ANALYSIS AND COMPARISON OF EIGENSPACE-BASED FACE RECOGNITION APPROACHES

PABLO NAVARRETE and JAVIER RUIZ-DEL-SOLAR

Department of Electrical Engineering, Universidad de Chile, CHILE

Email: pnavarre, jruizd@cec.uchile.cl

Different eigenspace-based approaches have been proposed for the recognition of faces. They differ mostly in the kind of projection method been used and in the similarity matching criterion employed. The aim of this paper is to present a comparative study between some of these different approaches. This study considers theoretical aspects as well as experiments performed using a face database with a few number of classes (Yale) and a face database with a large number of classes (FERET).

Keywords: PCA, Fisher Linear Discriminant, Evolutionary Pursuit, Eigenfaces, Fisherfaces, FERET.

1. Introduction

Among the most successful approaches used in face recognition we can mention *eigenspace-based* methods, which are mostly derived from the *Eigenface*-algorithm[9]. These methods project the input faces onto a reduced dimensional space where the recognition is carried out, performing a holistic analysis of the faces. Different eigenspace-based approaches have been proposed. They differ mostly in the kind of projection/decomposition method been used and in the similarity matching criterion

employed, which show different statistical properties. The aim of this paper is to present an independent, comparative study among some of these different approaches, focusing on the confirmation of the theoretical properties. We believe that to carry out an independent study is important, because comparisons are normally made using the own implementations of the research groups that have proposed each method (e.g. in FERET contests), which does not consider completely equal working conditions (e.g. exactly the same preprocessing steps). Very often, more than a comparison between the capabilities of the methods, a contest between the abilities of the research groups is performed. Additionally, not all the possible implementations are considered (e.g. p projection methods with q similarity criteria), but only the ones that some groups have decided to use.

This comparative study considers the use of three different projection methods (Principal Component Analysis, Fisher Linear Discriminant and Evolutionary Pursuit) and five different similarity-matching criteria (Euclidean-, Cosines- and Mahalanobis-distance, Self-Organizing Map and Fuzzy Feature Contrast). The pre-processing aspects of these approaches (normalization, illumination invariance, geometrical invariance, etc.) are not going to be addressed in this study. The mentioned methods are described in section 2, and the comparative study is presented in section 3. In section 4 some conclusions are given.

2. Eigenspace-based Approaches

Eigenspace-based approaches approximate the face vectors (face images) with lower dimensional feature vectors. The main supposition behind this procedure is that the face space (given by the feature vectors) has a lower dimension than the image space (given by the number of pixels in the image), 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 face database is created and the *projection matrix*, the one that achieve the dimensional reduction, is obtained from all the database face images. In the off-line phase, also calculated are the *mean face* and the reduced representation of each database image. These representations are the ones to be used in the recognition process.

2.1. General Approach

Figure 1 shows the block diagram of a generic eigenspace-based face recognition system. A preprocessing module transforms the face image into a unitary vector and then performs a subtraction of the mean face ($\bar{\mathbf{x}}$). After that, the resulting vector $\mathbf{x}$ is projected using the projection matrix $\mathbf{W} \in R^{N \times m}$ that depends on the eigenspace method been used (see section 2.2). This projection corresponds to a dimensional reduction of the input, starting with vectors $\mathbf{x}$ in R^N (when N is the dimension of the image vectors) and obtaining projected vectors $\mathbf{q}$ in R^m with $m<N$ (usually $m<<N$). The *Similarity Matching* module compares the similarity of the reduced representation of the query face vector $\mathbf{q}$ with the reduced vectors $\mathbf{p}^k$, $\mathbf{p}^k \in R^m$ that represent the faces in the database. By using a given criterion of similarity (see section 2.3), this module determines the most similar vector $\mathbf{p}^k$ in the database. The class of this vector is the result of the recognition process, i.e. the identity of the face. In addition, a *Rejection System* for unknown faces is used if the similarity matching measure is not good enough (see description in [2]).

