances, we repeated our experiments for ten random partitions of the data set and reported the average rank order statistics. Figure 3 shows an example of face images from the Yale data set. The Yale data set is publicly available at http://cvc.yale.edu/projects/yalefaces/yalefaces.html. This data set was used to simulate a "real" environment with notable lighting and facial expression variations.

The MIT data set was taken at the MIT Media Laboratory. The subjects are 16 males (mostly grad students). Each subject sat on a couch and was digitised 27 times, varying the head orientation, the lighting and

Fig. 2 Examples of face images from the AT&T data set

Fig. 3 Examples of face images from the Yale data set

the scale (camera zoom). The variations in lighting, etc. are not precisely calibrated, and no effort was made to keep the subjects from moving in between pictures. In our experiments, we used the images based on the scale variations, thereby, obtaining three data sets, one for each head rotation (144 face images per data set). Figure 4 shows an example of face images from the MIT data set. The MIT data set is publicly available at ftp://whitechapel.media.mit.edu/pub/images/face-images/. This data set was used to simulate a "real" environment for face recognition purposes, with scale and lighting variations. Large-scale variations make poor sense for verification purposes, as the environmental conditions are usually well controlled for identity verification applications.

Results for face verification task

After computing the PCA and LDA matrices on the training set, we computed the DET curves using the genuine and impostor matching distances on the test set. For each combination performed on the AT&T data set, we computed 200 genuine matching distances and 7,800 impostor matching scores. For each combination performed on the Yale data set, we computed 95 genuine matching distances and 1,260 impostor matching distances. With the term "genuine matching distances", we refer to the distances computed by comparing faces of the same identity; with the term "impostor distances", we refer to the distances computed by comparing faces of different identities.

We compared our algorithms with two classical fusion strategies called "AND" and "OR" [5]. Claimed identity is accepted when both experts accept it ("AND"), or when at least one accepts it ("OR").

For AT&T and Yale data sets, performances are reported in terms of the average DET curves [35] of the individual algorithms (PCA and LDA) and their fusion with the strategies described in Sect. 4. The DET curves show the false rejection rate (percentage of rejected genuine users) as a function of the false acceptance rate (percentage of accepted impostors) for all threshold values.

For the AT&T data set, Fig. 5 shows the average DET curves. These curves point out that the fusion of PCA and LDA outperforms the best individual verification algorithm. In addition, the proposed fusion algorithms perform better than the classical AND/OR rules. The "mean" algorithm provided the best performance. It is worth noting that the AT&T data set represents an "ideal" work condition for verification purposes. As an example, the individual PCA recogniser worked well too. However, results show that the fusion allows improving performances.

For the Yale data set, Fig. 6 shows the average DET curves. Even in this case, the fusion of PCA and LDA outperforms the best individual verification algorithm and the classical AND/OR rules. In this case, the different fusion rules exhibit similar performances. We think that the environmental variations in the Yale data set can be considered as "noisy" information that limits the effectiveness of the feature extraction performed by the PCA and LDA. In this case, there is no fusion rule much better than the others. We believe that the strong environmental variations among each session of experiments, as can be seen in Fig. 3, limited the performance differences among fusion rules. It is worth noting that such variations of expressions and lighting are not usually permitted in the controlled environments used for personal identity authentication. However, the results pointed out that fusion improved performances with respect to the best individual recogniser.

Results for face recognition task

Figure 7 shows the rank order statistics, i.e. the percentage accuracy that can be achieved by considering the M faces of the database nearest to the given input face. The input face is considered as correctly recognised if the right identity is associated to one of the first M faces. Such curves measure the system reliability—very important for video-surveillance applications in uncontrolled environments. The combination of PCA and LDA gives a sharp improvement in the performance, and a better identification reliability.

