Towards a Generalized Eigenspace-based Face Recognition Framework

Javier Ruiz del Solar and Pablo Navarrete
Department of Electrical Engineering, Universidad de Chile.
Email: {jruizd, pnavarre}@cec.uchile.cl

Abstract. Eigenspace-based approaches (differential and standard) have shown to be efficient in order to deal with the problem of face recognition. Although differential approaches have a better performance, their computational complexity represents a serious drawback. To overcome that, a *post-differential approach*, which uses differences between reduced face vectors, is here proposed. The mentioned approaches are compared using the Yale and FERET databases. Finally, a generalized framework is also proposed.

1. Introduction

Face Recognition is a highly dimensional pattern recognition problem. Even low-resolution face images generate huge dimensional spaces (20,000 dimensions in the case of a 100x200 pixels face image). In addition to the problems of large computational complexity and memory storage, this high dimensionality makes very difficult to obtain statistical models of the input space using well-defined parametric models. However, the intrinsic dimensionality of the face space is much lower than the dimensionality of the image space, since faces are all similar in appearance and posses significant statistical regularities. This fact is the starting point of the use of eigenspace methods to reduce the dimensionality of the input face space, which is the subject to be studied in this paper.

The main task of face recognition is the identification of a given face image among all the faces stored in a database. The result of identification corresponds to the subject that shows the most similar features to the ones of the requested face image. In this way, it is necessary to define a similarity function $S(\mathbf{x},\mathbf{y})$ to quantify the likeness between two face feature vectors $\mathbf{x}$ and $\mathbf{y}$. Furthermore, in most face image databases exist only a few number of images per subject (usually one) and then there is not a reliable statistical knowledge about each subject or class. Thus, the immediate problem consists in finding the nearest-neighbor of the requested face image among all the face images in the database, working in a metric space defined by the similarity function. Two approaches can be utilized to implement that: (i) feature matching, in which local features (eyes, mouth, etc.) are used to compute the feature vector, and (ii) template matching, in which the whole face image is considered as the feature vector, considering each pixel value as a vector component. This last approach works but is not efficient. The huge and redundant number of components demands the application of dimensional reduction methods, leading to the eigenspace-based approaches in which we are focused.

Standard eigenspace-based approaches project input faces onto a dimensional reduced space where the recognition is carried out. In 1987 Sirovich and Kirby used Principal Component Analysis (PCA) in order to obtain a reduced representation of face images [9]. Then, in 1991 Turk and Pentland used the PCA projections as the

feature vectors to solve the problem of face recognition, using the euclidean distance as the similarity function [10]. This system was the first eigenspace-based face recognition approach and, from then on, many eigenspace-based systems have been proposed using different projection methods and similarity functions. A differential eigenspace-based approach was proposed in 1997 by Pentland and Moghaddam [4], and it allows the application of statistical analysis in the recognition process. The main idea is to work with differences between face images, rather than with face images. In this way the recognition problem becomes a two-class problem, because the so-called "differential image" contains information of whether the two subtracted images are of the same class or different classes. In this case the number of training images per class increases so that statistical information becomes available. The system proposed in [4] used Dual-PCA projections and a Bayesian classifier. Following the same approach, a system using single PCA projections and a Support Vector Machine (SVM) classifier is here outlined.

In several comparisons that we have done between these two different approaches, we have realized that the "differential" approaches work better than the standard ones. However, in the differential case all the face images need to be stored in the database, which slow down the recognition process. This is a serious drawback in practical implementations. To overcome this drawback a so-called post-differential approach is here proposed. Under this new approach, differences between reduced face vectors are used instead of differences between face images. This allows a decreasing of the computations and storage required (only reduced face vectors are stored in the database), without losing the recognition performance of the differential approaches.

This paper is structured as follows. In section 2 the mentioned eigenspace-based recognition approaches are described. In this section is also proposed a generic framework in which the standard and the differential approaches can be included. In section 3 are presented some simulation results of recognition using the Yale and FERET databases, which allows to compare the different approaches. Finally, some conclusions of this work are given in section 4.