2.2. Projection/Decomposition Methods

Principal Components Analysis - PCA

PCA is a general method to identify the linear directions in which a set of vectors are best represented and after that to make a dimensional reduction of them. In 1987 Sirovich and Kirby[7] used Principal Component Analysis (PCA) in order to obtain a reduced representation of face images. Then, in 1991 Turk and Pentland[9] used the PCA projections as the feature vectors to solve the problem of face recognition, using the Euclidean distance as the similarity function. In order to obtain the eigenfaces, we first need to obtain the projection axes having the largest variance of the projected face images. Then, we repeat this procedure in the orthogonal space that is still uncovered, until we realize

that there is no more variance to take into account. The theoretical solution of this problem is well

known and is obtained by solving the eigensystem of the correlation matrix $\mathbf{R} \in R^{N \times N}$:

$$\mathbf{R} = \mathrm{E}\left\{(\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^{\mathrm{t}}\right\} \tag{1}$$

where $\mathbf{x}$ represents the normalized image vector, $\overline{\mathbf{x}}$ is the mean face image, and N is the original

vector image dimension. The eigenvectors of this system represent the projection axes or eigenfaces,

and the eigenvalues represent the projection variance of the corresponding eigenface. Then by sorting

the eigenfaces in descending order of eigenvalues we have the successive projection axes that solve

our problem.

The main problem is that $\mathbf{R} \in R^{N \times N}$ is too big for a reasonable practical implementation and then

we need to estimate the correlation matrix using a database with NT face images (the training set), by

taking the corresponding averages in the training set. Let $\mathbf{X} = \left[(\mathbf{x}^1 - \overline{\mathbf{x}})(\mathbf{x}^2 - \overline{\mathbf{x}}) \cdots (\mathbf{x}^{NT} - \overline{\mathbf{x}})\right]$ be the

matrix of the normalized training vectors. Then, the $\mathbf{R}$ estimator will be given by $\mathbf{R} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$. We could

say that the number of eigenfaces must be less than, or equal to, NT, because with NT training images

all the variance must be projected into the hyperplane subtended by the training images. In other

words the rank of $\mathbf{R}$ is less than, or equal to, NT. The correlation matrix could have several null or

negligible eigenvalues depending on the linear dependence of the vectors in the training set. In

addition, the eigensystem of $\mathbf{X}^{\mathrm{T}}\mathbf{X} \in R^{NT \times NT}$ has the same non-zero eigenvalues of $\mathbf{R}$, because

$\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{X}\,\mathbf{v}^k = \lambda_k\,\mathbf{X}\,\mathbf{v}^k$ represents both systems at the same time.

Now we can solve the reduced eigensystem of $\mathbf{X}^{\mathrm{T}}\mathbf{X} \in R^{NT \times NT}$. The corresponding eigenvalues are

the same eigenvalues of the original system, and the eigenfaces are represented by $\mathbf{w}^k = \mathbf{X}\,\mathbf{v}^k$, and to

normalize they must be divided by $\sqrt{\lambda_k}$.

To improve the dimensional reduction, it is recommendable to neglect the components with small

projection variances. If we ignore some components, the mean square error of the given representation

is the sum of the eigenvalues not used in the representation. Therefore, a good criterion would be to

choose only m components, such that the *normalized Residual Mean Square Error*[8]:

$$\text{RMSE}(m) = \sum_{k=m+1}^{N} \lambda_k \Bigg/ \sum_{k=1}^{N} \lambda_k \tag{2}$$

is less than, say, 5%.

Fisher Linear Discriminant - FLD

FLD searches for the projection axes on which the face images of different classes are far from each other (similar to PCA), and at the same time images of a same class are close to each other. In order to define the mathematical structure under FLD, first we define the parameter $\gamma(\mathbf{u})$ to be maximized on the successive projection axes as:

$$\gamma(\mathbf{u}) = \frac{s_b(\mathbf{u})}{s_w(\mathbf{u})} \tag{3}$$

where $\mathbf{u}$ represents any unitary projection vector in the image space, and $s_b(\mathbf{u})$ and $s_w(\mathbf{u})$ are given by:

$$s_b(\mathbf{u}) = \sum_{i=1}^{NC} P(C_i)\left\{\left(\mathbf{m}^{(i)} - \mathbf{m}\right)\mathbf{u}\right\}^2, \tag{4}$$

