Face Verification Advances Using Spatial Dimension Reduction Methods: 2DPCA & SVM

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

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

Abstract. Spatial dimension reduction called Two Dimensional PCA method has recently been presented. The application of this variation of traditional PCA considers images as 2D matrices instead of 1D vectors as other dimension reduction methods have been using. The application of these advances to verification techniques, using SVM as classification algorithm, is here shown. The simulation has been performed over a complete facial images database called FRAV2D that contains different sets of images to measure the improvements on several difficulties such as rotations, illumination problems, gestures or occlusion.
The new method endowed with a classification strategy of SVMs, seriously improves the results achieved by the traditional classification of PCA & SVM.

1 Introduction

Improving security and developing new smart environments are some of the key points in which biometry plays a most relevant role. Recent studies [1] have shown that technology is in very early stages of development to perform surveillance tasks at critical locations. However, simulations or real tests are crucial to obtain the required feedback in order to improve in the right direction.

Most of current face verification systems [2] relay not only in one algorithm but in an optimal ensemble of different methods that improve the global result. Classical dimension reduction methods, like Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA), and Gabor Filters [3], as well as other improved variations, like Independent Component Analysis (ICA) [4] and Kernel Principal Component Analysis (KPCA), are jointed and combined with classification algorithms, like Neural Networks and Support Vector Machines (SVM) [3,5]. The improvement of a single method implies the global improvement of the system.

Dimensional reduction methods applied to face verification tasks obtain a feature set for each image. Efficient feature extraction and selection schemes

F. Roli and S. Vitulano (Eds.): ICIAP 2005, LNCS 3617, pp. 978–985, 2005.

are crucial for a successful face verification system. Robustness in the feature extraction process to pose and light variations is aimed.

New advances on PCA method called Two-Dimensional PCA [6,7] have shortly been presented, and preliminar experiments and junctions of this new method with SVM are the focus of this work. Experiments are performed over a wide set of subjects, joined in a facial database of images, that combines different pose, gesture and illumination variations, which allow the measurement of the advances.

2 Feature Extraction

The PCA related feature extraction techniques require that 2D face images are transformed into a 1D row vector to then perform the dimension reduction [4]. The resulting image vectors belong to a high-dimensional image vector space where covariance matrices are evaluated with a high associated computational cost.

Recently, a Two-Dimensional PCA method (2DPCA) has been developed for bidimensional data feature extraction. 2DPCA is 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 orthornormal 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 [8]. 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, $\sum_1^d \lambda_i / \sum_1^N \lambda_i$, being $\lambda_1 > \lambda_2 > \cdots > \lambda_N$ the eigenvalues, 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.

$$Y_{d\times 1} = A^T_{d\times h\cdot w} \cdot I_{h\cdot w\times 1} \tag{1}$$

2.2 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 neighbors 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 by the following transformation [6,7],

$$Y_{h\times 1} = I_{h\times w} \cdot X_{w\times 1}. \tag{2}$$

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, $\overline{Y} = \overline{I} \cdot X$, being $\overline{I}$ the mean image of the training set, is taken into account.

$$\begin{aligned} S_X &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}(Y_i - \overline{Y})(Y_i - \overline{Y})^T \\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}[(I_i - \overline{I})X][(I_i - \overline{I})X]^T \end{aligned} \tag{3}$$

The maximization of the total scatter of projections is chosen as the criterion to select the vector X. 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 (3) the following expression is obtained,

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

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

$$G = \frac{1}{N}\sum_{i=1}^{N}(I_i - \overline{I})^T(I_i - \overline{I})]. \tag{5}$$

The optimal projection axis X_{opt} is the unitary vector that maximizes (4), which corresponds to the eigenvector of G of largest associated eigenvalue.

