MIXED 2D-3D INFORMATION FOR POSE ESTIMATION AND FACE RECOGNITION[1]

Antonio Rama, Francesc Tarres**
{alarma, tarres}@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 depth which makes 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 such as 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. Experiments have been carried out on the public UPC face database which is composed of a total of 621 face images of several persons taken from different views and illuminations

1. INTRODUCTION

Although automatic recognition of human faces has been widely investigated, it is still an open 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 demonstrated the robustness of Partial Principal Component Analysis (P^2CA) towards pose variations from -90º to +90º. It was also shown that this method can be used as a 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 results are shown in the next section and finally, conclusions and future work are presented in section 5.

2. FACE RECOGNITION AND POSE ESTIMATION USING PARTIAL PCA

2.1. P^2CA fundamentals

The objective of P^2CA is to implement a mixed 2D-3D method, where either 2D (pictures or video frames) or 3D data (180º 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 180º cylindrical texture images, as the ones shown in Fig 1, are available for the training stage.

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. P^2CA 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

[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.

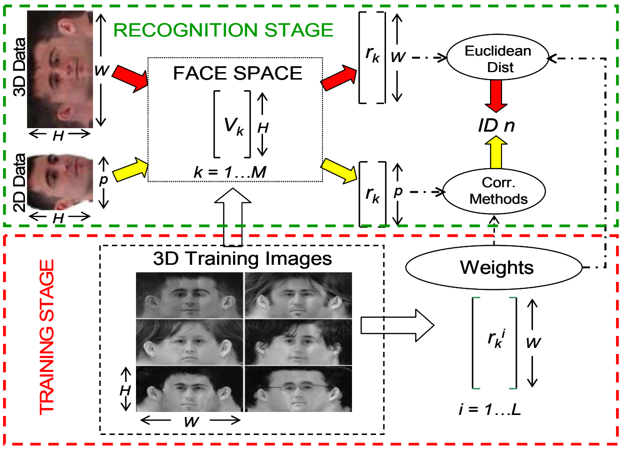

Fig.1 P²CA: General Block Diagram

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:

$$\mathbf{r}_k^{\,i} = \mathbf{A}_i^T \cdot \mathbf{v}_k \qquad k = 1,..,M \tag{1}$$

where A_i^T is the transpose of the 3D training texture image which represents individual i and $\boldsymbol{v}_k$ are the M optimal projection vectors that maximize the energy of the projected vectors $\boldsymbol{r}_k^i$ averaged through the whole database (weights or coefficients used for the recognition). Each vector $\boldsymbol{r}_k^i$ 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. Summarizing, each identity of the database will be represented by a WxM matrix of features. 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. The complete process is illustrated in Fig 1, and 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 $\boldsymbol{r}_k$. 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 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 $\boldsymbol{r}_k$ 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 $\boldsymbol{r}_k^i$ computed during the training stage. Therefore, the measure proposed below can be used to identify the partial available information (p components) through the vectors r_k^i:

$$\min_{(i,j)}\left\{\sum_{k=1}^{M}\sum_{l=1}^{p}\left(r_k(l)-r_k^{\,i}(l+j)\right)^2\right\}$$
$$i = 1,..,L; \quad j = 0,..,W-p \tag{2}$$

2.3. Pose Estimation using P²CA

Since the technique makes use of 180º 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) as illustrated in Fig 2. The example shows how index j (minimum distance between the features extracted for the test face and the ones corresponding to a reference image) is located in completely different ranges when using a frontal view or a profile image. Following this procedure, the index j is coarsely quantized in 9 ranges with are related to 9 possible angles (0º, ±30º, ±45º, ±60º and ±90º). In our previous work [7], the reference image used for estimating the pose was the one that presented the minimum error after performing face recognition. 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, since the objective of this work is to refine the recognition accuracy with the pose information, the estimation of the pose should require as low computational cost as possible. In the case of using the training image with the minimum error, then, the complete face recognition process will be performed twice. For this reason, one improvement with respect to the previous work is to use the

