

USING PARTIAL INFORMATION FOR FACE RECOGNITION AND POSE ESTIMATION¹

Antonio Rama*, Francesc Tarres*
{alrama, tarres}@gps.tsc.upc.es

Davide Onofrio⁺, Stefano Tubaro⁺
{d.onofrio, tubaro}@elet.polimi.it

*Dept. Teoria del Senyal i Comunicacions - Universitat Politècnica de Catalunya, Barcelona, Spain

⁺Dipartimento di Elettronica e Informazione - Politecnico di Milano, Meiland, Italy

ABSTRACT

The main achievement of this work is the development of a new face recognition approach called Partial Principal Component Analysis (P²CA), which exploits the novel concept of using only partial information for the recognition stage. This approach uses 3D data in the training stage but it permits to use either 2D or 3D data in the recognition stage, making the whole system more flexible. Preliminary experiments carried out on a multi-view face database composed of 18 individuals have shown robustness against big pose variations obtaining higher recognition rates than the conventional PCA method. Moreover, the P²CA method can estimate the pose of the face under different illuminations with accuracy of the 96.15% when classifying the face images in 0°, ±30°, ±45°, ±60° and ±90° views.

1. INTRODUCTION

Automatic face recognition of people is a challenging problem which has received much attention during the last three decades; but, up to date, there is no technique that provides a robust solution to all situations. Among all the challenges implied to face recognition, illumination and pose are the most difficult ones. Any of these problems can change the appearance of a face and cause serious performance degradation in a face recognition system.

Some of the new face recognition strategies tend to overcome both challenges from a 3D perspective. Most of these methods try to reconstruct 3D face models either from multiple images of the same person acquired from a multi camera system or directly obtained from 3D devices like 3D lasers or 3D scanners. The main advantage of using 3D data is that 360° (usually only 180°) texture and depth information are available for face recognition. On the other hand, the main disadvantage is that the accuracy of 3D reconstruction algorithms is proportionally related

to the acquisition parameters; thus, a controlled scenario with all its components well calibrated and synchronized is required. These controlled conditions can be available in the training stage, but on the contrary, most of the surveillance and control access applications present uncontrolled scenarios where only one normal 2D picture of the subject is available. Moreover, all these methods are, usually, computationally very expensive and sometimes the setup cost is very high due to special 3D equipment or multi-camera systems.

Therefore, our research is focussed on the development of a new flexible algorithm which may combine the possibility of processing faces using either 3D or 2D data. It is assumed that in the training stage there is no time or setup cost restriction so that 3D data will be available for the description of the face images on the database. This concept of comparing 2D with 3D data can be also seen as using only partial information which has to be projected in the whole space. Since this technique is based on this concept and on Principal Component Analysis [1], it has been called *Partial Principal Component Analysis* (P²CA).

The rest of the paper is organized as follows. In section 2 an overview of the P²CA algorithm is given and the mathematical procedure is formulated in detail. Section 3 describes the UPC face database and the morphing process that has been used to estimate a 180° texture representation of the faces in cylindrical coordinates. Section 4 summarizes the main results obtained in our experiments making special emphasis in pose estimation. Finally, section 5 contains the conclusions together and the future research.

2. P²CA: AN OVERVIEW

2.1. Algorithm

The objective of P²CA is to implement a mixed 2D-3D method, where either 2D (normal pictures) or 3D data (180° texture images in cylindrical coordinates) can be

¹ACKNOWLEDGMENT: The work presented was developed within VISNET, a European Network of Excellence (<http://www.visnet-noe.org>), funded under the European Commission IST FP6 programme.

used in the recognition stage. However, the method requires a cylindrical representation of the 3D face data for training and including new individuals in the database.

The 3D face information required in the training stage could be obtained by means of 3D laser scanners, by binocular/trinocular camera sets using stereoscopic techniques or simply by a cylindrical scanning of the face with a single camera mounted on a moving structure. In this paper, however, the main objective of this paper was to show the validity of the method and a simpler approach based on manual morphing of 2D pictures taken from different views has been followed.