Usually, as well as in traditional PCA, a proportion of retained variance is 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 orthornormal 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

$$V_{h\times d} = I_{h\times w} \cdot U_{w\times d}. \tag{6}$$

A set of projected vectors, $Y_1, Y_2, \ldots, Y_d$, are obtained and in 2DPCA each principal component is a vector instead of an scalar as in traditional PCA. A feature matrix $V_{h\times d}$ is produced containing the most discriminating features of image I.

2.3 Image Reconstruction

In both methods, PCA and 2DPCA, a reconstruction of the images from the features is possible. An approximation of the original image with the retained information determined by d is obtained.

$$\begin{array}{ll} \widetilde{I}_{h\cdot w\times 1} = A_{h\cdot w\times d}\cdot Y_{d\times 1} & \text{PCA image reconstruction.} \\ \widetilde{I}_{h\times w} \ \ = V_{h\times d}\cdot U^T_{d\times w} & \text{2DPCA image reconstruction.} \end{array} \tag{7}$$

3 Classification with SVM

SVM is a method of learning and separating binary classes [9], it is superior in classification performance and is a widely used technique in pattern recognition and especially in face verification tasks [5].

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

$$(\omega \cdot y) + b = 0, \quad \omega \in \mathbb{R}^n, b \in \mathbb{R}. \tag{8}$$

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 [10].

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 negative label associated to all the other classes.

3.1 Facial Verification Using SVM

In our experiments a group of images from every subject is selected as the training set and a disjoint group of images is selected as the test set. The training set is used in the feature extraction process through PCA and 2DPCA. Then, the training images are projected onto the new orthornormal axes and the feature vector (PCA), or vectors (2DPCA), are obtained. For each subject the required SVMs are trained.

For both methods, PCA and 2DPCA, the same amount of retained variance has been fixed, giving as result different values of d (number of considered principal components) for each reduction method. 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. On the other hand, for feature vectors obtained from 2DPCA, each image generates a set of projected vectors, $V_{h\times d} = [Y_1, Y_2, \ldots, Y_d]$, and consequently for each subject d SVMs need to be trained, one for each feature vector Y_i.

Once the SVMs are trained, images from the test set are projected onto the eigenspace obtained from the training set. 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 feature vectors, the output is compared with the known label of every test image. However, for the ensemble of SVMs obtained from the 2DPCA feature vectors, the d outputs are combined through a weighted mean. Every output is weighted with the amount of variance explained by its dimension, that means that each output will be taken in 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. Results are then described by using Receiver Operating Curve, ROC curve, as there are four possible experiment outcomes: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). 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, are required.

4 Images Database: FRAV2D

The Face Recognition and Artificial Vision[1] group (FRAV) at the Universidad Rey Juan Carlos, has collected a quite complete set of facial images for 109 subjects. All the images have been taken under controlled conditions of pose and illumination. 32 images were taken of each subject, being 12 frontal, 4 preforming a 15° rotation, 4 performing a 30° rotation, 4 with zenithal instead of diffuse illumination, 4 performing different gestures and 4 occluding parts of the face. A partial group of this database is freely available for research purposes.

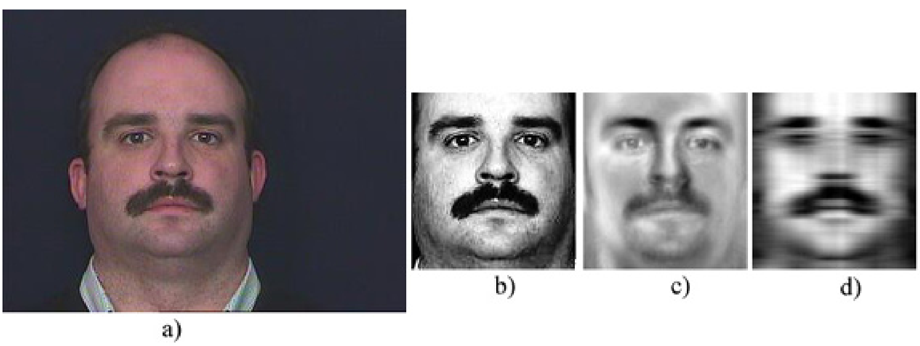

Fig. 1. a) One of the original frontal images in the FRAV2D database. b) Automatically selected window containing the facial expression of the subject in equalized gray scale. c) and d) Reconstructed images (7), for the 60% of retained variance, from PCA and 2DPCA projections respectively.

