FACE RECOGNITION USING A FAST MODEL SYNTHESIS FROM A PROFILE AND A FRONTAL VIEW[1]

Antonio Rama, Francesc Tarrés

Technical University of Catalonia, Barcelona, Spain
{alrama, tarres}@gps.tsc.upc.edu

ABSTRACT

In our previous work we presented a new 2D-3D mixed face recognition scheme called Partial Principal Component Analysis (P^2CA) [1]. The main contribution of P^2CA is that it uses 3D data in the training stage but it accepts either 2D or 3D information in the recognition stage. We think that 2D-3D mixed approaches are the next step in face recognition research since most of surveillance or access control applications only dispose of a single camera which is used to acquire a single 2D texture image. Nevertheless, one of the main problems of our previous work was the enrollment of new persons in the database (*gallery set*) since a total of five different pictures are needed for getting the 180º texture maps (manual morphing). Thus, this work is focused on the automatic and fast creation of those 180º texture maps from only two images (frontal and profile views). Preliminary results show that there is not a significant degradation of the recognition accuracy when using this automatically and synthetically created gallery set instead of the one created by morphing the five views manually.

Index Terms— 2D+3D Face Recognition, mixed schemed, partial information, P^2CA, modeling

1. INTRODUCTION

Some of the new face recognition strategies tend to solve the face recognition problem from a 3D perspective [2-6]. The 3D information (depth and texture maps) corresponding to the surface of the face may be acquired using different alternatives: A multi camera system (stereoscopy), structured light, range cameras or 3D laser and scanner devices. The main advantage of using 3D data is: On the one hand, depth information does not depend on illumination; and on the other hand, complete (180º) texture maps may incorporate information from all possible views making the approach more robust towards pose variations. We will classify all these 3D approaches into two different philosophies: The first one would correspond to all 3D approaches that require the same data format in the training and in the test stage [2-4]. The second philosophy would enclose all approaches that take advantage of the 3D data during the training stage but then use 2D data in the recognition stage. It should be remarked that the *multimodal methods* presented in [2, 3] are enclosed in the first category. Although they can use only depth, only texture or a combination of both modalities, they need that if *frontal views* have been used during the training stage then a depth and/or intensity *frontal image* is also available in the recognition stage. Approaches of the first category report even better results [4] than of the second group; however, they present the main drawback that the acquisition conditions and elements of the test scenario should be well synchronized and controlled in order to acquire accurate 3D data. Thus, they are not suitable for surveillance applications or control access points where only one "normal" 2D texture image (from any view) acquired from a single camera is available.

The second category encloses model-based approaches [5, 6]. For instance, in [6] a 3D face model for each person of the database is constructed by integrating several 2.5D face scans from different viewpoints. The authors use the term of 2.5D scan to emphasize that the obtained range and color images correspond only to a part of the face and not the complete 180º representation (complete 3D model). During the recognition stage they fit the *test 2.5D scan* to each face model in the database using an Iterative Closest Point (ICP). Once they have matched the surface, they generate synthetic 2D texture images under the estimated pose view for the 30 persons of the database with the best surface matching. Finally, they create a 2D LDA face space with these synthetic training images where the texture data of the input 2.5 scan is projected. They treat texture and depth as two experts that they fuse their opinions to get the recognized ID. Vetter et al. [5] overcome the previous approach in the sense that during the recognition stage only a 2D texture image is used. A generic morphable 3D face model is built using 200 different 3D scans (depth plus

[1] ACKNOWLEDGMENT: The work presented was developed within VISNET II, a European Network of Excellence funded under the European Commission IST FP6 programme