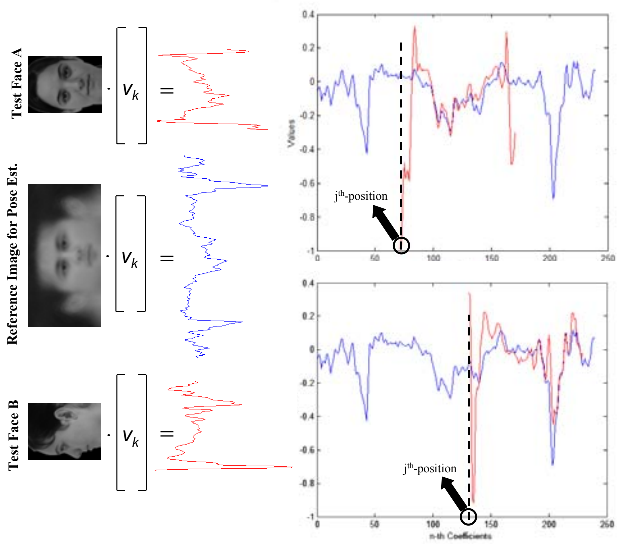

Fig.2 Pose Estimation using P²CA

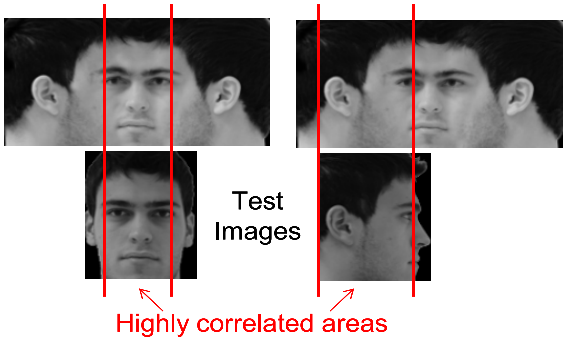

Fig. 3 Useful part of the test face

average face of the training ensemble as the reference image for estimating the pose (Fig.2). The experimental results, using the UPC FaceDatabase described in section 4.1, have shown that it is possible to estimate the angle view of the faces with 92.96% accuracy when using the average face.

3. POSE ESTIMATION AS A PREPROCESSING STEP TO FACE RECOGNITION

In this section an efficient scheme for joint pose and face recognition is presented. This procedure takes advantage of the pose estimation for refining the recognition results. The main idea is illustrated in Fig 3. When normalizing the 2D images, there is part of the test face that is poorly correlated with the 180º 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 is possible to reduce the recognition errors if the pose is estimated in a first pre-processing step and, as indicated in Fig 4, the most adequate part of the face is selected before performing the recognition stage.

During the training stage, P^2CA is performed on the training ensemble for creating the face space and for extracting the relevant features (weights) of each person of our database. During this process, the average 3D face image has been also calculated and saved as the reference image for the pose estimation block. After this training stage, the test face is projected into the face space using the P^2CA technique. The test features are correlated (Eq. 2) with the features of the average face image in order to estimate the pose view as depicted in Fig 2. 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. Note that Table 1 has been calculated taking into account that the resolution of the test face images is

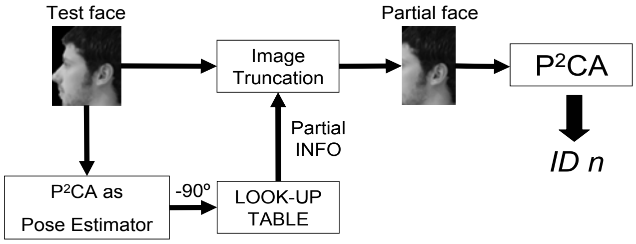

Fig. 4 New Scheme for the test stage using Pose Estimation as pre-processing step

122x100 pixels and it should be recalculated if the width of the test images changes. Finally, P^2CA is performed on the truncated face image (partial face in Fig 4) to recognize the identity of the person.