It is worth noting that the best combination result for rank = 1 (97.3%) is comparable with that reported in [16] and [20]. In [16], a 97.5% accuracy is reported, but it is averaged only on five runs; in [20], a 99.5% accuracy is reported, but with a rejection rate of 0.5%. It is also

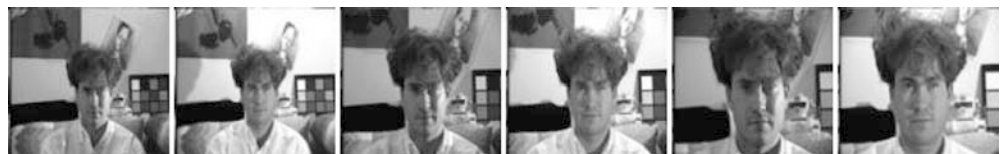**Fig. 4** Examples of face images from the MIT data set

Fig. 5 For the AT&T data set, the average DET curves of the individual algorithms (PCA and LDA) and their fusion with different strategies are depicted. In order to assess verification performances, we used all possible combinations of five images out of ten to generate the training set. Each DET curve refers to the average of the 252 cases considered
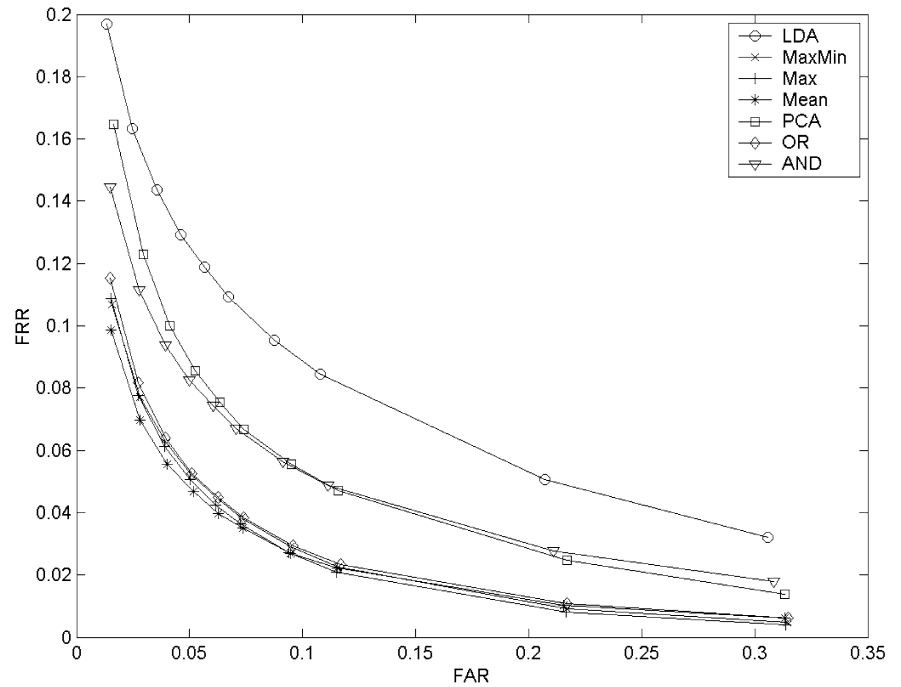

worth noting that the approaches in [16, 20] are more complex than ours; in [16], the combination of 500 face recognisers is necessary to achieve the above identification accuracy, whilst in [20], the individual algorithms (LVQ and RBF) are intrinsically more complex than our minimum Euclidean distance approach.

The identification performance improvement points out a certain complementarity between the PCA and the LDA. In this case, such complementarity is also suggested by the average correlation coefficient that we

computed between the d^{PCA} and the d^{LDA} vectors. A low value was obtained: 0.39. We think that the PCA and LDA are weakly correlated thanks to the good quality of the images in terms of pose and lighting conditions. However, high correlation among such vectors does not necessarily implies an ineffectiveness of the fusion [14].

Whilst the AT&T data set is characterised by small variations of pose and lighting, the Yale data set is characterised by strong variations of expression and lighting. This task is, therefore, more complex and the