texture). During the training stage, all the 3D scans are first aligned to a reference face F_o using an optic flow algorithm. Afterwards, each scan (mesh of 75,972 vertices) is parameterized to cylindrical coordinates first, and then to a Cartesian representation of the shape and texture ($\boldsymbol{S}_i$ and $\boldsymbol{T}_i$). Finally, those are used to construct the morphable face model; concretely, the model is composed of the texture eigenvectors ($\boldsymbol{t}_i$) and the shape eigenvectors ($\boldsymbol{s}_i$) computed when applying Principal Component Analysis to the $\boldsymbol{T}_i$ and $\boldsymbol{S}_i$ Cartesian representations. Given a certain 2D texture image I, the fitting procedure is based on the minimization of the reconstruction error (difference between the synthetic image I_{model} created by the morphable 3D model and the input image I). This fitting procedure depends on the following parameters: α_i (texture projection coefficients), β_i (depth projection coefficients) and a total of 22 different rendering parameters. This approach has been evaluated with 10 face recognition systems in the Face Recognition Vendor Test 2002 and for 9 out of 10 systems the 3D morphable model and fitting procedure improved performance on nonfrontal faces substantially.

Nevertheless, all model-based face recognition approaches present the main drawback of a high computational burden required to fit the images to the 3D models. In the first case [6] the fitting process of the surface of the 2.5D scan with each face model lasts 30 seconds whereas in [5] 4.5 minutes are needed to adjust the 2D color image to the generic morphable 3D face model. Both times are computed on a workstation with a Pentium IV 2GHz processor.

Model-based methods are more flexible and follow the concept of 2D-3D mixed face recognition schemes; i.e. they are trained with 3D data but then they used only partial information during the recognition (2D or 2.5D information). However, the main problem of those methods is the difficulty and computational complexity of adjusting the image to a model. Recently, we have presented a novel approach called *Partial Principal Component Analysis* (*P^2CA*) [1] which could be enclosed in the second category of the 2D+3D mixed schemes although is not a model-base approach. The main advantage in comparison with the model-based approaches is its low computational complexity since *P^2CA* does not require any fitting process. However, one of the main problems of our previous work was the enrollment of new persons in the database (gallery set) since a total of five different images are needed for getting the 180º texture map. Thus, this work is focused on the automatic creation of 180º texture maps from only two images (frontal and profile views).

The rest of the paper is organized as follows. In section 2 a proposed approach based on PCA and P^2CA for creating texture maps from one single image is formulated in detail. Section 3 describes how the previous algorithm can be extended in order to use two pictures instead of only one, whereas section 4 evaluates the usefulness of these automatic 180º texture images for face recognition. Finally, section 5 contains the conclusions together with the future research.

2. EXTENDED FACE SPACE

2.1. Objective of this work and previous considerations

P^2CA [1] is a novel face recognition approach based on 2DPCA that enables the projection and recognition of 2D texture images on a 3D data trained system. The main problem of our system is that the gallery set is composed of 180º cylindrical texture images as the ones presented in **Fig 1** which has been manually created by morphing a total of five different views of the person. The reference points used for aligning the five images have been the two eyes and the two ears. So, one of the main drawbacks of *P^2CA* is the necessity of these five different views images if a new person should be enrolled on the database. For this reason, in the next section a new proposal is presented in order to get automatically this kind of 180º texture images from one or two different image views.

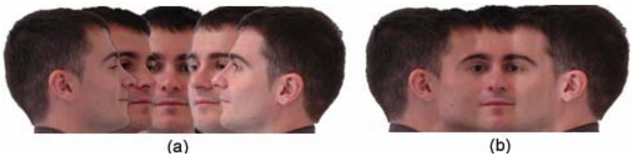

Fig 1 (a) Set of images used for the creation of the training data; (b) Example of a 180º texture training image

2.2. Proposed approach

The proposed approach is based on the creation of an extended face space $\boldsymbol{V_k}$ ($k=1...M$) by applying PCA to a total of 75 180º texture images like the one depicted in Fig 1 which correspond to 25 different persons under 3 different illuminations. The Eigenfaces obtained are the ones represented in Fig 2. We call it extended face space because the size of each *eigenface* is greater than the resolution of the images used in the recognition. Now our objective is to project conventional 2D images to the extended face space in order to create a 180º cylindrical representation that can be used in the gallery set. One feature of Eigenfaces is that they maintain spatial relationship although decorrelating the

