Comparison of Novel Dimension Reduction Methods in Face Verification

Licesio J. Rodríguez-Aragón, Cristina Conde, and Enrique Cabello

Universidad Rey Juan Carlos, c\ Tulipán, s/n,
E-28933, Móstoles, Madrid, Spain
{licesio.rodriguez.aragon, cristina.conde, enrique.cabello}@urjc.es
http://frav.escet.urjc.es

Abstract. The problem of high dimensionality in face verification tasks has recently been simplified by the use of underlying spatial structures as proposed in the Two Dimensional Principal Component Analysis, the Two Dimensional Linear Discriminant Analysis and the Coupled Subspaces Analysis. Besides, the Small Sample Size problem that caused serious difficulties in traditional LDA has been overcome by the spatial approach 2DLDA. The application of these advances to facial verification techniques using different SVM schemes as classification algorithm is here shown. The experiments have been performed over a wide facial database (FRAV2D including 109 subjects), in which only one interest variable was changed in each experiment: illumination, pose, expression or occlusion. For training the SVMs, only two images per subject have been provided to fit in the small sample size problem.

1 Introduction

Face verification is a high dimensional pattern recognition problem. Even low-resolution facial images generate huge dimensional feature spaces. Many face recognition techniques have been developed over the past few decades. Some of them are based on dimension reduction methods such as Eigenface method [1], using principal component analysis to obtain a most expressive subspace for face representation. Others are based on Fisher Face method [2], that uses linear discriminant analysis to obtain the most discriminant subspace. These methods as other improved variations treat input images as vectors. Some recent works have begun to treat images as two dimensional matrices.

Another problem present in face recognition tasks is learning in the small sample size problem [3], what is of great practical interest. In face recognition it is not easy to collect a database with a large number of images of each individual. Furthermore, it is often difficult to collect training or testing data to cover all possible variations as illumination, pose, expression, occlusion, or even age.

Linear discriminant analysis (LDA) approach is theoretically one of the best classification methods as it distinguishes between within and between class variations. However, in the traditional approach [2] after transforming input images into vectors the within-class scatter matrix is almost singular, as the number

A. Campilho and M. Kamel (Eds.): ICIAR 2006, LNCS 4142, pp. 305–316, 2006.

of training samples is very small compared with the dimension of the image vector. Various schemes have been proposed to solve that problem. The recent approach of treating images as matrices instead of vectors solve in a simple way the singularity of the within class scatter matrix.

The analysis of these new methods over a wide set of subjects, joined in a facial database of images, allows to measure the improvements of the new proposed dimension reduction methods based on matrices over individual variations such as illumination, pose, expression, occlusion. A set of different experiments with only two training images per subject to fit the small sample size problem has been carried out. Different classification strategies using SVMs have been designed and used along our experiments.

2 Feature Extraction

Traditional feature extraction techniques require that 2D face images are vectorized into a 1D row vector to then perform the dimension reduction [4,5]. The resulting image vectors belong to a high-dimensional image vector space where covariance matrices are evaluated with a higher associated computational cost than in the following proposed methods.

Recently, Two-Dimensional PCA (2DPCA), Two-Dimensional LDA (2DLDA) and Coupled Subspace Analysis (CSA) have been developed for bidimensional data feature extraction. These methods are based on 2D matrices rather than 1D vectors, preserving spatial information.

2.1 Principal Component Analysis

Given a set of images $I_1, I_2, \ldots, I_N$ of height h and width w, PCA considers the images as 1D vectors in a $h \cdot w$ dimensional space. The facial images are projected onto the eigenspace spanned by the leading ortonormal eigenvectors, those of higher eigenvalue, from the sample covariance matrix of the training images. Once the set of vectors has been centered, the sample covariance matrix is calculated, resulting a matrix of dimension $h \cdot w \times h \cdot w$. It is widely known that if $N \ll h \cdot w$, there is no need to obtain the eigenvalue decomposition of this matrix, because only N eigenvectors will have a non zero associated eigenvalue [1], the first eigenvector corresponding to the mean face. The obtention of these eigenvectors only requires the decomposition of an $N \times N$ matrix, considering as variables the images, instead of the pixels, and therefore considering pixels as individuals.