2. Eigenspace-based Face Recognition

2.1 Standard Eigenspace Approaches

Fig. 1 shows the block diagram of a generic, standard eigenspace-based face recognition system. Standard eigenspace-based approaches approximate the face vectors (face images) by lower dimensional feature vectors. The main supposition behind this procedure is that the face space has a lower dimension than the image space, and that the recognition of the faces can be performed in this reduced space. These approaches consider an off-line phase or training, where the *projection matrix*, the one that achieve the dimensional reduction, is obtained using all the database face images. In the off-line phase are also calculated the *mean face* and the reduced representation of each database image. These representations are the ones to be used in the recognition process. Among the projection methods employed for the reduction of dimensionality, we can mention: PCA [10], Linear Discriminant Analysis (LDA) [2], and Evolutionary Pursuit (EP) [3]. Among the similarity matching criteria employed for the recognition process, they have been used: Euclidean-, Cosine- and

Mahalanobis-distance, Self-Organizing Map (SOM) clustering, and Fuzzy Feature Contrast (FFC) similarity (see definitions in [6]). All this methods have been analyzed and compared in [6]. Under this standard eigenspace approach a *Rejection System* for unknown faces is implemented by placing a threshold over the similarity measure (see Fig. 1).

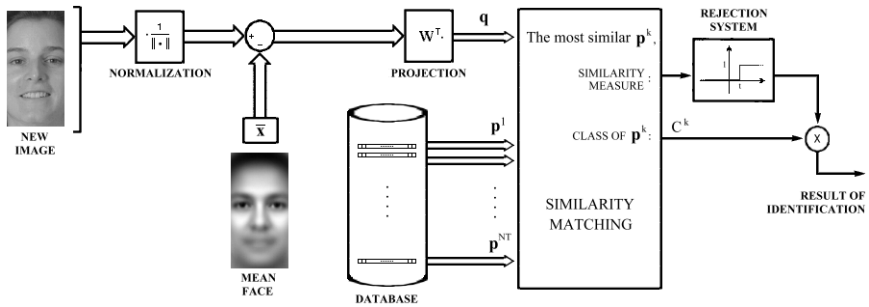

Fig. 1. Block diagram of a generic, standard eigenspace-based face recognition system.

2.2 Differential Eigenspace Approaches

Fig. 2 shows the block diagram of a generic, differential eigenspace-based face recognition system. In this approach the whole face images are stored in the database. Previously the face images are centered and scaled so that they are aligned. An input face image is normalized and subtracted from each database image. The result of each subtraction is called "differential image" $\mathbf{\Delta}$ in R^N and it is the key for identification. That because it contains information of whether the two subtracted images are of the same class or different classes. In this way the original problem of *NC* classes becomes a two-class problem. The so-called differential images are projected into a reduced space using a given projection method. Thus, each image is transformed into a reduced differential vector $\boldsymbol{\delta}$ in R^m. Thereafter the classification of the reduced differential vectors is performed. The result of each classification (S_i) is negative if the subtracted images (each $\boldsymbol{\delta}$) are of different classes and positive in other case. In order to determine the class of the input face image, the reduced vector with maximum classification value is chosen, and the class of its initial database image is given as the result of identification. The rejection system acts just when the maximum classification value is negative, i.e. it corresponds to the subtraction of different classes. Dual-PCA and Single-PCA projections have been used as projection methods. The Dual-PCA projections employ two projection matrices: $\mathbf{W}_{\mathrm{I}} \in R^{N\times m_{\mathrm{I}}}$ for intra-classes $\mathbf{\Omega}_{\mathrm{I}}$ (subtractions within equal classes), and $\mathbf{W}_{\mathrm{E}} \in R^{N\times m_{\mathrm{E}}}$ for extra-classes $\mathbf{\Omega}_{\mathrm{E}}$ (subtractions between different classes). Dual-PCA projections are employed together with a Bayesian classifier in order to perform the classification of the differential images [4]. Single-PCA projection employs a single projection matrix $\mathbf{W} \in R^{N\times m}$ (standard PCA) that reduces the dimension of the differential face images, and it is used together with a SVM classifier in order to perform the classification.