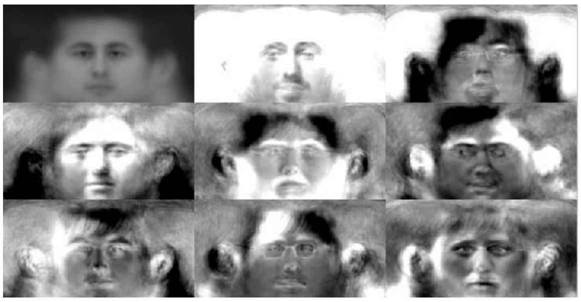

Fig 2 Mean face + 8 first Eigenfaces of the extended face space

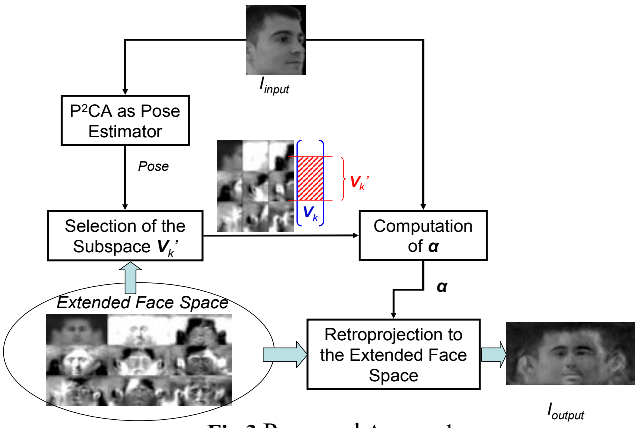

Fig 3 Proposed Approach

training data. This is the reason, why the Eigenfaces look like strange faces. Thus, the problem can be formulated as: Given a 2D image I_{input}, search the most suitable subspace (***V**ₖ**’***) from the extended face space (***V**ₖ*) so that the reconstruction error of the input image is minimized. In fact, this subspace (***V**ₖ**’***) can be easily computed if the pose of I_{input} is estimated since, as already stated, Eigenfaces retain spatial information. In [7] we have shown that P^{2}CA could be used also as a fast and robust pose estimator. So, we proposed the scheme depicted in **Fig 3** in order to get a 180º cylindrical training image from a 2D color image.

As shown in Fig 3, P^{2}CA estimates the pose as explained in [7] in order to determine which subspace $\boldsymbol{V_k}$***’*** should be selected from the extended face space. This block can be considered as a point registration process to fit the image to the model. Thus, the output of the “Selection of Subspace” block are *k* Eigenfaces with the same size as the resolution of the Input image I_{input} (stripped part of $\boldsymbol{V_k}$ in **Fig 3**). Once the subspace $\boldsymbol{V_k}$***’*** has been chosen, the next step is to compute the coefficients ***α*** that minimize the reconstruction error. The main problem now is that the input image can not directly be projected to the face subspace since there is not the certainty that this would minimize the reconstruction error. Instead, the following linear equation system has to be solved:

$$\alpha_1 \cdot \overline{v_1}' + \alpha_2 \cdot \overline{v_2}' + \ldots + \alpha_n \cdot \overline{v_n}' = \overline{I_{input}}$$

where $\overline{I_{input}}$ is the input image represented like a vector by concatenating each column. The solution is given by:

$\overline{\alpha} = \left[\overline{\overline{V'^T}} \cdot \overline{\overline{V'}}\right]^{-1} \cdot \overline{\overline{V'^T}} \cdot \overline{I_{input}}$ where $\overline{\overline{V'}} = \left[\overline{v_1}', \overline{v_2}', \ldots, \overline{v_n}'\right]$ is the matrix whose columns are the eigenvectors of the subspace. Finally, the computed coefficients ***α*** are ‘*retroprojected*’ to the extended face space so that the reconstructed 180º coordinate image (in vector form) is computed as: $I_{180^o} = \sum_{k=1}^{M} \alpha_k \cdot \overline{v_k}$

being *M*, the total number of Eigenfaces used for the reconstruction. **Fig 4** shows three different examples of 180º coordinate texture images created from a frontal, a lateral and a profile view respectively using the proposed method presented in **Fig 3**.