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

The images are colored and 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).

5 Design of Experiments

The purpose of the following experiments is to confront the performance of the traditional PCA method to the new proposed 2DPCA method in the task of face verification through SVM. The retained variance for both methods has been fixed up to 60% resulting values of $d = 20$ for PCA and $d = 5$ for 2DPCA.

Each experiment has been performed for 100 randomly chosen subjects from the whole FRAV2D. In all the experiments, the train set for the extraction of the feature vectors and for the classifiers training is formed by 8 frontal images of each subject. Then, the classifiers have been tested over 5 different groups of images. Firstly, the 4 remaining frontal 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 tests set. The 4 15° images have been selected to measure the performance of the system to rotations. In the fourth experiment 4 images with gestures have been used. And finally, the 4 occluded images for each subject have formed the test set.

Results for each experiment are presented as ROC curves, showing the compared performance of the verification process using PCA and 2DPCA. True positive rate (TP), that is the proportion of correct classifications to positive verification problems, and true negative rate (TN), that is the proportion of correct classifications to negative verification problems, are plotted (Fig. 2).

Besides, the equal error rate (EER), that is the value for which false positive rate (FP) is equal to false negative rate (FN), is presented for each experiment in Table 1.

Table 1. EER, values for which the rate of incorrect classification of positive verifications is equal to the rate of incorrect classification of negative verifications, for each dimension reduction method

Experiment	PCA	2DPCA
1) Frontal Images	1.8	1.0
2) Zenithal Illumination	6.6	2.1
3) 15° Rotated	27.3	18.3
4) Gesture Images	15.8	12.9
5) Occluded Images	34.7	29.5