Fig. 6 For the Yale data set, the average DET curves of the individual algorithms (PCA and LDA) and their fusion with different strategies are depicted. In order to assess verification performances, we used all possible combinations of five images out of ten to generate the training set. Each DET curve refers to the average of the 330 cases considered
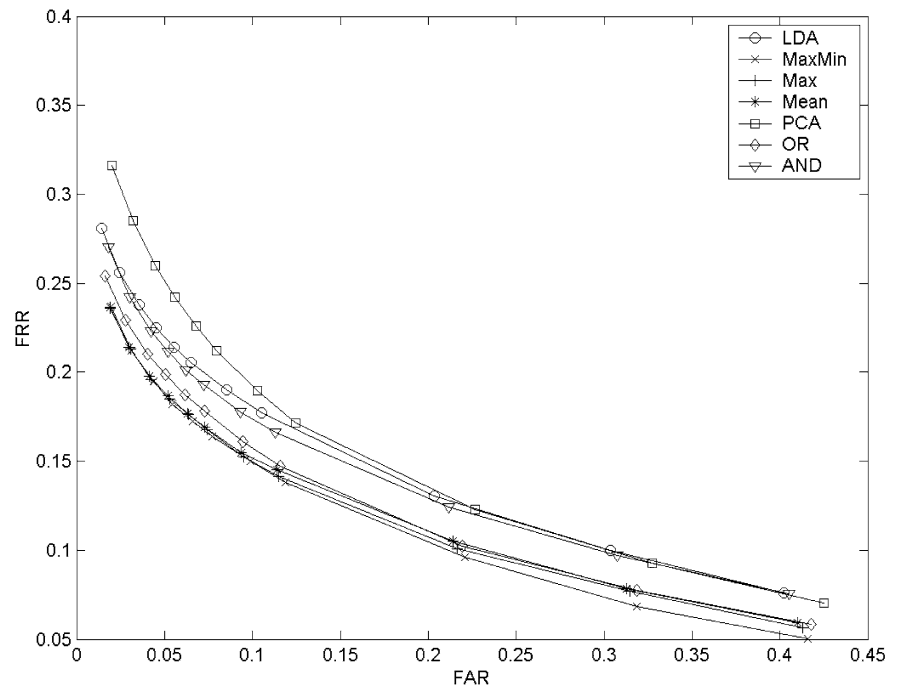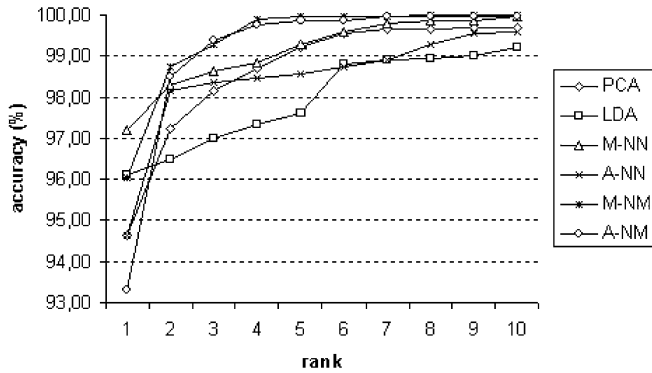
Fig. 7 Rank order statistics for the AT&T data set. Reported results show that the combination of PCA and LDA improves the reliability of the system

Fig. 9 Standard deviation of the identification accuracy on the test set as a function of the first *M* recalls from the training set for the AT&T data set. Reported results show that such standard deviations decrease more rapidly using the fusion of PCA and LDA, thereby, increasing the "stability" of the rank order statistics reported in Fig. 7

results are obviously worse, even if the number of identities is inferior.

Even in this case, the rank order statistics reported in Fig. 8 show the effectiveness of the decision fusion for improving the performance and the reliability of a face recognition system. It is worth noting that the use of a template condensing the various lighting and face expressions of the training set (nearest mean approach) helps to considerably increase the performance of the system (M-NM and A-NM curves of Fig. 8). On the other hand, for the AT&T data set, which exhibits small environmental variations, no significant performance differences are pointed out by using one template or multiple templates (Fig. 7).

Unfortunately, in this case, we could not compare our results with others, because no work reported in the literature used the Yale data set for combining multiple algorithms for face recognition.

It should be noted that the average correlation coefficient in this case is high: a value of 0.69 was obtained. In our opinion, PCA and LDA are correlated because of the lighting and face expression variations in the images. As said in Sect. 5.2, the above variations can be considered as "noisy" information that limits the goodness

of the feature extraction performed by PCA and LDA. However, the identification accuracy increase by our fusion algorithms shows that this high correlation can be partially overcome.