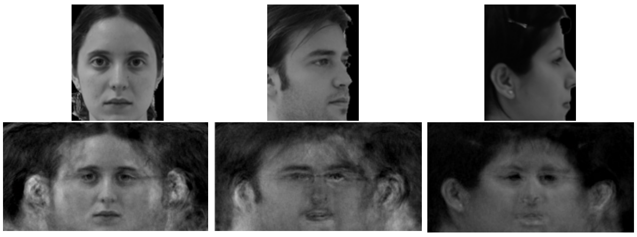

Fig 4 (Top) Input images. (Bottom) 180º Output

3. PROJECTING TWO IMAGES

After analyzing the results of **Fig 4** it can be observed that the resulting 180º texture images are very noisy, especially when using only one lateral view. Thus, the problem statement of Section 2 will be modified for using at least two images in the creation of I_{180^o} instead of one. Thus, the problem differs slightly: Now we have to compute the coefficients ***α*** that minimize the reconstruction error of both images at the same time:

$$\begin{aligned} e^2 &= \left\|\widetilde{I}_{input1} - I_{input1}\right\|^2 + \left\|\widetilde{I}_{input2} - I_{input2}\right\|^2 \\ &= \left\|\overline{\overline{V_1'^T}} \cdot \overline{\alpha} - I_{input1}\right\|^2 + \left\|\overline{\overline{V_2'^T}} \cdot \overline{\alpha} - I_{input2}\right\|^2 \end{aligned} \qquad \textbf{Eq 1}$$

where $\overline{\overline{V_1'^T}}$ and $\overline{\overline{V_2'^T}}$ are the subspace for *input image 1* and *input image 2* respectively. One possibility would be to fix those subspaces since the input images may always be a frontal and a profile view. However, by maintaining the Pose Estimator Block the system is less sensitive to small pose variations and it has the possibility of creating I_{180^o} from two views different from the frontal and profile ones.

If the quadratic error of Eq. 1 is minimized using the gradient, the coefficients can be obtained by the following expression:

$$\overline{\alpha} = \left[\overline{\overline{V_1'}} \cdot \overline{\overline{V_1'^T}} + \overline{\overline{V_2'}} \cdot \overline{\overline{V_2'^T}}\right]^{-1} \cdot \left[\overline{\overline{V_1'}} \cdot \overline{I_{input1}} + \overline{\overline{V_2'}} \cdot \overline{I_{input2}}\right]$$

Fig 5 shows the resulting 180º texture images obtained from the coefficients computed using the procedure explained above. The resulting pictures have improved their quality compared with the ones created when using only one single image. Nevertheless, the reconstructed pictures have some errors (noisy areas) due to the fact that the registration process (pose estimation block of Fig 3) can only cope with variations in the horizontal axis. For the frontal view images an eye detector based on Adaboost [8] can be used to correct rotations and vertical misalignments. However, the only constraint of this normalization (registration) process is that both eyes should be perfectly aligned at a fixed distance. Thus, errors in the registration of the profile view lead to noisy areas of the reconstructed 180º images.

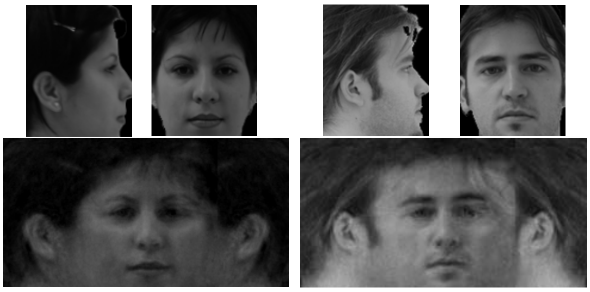

Fig 5 Creation of a 180º Output image using two views.

4. FACE RECOGNITION PERFORMANCE USING THE RECONSTRUCTED 180º IMAGES

Finally, the face recognition accuracy will be tested when using the 180º texture images created from two views by the proposed approach of Fig 3 as the gallery set.

The *training set* is composed of a total of 75 different 180º texture images as the one depicted in Fig 1 which correspond to a total of 25 subjects under 3 different illuminations (neutral one, a hard spotlight coming from 45º and a spotlight coming from the ceiling). This training set is used to compute the extended face space.

