

TWO IN ONE: JOINT POSE ESTIMATION AND FACE RECOGNITION WITH P²CA¹

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

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

Face recognition based on 3D techniques is a promising approach since it takes advantage of the additional information provided by the 3D world, i.e. information of depth and of all possible view angles of the face is available making the whole approach more robust against illumination and pose variations. However, these 3D approaches require the cooperation of the person to acquire accurate 3D data; thus, they are not appropriated for some applications like video surveillance or restricted area access points where only a 2D face image is disposable. In this paper, a novel approach is presented which takes advantage of 3D data in the training stage but only requires 2D data in the recognition stage. The proposed method can be used for both pose estimation and face recognition. Moreover, the estimation of the pose can be used as side information to improve the performance of the face recognition stage. The experiments have been performed on the public UPC face database which is composed of 23 persons with faces that vary from -90° to +90°.

1. INTRODUCTION

Although automatic recognition of human faces has been widely investigated by the research community, it is still an open and unresolved topic mainly because of three main challenges: Pose, illumination and expression variation. During the most of 30 years of research, improvements has been obtained step by step, and new and more robust approaches like PCA, LDA, LFA or EGM [1-4] have appeared. However, most of these techniques can only cope with small pose and illumination variations. Recently, some new techniques focused on 3D data have reported better results than traditional 2D algorithms [5,6]. This is due to the fact that depth is not dependent of illumination and pose. But the main problem is that the accuracy of 3D reconstruction algorithms meets two requirements: First, the cooperation

of the person that should be recognized and second, the calibration and synchronization of all the elements of the system. These requirements are not possible in most of the restricted area access and video surveillance applications where only an intensity 2D image of the face is available in the recognition stage. For this reason, in this paper the following problem is addressed: How can 2D and 3D information be integrated in a flexible face recognition system? A possible solution was presented in [7,8] where good results have demonstrated the robustness of Partial Principal Component Analysis (P²CA) towards pose variations from -90° to +90°. It was also shown that this method can be used as an accurate pose estimator. In this paper, a new scheme of this technique is presented that takes advantage of the pose estimator to improve the final recognition results.

The rest of the paper is organized as follows. In section 2, the fundamentals of the technique presented in [7,8] are reviewed, whereas in section 3 a new scheme that takes advantage of the pose estimator is proposed. The more salient current results are analyzed in the next section and finally, some conclusions and future work are presented in section 5.

2. FACE RECOGNITION AND POSE ESTIMATION USING PARTIAL PCA

2.1. P²CA fundamentals

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 used in the recognition stage. However, the method requires a cylindrical representation of the 3D face data for the training stage. In this paper, it is supposed that 180° cylindrical texture images, as the ones shown in Fig 1, are available for the training stage and they are considered as the 3D training data.

¹ ACKNOWLEDGMENT: 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.

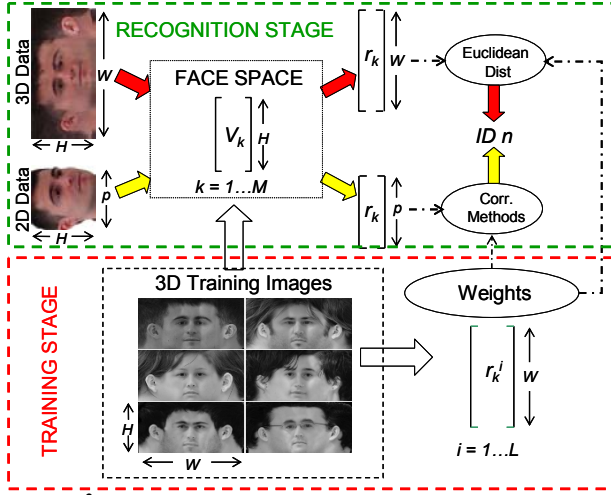


Fig 1 P²CA: General Block Diagram

The first step in a face recognition system is to characterize each identity of the database with an optimal set of features that should be as small as possible. P²CA reduces the dimensionality of the face images through the projection into a set of M optimal vectors v_k (face space) which are the eigenvectors of the Covariance Matrix of the training ensemble. Once the face space is created during the training stage, each individual i^{th} can be represented by the result of projecting the image data into the face space vectors:

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

where \mathbf{A}_i^T is the transpose of the 3D training 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^i averaged through the whole database (weights or coefficients used for the recognition).