Figures 9 and 10 show the standard deviation of the identification accuracy on the test set as a function of the first *M* faces (rank) from the training set of the AT&T and the Yale data sets, respectively. As well as the average identification accuracy, the standard deviation is related to the ten runs in our experiments. With the standard rank order statistics, which show the average identification accuracy as a function of the training set recalls, these curves help in pointing out further the benefits of fusion. In all cases, the standard deviation decreases more rapidly for the combined recognisers. This means that the rank order statistics related to the combined systems are more "stable". This result is particularly relevant for the Yale data set (Fig. 10), which exhibits large lighting and facial expression variations. On the other hand, the curves related to the AT&T data set confirm that, being such a data set
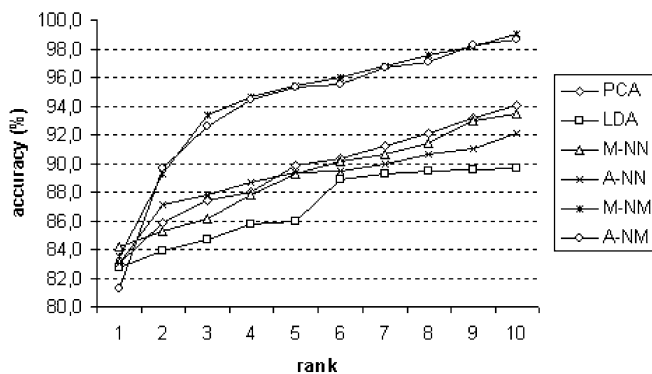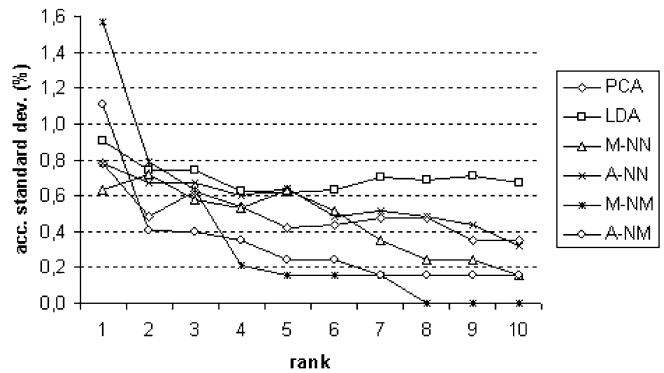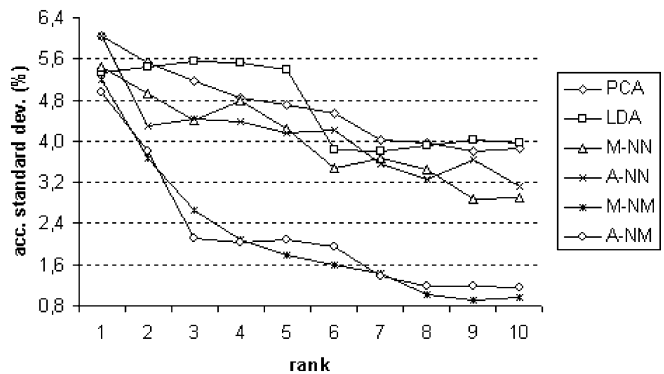

Fig. 8 Rank order statistics for the Yale data set. Reported results show that the fusion improves the performance of the individual algorithms

Fig. 10 Standard deviation of the identification accuracy on the test set as a function of the first *M* recalls from the training set for the Yale data set. Reported results show that such standard deviations decrease more rapidly using the fusion of PCA and LDA, thereby, increasing the "stability" of the rank order statistics reported in Fig. 8

subjected to small environmental variations, the use of one template embedding environmental variations (nearest mean approach) is substantially equivalent to the use of multiple templates (nearest neighbour approach).

Finally, Figs. 11and 12 report the rank order statistics and the relative standard deviations for the MIT data set. For these experiments, images nearest and farthest from the camera were used for training, and the intermediate images were used for testing (Fig. 4). We averaged the results for each of the three admitted head rotations. It should be noted that the individual recognisers performed very bad, especially the PCA-based one. In this case, the NM-based classifiers exhibited the best performance: it is worth noting the improvement of about 15% over the best individual classifier (LDA). The effectiveness of the NM approach could be explained with the expressive power of the average image, that is, the template image for each class, in both PCA and LDA spaces. We hypothesise that (a) the template has also been able to effectively represent the images acquired at an intermediate degree of scale, and (b) the template also smoothed the small background variations due to the scale variability (see Fig. 4). Consequently, fusion by NM-based decision increases the recognition robustness with respect to scale variations. This property is very interesting because, in many recognition applications, such as video-surveillance, the system must make a decision on faces acquired at scale degrees that are "intermediate" with respect to those for which the system has been trained. If this property was confirmed by further experiments, then we could try to better exploit it for handling the scale variations problem.