Let us present the main idea of P²CA. As in PCA the dimensionality of the face images is reduced through the projection into a set of M optimal vectors. The vectors representing the i^{th} individual are obtained as

$$\mathbf{r}_k^i = \mathbf{A}_i^T \cdot \mathbf{v}_k \quad k = 1, \dots, M \quad (1)$$

where \mathbf{A}_i^T is the transpose of the texture image representing individual i and \mathbf{v}_k are the M optimal projection vectors that maximize the energy of the projected vectors \mathbf{r}_k averaged through the whole database. The main difference with conventional PCA is that the whole image is represented as a 2D matrix instead of a 1D vector arrangement representing the image. Clearly, the optimization problem for finding the projection vectors has to be reformulated as will be discussed in section 2.2.

The projection described in Eq. (1) is depicted in Fig 1. The image of the subject i is represented by the M vectors \mathbf{r}_k^i . Each vector \mathbf{r}_k^i has n components where n is the dimension of the matrix \mathbf{A}_i in the horizontal direction (vertical when transposed).

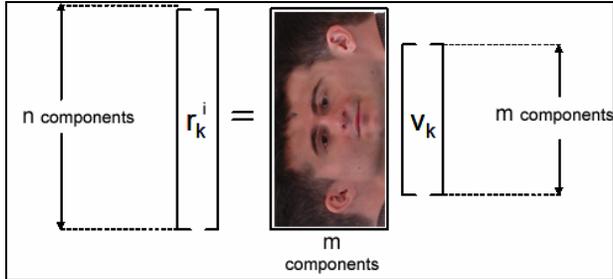


Fig 1 Description of a cylindrical coordinate image by means of projection vectors (training stage)

If complete 3D data of the individual is available the recognition stage is simple. In fact, it is only necessary to convert the 3D data to cylindrical coordinates and compute the resulting M vectors \mathbf{r}_k . The best match is found for the person i that minimizes the Euclidean distance:

$$\min_i \left\{ \xi_k = \sum_{k=1}^M \sum_{l=1}^n (r_k(l) - r_k^i(l))^2 \right\} \quad i = 1, \dots, L \quad (2)$$

where L represents the number of persons in the database.

The main advantage of this representation scheme is that it can also be used when only partial information of the individual is available. Consider, for instance, the situation depicted in Fig.2, where it is supposed that only one 2D picture of the individual is available. Each of the 2D pictures of the subject (frontal and lateral) show a high correlation with the corresponding area of the cylindrical representation of the 3D image.

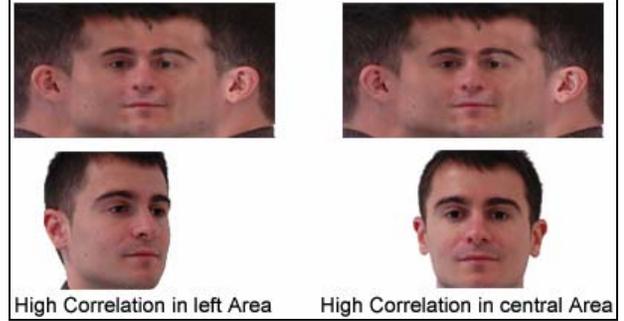


Fig 2 Comparing 2D pictures with a cylindrical representation of the subject

The procedure for extracting the information of a 2D picture is illustrated in Fig.3.

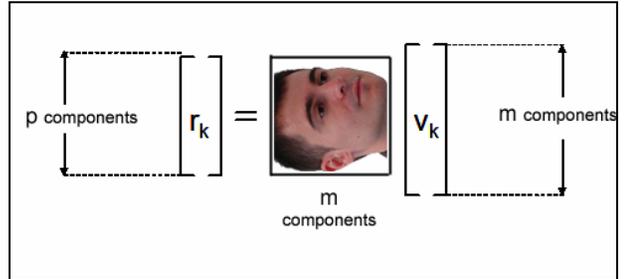


Fig 3 Projection of a partial 2D picture through the vector set \mathbf{v}_k (recognition stage)