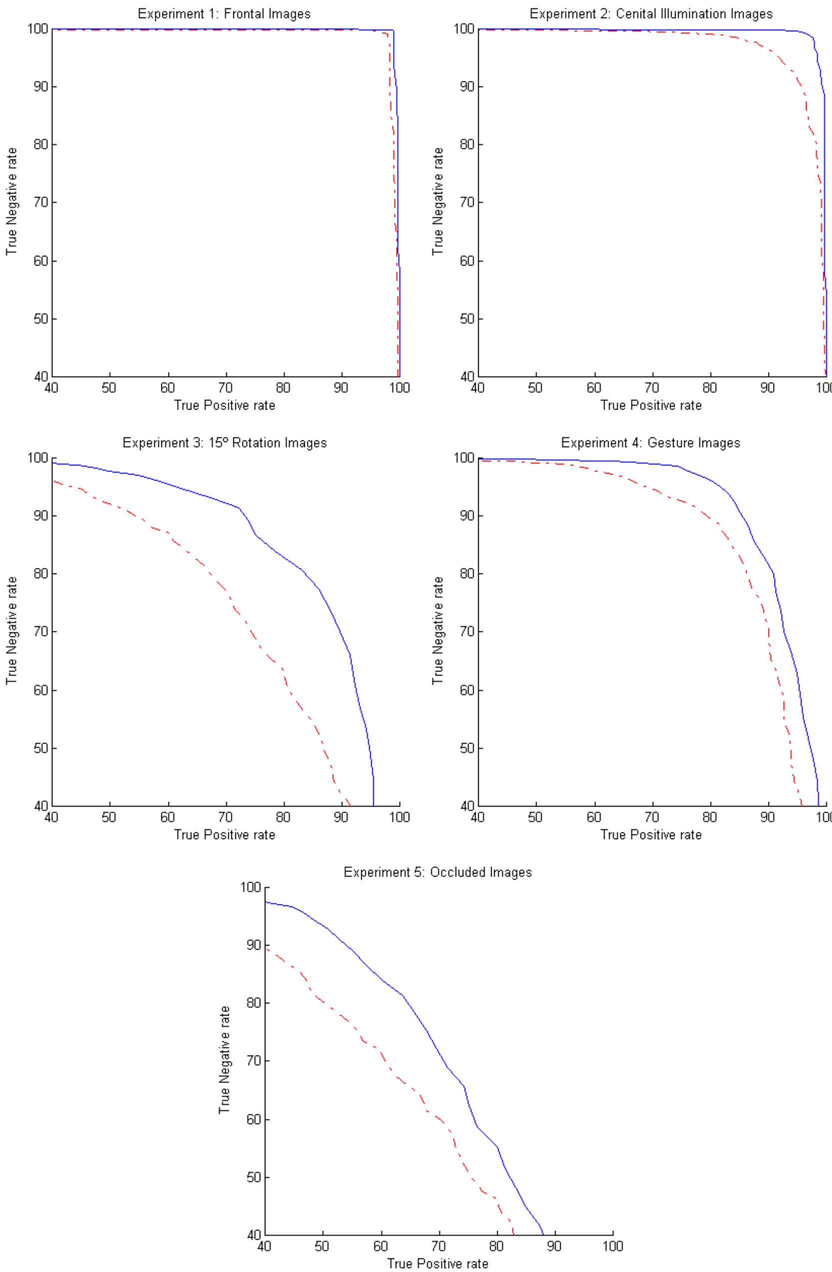

Fig. 2. ROC curve for each experiment, with TP rate in abscises and TN rate in ordinates. The performance of both classifying strategies PCA, dashed line, versus 2DPCA, solid line, is shown.

6 Conclusions

From the ROC curves (Fig. 2) and the EER values presented (Table 1) for each experiment and method, better results are achieved for the spatial reduction method 2DPCA. This improvement is evidently present in all the experiments

Problems merged from illumination and gestures are more accurately solved with this improvement of the dimension reduction and the classification strategy used. Serious improvements have also been done as far as rotated and occluded images are concerned. The good performance reached by traditional PCA in frontal images has even been improved.

The ensemble of the here proposed strategy with other methods and techniques used in face verification tasks will be an aim to be reached in the future developments.

Acknowledgments

Authors would like to thank César Morales García for his enthusiastic work. Also thanks must be given to every one that offered his help to join FRAV2D data base. This work has been partially supported by URJC grant GVC−2004 − 04.

References

1. Bowyer, K. W.: Face recognition technology: security versus privacy. IEEE Technology and society magazine. **Spring** (2004) 9–20
2. Messer, K., Kittler, J., Sadeghi, M., et al. : Face authentification test on the BANCA database. Proceedings of the International Conference on Pattern Recognition. (2004) 523–532.
3. Pang, S., Kim, D. and Bang, S. Y.: Memebership authentication in the dynamic group by face classification using SVM ensemble. Pattern Recognition Letters **24** (2003) 215–225.
4. Kim, T., Kim, H., Hwang, W. and Kittler, J.: Independent Component Analysis in a local facial residue space for face recognition. Pattern Recognition. **37** (2004) 1873–1885.
5. Fortuna, J. and Capson, D.: Improved support vector classification using PCA and ICA feature space modiffication. Pattern Recognition **37** (2004) 1117–1129
6. Yang, J. and 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 and 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. Turk, M. and Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neurosicience. **3** (1991) 71–86.
9. Cortes, C. and Vapnik, V.: Support vector network. Machine Learning. **20** (1995) 273–297.
10. Joachims, T.: Making large scale support vector machine learning practical. In: Advances in Kernel Methods: Support Vector Machines. MIT Press, Cambridge, MA.