Scalability of the proposed system

The proposed approach requires that the memory of our face recognition system must contain:

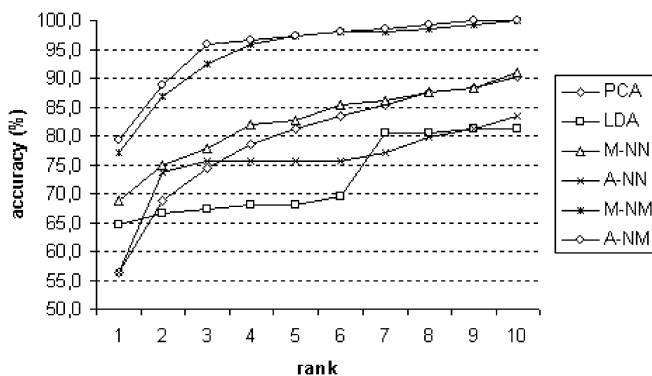1. The transformation matrices of PCA and LDA for coding the input image

Fig. 11 Rank order statistics for the MIT data set. Reported results point out that the fusion can handle the scale variations, thereby, improving performance with respect to the best individual classifier
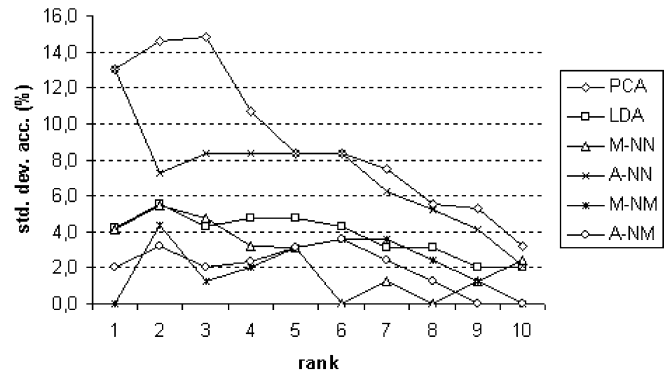

Fig. 12 Standard deviations of the rank order statistics for Fig. 11. Reported results point out that the rank order statistics related to the fusion algorithms are more "stable" than those related to the individual algorithms

2. The templates in the PCA and LDA spaces

Therefore, as it is well-known for the case of appearance-based face recognition systems, memory space problems could arise if a large-scale system should be deployed for a complex real application. In order to perform a preliminary investigation of this issue, we performed, with the AT&T data set, PCA and LDA transformations by increasing the number of classes.

We considered no cropped face images, thereby, evaluating PCA and LDA transformations on the 10,304-dimensional original feature vector of the AT&T data set. Then, we increased the number of classes from 10 to 40, and evaluated the number of the F significant features (that is, the eigenfaces) used by the PCA. Such a number, F, has been computed according to the ratio:

$$F = \arg\min_f \left(\frac{\sum_{i=1}^{f} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geqslant 0.95 \right) \qquad (16)$$

with λ_i being the i-th PCA-ordered eigenvalue and d the original face space dimensionality. Equation 16 is the standard way to select the significant features generated by the PCA [24]. Figure 13 shows the number of features as function of the number of classes for PCA and LDA. It is worth noting that, for LDA, the number of features is always fixed to $C-1$, with C being the number of classes. Figure 13 also points out that the number of features of PCA seems to be linearly dependent on the number of classes. Therefore, the total storage memory of PCA and LDA, by assuming that each feature value is a double-precision real value of four bytes, can be simply approximated. Using 40 classes, the PCA and LDA transformation matrices and the 200 templates (five templates for each class) require about 5 MB. By hypothesising that the number of classes increases to 200, therefore, requiring 1000 templates, the predicted number of PCA features should be about 573, according to Fig. 13, under the same environmental conditions of the
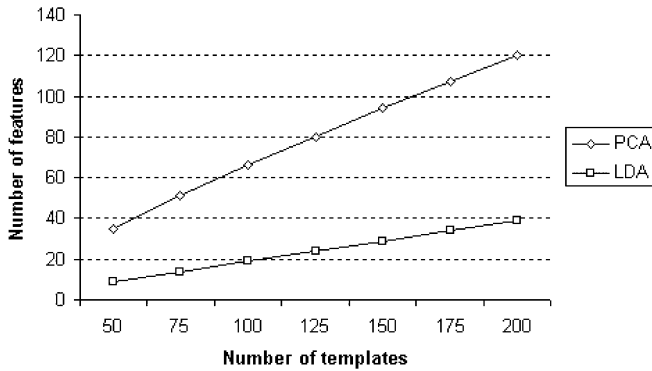
Fig. 13 Number of features required by PCA and LDA when the number of templates increases. In both cases, such a number can be approximated as being linearly dependent on the number of templates