Each vector \mathbf{r}_k^i has W components where W is the dimension of the 3D training image (A_i) in the horizontal direction (vertical when transposed). These vectors are the extracted features (weights in Fig 1) that will be stored in the system during the training stage and used later in the recognition stage. 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. A more detailed explanation of the mathematics related to P²CA can be found in [8,9].

2.2. Face Recognition using P²CA

In the recognition stage (upper box in Fig 1), two different cases can occur depending on the nature of the test face. If complete 3D data of the individual is available the recognition stage is quite obvious. 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 individual i that minimizes the Euclidean distance.

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 second situation depicted in Fig 1, where it is supposed that only one 2D picture of the individual is available. 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. Therefore, the measure proposed below can be used to identify the partial available information (p components) through the vectors \mathbf{r}_k^i :

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

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

2.3. Pose Estimation using P²CA

Since the technique makes use of 180° texture images for the training stage, it can also be used to estimate the pose of the face using the index j that minimizes Eq (2). The index j is coarsely quantized to 9 possible angles (0°, ±30°, ±45°, ±60° and ±90°) [7]. This procedure presented two main drawbacks: First, if the recognized ID (training image with the minimum error) was a false match, then possible error in the estimation of the pose would occur. Second and more important, the increase of the computational cost of the algorithm. Thus, it is advisable to use a new method that may use only one 3D training image in the pose estimation block. The proposed solution is to use the average face of the training ensemble. Section 4.3 explains the results for the pose estimation block.

3. POSE ESTIMATION AS A PREPROCESSING STEP TO FACE RECOGNITION

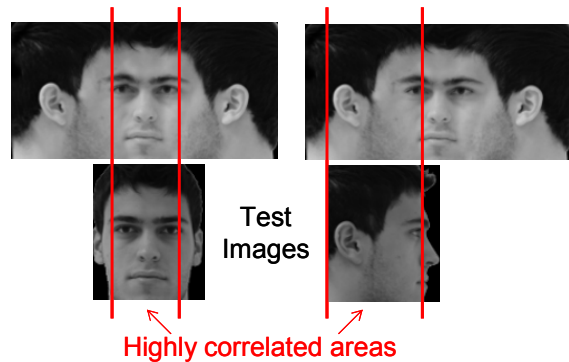


Fig 2 Useful part of the test face

In this paper, a new scheme is presented that takes advantage of the pose estimation for refining the recognition results. The main idea is illustrated in Fig 2. When normalizing the 2D images, there is part of the test face that is poorly correlated with the 180° training image (part outside the two lines). If this part is eliminated, the recognition may be better than the one obtained when using the complete face, since this part is not included in the training ensemble and it may lead to recognition errors.

Under this consideration, it may be possible to reduce the recognition errors if the pose is estimated in a first pre-processing step and depending on the result the most adequate part of the face is selected before performing the recognition stage as illustrated in Fig 3.

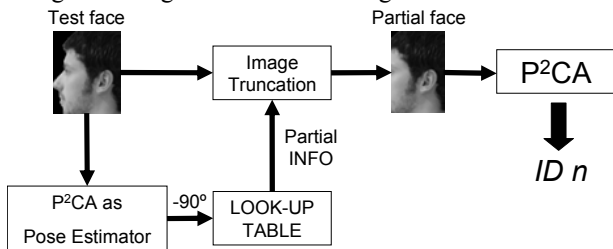


Fig 3 New Scheme using Pose Estimation as pre-processing step for refining the recognition results

First the test face is analyzed using P²CA in order to estimate between 9 different pose views as explained in [7]. Once the pose has been estimated, the result is used to determine through a Look-Up-Table which part of the face (only in the width direction) is the most suitable for the recognition stage. This Look-Up-Table has been empirically created by maximizing the recognition results of 5 identities of the UPC face database for each pose independently. Table 1 shows the partial information that maximized the recognition results for each pose.