The *gallery set A* is composed of a total of 20 different 180º texture images which have been created using a frontal and a right profile view from 20 different persons using the method proposed in this work. These 20 identities are different from the persons included in the training set.

Gallery Set B is composed also of 20 180º texture images corresponding to the same 20 identities of *gallery set A*. The main difference is that now the images have been created by morphing five pictures as shown in Fig 1.

The *probe set* contains a total of 27 different pictures from each subject of the gallery set which correspond to 9 different pose views (0º, ±30º, ±45º, ±60º and ±90º) under the three different illuminations [9]. In order to make a comparative analysis P^2CA is chosen [1] as the face recognition method. The *training* and *probe set* are the same in both cases and only the gallery set changes (*A* or *B*).

Table 1 summarizes the face recognition accuracy when using P^2CA on both *gallery sets*. It shows a decreasing on the recognition accuracy when using *Gallery Set A*. This was something expected since as already mentioned the reconstructed 180º images are still noisy in some zones. So, the main false recognition errors were due to the profile or semi-profile views. Moreover, since a right profile view is used, there is the presence of more noise in the left side of the 180º texture image. Thus, more false recognition errors occur for left variations of the pose. A possible solution would be the use of two profiles (left and right) and a frontal image in the creation of the images or to implement a more robust registration procedure for the profile views. Nevertheless, results seem to be promising since our final objective is the face recognition performance and not the creation of 3D real-looking face models. **Table 1** shows only slight degradation (less than 3.5%) between recognition rates using both galleries.

	Recognition Rate (P^2CA)
Gallery Set A	74.7%
Gallery Set B	78%

Table 1 Comparative results for both Gallery Sets

5. CONCLUSIONS AND FUTURE WORK

After analyzing the first results, the method proposed here could be foreseen as a simplification of the morphable model of Vetter&Blanz [5] since only the texture face space is computed. Nevertheless, no complex fitting procedure is necessary like in [5,6] and only a fast pose estimator block based on P^2CA [1,7] is used to select the most suitable subspace from the complete extended face space. However, the results of the computed images present still some error due to registration problems of the profile view; and this issue introduces some degradation in the face recognition stage. The next step would be making a more precise alignment of the profile views (not only eyes but also other features like ears), computing the extended face space with more training images (75 from 25 persons is very limited if it is compared with the 200 3D scans of 200 different persons in [5]) and also adding depth information to create not only an extended face space for texture but also for shape. Additionally, a deeper analysis on how pose estimation errors influence the final 180º images should be made.

6. REFERENCES

[1] A. Rama, and F. Tarrés, “P^2CA: A new face recognition scheme combining 2D and 3D information”, *in IEEE International Conference on Image Processing*, Genoa, Italy, September 2005

[2] Y. Wang, C. Chua, and Y. Ho, “Facial Feature Detection and Face Recognition from 2D and 3D images”, *Pattern Recognition Letters*, Vol 23, pp 1191-1202, 2002

[3] K. I. Chang, K. W. Bowyer, and P. J. Flynn, “An Evaluation of Multimodal 2D+3D Face Biometrics”, *in IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.27, April 2005

[4] A. M. Bronstein, M. M. Bronstein, R. Kimmel, “Three-dimensional face recognition „ *in International Journal of Computer Vision* Vol.64/1, pp. 5-30, August 2005

[5] V. Blanz, and T. Vetter, “Face Recognition based on fitting 3D morphable model”, *in IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(9):1063-1074, 2003

[6] X. Lu, and A.K. Jain, “Integrating Range and Texture Information for 3D Face Recognition”, *in Proc. IEEE WACV*, Breckenridge, Colorado 2005

[7] A.Rama, F.Tarres, D.Onofrio, and S.Tubaro, “Mixed 2D-3D Information for pose estimation and face recognition” *in IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, May 14th- 19th, 2006

[8] Viola P., Jones M.: *Rapid Object Detection using a Boosted Cascade of Simple Features,* Computer Vision and Pattern Recognition, 2001

[9] “UPC Face Database” in http://gps-tsc.upc.es\GTAV\ResearchAreas\GTAVDatabase.htm