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

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

4. EXPERIMENTAL RESULTS

4.1. Description of the database

The UPC face database [10] contains a total of 621 images corresponding to 23 persons with 27 pictures per person acquired under different pose views (0º, ±30º, ±45º, ±60º and ±90º) and three different illuminations (environment or natural light, strong light source from an angle of 45º, and finally an almost frontal mid-strong light source). The images have been normalized to an output resolution of 122x100 pixels. The 180º cylindrical training images has been created by manually morphing five images (0º, ±45ºand ±90º) that have been acquired in a different session as the rest of the pictures under environmental light conditions.

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.

4.2. Recognition accuracy of the new scheme

Fig 5 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 P^2CA alone, improving the recognition in a 3% in the maximum of the curves, which correspond to a dimension of the face space of 21 eigenvectors, P^2CA

reveals a 88.89% recognition accuracy whereas P^2CA 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. 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 6 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 +45º. 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 +60º pose view the new scheme presents a lower recognition rate. Again, the explanation is that the hard illumination coming from +45º makes the estimation block fail in choosing the best part of the face.

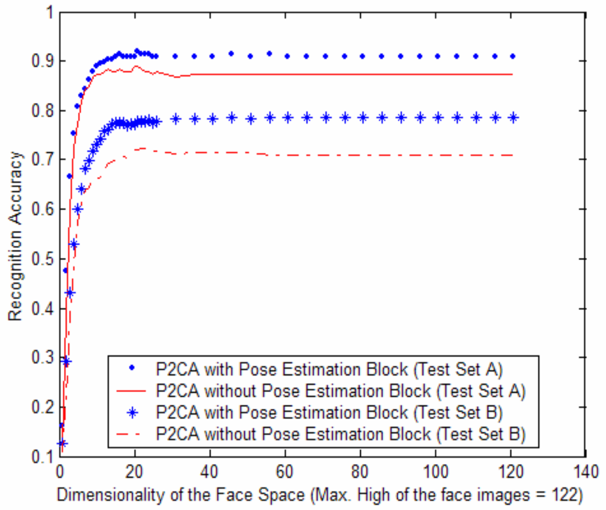

Fig.5 Recognition Rate Vs Number of Eigenvectors

5. CONCLUSIONS AND 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 P^2CA 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.

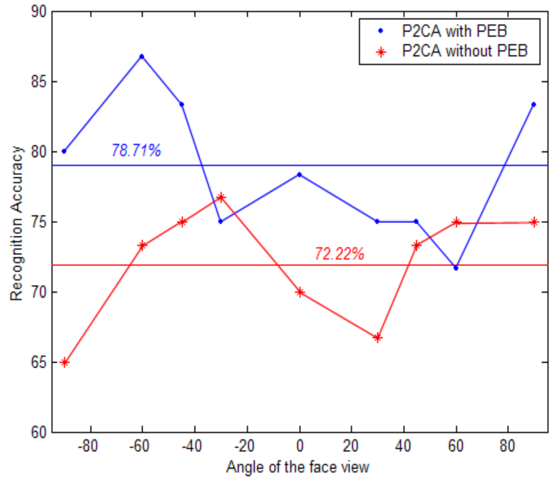

Fig. 6 Recognition Rate Vs the 9 different pose views

6. REFERENCES

[1] M. A. Turk, A. P. Pentland, "*Face recognition using eigenfaces*", Proc. of the IEEE Comp. Soc. Conf. on CVPR, pp. 586-591, 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. Malsburg, R.P. Wurtz, W. Konen. "*Distortion Invariant Object Recognition in the Dynamic Link Architecture*" IEEE Trans. on Comp. Vol. 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. Hesher, A. Srivastava, and G. Erlebacher, "*A Novel Technique for Face Recognition using range imaging*", in Proc. 7th Int. Symp. on SP 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, "*P^2CA: 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\GTAVDatabase.htm

[11] D. Onofrio, A. Sarti, and S. Tubaro, "*Area Matching Based on Belief Propagation with Applications to Face Modeling*", in *ICIP*, Singapore 2004