Fusion of LDA and PCA for Face Recognition

Gian Luca Marcialis and Fabio Roli
Department of Electrical and Electronic Engineering - University of Cagliari
Piazza d'Armi - 09123 Cagliari (Italy)
{marcialis,roli}@diee.unica.it

Abstract. Although many approaches for face recognition have been proposed in the last years, none of them can overcome the main problem of this kind of biometrics: the huge variability of many environmental parameters (lighting, pose, scale). Hence, face recognition systems can achieve good results at the expense of robustness. In this work we describe a methodology for improving the robustness of a face recognition system based on the "fusion" of two well-known statistical representations of a face: PCA and LDA. Experimental results that confirm the benefits of fusing PCA and LDA are reported.

1 Introduction

In the last years, Face Recognition [1] has become one of the most challenging task in the pattern recognition field. The recognition of faces is very important for many applications: video-surveillance, retrieval of an identity from a data base for criminal investigations and forensic applications.

The face is considered a good biometric for many reasons: the acquisition process is non-intrusive and does not require collaboration of the subject to be recognised. The acquisition process of a face from a scene is simpler and cheaper than the acquisition of other biometrics as the iris and the fingerprint. On the other hand, many problems arise, because of the variability of many parameters: face expression, pose, scale, lighting, and other environmental parameters.

For this reason, we can subdivide the applications which involve face recognition in two fields: applications in a controlled environment and applications in a uncontrolled environment. The first kind of applications refers to the problem of "identity authentication": a subject submits to the system its face (frontal and/or profile view) and he declares his identity. The aim of the system is to verify the matching between the claimed identity and the given biometric. This kind of applications is typical for internet transactions, driver's licenses, access to limited areas.

The second kind of applications refers to the problem of "recognition of an identity in a scene", and it is typical for video-surveillance applications. A system that automatically recognises a face in a scene, first detects it and normalises it with respect to the pose, lighting and scale. Then, the system tries to associate the face to one or more faces stored in its database, and gives the set of faces that are considered as "nearest" to the detected face. This problem is much more complex than the "verification" problem, and it requires more computational resources and very robust algorithms for detection, normalisation and recognition. Usually, each of these three problems is so complex that it must be studied separately. In this work, we propose some algorithms only *for the face recognition problem.* Our methodology is based on the fusion of multiple recognisers for improving the performance of the best individual one.

The paper is structured as follows. In section 2 we briefly discuss the state of the art in face recognition systems and the role of multiple classifier systems. In section 3 we present our methodology by describing the features and the combination rules that we used. In section 4 we present experimental results and in section 5 we draw some preliminary conclusion and we point out further investigations.

2 Face recognition systems and multiple classifiers

Many face recognition systems have been proposed in the last years. Each of them is based on a particular representation of a face. To the best of our knowledge, we can identify two kinds of approaches: the "appearence-based" approaches, in which the face image is viewed as a feature vector, and the structural approaches, in which a deformable model like a graph is used for face representation.

Methods of the first kind try to reduce the dimensionality of the original face space with respect to a certain criterion. Because of the huge dimensionality of a face image, it may contain redundant or noisy information, and its processing requires a high computational cost. A feature reduction is performed by applying some standard algorithms of pattern recognition. The most known approach is the PCA representation or "eigenface" approach, proposed by Turk and Pentland [2]: the face image is projected in a space in which the correlation among the components is zero. This space transformation is called "Karhunen-Loeve transform". Another "appearence-based" approach is the LDA representation or "fisherface" approach, proposed by Kriegmann et al. [3]: the face image is projected in the Fisher space, in which the variability among the face-vectors of the same class is minimised and the variability among the face-vectors of different classes is maximised. We discuss both "eigenface" and "fisherface" approaches in section 3.

A well-known structural approach is the "elastic bunch graph method" [4] that refers to the dynamic link architectures [5] proposed by Wiskott et al. Briefly, a set of reference points, named "fiducial points", is selected in the face. Each fiducial point is a node of a full connected graph, and it is labelled with the Gabor filters responses applied to a window around the fiducial point. Each arch is labelled with the distance between the correspondent fiducial points. Recognition is performed by an elastic matching between two graphs.

Other face recognition systems have been proposed, but none of them overcomes the limits due to the large variability of many environmental parameters: pose, lighting, scale, face expression, some kind of forgery in the subject appearence (eg. the beard). In particular, the main problems of lighting, pose and scale are still in a research phase. Hence, a single face recognition system can usually achieve a good performance at the expense of robustness and reliability.