Once the first d eigenvectors are selected and the proportion of the retained variance fixed, a projection matrix A is formed with $h \cdot w$ rows and d columns, one for each eigenvector. Then a feature vector $Y_{d\times 1}$ is obtained as a projection of each image $I_{h\cdot w\times 1}$, considered as a 1D vector, onto the new eigenspace.

2.2 Linear Discriminant Analysis

The previous method maximizes the total scatter retained by the fixed dimension. Information provided by the labels of the set of images, $I_1, I_2, \ldots, I_N$, is not used. Linear Discriminant Analysis shapes the scatter in order to make it more reliable for classification. Traditional Linear Discriminant Analysis uses this information to maximize between-class scatter whereas within-class scatter is minimized simplifying the classification process and focusing the problem in a more reliable way.

As images are transformed into a 1D vector, the method faces the difficulty that the within-class scatter matrix, of dimension $h \cdot w \times h \cdot w$, is always singular as the number of images N of the set is usually much lower than the number of pixels in an image. An initial projection using PCA is done to a lower dimensional space so that the within-scatter matrix is non singular. Then applying the standard Fisher Linear Discriminant Analysis, the dimension is finally reduced [2].

2.3 Two-Dimensional Principal Component Analysis

The consideration of images $I_{h\times w}$ as 1D vectors instead as 2D structures is not the right approach to retain spatial information. Pixels are correlated to their neighbours and the transformation of images into vectors produces a loss of information preserving the dimensionality. On the contrary, the main objective of these methods is the reduction of dimensionality and the least loss of information as possible.

The idea recently presented as a variation of traditional PCA, is to project an image $I_{h\times w}$ onto X^{PCA} by the following transformation [6,7],

$$Y_{h\times 1} = I_{h\times w} \cdot X^{PCA}_{w\times 1}. \tag{1}$$

As result, a h dimensional projected vector Y, known as projected feature vector of image I, is obtained. The total covariance matrix S_X over the set of projected feature vectors of training images $I_1, I_2, \ldots, I_N$ is considered. The mean of all the projected vectors, $\overline{Y} = \overline{I} \cdot X^{PCA}$, being $\overline{I}$ the mean image of the training set, is taken into account.

$$\begin{aligned} S_X &= \tfrac{1}{N} \textstyle\sum_{i=1}^{N} (Y_i - \overline{Y})(Y_i - \overline{Y})^T \\ &= \tfrac{1}{N} \textstyle\sum_{i=1}^{N} [(I_i - \overline{I})X][(I_i - \overline{I})X]^T \end{aligned} \tag{2}$$

The maximization of the total scatter of projections is chosen as the criterion to select the vector X^{PCA}. The total scatter of the projected samples is characterized by the trace of the covariance matrix of the projected feature vectors. Applying the criterion to (2) the following expression is obtained,

$$J(X) = tr(S_X) = X^T[\frac{1}{N}\sum_{i=1}^{N}(I_i - \overline{I})^T(I_i - \overline{I})]X. \tag{3}$$

What is known as image covariance matrix S defined as a $w \times w$ nonnegative matrix can be directly evaluated using the training samples,

$$S = \frac{1}{N}\sum_{i=1}^{N}[(I_i - \overline{I})^T(I_i - \overline{I})]. \tag{4}$$

The optimal projection axis X^{PCA} is the unitary vector that maximizes (3), which corresponds to the eigenvector of S of largest associated eigenvalue.

