Partial PCA in Frequency Domain

Antonio Rama[1], Jürgen Rurainsky[2], Francesc Tarrés[1], Peter Eisert[2]
[1] Technical University of Catalonia (UPC)
[2] Fraunhofer HHI
E-mail: {*tonirama, tarres}@gps.tsc.upc.edu*
E-mail: {*rurainsky, eisert}@hhi.fraunhofer.de*

Abstract – *Partial Principal Component Analysis has demonstrated to be a robust face recognition approach a also for big pose variations. The main idea behing P^2CA is to use 3D (2.5D) data during the training stage but then it can use a single 2D image in the recognition stage, in order to be useful for pose variations. Nevertheles this approach present two major drawbacks: First one is that the training data should be well aligned for obtaining good results; the second inconvenient is computational cost since the face space is computed maintining the spatial dimensionality of the images. Thus, in this paper, we address both problems. First we propose a two step procedure for a more accurate alignment of the 2.5D training texture maps. This alignment is done by applying a global and local affine transformation to all texture maps, while the local alignment is aplied to some facial feature using a triangulation mesh. The computational cost is also reduced by applying P^2CA in the frequency domain and perform the normalized cross correlation as similarity measure. Introducing these two improvements the recognition accuracy increases a 4% due to the alignment procedure and the computational time is reduced between 10 and 100 times (depending on the number of eigenvectors used) due to the application of P^2CA in the frequency domain.*

Keywords – *Partial Principal Component Analysis, Face Recognition, face feature alignment, texture maps*

1. INTRODUCTION

Face recognition based on 3D techniques is a promising approach since it takes advantage of textured shape data [1]. Using these kind of depth and 180º multi-view texture maps is supposed to increase the robustness towards the two main challenges in Face Recognition: Pose and illumination robustness. Nevertheless, the robustness of 3D approaches is directly related with the accuracy of the acquired data. This issue represents the main drawback of 3D face research since 3D data should be acquired under highly controlled conditions and in most cases depends on the collaboration of the subject to be recognized. Thus, in applications such as surveillance or control access points, this kind of 3D data may not be available during the recognition process. This leads to a new paradigm using some mixed 2D-3D face recognition systems where 3D data is used in the training but either 2D or 3D information can be used in the recognition depending on the scenario. Following this concept, we have presented a novel technique called Partial Principal Component Analysis (P^2CA) [2]. P^2CA is a face recognition approach based on 2DPCA [3] that intends to cope with big pose variations. This approach makes use of 180º cylindrical texture maps for training the system but then only images acquired from a single, normal camera are used for the recognition. This kind of texture training maps (2.5D data) provides pose information from different views. One problem of this approach is that the cylindrical texture images used for training this FR system have been globally aligned by considering the position of the eyes, and thus, the face space created is a bit noisy [4]. The other problem is that P^2CA maintains the spatial relationship of the image after the dimensionality reduction; thus, when an image is projected to the face space, the number of coefficients is large compared with e.g. conventional Eigenfaces [5]. Additionally, the similarity measure proposed in [2] is based on correlation methods leading to an important increase of the computational time in the identification process. Thus, the two main contributions presented in this paper are: (a) An accurate approach based on wire grid models for the alignment of the 180º texture images; (b) a reformulation of the P^2CA method in the frequency domain to speed up the comparison between the test and the gallery images. This reduction of the computational time allows us to introduce the Normalized Cross Correlation (NCC) measure for performing the recognition instead of using the minimum differences measure [2].

The rest of the paper is organized as follows. In section 2, a brief overview of the P^2CA technique is presented remarking the two drawbacks that are addressed in this paper. Section 3 is divided into two parts: One describes the proposed alignment method with some results, whereas the other explains the formulation of P^2CA in the frequency domain. Section 4 covers the experiments performed and the obtained improvements in the recognition rate and in the computational time. Finally, section 5 contains the conclusions and future research.

2. OVERVIEW OF P²CA APPROACH

The objective of P²CA is to implement a mixed 2D-2.5D method (2D-3D when using texture and depth maps [6]), where either 2D (pictures or video frames) or 2.5D data (180º texture images in cylindrical coordinates) can be used in the recognition stage. However, the method requires a cylindrical representation of the 2.5D face data for the training stage. Like in the majority of face recognition methods, in P²CA the dimensionality of the face images is reduced through the projection into a set of M optimal vectors which composed the so called *feature space* or *face space*. The vectors representing the i^{h} individual are obtained as:

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

where A_i^T is the HxW image representing individual i transposed, 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. These vectors could be interpreted as unique signatures that identify each person. The projection described in equation (1) is depicted in Fig.1. The (training) texture map of the subject i is represented by the M vectors $\mathbf{r}_k^i$. Each vector $\mathbf{r}_k^i$ has W components where W is the width of the matrix A_i.