Pose Angle	Partial INFO	Pose Angle	Partial INFO	Pose Angle	Partial INFO
-90°	0 - 50	-30°	0 - 70	+45°	20 - 90
-60°	10 - 40	0°	20 - 60	+60°	30 - 90
-45°	0 - 60	+30°	40 - 90	+90°	40-100

Table 1 Most suitable part of the face (0-100 will mean the complete face image).

Finally, after truncating the test image, P²CA is performed to recognize the identity of the person.

4. EXPERIMENTAL RESULTS

4.1. Description of the face images

The UPC face database [10] contains a total of 23 persons with 27 pictures per person which correspond to different pose views (0°, ±30°, ±45°, ±60° and ±90°) 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). The images have been normalized to an output resolution of 122x100 pixels. The 180° cylindrical training images has been created by manually morphing five images (0°, ±45° and ±90°) that have been acquired in a different session as the rest of the pictures under environmental light conditions.

4.2. Experiments description

From the 621 available test images, two different test sets have been created. The set A corresponds to the 207 images acquired under environmental illumination conditions and it will be used to test the robustness of the algorithm towards pose variations, whereas the test set B is composed of all the images of the database in order to analyze the pose and the illumination problem together.

Two different experiments have been carried out to evaluate and test the performance of the P²CA technique. First, the pose estimation block using the test set B is analyzed by comparing the pose estimation results using as reference: (a) the identity that has been recognized after performing P²CA, and (b) the average face image of the training ensemble. The second experiment compares in detail the previous face recognition results with the ones obtained using the new scheme depicted in Fig 3.

4.3. Pose Estimation Block

Table 2 shows the pose estimation results for each of the two different procedures mentioned before using test set B. It can be concluded, that using the average face image is the method that presents a better accuracy-complexity relationship, because although it does not reach the 93.07% pose estimation rate of (a), the computational cost is reduced a factor equal to the number of persons in the database. Hence, the average face image is used in the estimation block of the general diagram block of Fig 3 for the subsequent experiments.

	Recognized ID	Average face
Pose Accuracy	93.07%	92.96 %

Table 2 Comparison of the two different methods for estimating the pose

4.4. Recognition accuracy of the new scheme

Fig 4 represents the recognition rate as a function of the face space dimension. From the results, it can be concluded that in all the points of the curve the new scheme presents better results than the P²CA alone, improving the recognition in a 3% more or less. In the maximum of the curves, which correspond to a dimension of the face space of 21 eigenvectors, P²CA reveals a

88.89% recognition accuracy whereas P²CA with pose estimation as pre-processing step shows a 91.91% when utilizing test ensemble A. Both rates decrease when using the test set B that encloses three different illuminations, till a maximum of 72.22% and 78.71% respectively.