2.4 Two-Dimensional Linear Discriminant Analysis

The idea presented as 2DPCA, has been upgraded to consider the class information [8,9]. Suppose there are L known pattern clases having M samples for each class, $N = L \cdot M$. The idea is to project each image as in (1), but to obtain X^{LDA} with the information provided by the classes. The covariance over the set of images can be decomposed into between-class and within-class. The mean of projected vectors as in 2DPCA as well as the mean of projected vectors of the same class $\overline{Y^j} = \overline{I^j} \cdot X^{LDA}$, being $\overline{I^j}$ the mean image of the class $j = 1, \ldots, L$, are taken into account.

$$\begin{aligned} S_{XB} &= \textstyle\sum_{j=1}^{L} M(\overline{Y^j} - \overline{Y})(\overline{Y^j} - \overline{Y})^T \\ &= \textstyle\sum_{j=1}^{L} M[(\overline{I^j} - \overline{I})X][(\overline{I^j} - \overline{I})X]^T \end{aligned} \tag{5}$$

$$\begin{aligned} S_{XW} &= \textstyle\sum_{j=1}^{L} \sum_{i=1}^{M} (Y_i^j - \overline{Y^j})(Y_i^j - \overline{Y^j})^T \\ &= \textstyle\sum_{j=1}^{L} \sum_{i=1}^{M} [(I_i^j - \overline{I^j})X][(I_i^j - \overline{I^j})X]^T \end{aligned} \tag{6}$$

The objective function maximized in this case to select X^{LDA} is considered a class specific linear projection criterion, and can be expressed as

$$J(X) = \frac{tr(S_{XB})}{tr(S_{XW})}. \tag{7}$$

The total between and within covariances are defined as $w \times w$ nonnegative matrices and can be directly evaluated.

$$S_B = \sum_{j=1}^{L} M[(\overline{I^j} - \overline{I})][(\overline{I^j} - \overline{I})]^T; \quad S_W = \sum_{j=1}^{L} \sum_{i=1}^{M} [(I_i^j - \overline{I^j})][(I_i^j - \overline{I^j})]^T \tag{8}$$

Both matrices are formally identical to the corresponding traditional LDA, and by maximizing (7) the within-class scatter is minimized whereas the between-class scatter is maximized, giving as result the maximization of discriminating information. The optimal projection axis X^{LDA} is the unitary vector that maximizes (7), which corresponds to the eigenvector of $S_B \cdot S_W^{-1}$, of largest associated eigenvalue.

2.5 Coupled Subspace Analysis

The previous 2DPCA and 2DLDA methods consider the projection of an image onto a vector X^{PCA} or X^{LDA} respectively. As result, a vector of the same dimension as the image height, h, is obtained (1). Recently a new approach has been presented to reconstruct the original image matrices with two low dimensional coupled subspaces, in the sense of least square error [10]. These two subspaces encode the row and column information of the image matrices.

Let us denote Y_i, of height h' and width w', as the lower dimensional matrix representation of sample I_i, $i = 1, \ldots, N$, derived from two projection matrices $B_{h \times h'}$ and $C_{w \times w'}$,

$$Y_{h' \times w'} = B^T_{h' \times h} \cdot I_{h \times w} \cdot C_{w \times w'}. \tag{9}$$

The matrices B and C are chosen as those that best reconstruct the original images from the projections, in the sense of least square error satisfying the following optimal matrix reconstruction criterion,

$$(B^*, C^*) = \underset{B,C}{\operatorname{argmin}} \sum_i \|B \cdot Y_i \cdot C^T - I_i\|^2_F. \tag{10}$$

Being $\| \cdot \|_F$ the Frobenius norm of a matrix.

The objective function has no closed form and to obtain a local optimal solution, an iterative procedure has been presented. C^* is obtained for a given B and, in an iterative way, B^* is obtained for a given C. By iteratively optimizing the objective function with respect to B and C, respectively, we can obtain a local optimum of the solution. The whole procedure is called Coupled Subspace Analysis [10].

That Coupled Subspace Analysis is connected with traditional PCA and 2DPCA has been shown. Principal component analysis is a special case of CSA algorithm with $w = 1$ and 2DPCA is a special case of CSA algorithm with fixed $B = Id$.