$$s_w(\mathbf{u}) = \sum_{i=1}^{NC} P(C_i) E\left[\left\{\left(\mathbf{x}^{(i)} - \mathbf{m}^{(i)}\right)\mathbf{u}\right\}^2\right] \tag{5}$$

where $\mathbf{m}$ is the global mean vector, $P(C_i)$ is the probability associated to class C_i, $\mathbf{m}^{(i)}$ is the average vector of C_i and $\mathbf{x}^{(i)}$ is a vector associated to class C_i. In this way, $s_b(\mathbf{u})$ measures the separation between the individual class means with respect to the global mean face, and $s_w(\mathbf{u})$ measures the separation between vectors of each class with respect to their own class mean. Alternatively, when scatter matrices are defined by:

$$\mathbf{S}_b = \sum_{i=1}^{NC} P(C_i)\left(\mathbf{m}^{(i)} - \mathbf{m}\right)\left(\mathbf{m}^{(i)} - \mathbf{m}\right)^{\mathrm{T}} \tag{6}$$

$$\mathbf{S}_w = \sum_{i=1}^{NC} P(C_i) E\left[\left(\mathbf{x}^{(i)} - \mathbf{m}^{(i)}\right)\left(\mathbf{x}^{(i)} - \mathbf{m}^{(i)}\right)^{\mathrm{T}}\right], \tag{7}$$

5

then:

$$\gamma(\mathbf{u}) = \frac{\mathbf{u}^{\mathrm{T}} \mathbf{S}_b \mathbf{u}}{\mathbf{u}^{\mathrm{T}} \mathbf{S}_w \mathbf{u}} \quad . \tag{8}$$

At this point it is not difficult to demonstrate that the solution of our maximization problem is the solution of the generalized eigensystem[1]:

$$\mathbf{S}_b \mathbf{w}^k = \lambda_k \mathbf{S}_w \mathbf{w}^k . \tag{9}$$

Here $\mathbf{w}^k$'s are the fisherfaces and λ_k's are the successive γ parameters associated with each fisherface. The system given by (9) is conventionally solved by writing $\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{w}^k = \lambda_k \mathbf{w}^k$. Notice that this eigensystem does not have orthogonal eigenvectors because $\mathbf{S}_w^{-1} \mathbf{S}_b$ is not symmetric in general, so the fisherfaces would not be an orthogonal projection set. Other implementation problems are: the matrices $\mathbf{S}_b$ and $\mathbf{S}_w$ are too big, and also $\mathbf{S}_w$ could be singular and then non-invertible. An easy way to solve these two problems is to use a PCA decomposition previous to the FLD procedure. Then, the size of the scatter matrices would be small enough and depending on the criterion for dimensional reduction used, $\mathbf{S}_w$ will became non-singular. In this case, the eigensystem (9) will give reduced eigenvectors $\mathbf{v}^k$, that need to be transformed into the true eigenvectors $\mathbf{w}^k$ using $\mathbf{W}_{FF} = \mathbf{V} \mathbf{W}_{EF}$, where $\mathbf{W}_{EF}$ and $\mathbf{W}_{FF}$ are the PCA and Fisher projection matrices, respectively.

With two classes we need only one γ to describe the between/within class scatter, then the projection space might have just one dimension. In the general case with *NC* classes, we need *NC-1* values of γ, and this corresponds to the dimension of the projection space. Given that (6) is the sum of *NC* matrices of rank one or less, *NC-1* of them are independent, and this gives *NC-1* possible values of non-zero γ's in (8) [1]. Thereafter we could adjust the criterion for the Fisher-space dimensional reduction. In analogy with PCA a good criterion would be to choose only *m* components ($m \leq NC\text{-}1$) depending on the normalized *Residual Fisher Parameter*:

$$\text{RFP}(m) = \sum_{k=m+1}^{NC-1} \lambda_k \bigg/ \sum_{k=1}^{NC-1} \lambda_k \ . \tag{10}$$

Considering m given by a very small RFP, say RFP(m)<10^{-5}%, will mean that the *NC-1* components are chosen in practice. The advantage of FLD against PCA is that the information kept in the dimensional reduction is better for recognition purposes. However, there are some drawbacks. For example, FLD uses the particular class information, so it is recommended to have many images per class in the training process, or at least a good characterization of each one. In other words, in PCA the convergence of the **R** estimator depends mostly on the total number of target images *NT*, but in FLD the convergence of the scatter matrices estimators depends also on the numbers of target images per class. The main drawback of FLD is that it could be over-adjusted on the target images, and then the recognition system may have lack of good generalization that may be reflected in the resulting system's recognition rate.