In order to improve the performance and robustness achievable by individual recognisers, the use of multiple classifier systems (MCSs) has been recently proposed. MCSs are currently a very active research field [6-8]. Multiple classifiers systems cover a wide spectrum of applications: handwritten character recognition, fingerprint classification and matching, remote-sensing images classification and so on. The effectiveness of this approach is documented by many experimental results [6-8].

Approaches for improving the performance and the robustness of a single face recognition system based on MCSs have also been proposed: Achermann and Bunke [9] present three recognisers based on frontal faces and profile. The outcome of each expert, represented by a score, i.e. a level of confidence about the decision, is combined with simple fusion rules (majority voting, rank summation, Bayes's combination rule); Lucas [10] uses a n-tuple classifier for combining the decisions of experts based on subsampled images; Tolba [11] presents a simple combination rule for fusing the decisions of a RBF network and the decisions of a LVQ network. The works by

Lucas and Tolba can be considered as the state of the art about multiple classifiers applied to face recognition.

3 The proposed methodology

In this section we present our methodology for fusing two appearence-based (or statistical) approaches to face recognition: the PCA representation ("eigenface" approach) and the LDA representation ("fisherface" approach). We already applied the fusion of LDA and PCA in the field of the face verification with good results [12]. Figure 1 shows the overview of the proposed method. It is composed of the following steps:

- representation of the face according to the PCA and the LDA approaches;
- the distance vectors d^{PCA} and d^{LDA} from all the N faces in the database are computed;
- for the final decision, these two vectors are combined according to a given combination rule. We propose two algorithms for the fusion phase: the K-Nearest Neighbours and the Nearest Mean.

In the following subsection we briefly describe the theoretical framework of the two face representations.

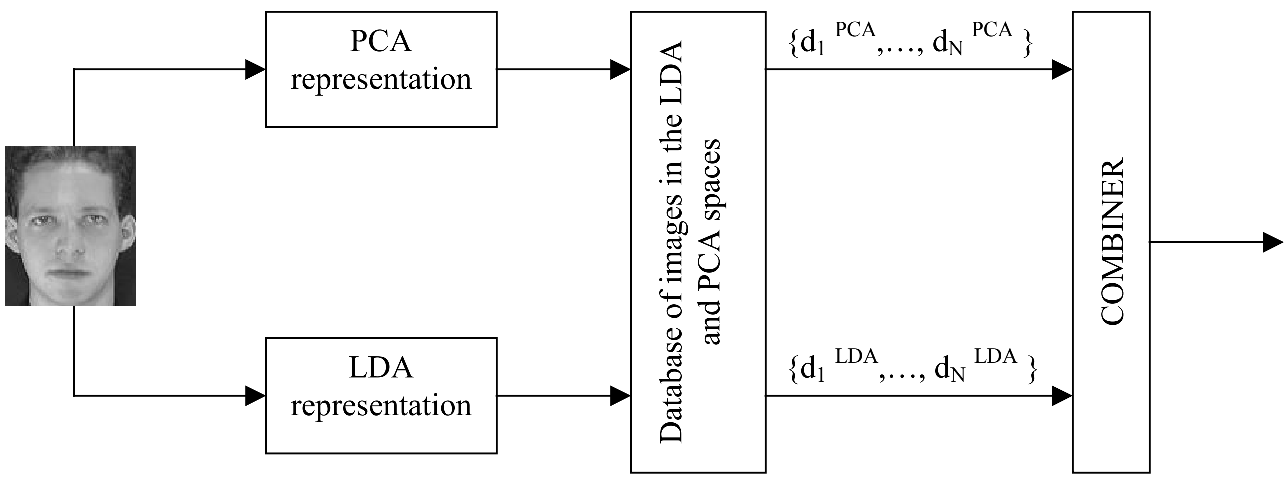

Figure 1: Overview of our fusion methodology

3.1 PCA and LDA representation for Face Recognition

Let X be a *d*-dimensional feature vector. In our case, *d* is equal to the number of pixel of each face image. The high dimensionality of the related "image space" is a well-known problem for the design of a good verification algorithm. Therefore, methods for reducing the dimensionality of such image space are required. To this end, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are widely used.