AT&T data set. The number of LDA features is 199. Therefore, the total storage memory should be about 25 MB. This value appears to be reasonable for a modern face verification/recognition system. Accordingly, the proposed system seems to be suitable for face recognition systems of medium scale, in terms of storage memory requirements.

It is worth noting that we approximated such memory requirements using a data set exhibiting small environmental variations. Although we noticed that the storage memory requirements were about the same as using the Yale data set and the MIT data set, we think that a larger data set is needed to obtain reliable predictions when the environmental variations increase.

Conclusions

So far, individual face recognition algorithms can achieve good performances, but only for very controlled environments. Recently, the fusion of multiple algorithms has been proposed to address the issues related to the environmental variations. Some works have been reported, but, in most of the cases, the problem was addressed by using multi-modal approaches where face, fingerprints and other biometrics are combined. Very few works proposed mono-modal systems based only on the face.

In our opinion, there is one good reason to investigate the potential of the face-based mono-modal systems. This approach limits the cost of the implementation of the biometric system because it does not require multiple sensors and other kinds of dedicated hardware. Therefore, mono-modal approaches should be better investigated.

To the best of our knowledge, none of the previous face-based mono-modal systems proposed the combination of PCA and LDA. Maybe the apparent strong correlation between PCA and LDA, especially when PCA is applied before LDA, discouraged such fusion.

However, in the authors' opinion, no work proved the superiority of a method with respect to the other for all the environmental conditions. In addition, results reported in the literature suggest that PCA and LDA could have a certain degree of complementarity. Hence, the fusion of PCA and LDA could improve the performance with respect to the best individual classifier. Moreover, such a fusion could handle the differences in the behaviour of the individual algorithms with respect to the environmental condition variability. In other words, it could improve the robustness of the whole biometric system.

We believe that our main contribution to the state-of-the-art was to show, by experiments, that the fusion of PCA and LDA with simple rules allows improving the performance and the system robustness to environmental conditions variability. Performance improvement has been shown both for recognition and verification tasks. Experimental results were carried out on three data sets, which represent three environment conditions: "ideal" conditions, with a very limited range of variability of the environmental parameters (AT&T data set); "real" conditions with large variability of lighting and expression (Yale data set); "real" conditions, with variability of scale and lighting (MIT data set). The results pointed out the improvement of the performance and robustness of the system with respect to the best individual recogniser, under the above environmental conditions.

Although definitive conclusions cannot be drawn on the basis of the limited set of experiments above, we think that the reported results suggest that the fusion of LDA and PCA for face verification and recognition is worthy of further theoretical and experimental investigations.

Originality and contributions

In this work, we investigated and assessed, by experimentation, the fusion of two appearance-based face recognition algorithms, namely, principal component analysis (PCA) and linear discriminant analysis (LDA). Experiments were carried out in order to simulate different environmental conditions; "ideal" conditions, with a very limited range of variability of environmental parameters, and "real" conditions with large variability of lighting, scale and facial expression. The results pointed out the improvement in performance and robustness with respect to the best individual face recogniser. The benefits of the PCA and LDA fusion has been shown for both face verification (through the detection-error trade-off curves) and recognition (through the rank order statistics).

To the best of our knowledge, no previous work investigated in detail the usefulness of fusing PCA and LDA. Accordingly, we believe that our main contribution was to show that, by the fusion of PCA and LDA, it is possible to improve the performance of the

best individual face recogniser. Moreover, we showed that such a fusion allows an improvement in the robustness of the system, as pointed out by the results obtained with variable environmental parameters, such as lighting, expression and scale.