Evolutionary Pursuit - EP

The eigenspace-based adaptive approach proposed in [4], searches for the best set of projection axes in order to maximize a fitness function, measuring at the same time the classification accuracy and generalization ability of the system. Because the dimension of the solution-space of this problem is too large, it is solved using a specific kind of Genetic Algorithm called *Evolutionary Pursuit* (EP). In the proposed representation, we are going to call the projection axes (image vectors) as EP-faces.

In order to obtain the EP-faces first an initial dimensional reduction is done using PCA, and then the Whitening Transformation (equivalent to a Mahalanobis metric system, see 2.3) is used. In the Whitened-PCA space several rotations between pair of axes are applied and a subset of the rotated Whithened-PCA axes is selected. This transformation is coded using a chromosome representation. The chromosome structure uses 10 bits for each angle α_k (between 0 and $\pi/2$), and a number of bits a_i equal to the number of Whithened-PCA components (m) in order to select a subset of axes. Notice that the number of possible rotations between axes would be $m\,(m-1)/2$, so the number of bits for

each chromosome is $5 m (m-1) + m$, and the size of the genome-space (too large to search it exhaustively) is $2^{5 m (m-1)+m}$.

Each chromosome represents certain projection system. Then in order to evaluate this system the following fitness function is used:

$$\zeta(\alpha_k, a_i) = \zeta_a(\alpha_k, a_i) + \lambda \, \zeta_s(\alpha_k, a_i) \tag{11}$$

$\zeta_a(\alpha_k, a_i)$ measures the accuracy, $\zeta_s(\alpha_k, a_i)$ measures the generalization ability, and λ is a positive constant that determines the importance of the second term against the first one. The generalization ability is computed as (remember $\mathbf{S}_b$ given by (6)):

$$\zeta_s(\alpha_k, a_i) = \sqrt{\sum_{i=1}^{NC} \left(\mathbf{m}^{(i)} - \mathbf{m} \right)^{\mathrm{T}} \left(\mathbf{m}^{(i)} - \mathbf{m} \right)}. \tag{12}$$

Here $\mathbf{m}$ is the global mean and $\mathbf{m}^{(i)}$ is the mean of the class C_i. Although the accuracy measure $\zeta_a(\alpha_k, a_i)$ proposed in [4] is the recognition rate of training face images as the top choice, in our implementation we used the top 2 identity because with the top choice it was too easy to obtain a 100% recognition rate.

As usual, in order to find the maximal value of the fitness function, a random set of chromosomes is randomly set and the GA procedure iterates until the system converges. The operators used to create a new set of chromosomes per iteration are: proportionate selection (pre-selection of parents in proportion to their relative fitness); two-point crossover (exchange the selection between the crossover points); and fixed probability mutation (each bit of the chromosome is given a fixed probability of flipping). Details of this procedure can be found in [4].

2.3. Similarity Matching Methods

The main objective of similarity measures is to define a value that allows the comparison of feature vectors (reduced vectors in eigenspace frameworks). With this measure the identification of a

new feature vector will be possible by searching the most similar vector into the database. This is the well-known nearest-neighbor method. One way to define similarity is to use a measure of distance, $d(\mathbf{x},\mathbf{y})$, in which the similarity between vectors, $S(\mathbf{x},\mathbf{y})$ is inverse to the distance measure. In the next sub-sections two distance measures are shown (Euclidean and Mahalanobis) and three similarity measures (Cosine, SOM and FFC).