Principal Component Analysis [2, 13] is defined by the transformation:

$$y_i = W^t x_i \quad \textbf{(1)}$$

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

$$\lambda e_i = S e_i \quad \textbf{(2)}$$

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

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

This transformation is called Karuhnen-Loeve transform. It defines the d-dimensional space in which the covariance among the components is zero. In this way, it is possible to consider a small number of "principal" components exhibiting the highest variance (the most expressive features). In the face space, the eigenvectors related to the most expressive features are called "eigenfaces".

The Linear Discriminant Analysis (also called Fisher Discriminant Analysis) [3,13] is defined by the transformation:

$$y_i = W^t x_i \quad \textbf{(4)}$$

The columns of W 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:

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

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

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

The eigenvectors of LDA are called "fisherfaces". 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 nonzero eigenvectors. c-1 being the upper bound of the discriminant space dimensionality. We need d+c samples at least to have a nonsingular S_w. It is impossible to guarantee this condition in many real applications. Consequently, an intermediate transformation is applied to reduce the dimensionality of the image space. To this end, we used the PCA transform.

3.2 Fusion of LDA and PCA for Face Recognition

Many works analysed the differences between these two techniques (see in particular [3]), but no work 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 PCA are not so correlated as one can think, as the LDA transformation applied to the principal components can generate a feature space significantly different from the PCA one. Therefore, the fusion of LDA and PCA for face recognition and verification is worth of theoretical and experimental investigation.

We propose two kind of approaches to fuse PCA and LDA face representations: the K-Nearest Neighbour approach (KNN) and the Nearest Mean approach (NM).

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

The second step is to compute a *combined distance vector d* that must contain both PCA and LDA informations. To this aim, we followed two ways:

- First way, we obtained the combined distance vector by computing the mean vector:

$$d = \left\{ \frac{d_1^{PCA} + d_1^{LDA}}{2}, \ldots, \frac{d_N^{PCA} + d_N^{LDA}}{2} \right\} \quad \textbf{(7)}$$

- Second way, we obtained the combined distance vector by appending the d^{PCA} vector and d^{LDA} vector:

$$d = \left\{ d_1^{PCA}, \ldots, d_N^{PCA}, d_1^{LDA}, \ldots, d_N^{LDA} \right\} \quad \textbf{(8)}$$

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

After computing and ordering the combined distance vector *d*, we follow the KNN decision: *the most frequent identity among the first K components of d is selected*. If the combined distance vector follows eq. (7), we call our algorithm “M-KNN” or “Mean-KNN”; if it follows eq. (8), we call our algorithm “A-KNN” or “Append-KNN”.

In the case of the NM approach, we first compute a template for each identity in the database. We selected the average image for both PCA and LDA representations. Consequently, our distance vectors d^{PCA} and d^{LDA} are composed by C components instead of N. These vectors are combined according to eq. (7) or (8). The identity associated to the smallest combined distance is selected. The related algorithms are called “Mean-NM” or “M-NM” and “Append-NM” or “A-NM”.

4 Experimental results

In this section we describe our experiments on two well-known face datasets: the AT&T and the Yale datasets. Whenever possible, we compared our results with other presented in the literature. A problem in this field is that there are no standard data sets.

4.1 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 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 5 images per class (200 images), and a test set, made up of 5 images per class (200 images). In order to assess recognition performance, we repeated our experiment for ten random partitions of the data set.

Reported results refer to the average performance in such ten runs. AT&T data set is publicly available at the URL http://www.cam-orl.co.uk/facedatabase.html.

The Yale data set is made up of 11 images per 15 classes (165 total images). Each face is characterised by different facial expressions or configurations: center-light, with/without glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. The data set was subdivided into a training set, made up of 5 images per class (75 images), and a test set, made up of 6 images per class (90 images). We repeated our experiments for ten random partitions of the data set and reported the average performance. Yale data set is publicly available for research aims at the URL http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

In both data sets the face images do not need any pre-processing phase, such as re-scaling, rotation or normalisation.

4.2 Results on the AT&T data set

Table 1 reports the results on the AT&T data set.

Table 1: Percentage accuracy values on the AT&T data set.

Single classifiers		**Combined classifiers**			
PCA	**LDA**	**A-KNN**	**M-KNN**	**A-NM**	**M-NM**
94.65%	96.05%	95.90%	**97.25%**	93.30%	96.05%

The average number of principal components for the PCA representation has 119, while we used all 39 components for the LDA representation.