3 Projection and Reconstruction

In 2DPCA and 2DLDA, as in traditional PCA, a proportion of retained variance can be fixed, $\sum_1^d \lambda_i / \sum_1^w \lambda_i$, where $\lambda_1 > \lambda_2 > \cdots > \lambda_w$ are the eigenvalues and $X_1, X_2, \ldots, X_d$ are the eigenvectors corresponding to the d largest eigenvalues.

Once d is fixed, $X_1, X_2, \ldots, X_d$ are the ortonormal axes used to perform the feature extraction. Let $V = [Y_1, Y_2, \ldots, Y_d]$ and $U = [X_1, X_2, \ldots, X_d]$, then

$$V_{h \times d} = I_{h \times w} \cdot U_{w \times d}. \tag{11}$$

A set of projected vectors, $Y_1, Y_2, \ldots, Y_d$, are obtained for both methods. Each projection over an optimal projection vector is a vector, instead of a scalar as in traditional PCA. A feature matrix $V_{h \times d}$ for each considered dimension reduction method is produced, containing either the most amount of variance, or the most discriminating features of image I.

In CSA the projection is performed through the optimal projection matrices $B^*_{h \times h'}$ and $C^*_{w \times w'}$ as in (9). As result, the extracted features form a lower dimensional matrix of height h' and width w'.

3.1 Image Reconstruction

In these dimension reduction methods, a reconstruction of the images from the features is possible. An approximation of the original image with the retained

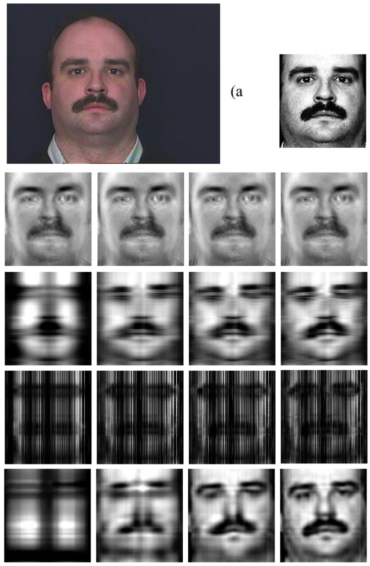

Fig. 1. a) One of the original frontal images in our database. b) Automatically selected window containing the facial expression of the subject in equalized gray scale. From left to right: c) reconstructed images (12), for $d = 30, 90, 150, 210$, from PCA projection. d) and e) reconstructed images (12), for $d = 2, 4, 6, 8$, from 2DPCA and 2DLDA projections respectively. f) reconstructed images (12, for $h' = w' = 3, 9, 15, 21$, from CSA projection.

information determined by d for PCA, 2DPCA and 2DLDA or (h', w') for CSA, is obtained.

$$\begin{aligned} \widetilde{I}_{h \cdot w \times 1} &= A_{h \cdot w \times d} \cdot Y_{d \times 1} \quad \text{PCA image reconstruction.} \\ \widetilde{I}_{h \times w} &= V_{h \times d} \cdot U^T_{d \times w} \quad \text{2DPCA or 2DLDA image reconstruction.} \\ \widetilde{I}_{h \times w} &= B_{h \times h'} \cdot Y_{h' \times w'} \cdot C^T_{w' \times w} \quad \text{CSA image reconstruction.} \end{aligned} \tag{12}$$

Some samples of this reconstruction process are shown in Fig. 1.

4 Classification with SVM

SVM is a method of learning and separating binary classes [11]; it is superior in classification performance and is a widely used technique in pattern recognition and especially in face verification tasks [12].

Given a set of features $y_1, y_2, \ldots, y_N$ where $y_i \in \mathbb{R}^n$, and each feature vector associated to a corresponding label $l_1, l_2, \ldots, l_N$ where $l_i \in \{-1, +1\}$, the aim of a SVM is to separate the class label of each feature vector by forming a hyperplane

$$(\omega \cdot y) + b = 0, \quad \omega \in \mathbb{R}^n, b \in \mathbb{R}. \tag{13}$$

The optimal separating hyperplane is determined by giving the largest margin of separation between different classes. This hyperplane is obtained through a minimization process subjected to certain constrains. Theoretical work has solved the existing difficulties of using SVM in practical application [13].