Euclidean Distance

$$d(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^{\mathrm{T}}(\mathbf{x}-\mathbf{y})} \ . \tag{13}$$

Cosine Distance

$$S(\mathbf{x},\mathbf{y}) = \cos(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \ . \tag{14}$$

Mahalanobis Distance

$$d(\mathbf{x},\mathbf{y}) = (\mathbf{x}-\mathbf{y})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{x}-\mathbf{y}) ; \qquad \mathbf{R}: \text{correlation matrix.} \tag{15}$$

From a geometrical point of view this distance has a scaling effect in the image space. Taking into consideration the face image subset, directions in which a greater variance exist are compressed and directions in which a smaller variance exist are expanded. It can be proved that in the PCA space the Mahalanobis distance is equivalent to the Euclidean distance, weighting each component by the inverse corresponding eigenvalue (see demonstration in [5]), and it is often called Whitening (PCA) Transformation.

SOM Clustering

Self-Organizing Maps (SOM) [3] are used as an associative network to match the projected query face with the corresponding projected database face (see Figure 2). The use of a SOM to implement

this module could improve the generalization ability of the system. The SOM approach uses reference vectors $\mathbf{m}_i$, the so-called SOM codewords, to approximate the probability distribution of the faces in a 2D map. In the training phase of the SOM a clustering of the reduced vector faces is carried out. Thereafter the SOM is transformed in an associative network by labeling all of its nodes. To find the reference vectors, first the best matching node is found.

After that a *labeling* algorithm allows to associate a fixed face identity for each node in the map. The label of the node associated with a new vector will be the result of identification. Therefore, in this case the similarity measure is given by the Euclidean distance between the new vector and the best matching node.

Fuzzy Feature Contrast – FFC

$$S(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{m} min\{\mu_i(\mathbf{x}),\mu_i(\mathbf{y})\} - \alpha \sum_{i=1}^{m} max\{\mu_i(\mathbf{x}) - \mu_i(\mathbf{y}),0\} - \beta \sum_{i=1}^{m} max\{\mu_i(\mathbf{y}) - \mu_i(\mathbf{x}),0\} \qquad (16)$$

where $\mu_i(\mathbf{x})$ is a membership function associated with the i-component of vector $\mathbf{x} \in R^m$.

This similarity measure, originally proposed in [6], is a fuzzy implementation of the Feature Contrast model from Tversky. The first sum measures the common features (intersection) and the others represent the distinctive features (difference in the two possible ways). The positive parameters α and β adjust the contrast of the three kind of features. By choosing $\alpha \neq \beta$, it is possible to introduce asymmetries in the subject-referent comparison. This model assumes that all features are independent. This can be assumed in PCA and WPCA, but not in FLD and EP. In our implementation we normalize each feature of PCA (in WPCA it is not necessary) and we chose $\mu_{i(\mathbf{x})}$ linear between –1 and 1, with x_i normalized.

3. Comparison among the approaches

This section is focused in the comparison of *eigenspace*-based face recognition systems using the projection methods and similarity measures previously presented. Two different face image databases

were used in order to test the recognition ability of each system: the Yale University Image Database and FERET. For further analysis, the most important properties of each database are: the total number of images, the number of images per class and the statistical representativeness of the given sets of images. In this sense the Yale Database is a small database (150 images) in comparison with FERET (762 selected images), but the Yale Database has more images per class (10) than FERET (4 frontal images) and it contains more representative images for each class (center-light, w/glasses, happy, left-light, w/no glasses, right-light, sad, sleepy, surprised, and wink). As it was mentioned in section 2, the statistics estimated by PCA should be more reliable than the ones of FLD, specially when few images per class are used. Nevertheless the performance of FLD can still overcome PCA using few images per class, subject to the statistical representativeness of the given sets of images. The main purpose of EP is to take advantage of the generalization ability of PCA and the discriminant ability of FLD. The similarity measures SOM and FFC are expected to improve the generalization ability in the recognition process. The generalization ability of a given system should be manifested in the experiments as the ability to maintain a recognition rate when reducing the number of images used in the training process.

3.1. Experiments using the Yale Face Image Database

In order to compare the described methods we have first made several experiments 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 experiments using different kind of representations and similarity matching methods. For each experiment we used a fixed number of training images, using the same type of images per class, according to 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 images not used for training are used for testing.

We can see that the best results are obtained with the FLD representation, and the difference against other representations decreases when the number of training images per class decreases, showing that the FLD discrimination ability strongly depends on the number of training images per class. The best results are almost always obtained with FLD-cosine. The systems that seem to be as efficient as FLD-cosine are the SOM based.