In this case, the M vectors \mathbf{r}_k representing the 2D picture, have a reduced dimension p . However, it is expected that these p components will be highly correlated with a section of p components in the complete vectors \mathbf{r}_k^i computed during the training stage. Thus, the measure proposed below can be used to identify the partial available information (p components) through the vectors \mathbf{r}_k^i :

$$\min_{(i,j)} \left\{ \sum_{k=1}^M \sum_{l=1}^p (r_k(l) - r_k^i(l+j))^2 \right\} \quad (3)$$

$$i = 1, \dots, L; \quad j = 0, \dots, n-p$$

2.2. Obtaining the projection vectors

The set of vectors which maximize the projection of Eq. (1) may be obtained as the solution to the following optimization problem: Find \mathbf{v}_k for $k=1, \dots, M$ such that $\xi = \sum_k \sum_n (r_k^n)^T \cdot r_k^n$ is maximum, where \mathbf{r}_k^n is defined as the projection of image n through the vector \mathbf{v}_k and n

accounts for the number of images in the training set. The function to be maximized may be expressed as:

$$\xi = \sum_k \sum_n (A_n^T v_k)^T \cdot (A_n^T v_k) = \sum_k v_k^T \left(\sum_n A_n \cdot A_n^T \right) \cdot v_k$$

which states that the optimum projection vectors may be obtained as the eigenvectors associated to the M largest eigenvalues of the $m \times m$ positive definite matrix C_s

$$C_s = \sum_n A_n \cdot A_n^T$$

This vector set will be used for feature extraction and recognition from partial information:

$$\{v_1, \dots, v_M\}$$

The procedure for feature extraction from an intensity image A consists in projecting the transposed image through every eigenvector:

$$r_k = A^T \cdot v_k \quad k = 1, \dots, M$$

Therefore, a total of M feature vectors are available, with n (width) components each, for the image. The image has been compressed to a total of $n \times M$ scalars with M always being smaller than m .

When a complete image sample A ($m \times n$) is available, the recognition stage is straightforward. First, the projected vectors of the sample image are computed using the previous equation and then, the best match is found as the individual i whose representing vectors minimize the Euclidean distance using equation (2).

The procedure is quite different from conventional PCA. Certainly, in PCA a scalar number is obtained when the vector image is projected to one eigenvector, whereas in P²CA, an n -dimensional vector (r_k) is obtained, when the image (in matrix form) is projected to an eigenvector. It can seem that the P²CA approach demands more computational cost because it uses vectors instead of numbers to represent the projections. However, the number of eigenvectors $\{v_k\}$ needed in P²CA for an accurate representation is much lower than in PCA.

It should be mentioned that the mathematical theory behind this approach is similar to one recent method which has extended the conventional PCA method [1] from 1D to 2D; this technique was called 2DPCA [2].

2.3. Recognition of an individual from partial information

The most outstanding point of this procedure is that the image projected in the n -dimensional space does not need to have dimension $m \times n$ during the recognition stage so that partial information can be used. It is possible to use a reduced $p \times n$ ($p < m$) image which is projected to a smaller subspace. If only partial information is used, a classification method is needed to compare the partial

projection with the data in the whole space. In this case, it is not possible to use nearest neighbour classifier like in conventional PCA and correlation of partial difference methods like the criteria defined in (3) have to be applied.

3. DATABASE DESCRIPTION

3.1. UPC Face Database Description

The majority of available face databases are focused on frontal views; therefore, the proposed P²CA method has been tested on an own database which has been created at the *Universitat Politècnica de Catalunya* [4] with the main purpose of testing the robustness of face recognition algorithms against strong pose and illumination variations. This database includes a total of 18 persons with 27 pictures per person which correspond to different pose views (0° , $\pm 30^\circ$, $\pm 45^\circ$, $\pm 60^\circ$ and $\pm 90^\circ$) under three different illuminations (environment or natural light, strong light source from an angle of 45° , and finally an almost frontal mid-strong light source).

3.2. 180° Texture Training Images

Since the work presented here is focused on face recognition and face pose estimation and not on the 3D reconstruction, a manual and simple method has been used for the generation of the cylindrical 180° 3D texture images which will be used as 3D information for the training stage. From the 9 different views under a natural illumination only five of them have been selected (frontal, $\pm 45^\circ$ and $\pm 90^\circ$ views) and morphed together. Some examples of the training images are illustrated in Fig. 4. The eyes of the frontal image are used to fuse that view with the $\pm 45^\circ$ views, whereas the ears are the junction point for the combination of the $\pm 45^\circ$ and $\pm 90^\circ$ face images. One 180° 3D texture image is constructed for each person and used for the training phase together with their mirror versions.



Fig 4 Examples of a 180° texture training image (122x240)

3.3. Normalization of the Test Images

A total of 486 grayscale 2D images have been used for the recognition stage (9 views x 3 illuminations x 18 persons). A scale factor has been computed using the frontal view picture of each person and each illumination and has been applied to all non-frontal pictures. Fig. 5 shows some images used for testing the performance of P²CA.