As SVM is a binary classifier, a *one vs. all* scheme is used. For each class, each subject, a binary classifier is generated with positive label associated to feature vectors that correspond to the class, and with negative label associated to all the other classes.

4.1 Facial Verification Using SVM

In our experiments and in order to fit in the small sample size problem [3], the same two frontal and neutral images of every subject are selected as the training set for every experiment. A disjoint group of images, all of them affected by the same perturbation, is selected as the test set. The training set is used in the feature extraction process through PCA, 2DPCA, 2DLDA and CSA. Then, the training images are projected onto the new ortonormal axes and the feature vector (PCA), vectors (2DPCA,2DLDA), or low dimensional matrix (CSA) are obtained. The required SVMs are trained for each subject.

Several strategies have been used to train and combine the SVMs. When training and classifying PCA features, each image generates one feature vector $Y_{d\times 1}$ and one SVM is trained for each subject, with its feature vectors labelled as $+1$ and all the other feature vectors as -1.

For feature vectors obtained from 2DPCA and 2DLDA, each image generates a set of projected vectors, $V_{h\times d} = [Y_1, Y_2, \ldots, Y_d]$, and three different strategies have been considered. First strategy generates a unique feature vector through a concatenation of d projected vectors, then one SMV is trained for each subject as in PCA. The second and third approaches consider d projected vectors and consequently for each subject d SVMs are trained, one for each feature vector. These d outputs are then combined to produce a final classification output, first through an arithmetic mean and secondly through a weighted mean.

On the other hand, applying CSA produces a low dimensional matrix $Y_{h'\times w'}$ for every image. This feature matrix is then transformed into a vector, $Y_{h'\cdot w'\times 1}$, and as in PCA one SVM is trained for each subject, with its features labelled as $+1$ and all the features belonging to others subjects labelled as -1.

Once the SVMs are trained, each image from the test set is projected obtaining the corresponding features for each dimension reduction method (1,9). The features of the test set are classified through the SVMs to measure the performance of the generated system.

For the SVM obtained from the PCA, from the concatenation strategy of 2DPCA, 2DLDA and the CSA feature vectors, the output is compared with the known label of every test image. However, for the ensemble of SVMs obtained from the 2DPCA and 2DLDA feature vectors, the d outputs are combined whether through an arithmetic or a weighted mean. Arithmetic approach combines the d outputs through an arithmetic mean. At weighted approach, every output is weighted with the amount of variance explained by its dimension, which means that each output will be taken into account proportionally to the value of the eigenvalue associated to the corresponding eigenvector: $\lambda_i / \sum_{j=1}^{d} \lambda_j$ is the weight for the $i-$SVM, $i = 1, 2, \ldots, d$.

To measure the system performance, a cross validation procedure is carried out. Four possible experiment outcomes are possible: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). The Equal Error Rate, EER, that is the value for which false positive rate (FP) is equal to false negative rate (FN) is a valuable reference of the performance of the system. The system threshold can then be adjusted to more or less sensitiveness, but in order to achieve fewer errors new and better methods, like 2DPCA, 2DLDA and CSA are required.

5 Design of Experiment

Our research group has collected FRAV2D, a quite complete set of facial images including 109 subjects. All the images have been taken under controlled conditions of pose, expression and illumination. 32 images of each subject were taken, being 12 frontal, 4 preforming a 15° rotation, 4 performing a 30° rotation, 4 with zenithal instead of diffuse illumination, 4 performing different facial expressions and 4 occluding parts of the face. A partial group of this database as well as other facial databases are freely available under demand for research purposes[1].

They are colored images of size 240×320 pixels with homogeneous background color. A window of size 140×130 pixels containing the most meaningful part of the face, has been automatically selected in every image and stored in equalized gray scale. That is the information that will be analyzed through the dimension reduction and classification methods (Fig. 1).