The best results using FFC were obtained employing an asymmetric subject-referent comparison: $\alpha=0.5$ $\beta=5$. This means that in the question "how is the subject face similar to the referent face?" the answer focuses more on the features of the referent (the unknown face). The generalization ability of the systems is not well measured in these experiments because the number of selected axes is about the same as the number of classes (15). That affects the EP representation method as well as the FFC and SOM similarity matching methods.

3.2. Experiments using FERET

In order to test the described methods using a large database, we made several experiments using FERET database[10]. 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 of centering and scaling images so that eyes position is kept in the same relative place. In Table 2 we show the results of experiments using different kind of representations and similarity matching methods. In this table the SOM similarity measure was not included because in these tests the number of classes (254) is much larger than the number of images per class (2 or 3), and the training process is very difficult. It is very unlikely that new images will be correctly classified in a SOM net in which very few nodes are labeled with the correct identity.

We can see that the best experimental results are obtained with the FLD representation, which is similar to the case when a small database is used. One of the main differences with the case of small databases is that the results obtained using FFC similarity measure considerably increases its

recognition rate. The best results for FFC similarity measure were obtained employing an asymmetric subject-referent comparison: α=1, β=2 (in this way the answer still focus more on the features of the referent). This confirms our previous hypothesis that large number of classes will force generalization ability on this system. Also we could see that FLD-cosine works better than other systems, showing us that the FLD dimensional reduction works better than other projection methods. We can also see that the EP-systems always performed worse than the FLD systems. At the same time, we find that when the number of axes selected is of the same order of the number of classes, then the accuracy pursuit (mostly dependent on the top 2 match) seems to fail and then FLD exhibits some advantages. The worst results seem to be obtained with Whitening-PCA-Euclidean and Whitening-PCA-SOM. We also see that the changes in the norms of the vectors seem to confuse the recognition ability.

3.3. Computational Complexity

Another important issue in comparing the different projection methods is the computational cost of the training processes. In PCA the computational cost is mainly due to the process of determining $\mathbf{R}$, which is $O(NT^2 \cdot N)$, and for solving the eigensystem, which requires $O(NT^3)$ computation. If we suppose that the number of training images NT is much smaller than the number of pixels per image N, then the computational cost of PCA is just the cost of determining $\mathbf{R}$, i.e., $O(NT^2 \cdot N)$. In our implementation of FLD we require the computation of PCA to reduce the dimension to m_1 ($m_1 < NT$), so an additional cost of $O(m_1^2 \cdot NC)$ is involved due to the process of determining the scatter matrices, and the cost for solving the general eigensystem is $O(m_1^3)$. Nevertheless the additional cost in FLD is usually much smaller than the cost of PCA, $O(NT^2 \cdot N)$. Finally, EP requires much more computations because this process must iterate until a given convergence criterion is accomplished. The computational complexity of EP strongly depends on the implementation but it usually takes more than 10 times the time of the PCA procedure.

The computational cost of on-line operation is mainly required by the comparison with database vectors, which is $O(NT \cdot m)$, except when the SOM-based similarity measure is used. In this case the cost is $O((\text{number of nodes}) \cdot m)$. The numerical stability of different methods depends mostly of the numerical algorithms used for solving the eigensystems. Either in PCA or FLD this is not a critical problem because it always involves the management of symmetric matrices.

In the experiments performed in this work the computation time of FLD was very similar to that of PCA when using Yale Database and it took approximately 1.4 the time of PCA when using FERET Database. The computation time of EP was about 40 times more than that of PCA when using Yale or FERET Database. The computation time of on-line operation was very similar for all the similarity measures presented here.

4. Conclusions

Eigenspace-based approaches have shown to be efficient in order to deal with the problem of face recognition. Although these methods have proved their goodness in several works, we believe that there is a necessity of a general evaluation to understand their differences and similarities. In this sense our goal was to provide a unified vision over the most thriving eigenspace approaches.