2.3 Post-Differential Eigenspace Approaches

Fig. 3 shows the block diagram of the generic, post-differential eigenspace-based face recognition system here proposed. In this approach only the reduced face images are stored in the database. Input face images are normalized and then projected into a reduced space using a given projection method (we consider only Single-PCA projections). Thereafter, the new reduced face image is subtracted from each database reduced face image. The result of each subtraction is called "post-differential image" $\boldsymbol{\delta}$ in R^m. This vector contains information of whether the two subtracted vectors are of the same class or different classes, and then it works in the same way as the "differential images" once projected on the reduced space. The classification module performs the classification of the post-differential vectors. The class of the reduced database vector that has the maximum classification value gives the class of the initial input face image. If the projection module does not significantly change the topology of the differential-image space, then the differential and post-differential approaches should have very similar recognition rates. The rejection system acts just when the maximum classification value is negative, i.e. it corresponds to the subtraction of different classes. We have implemented two different systems that follow this approach, and they are described in the following subsections.

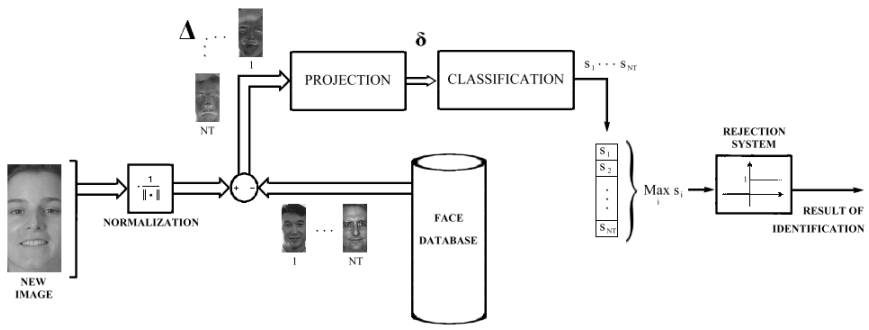

Fig. 2. Block diagram of a generic, differential eigenspace face recognition system.

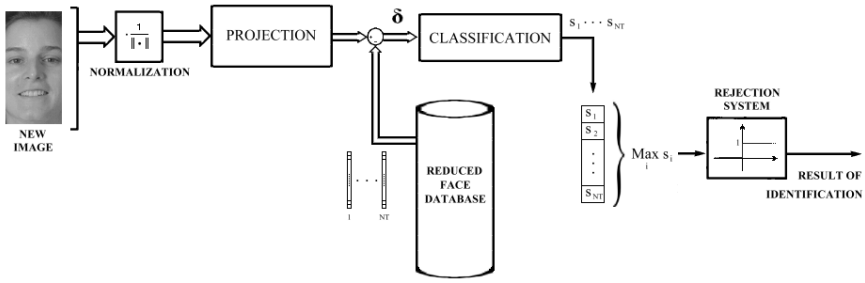

Fig. 3. Block diagram of a generic, post-differential eigenspace face recognition system.

SVM Classification

SVM in its simplest form, linear and separable case, is defined as the hyperplane that separates the vector sets belonging to different classes with the maximum distance to its closest samples, called support vectors. The problem is solved using a particular Lagrange formulation in which the problem is reduced to the computation of Lagrange multipliers. SVM in its general form, non-linear and non-separable, is very similar to its simplest form. Non-separable cases are considered by adding an upper bound to the Lagrange multipliers [7], and non-linear cases are considered by replacing all the dot products like $\mathbf{x}\cdot\mathbf{y}$, by a so-called kernel function $\mathrm{K}(\mathbf{x},\mathbf{y})$. As a classification system we are using SVM over the differential reduced training vectors $\boldsymbol{\delta}\in R^m$. Thus, the system to be solved corresponds to the following [1]:

$$\underset{\alpha_i}{\text{Max}}\ \mathrm{L_D}(\alpha_i) = -\tfrac{1}{2}\sum_{i=1}^{NT}\sum_{j=1}^{NT} y_i y_j\, \alpha_i\alpha_j\, \mathrm{K}(\boldsymbol{\delta}_i,\boldsymbol{\delta}_j) + \sum_{i=1}^{NT}\alpha_i \tag{1}$$

$$\text{subject to}: \qquad 0\le\alpha_i\le C \quad i=1\ldots\mathrm{NT}; \qquad \sum_{i=1}^{NT} y_i\,\alpha_i = 0\ .$$