The purpose of the following experiments is to confront the performance of the dimension reduction methods (PCA, 2DPCA, 2DLDA and CSA) in the task of face verification through SVM as previously mentioned. PCA is presented as a base line method to evaluate the improvements.

Each experiment has been performed for the 109 subjects from our database. In all the experiments, the train set for the extraction of features and for the classifiers training is formed by the same 2 frontal and neutral images of each subject, in order to fit in the small sample size problem. Then, the classifiers have been tested over 5 different groups of images. Firstly, the 10 remaining frontal and neutral images for each subject have been used to perform the cross validation process. In a second experiment, the 4 images obtained with zenithal illumination have formed the test set. The 4 15° turn images have been selected to measure the performance of the system to pose variations. In the fourth experiment 4 images with expressions changes have been used. And finally, 4 occluded images for each subject have formed the test set.

[1] http://frav.escet.urjc.es

Tests, varying the dimensions of the different feature spaces, have been carried out for the four dimension reduction methods. For PCA, experiments have been performed for values $d = 30, 60, 90, 120, 150, 180, 210$. For 2DLDA and 2DPCA, values $d = 1, 2, 3, \ldots, 10$ under the three different classification strategies. And for CSA, we have considered low dimensional square matrices of sizes $h' = w' = 3, 6, 9, 12, 15, 18$.

For each dimension reduction method and for each experiment, we present the evolution of the EER for different values of d or (h', w') (Figs. 2, 3).

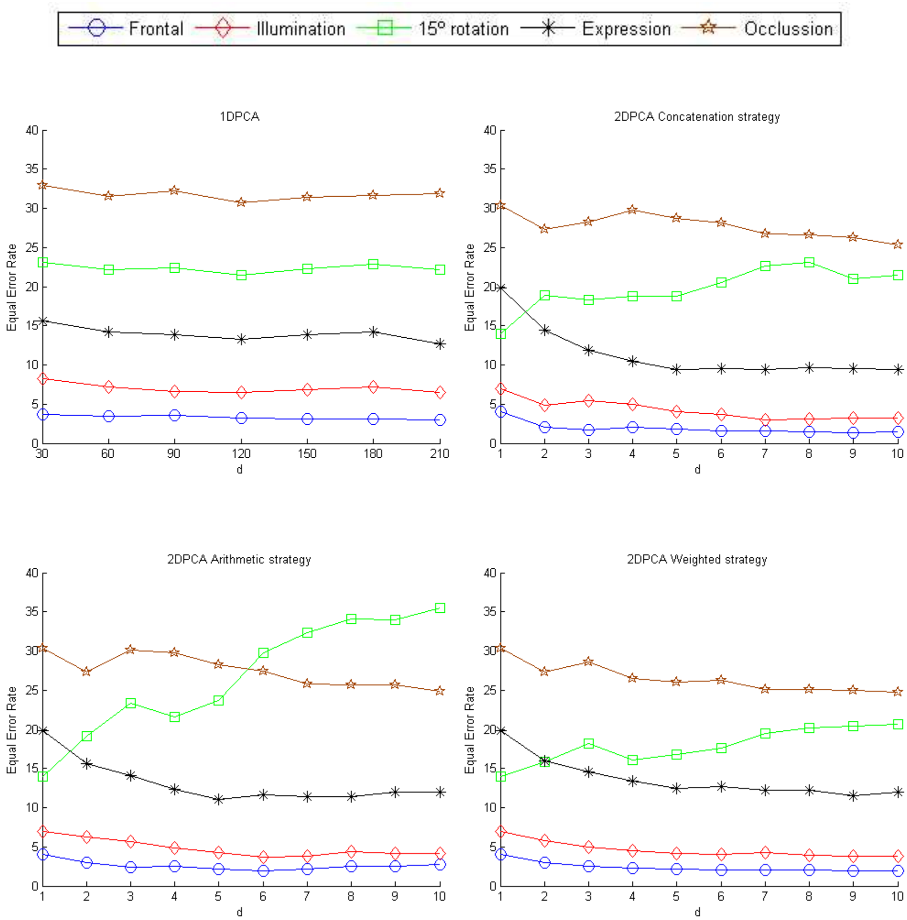

Fig. 2. Evolution of the EER for the PCA and 2DPCA methods for the five experiments. Three strategies are used with 2DPCA: concatenation, arithmetic and weighted. For PCA dimension reduction method $d = 30, 60, 90, 120, 150, 180, 210$. For 2DPCA dimension reduction method $d = 1, 2, \ldots, 10$.