We made an extensive analysis of different eigenspace-based approaches, separating the representation problem from the similarity matching method employed, and using databases with small and large number of classes. We also used the Whitening Transformation as a Mahalanobis metric system before the initial PCA processing, in order to match PCA and FLD against EP. In terms of recognition rates we saw that the FLD projection method and the cosine similarity matching work better, and that the whitening-cosine based methods seem to keep their recognition ability with fewer target images. Even if the EP projection method and the FFC similarity measure did not show the best recognition abilities, they showed a good performance and a good generalization ability in the tests using FERET database, in which the other systems could not overcome the adjustment of few projection axes (in relation to the number of classes). It must be noted that in these experiments a

problem persists when using FFC because of the non-complete supposition of independence between features. As future work we are going to test other projection methods, like Independent Component Analysis (ICA) and Kernel-PCA (KPCA), to solve this problem.

An important aspect of this work was to change the standard similarity matching methods for a SOM. In this way the structure of the identification system (see Figure 1) changes because we do not need the database reduced representation anymore (see Figure 2). Now this information would be appropriately included in the SOM reference vectors (codewords). In our experiments we realize that this kind of identification system works as good as the standard ones, in terms of recognition rates with a small database. An interesting feature of this approach is the possibility to adapt itself to changes in the faces. This has a direct application in adaptive security access systems where the persons to be recognized would be continuously observed by the system. Specifically, when a person arrives the neural system will carry out the recognition, and after that the SOM will perform one training iteration and the labeling procedure of the nodes will be re-applied. In this way the SOM map will adapt itself to changes in the features of faces like beard, hair, or even age evolution. In this sense this system represents a robust and adaptive identification system.

As future work we are focusing on the extension of this study, considering other kind of eigenspace-based approaches. We are interested in exploring the so-called *differential approaches*, which allow the application of statistical analysis in the recognition process, by working with differences between face images rather than with face images.

Acknowledgements

References

1. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, 2001.

2. M. Golfarelli, D. Maio and D. Maltoni, "On the Error-Reject Trade-Off in Biometric Verification Systems," *IEEE Trans. Patt. Anal. Mach. Intell.* **19**, (1997) 786-796.

3. T. Kohonen, *Self-Organized Maps*, Springer Verlag, Berlin, 1997.

4. C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Trans. Patt. Anal. Mach. Intell.* **22**, (2000) 570-582.

5. P. Navarrete and J. Ruiz del Solar, "Eigenspace-based Recognition of Faces: Comparisons and a new Approach," *Proc. of the Int. Conf. on Image Analysis and Processing ICIAP 2001*, 42-47.

6. S. Santini and R. Jain, "Similarity Measures," *IEEE Trans. Patt. Anal. Mach. Intell.* **21**, (1999) 871-883.

7. L. Sirovich and M. Kirby, "A low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Amer. A* **4**, (1987) 519-524.

8. D. L. Swets and J. J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Patt. Anal. Mach. Intell.* **18**, (1996) 831-836.

9. M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience* **3**, (1991) 71-86.

10. P. J. Phillips, H. Wechsler, J. Huang and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *J. Image and Vision Computing* **16**, (1998) 295-306.

11. Yale University Face Image Database, publically available for non-commercial use, http://cvc.yale.edu/projects/yalefaces/yalefaces.html

Javier Ruiz-del-Solar received his diploma in Electrical Engineering and the M.S. degree in Electronic Engineering from the Technical University Federico Santa Maria (Chile) in 1991 and 1992, respectively, and the Doctor-Engineer degree from the Technical University of Berlin in 1997. In March 1998 he became an Assistant Professor of Electrical Engineering at the University of Chile. His research interests include Biometric Systems, Face Recognition, Application of Soft-Computing technologies to Computer Vision and Pattern Recognition problems, Texture Analysis and Retrieval, and Robotics.

Pablo Navarrete received the B.S. degree in Physics from the University of Chile in 2000 and the Electrical Engineering diploma from the University of Chile in 2002. He is currently pursuing a Ph.D. in Electrical Engineering at Purdue University with major area in Communication and Signal Processing.