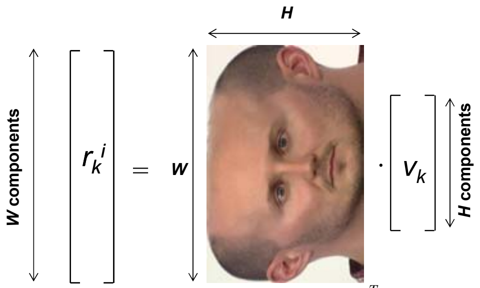

Fig. 1. Description of a texture map (A_i^T) by means of projection vectors using P²CA (training stage)

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. In this case, the M vectors $\boldsymbol{r}_k$ representing the 2D picture, have a reduced dimension W'. However, it is expected that these W' components will be highly correlated with a section of W' components in the complete vectors $\boldsymbol{r}_k^i$ computed during the training stage.

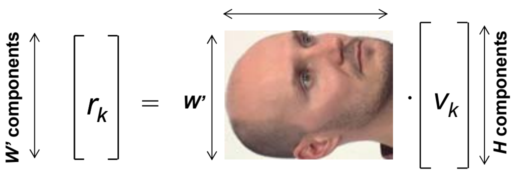

Fig. 2. Projection of a "partial" 2D image through the vector set $\mathbf{v}_k$ (recognition stage)

Therefore, the measure proposed below has been used to identify the partial available information (W' components) through the vectors $\boldsymbol{r}_k^i$ [2]:

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

with L being the total number of gallery images (subjects) of the database.

The Square Differences (SD) similarity measure of equation (2) has been proposed due to its relatively low computational cost. However, other measures have been used with better results for pattern matching like Normalized Cross Correlation (NCC) but at the expense of a higher computational cost:

$$\max_{(i,j)}\left\{\frac{\sum_{k=1}^{M}\sum_{l=1}^{W'}\left(r_k(l)\cdot r_k^{\,i}(l+j)\right)}{\sqrt{\sum_{k=1}^{M}\sum_{l=1}^{W'} r_k(l)^2 \cdot \sum_{k=1}^{M}\sum_{l=1}^{W'} r_k^{\,i}(l+j)^2}}\right\} \tag{3}$$
$$i = 1,..,L; \quad j = 0,..,W - W'$$

In section 3.1, the computation of P²CA in the frequency domain is formulated allowing us the introduction of the NCC for identification efficiently.

The other limitation of the method is that the training image (Fig.1) should be well aligned when computing the face space [2, 4]. If this is not the case, the correlation between test and the gallery images will be reduced after projecting them to the face space leading to a reduction of the face recognition accuracy. This is especially important for the relevant facial features since these are the key points used for the creation of the training texture maps. In section 3.2, a process for a more accurate alignment of these texture maps is presented.

3. IMPROVEMENTS IN P²CA APPROACH

3.1. P²CA in the frequency domain

In [8] the authors presented an efficient way of computing NCC (equation (3)) in the frequency domain. Based on this work [7] a new formulation of P²CA is presented.

The numerator of the NCC in equation (3) can be expressed again as:

$$c(i,j) = \sum_{k=1}^{M}\sum_{l=-\frac{(W'-1)}{2}}^{\frac{(W'-1)}{2}}\left(r_k(l)\cdot r_k^{\,i}(l+j)\right) \quad i = 1,..,L; \quad j = -\frac{W-W'}{2},..,\frac{W-W'}{2}$$

where for convenience we have accepted a change in the vector index l, choosing the zero coordinate in the center. In this case, we have to correlate vectors of L components with vectors of W' components in ½(W-W') lags. This condition can be implemented very efficiently in the frequency domain. So taking the Discrete Fourier Transform (DFT) of the inner summatory of previous equation:

$$S(u) = R_k(u)\cdot R^{i}{}_{k}{}^{*}(u) \tag{4}$$

The correlation between $\boldsymbol{r}_k$ and $\boldsymbol{r}_k^i$ has a total of W+W'-1 lags from which only W-W'-1 samples are interesting for the computation. Thus, we have to avoid that these W-W'-1 central samples do not present any time aliasing. For this reason it is necessary to compute the W-points-DFT of r_k^i and the W-points-DFT of $r^{zp}{}_k$ which consists on the zero padding version of r_k. Now taking equation (1), $R_k(u)$ can be expressed as follow:

$$R_k(u) = DFT\left[\mathbf{r}_k^{\,i}\right] = DFT\left[\mathbf{A}_i^{\,T} \cdot \mathbf{v}_k\right] = DFT2D\left[\mathbf{A}_i^{\,T}\right] \cdot \mathbf{v}_k' \qquad (5)$$

where $\mathbf{v}_k$' are the eigenvectors that minimize the energy projection of a given training set after applying two dimensional DFT. The demonstration of this statement have already been proved for the 1D case in [9] but the same parallelism can be followed if the images (A_i) are treated like matrixes.