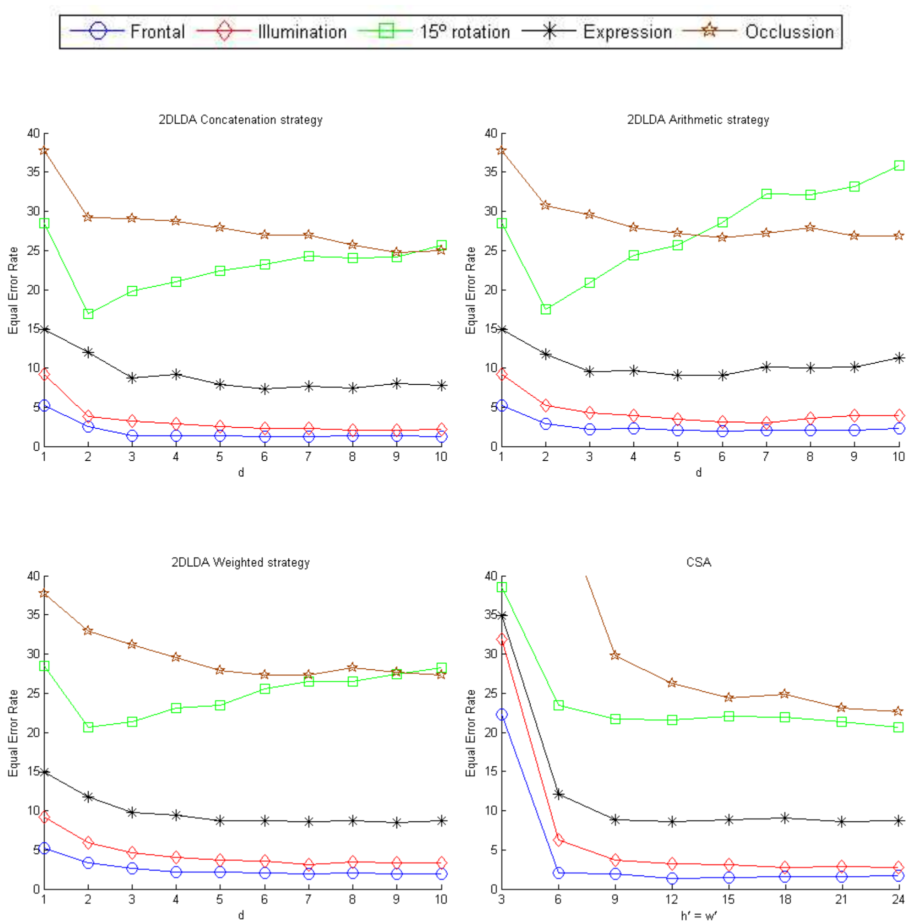

Fig. 3. Evolution of the EER for 2DLDA and CSA methods for the five experiments. As well as for 2DPCA, three strategies are used with 2DLDA: concatenation, arithmetic and weighted. For 2DLDA dimension reduction method $d = 1, 2, \ldots, 10$. For CSA, dimension reduction is perfomed for values $h' = w' = 3, 6, 9, 15, 18, 21, 24$.

6 Conclusions

Better results for spatial dimension reduction methods are evidently achieved as shown in EER curves (Figs. 2 and 3) and values presented in Table 1 for each experiment and method. This improvement is present throughout all the experiments.