Then, classification rule will be $S(\boldsymbol{\delta}) = \sum \alpha_i\, y_i\, \mathrm{K}(\boldsymbol{\delta}_i,\boldsymbol{\delta}) + b$, in which the parameter b is given by the following expression:

$$b = y_k - \sum \alpha_i\, y_i\, \mathrm{K}(\boldsymbol{\delta}_i,\boldsymbol{\delta}_k)$$

for some k so that $\alpha_k > 0$ ($\boldsymbol{\delta}_k$: support vector). (2)

Bayes Classification

If we suppose a normal distributed pattern for both $\boldsymbol{\delta}_\mathrm{I}\in R^m$ ($\boldsymbol{\delta}\in\boldsymbol{\Omega}_\mathrm{I}$, intra-class) and $\boldsymbol{\delta}_\mathrm{E}\in R^m$ ($\boldsymbol{\delta}\in\boldsymbol{\Omega}_\mathrm{E}$, extra-class), then the likelihood of a given $\boldsymbol{\delta}$ will be [5]:

$$P(\boldsymbol{\delta}/\boldsymbol{\Omega}) = \frac{\exp\left[-\tfrac{1}{2}(\boldsymbol{\delta}-\overline{\boldsymbol{\delta}})^\mathrm{T}\,\mathbf{R}^{-1}(\boldsymbol{\delta}-\overline{\boldsymbol{\delta}})\right]}{(2\pi)^{\mathrm{m}/2}\,|\mathbf{R}|^{1/2}}\ , \tag{3}$$

with $\overline{\boldsymbol{\delta}}$ the mean differential image and $\mathbf{R}$ the correlation matrix, for a given set $\boldsymbol{\Omega}$ ($\boldsymbol{\Omega}_\mathrm{I}$ or $\boldsymbol{\Omega}_\mathrm{E}$). Thus we can compute the likelihood $P(\boldsymbol{\delta}/\boldsymbol{\Omega}_\mathrm{I})$ and $P(\boldsymbol{\delta}/\boldsymbol{\Omega}_\mathrm{E})$ in order to obtain the *a posteriori* probability using the Bayes rule:

$$P(\boldsymbol{\delta}\in\boldsymbol{\Omega}) = P(\boldsymbol{\Omega}/\boldsymbol{\delta}) = \frac{P(\boldsymbol{\delta}/\boldsymbol{\Omega})\,P(\boldsymbol{\Omega})}{P(\boldsymbol{\delta}/\boldsymbol{\Omega}_\mathrm{I})\,P(\boldsymbol{\Omega}_\mathrm{I}) + P(\boldsymbol{\delta}/\boldsymbol{\Omega}_\mathrm{E})\,P(\boldsymbol{\Omega}_\mathrm{E})} \tag{4}$$

Therefore a given $\boldsymbol{\delta}$ would be classified as an intra-class vector if $P(\boldsymbol{\Omega}_\mathrm{I}/\boldsymbol{\delta}) - P(\boldsymbol{\Omega}_\mathrm{E}/\boldsymbol{\delta}) > 0$. Using expression (4) the decision rule yields:

$$S(\boldsymbol{\delta}) = P(\boldsymbol{\delta}/\boldsymbol{\Omega}_\mathrm{I})\,P(\boldsymbol{\Omega}_\mathrm{I}) - P(\boldsymbol{\delta}/\boldsymbol{\Omega}_\mathrm{E})\,P(\boldsymbol{\Omega}_\mathrm{E})\ , \tag{5}$$

and for numerical stability the logarithm of this decision rule is computed.