About the authors

Gian Luca Marcialis received his M.S. degree and Ph.D. degree in Electronic and Computer Science Engineering from the University of Cagliari, Italy, in 2000 and 2004, respectively.

He is currently a Post Doc member of the Pattern Recognition and Applications Group of the Electrical and Electronic Engineering Department at the University of Cagliari. His research interests are in the fields of fusion of multiple classifiers for person recognition by biometrics.

Gian Luca Marcialis is a member of the Institute of Electrical and Electronic Engineers (IEEE), the International Association for Pattern Recognition (IAPR) and the Italian Association for Artificial Intelligence (AI*IA).

Fabio Roli obtained his M.S. degree, with honours, and Ph.D. degree in Electronic Engineering from the University of Genoa, Italy in 1988 and 1993, respectively. From 1988 to 1994 he was a member of the research group on Image Processing and Understanding of the Department of Biophysical and Electronic Engineering at the University of Genoa, Italy. He was an adjunct professor at the University of Trento, Italy in 1993 and 1994.

Since 1995, he has lectured at the Department of Electrical and Electronic Engineering at the University of Cagliari, Italy. He is a full professor of computer engineering and is in charge of the research activities of the department in the areas of pattern recognition and computer vision. His main area of expertise is the development of pattern recognition algorithms for applications like biometrics personal identification, video surveillance and intrusion detection in computer networks. Prof. Roli's current research activity is focused on the theory and applications of multiple classifier systems.

He has organised and co-chaired the four editions of the International Workshop on Multiple Classifier Systems (http://www.diee.unica.it/mcs). He has written various papers that have been published in journals and has given lectures and tutorials on the fusion of multiple classifiers. In his research fields, Prof. Roli has published more than one hundred papers at conferences and in journals. He regularly acts as a reviewer for international journals.

References

1. Jain AK, Bolle R, Pankanti S (eds) (1999) Biometrics: Personal identification in networked society. Kluwer Academic Publishers, Boston/Dordrecht/London
2. Wechsler H, Phillips JP, Bruce V, Folgeman Soulie F, Huang TS (eds) (1997) Face recognition – From theory to applications. ASI NATO series, vol 163, Springer, Berlin Heidelberg New York
3. Yang MH, Kriegman D, Ahuja N (2002) Detecting face images: A survey. IEEE T Pattern Anal 24(1):24–58
4. Ross A, Jain AK (2003) Information fusion in biometrics. Pattern Recogn Lett 24(13):2115–2125
5. Verlinde P, Druyts P, Chollet G, Acheroy M (1999) A multi-level data fusion approach for gradually upgrading the performances of identity verification systems. In: Dasarathy B (ed) Sensor fusion: Architectures, algorithms and application III, vol 3719. SPIE Press, Orlando, Florida, USA, pp 14–25
6. Achermann B, Bunke H (1996) Combination of classifiers on the decision level for face recognition. Technical report IAM-96–002, Institut für Informatik und angewandte Mathematik, Universität Bern, January 1996
7. Zhao WY, Chellappa R, Rosenfeld A, Philips PJ (2000) Face recognition: A literature survey. UMD CfAR technical report CAR-TR-948
8. Turk M, Pentland A (1991) Eigenfaces for face recognition. J Cognitive Neurosci 3(1):71–86
9. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE T Pattern Anal 19(7):711–720
10. Wiskott L, Fellous JM, Krüger N, von der Malsburg C (1997) Face recognition by elastic bunch graph matching. IEEE T Pattern Anal 19(7):775–779
11. Kotropoulos CL, Tefas A, Pitas I (2000) Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions. Pattern Recogn 33(12):1935–1947
12. Penev PS, Atick J (1996) Local feature analysis: A general statistical theory for object representation. Network–Comp Neural 7(3):477–500
13. de Vel O, Aeberhard S (1999) Line-based face recognition under varying pose. IEEE T Pattern Anal 21(10):1081–1088
14. Roli F, Kittler J (eds) (2002) Multiple classifier systems, Springer, Berlin Heidelberg New York LNCS 2364
15. Yu K, Yang XJ, Bunke H (1995) Face recognition by facial profile analysis. In: Bichsel M (ed) Proceedings of the international workshop on automatic face and gesture recognition, Zurich, Switzerland, June 1995, pp 208–213
16. Lucas SM (1997) The continuous n-tuple classifier and its application to face recognition. Electron Lett 33:1676–1678
17. Lin S, Kung S, Lin L (1997) Face recognition/detection by a probabilistic decision-based neural network. IEEE T Neural Networ 8(1):114–132
18. Lawrence S, Giles C, Tsoi A, Back A (1997) Face recognition: a convolutional neural network approach. IEEE T Neural Networ 8(1):98–113
19. Tolba AS, Abu-Rezq AN (2000) Combined classifier for invariant face recognition. Pattern Anal Appl 3(4):289–302
20. Roli F, Fumera G, Kittler J (2002) Fixed and trained combiners for fusion of unbalanced pattern classifiers. In: Proceedings of the 5th international conference on information fusion (IF 2002), Annapolis, Maryland (USA), 7–11 July 2002, pp 278–284
21. Tan T, Yan H (1999) Face recognition by fractal transformation. In: Proceedings of the IEEE ICASSP, Turkey, May 1999, pp 3537–3540