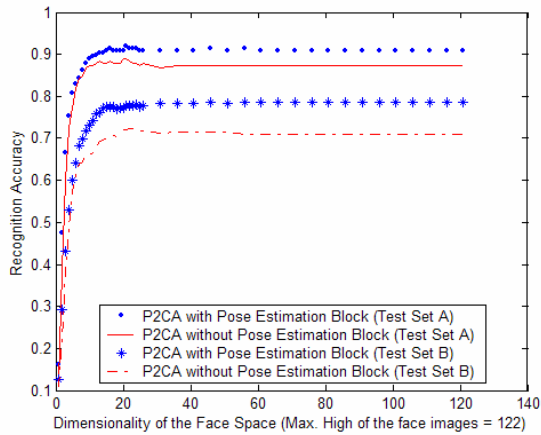


Fig 4 Recognition Rate Vs Number of Eigenvectors

Both schemes are further compared by testing each pose view separately, i.e. all test images of the ensemble B where organized depending on their pose. So, Fig 5 shows the recognition rate as a function of the pose view. The straight lines are the average recognition accuracy for both schemes. From the graphic, it can be concluded that the positive angles present lower recognition accuracy in the case of the new scheme in comparison with the negative angles. This can be explained since one third of the images of test ensemble B have been acquired under a strong light spot coming from +45°. Moreover, if the new results are compared with the previous ones, in the positive angles the recognition rate has not increase so much; in fact, in the +60° pose view the new scheme presents a lower recognition rate. Again, the explanation is that the hard illumination coming from +45° makes the estimation block fail in choosing the best part of the face.

5. FUTURE WORK

This paper improved the results of the 2D-3D face recognition scheme presented in [7,8] by introducing a pre-processing estimation block. Furthermore, a deeper analysis of the recognition accuracy has been presented. On one hand, the algorithm proves to be robust against big pose changes; but on the other hand, the main false recognition matches were due to strong illumination variations; Currently, a new multimodal P²CA face recognition system that integrates texture and depth data provided by a multicamera 3D reconstruction approach [11] is being developed in order to reduce the effects of illumination.

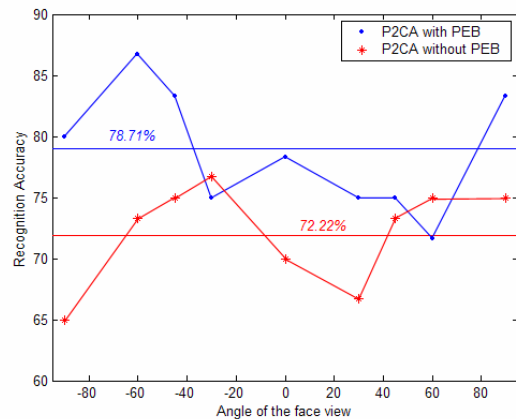


Fig 5 Recognition Rate Vs the 9 different pose views

6. REFERENCES

- [1] M. A. Turk, A. P. Pentland, "Face recognition using eigenfaces", Proc. of the IEEE Computer Society Conf. on CVPR, pp. 586-591, Maui, Hawaii 1991
- [2] R.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection" PAMI, vol 19, N. 7, July 1997
- [3] P. S Penev, and J.J. Atick, "Local Feature Analysis: a general statistical theory for object representation", Neural Systems, 7:477, 500, 1996
- [4] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, W. Konen. "Distortion Invariant Object Recognition in the Dynamic Link Architecture" IEEE Trans. on Computers archive Volume 42 , Issue 3 Pages: 300 – 311. 1993
- [5] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An Evaluation of Multimodal 2D+3D Face Biometrics", in IEEE Trans. on PAMI, Vol.27, pp 619-624, April 2005
- [6] C. Heshner, A. Srivastava, and G. Erlebacher, "A Novel Technique for Face Recognition using range imaging", in Proc. 7th Int. Symp. on Signal Processing and its applications 2003. Vol.2, pp 201-204
- [7] A.Rama, F.Tarrés, D.Onofrio, and S. Tubaro, "Using partial information for face recognition and pose estimation", in IEEE ICME, Amsterdam, July 6th-8th 2005
- [8] A.Rama, and F.Tarrés, "P²CA: A new face recognition scheme combining 2D and 3D information", in IEEE ICIP, Genoa, Italy, September 11th-14th 2005
- [9] J. Yang, D. Zhang, A.F. Frangi, and J.Yang, "Two-Dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition", PAMI, Jan. 2004
- [10] "UPC Face Database" in <http://gps-tsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm>
- [11] D. Onofrio, A. Sarti, and S. Tubaro, "Area Matching Based on Belief Propagation with Applications to Face Modeling", in ICIP, Singapore 2004