Fig 5 Examples of test images (122x100)

4. EXPERIMENTAL RESULTS

4.1. Face Recognition using P²CA

Two different sets of test images have been used to verify the performance of the P²CA algorithm. The first one is composed of 162 face images taken under the same illumination conditions as the images used for the construction of the 180° training images presented in Fig.4. The second test ensemble is composed of all the 486 face images available in the database. The robustness against pose variations can be tested with the first set, whereas the second set is useful to check the accuracy of the approach against pose and illumination simultaneously.

Moreover, to have a better idea of the robustness of P²CA in front of other strategies, the conventional PCA has also been implemented and tested using the same set of training and recognition images. In this case, the training set is composed of the five images of each subject (and the mirror versions) which have been used in the creation of the 180° texture training images of the P²CA.

Method \ Set	Only Pose	Pose & illumination
PCA	72.22%	60.37%
P ² CA	92.59%	75.51%

Table 1 Recognition Accuracy of PCA and P²CA

4.2. Pose Estimation using P²CA

Since the technique makes use of 180° texture images for the training stage and it is pose-invariant, it can be used to estimate the pose of the face using the j^{th} position that minimizes equation (3). If this parameter is between some predefined ranges the face image is supposed to be taken from a specific angle. The algorithm can estimate between the 9 angles (0°, ±30°, ±45°, ±60° and ±90°) because the range of the parameter j of equation (3) has been divided in 9 different bins as shown in Table 2; but if more bins are defined, more accurate view estimation can be done.

View	Pos of j	View	Pos of j	View	Pos of j
-90°	0-50	-30°	75-85	+45°	125-135
-60°	50-65	0°	95-105	+60°	135-145
-45°	65-75	+30	115-125	+90°	150-240

Table 2 Position Range of parameter j in equation (3) using the 27 pictures of one person as training data

Fig. 6 represents the evolution for the j parameter of equation (3) for the person with the minimum reconstruction error (recognized ID) for the first 5 face images presented in Fig. 5. For example, the first frontal image has a minimum in $j=100$ (pink curve) which agrees with Table 2; or the -90° profile images present a minimum in $j=43$ approximately (black and red curves).

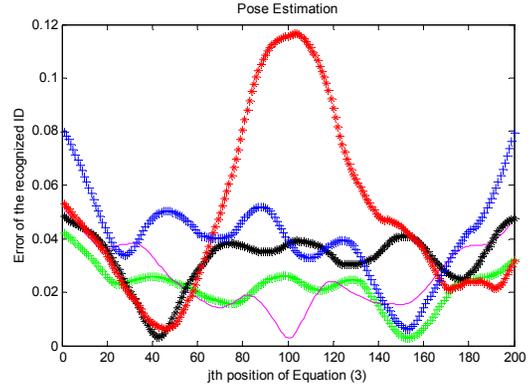


Fig 6 Pose Estimation using P²CA. The pink, red, black, green and blue curves correspond to the first 5 face images of the top line of Fig.5 respectively.

The angle view of the face of the 486 test images have been estimated as represented in Fig 6 by searching the minimum of the measure proposed in equation (3) with a *Pose Estimation Accuracy (PEA)* of the 96.15%.

5. CONCLUSIONS AND FUTURE WORK

This paper has presented a new 2D-3D face recognition scheme which shows robustness against pose variation. Thus, the method can also be used as a face pose estimator with a high reliability. That can be useful as a priori information for improving the recognition task. The main false recognition matches were due to strong illumination variations; thus, an equivalent *Partial Linear Discriminant Analysis* technique is currently under development. Another way to reduce the effects of illumination is to acquire the training data by more accurate methods like 3D laser scanners or trinocular camera sets using stereoscopic techniques to create complete 3D face models [3].

6. REFERENCES

- [1] M. A. Turk, A. P. Pentland, "Face recognition using eigenfaces", *Proc. of the IEEE CVPR* 1991
- [2] J. Yang, D. Zhang, A.F. Frangi, and J.Yang, "Two-Dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition", in *PAMI* 04
- [3] D. Onofrio, A. Sarti, and S. Tubaro, "Area Matching Based on Belief Propagation with Applications to Face Modeling", in *ICIP*, Singapore 2004
- [4] "UPC Face Database" in <http://gps-tsc.upc.es/GTAV>