In frontal, illumination and expression images, best performance is done by 2DLDA method with the concatenation strategy. For 15° rotated images, the best results are archived by 2DPCA for the lowest value of d. And for occluded images, the lowest error, though significantly high, is achieved by CSA.

Table 1. Best EER (in percentage), values obtained for each dimension reduction method in each experiment. In brackets: the dimension for which it is achieved. Concatenation strategy always obtains the best results for the different strategies in 2DPCA and 2DLDA.

Experiment	PCA (d)	2DPCA (d)	2DLDA (d)	CSA ($h' \times w'$)
1) Frontal Images	3.0 (210)	1.3 (9)	1.2 (6)	1.3 (12×12)
2) Zenithal Illumination	6.5 (210)	2.9 (7)	2.0 (9)	2.7 (18×18)
3) 15° Rotated	21.5 (120)	13.9 (1)	16.9 (2)	20.6 (24×24)
4) Expression Images	12.7 (210)	9.4 (10)	7.2 (6)	8.5 (12×12)
5) Occluded Images	30.7 (120)	25.2 (10)	24.6 (9)	22.6 (24×24)

Problems merged from illumination and facial expressions are more accurately solved with these novel dimension reduction methods. However, the classification strategies used (Arithmetic and Weighted), never improve the concatenation strategy. Serious improvements have also been done as far as rotated and occluded images are concerned. It is interesting that in these two special cases the best results are not achieved by 2DLDA (theoretically the best approach) as in the other approaches. The good performance reached by traditional PCA in frontal images has been seriously improved.

It is clear that the spatial dimension reduction methods are more reliable for the purpose of face verification, even under the small sample size problem (only 2 training images per subject), but deeper work has to be done to use all the information provided by the dimension reduction methods in order to achieve a more accurate verification. New algorithms based on matrix and tensor representations [14] are appearing, and further tests like the ones here performed need to be done to evaluate the improvements.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neurosicience **3** (1999) 71–86
2. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997) 711–720
3. Guo, G., Dyer, C.: Learning from examples in the small sample case: face expression recognition. IEEE Transactions on Systems, Man, and Cybernetics-Part B **35** (2005) 477–488
4. Pang, S., Kim, D., Bang, S.Y.: Memebership authentication in the dynamic group by face classification using SVM ensemble. Pattern Recognition Letters **24** (2003) 215–225
5. Cao, L., Chong, W.: Feature extraction in support vector machine: A comparison of PCA, KPCA and ICA. In: Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'OZ). Volume 2. (2002) 1001–1005
6. Yang, J., Yang, J.: From image vector to matrix: a straightforward image projection technique–IMPCA vs. PCA. Pattern Recognition **35** (2002) 1997–1999

7. Yang, J., Zhang, D., Frangi, F., Yang, J.: Two-dimmensional PCA: A new approach to apperance-based face representation and recognition. IEEE Transacctions on Pattern Recognition and Machine Intelligence **26** (2004) 131–137
8. Li, M., Yuan, B.: A novel statistical linear discriminant analysis for image matrix: two-dimensional fisherfaces. In: Proceedings of the International Conference on Signal Processing. (2004) 1419–1422
9. Chen, S., Zhu, Y., Zhang, D., Yang, J.: Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA. Pattern Recognition Letters **26** (2005) 1157–1167
10. Xu, D., Yan, S., Zhang, L., Liu, Z., Zhang, H.: Coupled subspace analysis. Technical Report MSR-TR-2004-106, Microsof Research (2004)
11. Cortes, C., Vapnik, V.: Support vector network. Machine Learning **20** (1995) 273–297
12. Fortuna, J., Capson, D.: Improved support vector classification using PCA and ICA feature space modiffication. Pattern Recognition **37** (2004) 1117–1129
13. Joachims, T.: Making large scale support vector machine learning practical. In: Advances in Kernel Methods: Support Vector Machines. MIT Press, Cambridge, MA (1998)
14. Yan, S., Xu, D., Yang, Q., Zhang, L., Tang, X., Zhang, H.: Discriminant analysis with tensor representation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 1. (2005) 526 – 532