2.4 Generalized Eigenspace Framework

The approaches previously presented can be thought as independent eigenspace-based systems. Nevertheless, the foundations of each approach are based on similar principles. Fig. 4 shows a generalized eigenspace-based face recognition framework from which all the previous approaches presented could be derived. The standard eigenspace approach is formed when the switches in Fig. 4 are set in the positions 1 and 3; the differential eigenspace approach is formed when the switches are set in the positions 2, 4 and 5; and the post-differential eigenspace approach is formed when the switches are set in the positions 1, 4 and 6. The main idea is that all the eigenspace approaches use a projection module that could work with original or differential images and, when differential approaches are being used, the differences could be computed before or after the dimensional reduction.

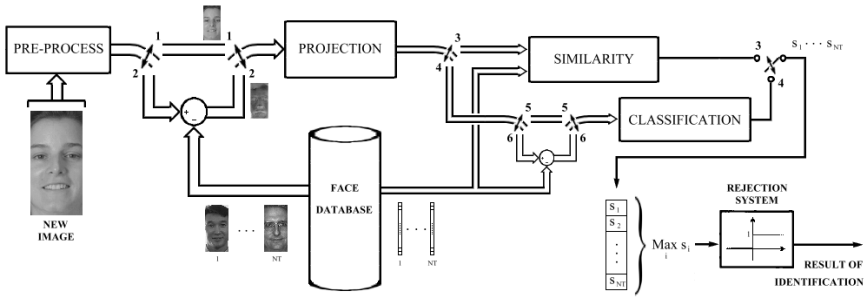

Fig. 4. Block diagram of a generalized eigenspace-based face recognition system.

3. Comparison among the approaches

3.1 Simulations using the Yale Face Image Database

In order to compare the described approaches we have first made several simulations using the Yale University - Face Image Database [11]. We used 150 images of 15 different classes. First, we preprocessed the images manually by masking them in windows of 100 x 200 pixels and centering the eyes in the same relative places. In table 1 we show the results of several simulations for standard approaches using different kind of representations and similarity matching methods. For each simulation we used a fixed number of training images, using the same type of images per class, according with the Yale database specification. In order to obtain representative results we take the average of 20 different sets of images for each fixed number of training images. All the images not used for training are used for testing. In tables 2 and 3 we show the results of several simulations using differential approaches. We used equal *a priori* probabilities for the Bayes-based methods, $P(\mathbf{\Omega}_{\mathrm{I}}) = P(\mathbf{\Omega}_{\mathrm{E}})$, and a penalty for non-separable cases $C = 0.01$ in the SVM classification method. The number of axes obtained with single-PCA was slightly smaller than the one obtained with standard PCA (shown in table 1). On the other hand, the number of axes obtained with dual-PCA was about the same for intra-class and extra-class images, and smaller than the number obtained with standard PCA. As it can be seen in these simulations, the differential and post-differential approaches

show a slightly better performance, which increases when a low number of training images per class (2) is used. That shows that both approaches have a better generalization ability than the standard one.

Tables 1, 2 and 3. Mean recognition rates using the Yale database and different numbers of training images per class, and taking the average of 20 different training sets. The small numbers are standard deviations. All results consider the top 1 match.

Table 1. Standard Eigenspace.