Table 1. Mean recognition rates using different numbers of training images per class, and taking the average of 20 different training sets. The small numbers are the standard deviation of each recognition rate. All results consider the top 1 match for recognition. Yale Database.

projection method	images per class	axes	euclidean	cos(·)	SOM	FFC	whitening euclidean	whitening cos(·)	whitening SOM	whitening FFC
PCA		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	6	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		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	5	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		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	4	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		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	3	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		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	2	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. Mean recognition rates for standard approaches using FERET database images with 3 images per class in the target set. All results consider the top 1 match for recognition.

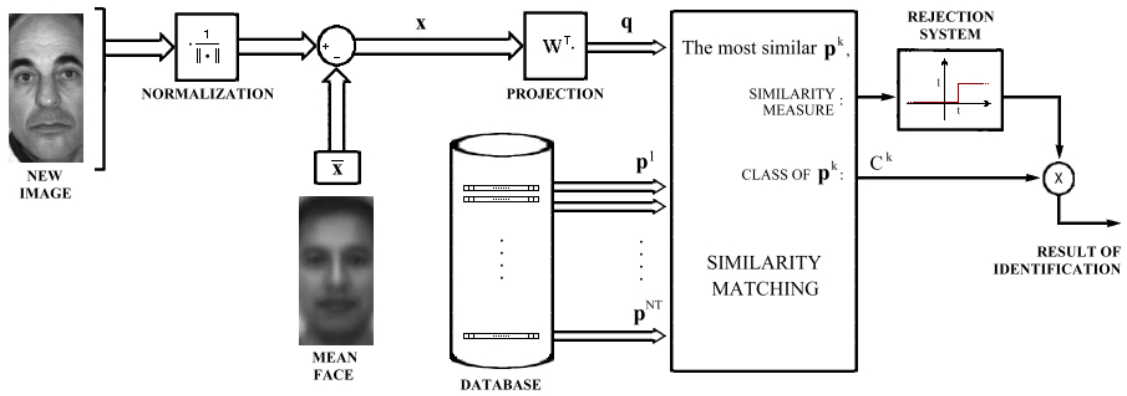| projection method	images per class	axes	euclidean	cos(·)	FFC	whitening euclidean	whitening cos(·)	whitening FFC
PCA		212	87.0	88.6	85.0	74.4	89.4	85.0
FLD	3	108	91.3	94.1	92.1	85.8	92.1	92.1
E.P.		115	91.0	93.1	91.0	-	-	-
PCA		180	81.9	83.7	80.7	62.3	86.0	80.7
FLD	2	73	79.5	88.2	85.2	79.5	88.2	85.2
E.P.		96	80.3	85.8	83.3	-	-	-

Figure Captions

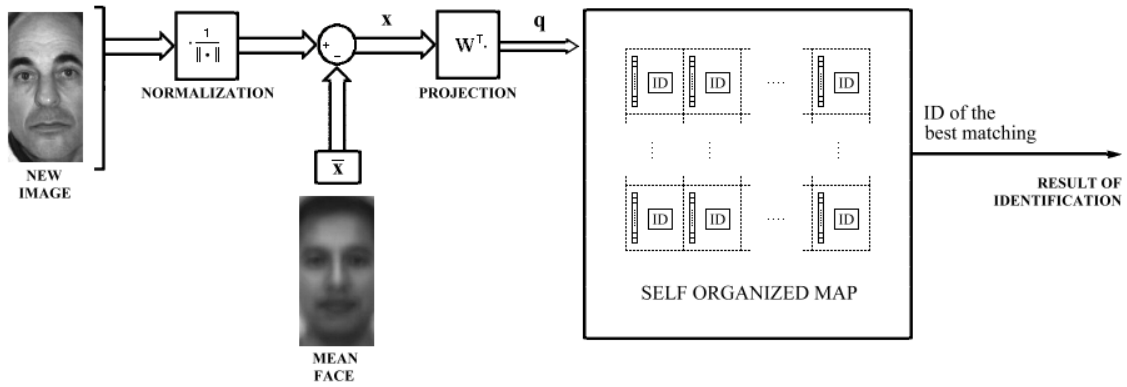Figure 1: Block diagram of a given eigen-space face recognition system.

Figure 2: Block diagram of an eigen-space face recognition system using SOM clustering.

Figure 1: Block diagram of a given eigen-space face recognition system.

Figure 2: Block diagram of an eigen-space face recognition system using SOM clustering.