It is worth noting that the best combination result is comparable with those reported in [8] and [9]. In [8] a 97.5% percentage accuracy is reported but it is averaged on five runs; in [9] a 99.5% percentage accuracy is reported, but at a rejection rate of 0.5% and without a good explanation of the experimental planning. Anyway, these results can be considered the state of the art in the field of multiple classifier systems for face recognition on the AT&T data set.

Figure 2 shows the so-called rank, i.e., the percentage accuracy that can be achieved by considering the first k nearest faces of the database nearest to the given test face. The rank is a reliability measure and it is very important for video-surveillance applications in uncontrolled environments. Even in this case, the combination of PCA and LDA gives a sharp improvement of the performance, and a better identification reliability and robustness.

Another measure of the favourable conditions for fusing PCA and LDA is given by the average correlation coefficient that we computed between the d^{PCA} and the d^{LDA} vectors. A very low value was obtained: 0.39. This suggests a strong complementarity of the information extracted by the PCA representation and the LDA representation. This confirms that these two approaches are not so correlated as one could think, even in the case of frontal views as in the AT&T data set.

4.3 Results on the Yale data set

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

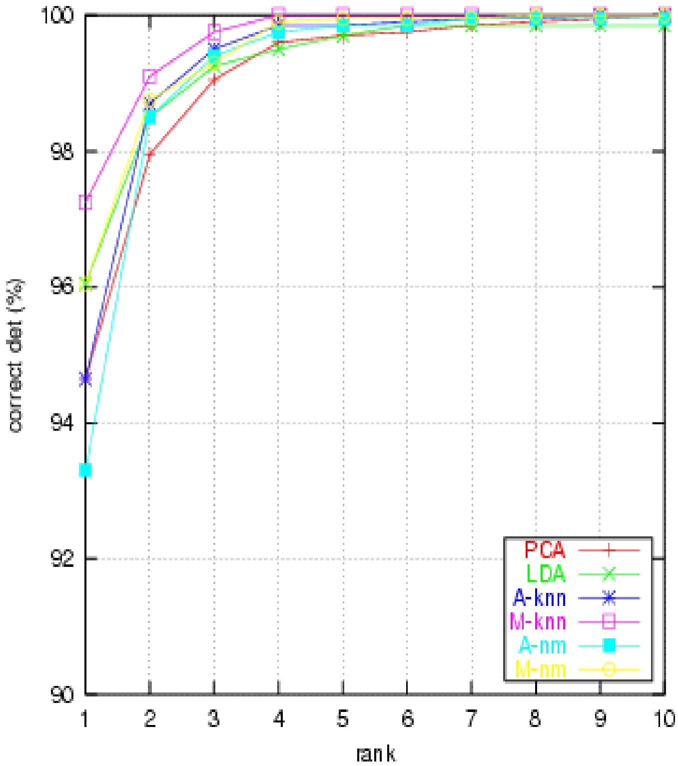

Figure 2: Rank-curves on the AT&T data set. Reported results show that the combination of PCA and LDA produces more reliable system.

Table 2 shows the percentage accuracy of our approaches on this data set.

Table 2: Percentage accuracy on the Yale data set.

Single classifiers		**Combined classifiers**			
PCA	**LDA**	**A-KNN**	**M-KNN**	**A-NM**	**M-NM**
83.00%	82.78%	**84.22%**	83.56%	83.56%	81.22%

Even in this case, the combination of PCA and LDA gives the best result. The gain is the same as for the AT&T data set (about 1.3%) but the final result is affected by the bad performance of PCA and LDA in this difficult task.

The average number of principal components is 33, while we used all 14 components for the LDA representation.

Unfortunately, in this case we cannot compare our results with others because no work reported in the literature used the Yale data set for combining multiple classifiers for face recognition. Even in this case, the rank-curves reported in fig.3 show the effectiveness of the decision combination for improving the reliability of a face recognition system.

It should be noted that the average correlation coefficient in this case is high: a value of 0.69 was obtained.