projection method	images per class	axes	euclidean	cos(·)	SOM	FFC	whitening euclidean	whitening cos(·)	whitening SOM	whitening FFC
PCA	6	56	87.9 (6.2)	86.0 (6.8)	84.6 (7.0)	77.1 (10.1)	64.7 (9.4)	79.3 (11.6)	64.7 (10.5)	77.1 (10.1)
FISHER		14	91.5 (6.6)	91.6 (6.5)	90.3 (6.7)	83.9 (9.3)	91.9 (5.8)	92.6 (5.6)	92.1 (6.2)	85.6 (8.3)
E.P.		15	81.2 (9.0)	85.3 (8.7)	83.7 (9.8)	77.2 (8.0)	-	-	-	-
PCA	5	34	88.7 (3.8)	87.1 (4.1)	86.0 (5.1)	78.5 (8.1)	69.5 (8.9)	83.2 (9.0)	66.1 (10.5)	78.5 (8.1)
FISHER		14	92.2 (5.7)	91.7 (6.2)	90.3 (6.4)	85.1 (9.1)	92.3 (4.7)	92.4 (5.7)	92.1 (5.3)	85.4 (8.5)
E.P.		13	84.1 (5.7)	87.7 (6.6)	86.7 (7.6)	78.7 (6.8)	-	-	-	-
PCA	4	46	87.3 (3.9)	86.7 (3.9)	84.8 (3.6)	77.6 (5.2)	72.9 (5.5)	84.4 (5.6)	66.7 (6.5)	77.6 (5.2)
FISHER		14	90.3 (4.5)	91.1 (5.0)	90.3 (4.4)	84.4 (5.9)	90.4 (4.2)	91.0 (4.4)	90.1 (4.7)	82.9 (5.7)
E.P.		18	83.6 (4.6)	86.9 (4.7)	85.0 (5.0)	74.7 (6.0)	-	-	-	-
PCA	3	35	86.6 (4.0)	85.4 (3.9)	82.0 (5.6)	77.9 (4.6)	75.0 (5.6)	84.8 (5.4)	67.4 (6.9)	77.9 (4.6)
FISHER		14	89.0 (3.6)	90.4 (4.0)	87.4 (4.0)	80.7 (6.3)	88.9 (3.1)	89.9 (3.9)	88.7 (3.9)	81.5 (3.4)
E.P.		14	81.1 (4.3)	86.9 (3.7)	82.5 (3.7)	75.9 (4.4)	-	-	-	-
PCA	2	26	82.7 (5.9)	80.8 (5.9)	76.2 (7.9)	71.1 (5.9)	75.6 (4.9)	82.1 (4.6)	60.8 (7.3)	71.1 (5.9)
FISHER		14	81.5 (5.6)	82.2 (5.8)	79.4 (5.8)	69.3 (8.6)	80.7 (4.7)	82.8 (4.9)	78.8 (5.8)	73.6 (6.2)
E.P.		14	77.8 (5.6)	81.2 (5.3)	76.0 (7.3)	70.0 (7.4)	-	-	-	

Table 2. Differential Eigenspace.

images per class	Dual PCA Bayes	SVM
6	93.5 (6.1)	94.1 (4.1)
5	93.3 (5.7)	92.5 (5.3)
4	90.9 (3.5)	91.3 (4.5)
3	90.0 (3.7)	89.6 (5.7)
2	84.7 (5.1)	86.9 (5.9)

Table 3. Post-differential Eigenspace.

images per class	Bayes	SVM
6	91.6 (6.2)	93.5 (4.8)
5	90.9 (6.5)	92.1 (4.5)
4	89.8 (5.1)	90.2 (4.1)
3	88.3 (4.8)	89.5 (6.5)
2	87.5 (5.5)	87.0 (5.8)

3.2 Simulations using FERET

In order to test the described approach using a large database, we made simulations using the FERET database [8]. We use a target set with 762 images of 254 different classes (3 images per class), and a query set of 254 images (1 image per class). Eyes' location is included in FERET database for all the images being used. Then the preprocessing consists in centering and scaling images so that eyes' position keeps in the same relative place. In table 4 we show the results of simulations for standard approaches using different kind of representations and similarity matching methods. In this table the SOM-based clustering was not included because in these tests the number of classes (254) is much larger than the number of images per class (3), and the training process is very difficult. In tables 4 and 5 we show the results of simulations using differential and post-differential approaches. The recognition rates of both approaches are better than almost all results using standard approaches, with the exception of the FLD-cosine and EP-cosine when 3 images per class were used for training. It must be noted that, when 2 images per class were used for training, the differential and post-differential approaches work better than all the standard ones. This fact shows again that differential approaches have a better generalization ability.

Tables 4, 5 and 6. Mean recognition rates for standard approaches using FERET. All results consider the top 1 match for recognition.