The test stage of the P^{2}CA in frequency domain will be summarized as follow:

- Given an image A_{test}, normalize this image (same as P^{2}CA in spatial domain [2]).
- Extend A_{test} with columns of zeros until we have $A^{zp}{}_{test}$ with W columns (same width as the training texture maps).
- Compute the DFT-2D of $A^{zp}{}_{test}$ and project the result to the face space $\mathbf{v}_k$' obtained during the training stage.
- Obtain the product between the frequency domain test coeffients and the frequency domain weights of each identity (equation (4)).
- Compute the IDFTof (4) for the M different coefficients vectors where only the W-W'-1 central samples should be considered because these are not affected by the time aliasing.
- Identify the identity of the database that gets a max value of the sum of the M IDFT vectors.

3.2. Local Alignment method

The importance of local aligned face features for recognition is described by Tsapatsoulis *et al.* [8]. There the local alignment is described as resizing in contrast to a complete affine transformation. In most of face recognition systems, the eye centers are used for the alignment of the images before analyzing them whereas other face features are usually not received attention although it has been demonstrated that these are as important as eyes centers for performing recognition [10]. In this paper, a two step alignment approach is proposed. First, a global alignment is applied to the training texture maps, which is based on 2D affine transformation. The parameters for the transformation are determined by using a reduced number of manually selected face feature locations. The average feature point location is used as reference and all transformations are calculated with respect to these data.

In order to achieve a better alignment result, the regions of the selected face features are aligned locally. The generic triangle mesh of Fig. 3 is adapted and placed at the global transformed face feature locations and the associated texture information for each triangle is extracted. The face feature locations are transformed to the desired position and the associated textured triangles will perform the texture map bilinear interpolation.

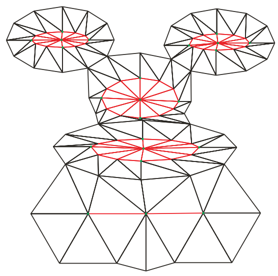

Fig.3. Adapted triangle mesh for the local alignment of selected face features (both eyes, nose, mouth and chin).

Using this approach leads to a better alignment of the local features as depicted in Fig. 4. Fig. 4a) corresponds to all the marked points on all training texture maps (only frontal features). Fig. 4b) represents the same points after applying to each image the 2D affine global transformation. And finally, Fig. 4c) are all the facial feature points after the local affine transformation for each feature, therefore the feature points shown in Fig. 4c do not match in one center. Results show that in each step the facial features are always more grouped. A more visual comparison between just global or global plus local aligned texture maps is given in Fig.5.

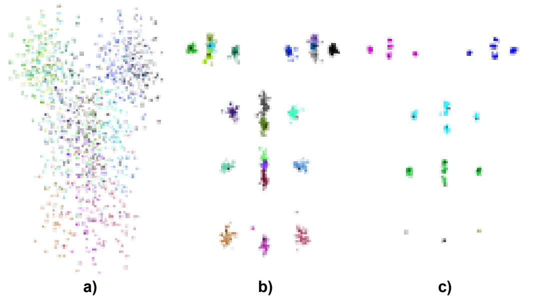

Fig. 4. Facial Feature localizations of the facial images a) before any transformation, b) after global transformation, and c) after local transformation using the wired mesh

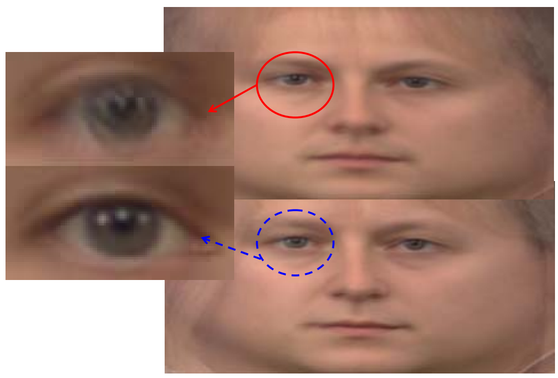

Fig. 5. Average image. Top: after global alignment. Bottom: after global plus local alignment.

4. EXPERIMENTAL RESULTS

4.1. Dataset and Experiment Description

Two different databases are used for obtaining the experimental results. The first one is the HHI face database which is composed of 10 different subjects. The HHI database contains one 180º texture map for each subject that have been aligned using only global alignment; and another texture map using the global and local alignment process described in Section 3.2. This texture maps will be used as the training and gallery sets. The test set is composed of a total of 9 different views (0º, ±6º, ±16º, ±25º and ±37º) for each subject (90 test images) acquired in a second session.