22. Messer K, Matas J, Kittler J, Luettin J, Maitre G (1999) XM2VTSdb: The extended M2VTS database. In: Proceedings of the 2nd international conference on audio- and video-based biometric person authentication, Washington, DC, March 1999

23. Luettin J, Matre G (1998) Evaluation protocol for the extended m2vts database (xm2vtsdb). Technical report IDIAP-COM 98–05, Dalle Mole Institute for Perceptual Artificial Intelligence,http://www.idiap.ch, July 1998

24. Marcialis GL, Roli F (2003) Fusion of face recognition algorithms for video-based surveillance systems, in multisensor surveillance systems: The Fusion Perspective. In: Foresti GL, Regazzoni C, Varshney P (eds) (2003) Kluwer Academic Publishers, Boston/Dordrecht/London, pp 235–250

25. Marcialis GL, Roli F (2002) Fusion of LDA and PCA for face verification. In: Tistarelli M, Bigun J, Jain AK (eds) (2002) Proceedings of the workshop on biometric authentication, Copenhagen, Denmark, June 2002, Springer, Berlin Heidelberg New York, LNCS 2359, pp 30–37

26. Duda RO, Hart PE, Stork DG (2001) Pattern Classification, Wiley, USA

27. Zhao W, Krishnaswamy A, Chellappa R, Swets D, Weng J (1998) Discriminant analysis of principal components for face recognition. In: Wechsler H, Phillips PJ, Bruce V, Soulie FF, Huang TS (eds) Face recognition: From theory to applications. Springer, Berlin Heidelberg New York, pp 73–85

28. Chen LF, Liao HYM, Ko MT, Lin JC, Yu GJ (2000) A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recogn 33:1713–1726

29. Goudail F, Lange E, Iwamoto T, Kyuma K, Otsu O (1996) Face recognition system using local autocorrelation and multiscale integration. IEEE T Pattern Anal 18(10):1024–1028

30. Fukunaga K (1990) Introduction of statistical pattern recognition, Academic Press, New York

31. Liu K, Cheng YQ, Yang JY, Liu X (1992) A generalised optimal set of discriminant vectors. Pattern Recogn 25(7):731–729

32. Yang J, Yang J (2003) Why can LDA be performed in PCA transformed space? Pattern Recogn 36(2):563–566

33. Kittler J, Li YP, Matas J (2000) On matching scores for LDA-based face verification. In: Mirmehdi M, Thomas B (eds) Proceedings of the British Machine Vision Conference 2000, Bristol, UK, September 2000, pp 42–51

34. Martinez AM, Kak AC (2001) PCA versus LDA. IEEE T Pattern Anal 23(2):228–233

35. Mansfield T, Wayman J (2002) Best practices in testing and reporting performance of biometric devices. For the Biometrics Working Group. Technical report issue 2 draft 9, Centre for Mathematics and Scientific Computing, National Physics Laboratory, Middlesex, UK, February 2002

36. Lu X, Wang Y, Jain AK (2003) Combining classifiers for face recognition, Proc. ISME, IEEE Int. Cont. on Multimedia & Expo, vol. III, pp. 13–16, Baltimore, MD, July 6–9, 2003