Table 4. Standard Eigenspace.

projection method	images per class	axes	euclidean	cos(·)	FFC	whitening euclidean	whitening cos(·)	whitening FFC
PCA	3	212	87.0	88.6	85.0	74.4	89.4	85.0
FISHER		108	91.3	94.1	92.1	85.8	92.1	92.1
E.P.		115	91.0	93.1	91.0	-	-	-
PCA	2	180	81.9	83.7	80.7	62.3	86.0	80.7
FISHER		73	79.5	88.2	85.2	79.5	88.2	85.2
E.P.		96	80.3	85.8		-	-	-

Table 5. Differential Eigenspace.

images per class	Axes Bayes	Axes SVM	Bayes	SVM
3	148 (i) / 156 (e)	186	92.6	92.7
2	106 (i) / 128 (e)	124	88.3	90.6

Table 6. Post-differential Eigenspace.

images per class	Axes Bayes	Axes SVM	Bayes	SVM
3	158	218	91.3	92.8
2	115	173	88.1	90.6

4. Conclusions

Eigenspace-based approaches have shown to be efficient in order to deal with the problem of face recognition. Although differential approaches have a better performance than the standard ones, their computational complexity represents a

serious drawback in practical applications. To overcome that, a *post-differential approach*, which uses differences between reduced face vectors was here proposed.

The three mentioned approaches were compared using the Yale and FERET databases. The simulations results obtained have shown that the two differential approaches have a better generalization ability than the standard one. This is probably because of the classification techniques used in the differential and post-differential approaches take advantage of additional statistical information. This property was decisive when a low number of training images per class (2) was used. In the simulations, the Bayes and SVM implementations do not show a significant difference in their performance, so both could be used for this kind of application. The simulation results show that the here proposed post-differential approach corresponds to a very practical solution in order to obtain a good recognition performance as well as a fast processing speed.

Eigenspace decompositions can be divided in generic (e.g. generic PCA) and specific. In a specific decomposition, the faces which need to be identified are those whose images are used when computing the projection matrix. This is not the case in a generic decomposition. In the here presented simulations only specific decompositions where used. As a future work we want to perform simulations using generic decompositions to go deeper in the comparison of the described approaches. We believe that in this case differential approaches will show an even better generalization ability.

Acknowledgements

Portions of the research in this paper use the *FERET* database of facial images collected under the *FERET* program.

This research was supported by the DID (U. de Chile) under Project ENL-2001/11, and by the join "Program of Scientific Cooperation" of CONICYT (Chile) and BMBF (Germany) under the project "Face Recognition and Signature Verification using Soft-Computing".

References

1. Burges C.J.C., “A tutorial on support vector machines for pattern recognition”, *Data Mining and Knowledge Discovery*, 2(2), pp. 121–167, 1998.
2. Fisher R.A., “The Use of Multiple Measures in Taxonomic Problems”, *Ann. Eugenics*, vol. 7, pp. 179-188, 1936.
3. Liu C., and Wechsler H., “Evolutionary Pursuit and Its Application to Face Recognition”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570-582, June 2000.
4. Pentland A., and Moghaddam B., “Probabilistic Visual Learning for Object Representation”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696-710, July 1997.
5. Duda R.O., Hart P.E., and Stork D.G., “Pattern Classification”, Second Edition, 2001.
6. Navarrete P., and Ruiz-del-Solar J., “Comparative study between different Eigenspace-based approaches for Face Recognition”, *Lecture Notes in Artificial Intelligence 2275*, AFSS 2002, Springer, 178 - 184.
7. Cortes C., and Vapnik V., “Support Vector Networks”, *Machine Learning*, 20, pp. 273-297, 1995.
8. Phillips P. J., Wechsler H., Huang J., and Rauss P., “The FERET database and evaluation procedure for face recognition algorithms”, *Image and Vision Computing J.*, Vol. 16, no. 5, 295 -306, 1998.
9. Sirovich L., and Kirby M., “A low-dimensional procedure for the characterization of human faces”, *J. Opt. Soc. Amer. A*, vol. 4, no. 3, pp. 519-524, 1987.
10. Turk M., and Pentland A., “Eigenfaces for Recognition”, *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
11. Yale University Face Image Database, publicly available for non-commercial use, http://cvc.yale.edu/projects/yalefaces/yalefaces.html .