The second database is the UPC database [11]. This database includes a test set of 30 persons with 9 pictures per person which correspond to different pose views (0º, ±30º, ±45º, ±60º and ±90º). Furthermore also a total of 30 different 180º texture maps have been created [2], aligned and used as the training and gallery ensemble.

The first database is used for testing the improvement of the recognition rate when using the alignment images with the proposed approach of section 3.2, whereas the second one is used for testing the computational time when performing the P^2CA approach in the frequency domain. Finally, both databases are used for comparing the results using the SD and the NCC measures.

4.2. Experimental Results

Table 1 summarizes the recognition accuracy for the two correlation methods proposed in Equations (1) and (2). From the results it is clear that using NCC improves the recognition rate results since this measure is more robust towards slight changes in illumination. For the UPC database the improvement is more visible since 10 out of the 30 identities have been enrolled on the database in a different session with slightly different illumination conditions.

Table 1. Recognition Accuracy

Dataset	SD	NCC
HHI	94.44%	96.66%
UPC	81.85%	89.63%

Table 2 presents the results for face recognition when using the alignment process described above. Although only 10 different persons are used in the experiments, results show that the local aligned images present a slight improvement in the recognition rate. The improvement of this rate has been obtained for the 0º and ±6º views since these enclose all the face features used for the alignment.

Table 2. Face Recognition Results using NCC

Dataset	FR (global alignment)	FR (global + local alignment)
HHI	93.33%	96.66%

Finally the computational time is analyzed when using the P^2CA approach in the spatial and in the frequency domain. Simulations have been run in MATLAB using a 2.0 GHz µP with 1GB of RAM. Table 3 shows the computational time for P^2CA in Spatial and in Frequency Domain depending on the total number of eigenvectors used for computing the face space. This time is computing for matching one image to the 30 enrolled persons. Results illustrate the importance of performing P^2CA in FD for higher dimensions since the computational time factor reduction is between 70 and 150.

Table 3. Computational Time for NCC measure

dim	P^2CA in SD	P^2CA in FD	factor
1	0.14 sec	0.011 sec	12
20	0.96 sec	0.042 sec	22
60	8.15 sec	0.114 sec	71
122	32.9 sec	0.222 sec	150

5. CONCLUSION AND FUTURE WORK

In this paper, two improvements for the P^2CA approach have been proposed: Firstly, a local alignment method of the training images, and secondly, a reformulation of the complete approach in the frequency domain. Both improvements lead to an increase in the recognition accuracy and a reduction of the computational time.

P^2CA in FD introduce new possibilities that should be analyzed: First, the dimension may be more reduced as in the spatial domain since the frequency features for computing the face space are more compacted; and second, illumination variations may be mitigated by eliminating some frequency components.

ACKNOWLEDGEMENT

The work presented was developed within VISNET II, a European Network Of Excellence funded under the EC IST FP6 programme.

REFERENCES

[1] K. Bowyer, K. Chang, and P. Flynn, "A Survey of Approaches to 3D and Multi-Modal 3D+2D Face Recognition," *IEEE ICPR*, 2004.

[2] A. Rama, and F. Tarrés, "P^2CA: A new face recognition scheme combining 2D and 3D information", *in IEEE ICIP*, Genoa, Italy, 2005

[3] J. Yang, D. Zhang, A.F. Frangi, and J.Yang, "Two-Dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition", *in IEEE PAMI*, Jan. 2004

[4] A. Rama, F. Tarrés, "Face Recognition Using A Fast Model Synthesis from A Profile and A Frontal View", *IEEE ICIP*, San Antonio, Texas, USA, 2007

[5] M. A. Turk, A. P. Pentland, "Face recognition using eigenfaces", *Proc. of the IEEE Comp. Soc. Conf. on CVPR*, pp. 586-591, Hawaii 1991

[6] D. Onofrio, A. Rama, F. Tarres, S. Tubaro, "P^2CA: How Much Information is needed", *IEEE ICIP*, Atlanta, USA, October 2006

[7] J.P. Lewis, "Fast Normalized Cross-Correlation," *Vision Interface,* 1995

[8] N. Tsapatsoulis, N. Doulamis, A. Doulamis, and S.Kollias, "Face extraction from non-uniform background and recognition in compressed domain" *IEEE ICASSP*, Seattle, WA, USA, May 1998

[9] M. Savvides, B. V. Kumar, and P. K. Khosla, "Eigenphases vs. Eigenfaces" *Int. Conf. On Pattern Recognition,* Washington DC, 2004

[10] A. Kouzani, and K. Sammut, "Quadtree principal component analysis and is application to facial expression classification" *in IEEE Int. Conf. on Systems, Man, and Cybernetics*, October 1999

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