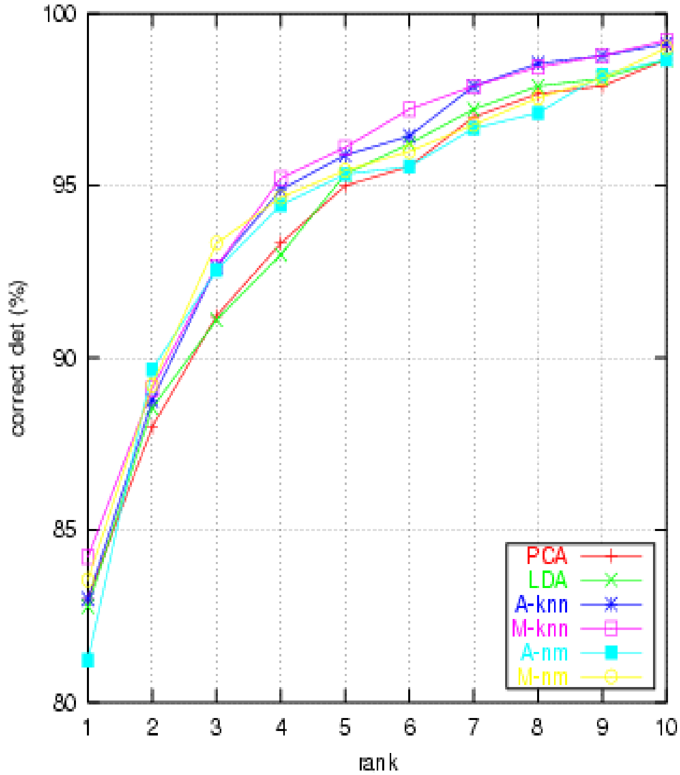

Figure 3: Rank-curves on the Yale data set. Reported results show that the combination produces more reliable performance

5 Conclusions

The fusion of two statistical approaches, namely PCA and LDA, for face representation and recognition have been investigated. Reported results confirm the benefits in fusing them with two kind of combination rules. In particular, for the AT&T data set these two representations proved to be complementary as shown by the correlation coefficient. We combined PCA and LDA with the KNN-based combination rule and the NM-based combination rule. In general, the performance of the KNN rule is much better than that of the NM rule: this should mean that the average template (that can be viewed as a low-pass filtering in the domain of the PCA and LDA spaces) reduces the available information. The rank-curves show that the reliability of the recognition always increases with respect to the best single approach.

Reported results are strongly dependent on the data set: a bit difficult task like the one presented by the Yale data set shows that the results on the single classifiers decrease dramatically. However, they can be increased using a combination rule.

On the basis of the reported results it is worth devoting further theoretical and experimental investigations to understand the behaviour of PCA and LDA in order to combine them.

References

[1] A. Jain, R. Bolle, S. Pankanti Eds, BIOMETRIC – Personal Identification in Networked Society, Kluwer Academic Publishers, Boston/Dordrecht/London (1999).

[2] M. Turk, and A. Pentland, Eigenfaces for Face Recognition, Journal of Cognitive Neuroscience, vol. 3 num.1 (1991) 71-86.

[3] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Trans. on PAMI, vol.19 num.7 (1997) 711-720.

[4] L. Wiskott, J.M. Fellous, N. Krüger, and C. von der Malsburg, Face Recognition by Elastic Bunch Graph Matching, IEEE Trans. on PAMI, vol.19 num.7 (1997) 775-779.

[5] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, and W. Konen, Distortion invariant object recognition in the dynamic link architectures, IEEE Trans. on Computers, vol.42, num.3 (1993) 300-311.

[6] J. Kittler, and F. Roli Eds, Proceedings of the First International Workshop on Multiple Classifier Systems, Springer LNCS 1857, Is Molas Golf Hotel Cagliari Italy (2000).

[7] J. Kittler and F. Roli Eds, Proceedings of the Second International Workshop on Multiple Classifier Systems, Springer LNCS 2096,Cambridge UK (2001).

[8] F. Roli and J. Kittler Eds., Proceedings of the Third International Workshop on Multiple Classifier Systems, Springer LNCS 2364, Grand Hotel Chia Laguna Cagliari Italy (2002).

[9] B. Achermann and H. Bunke, 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.

[10] SM. Lucas, Continuous n-Tuple Classifier and its Application to Real-time Face Recognition, IEE Proceedings of Visual Image and Signal Processing, vol. 145 num. 5 (1998) 343-348.

[11] AS. Tolba and AN. Abu-Rezq, Combined Classifier for Invariant Face Recognition, Pattern Analysis and Applications, vol. 3 num.4 (2000) 289-302.

[12] G.L. Marcialis and F. Roli, Fusion of LDA and PCA for Face Verification, Proceedings of the Workshop on Biometric Authentication, M. Tistarelli and J. Bigun Eds., Springer LNCS 2359, Copenhagen Denmark (2002).

[13] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, John Wiley & Sons, USA (2001).