P^2CA:
HOW MUCH FACE INFORMATION IS NEEDED?

Davide Onofrio, Antonio Rama+, Francesc Tarres+, Stefano Tubaro**
d.onofrio@elet.polimi.it, alrama@gps.tsc.upc.es, tarres@gps.tsc.upc.es, tubaro@elet.polimi.it

*Dipartimento di Elettronica e Informazione - Politecnico di Milano
+Department Teoria del Senyal i Comunicacions de la Universitat Politècnica de Catalunya

ABSTRACT[1]

Multimodal 2D+3D face biometrics commonly report that performance improves relative to that of a single modality. Complete 2D and 3D data can be available during training because they are acquired in a controlled scenario. However, in the evaluation scenario, only partial 2D and 3D data can be acquired and hence available for recognition. In this paper we present experimental results that determine how partial data contribute to the task of recognition using Partial Principal Component Analysis (P^2CA) algorithm in a multimodal scheme. From our results it seems that discrimination power on individuals is ascribed to different regions of the face if we consider 2D or 3D data.

1. INTRODUCTION

Most face recognition systems are based on the evaluation of 2D-intensity or colour images. Since the extraction of reliable features from 2D-images is difficult and is subjected to a variety of possible interpretation errors, due for example to changes in illumination conditions, the recognition accuracy of such systems is limited [5]. The use of additional 3D information improves the reliability of the recognition scheme because of the relative independence from illumination and head rotation. Various multimodal 2D+3D face recognition schemes are proposed in [2], [8], (for additional survey details on 2D+3D and 3D face recognition, see [1]).

[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.

In all of these 2D+3D studies the multimodal approaches are shown to outperform either mode alone [3].
The main disadvantage of using 3D data is that acquisition is related to a controlled scenario. These controlled conditions can be available in the training stage; however most of the surveillance and control access applications present uncontrolled scenarios where only partial information both on 2D picture of the subject and 3D data are available. This concept of comparing partial 2D and 3D data in the recognition stage, with complete information acquired in the training stage, leads us to another question: how partial 2D and 3D data contribute to the final recognition result?
It is clear that the answer will depend on many factors and on the recognition algorithm used. In this paper we present the recognition results based on the Partial Principal Component Analysis (P^2CA) [6][7]. Here the objective of using P^2CA is to implement a mixed 2D-3D method, where both partial 2D (texture) and partial 3D data (depth map) can be used in the recognition stage.
The remaining of the paper is organized as follows: Section 2 describes the main idea of P^2CA; Section 3 illustrates texture and depth data used in our experiments by the algorithm; Section 4 presents the experiments made with the P^2CA and results are illustrated in Section 5; Finally, Section 6 outlines the conclusions.

2. P^2CA: DESCRIPTION OF THE ALGORITHM

Like in the Principal Component Analysis (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 follows

$$\mathbf{r}_\mathbf{k}^\mathbf{i} = \mathbf{A}_\mathbf{i}^\mathbf{T} \cdot \mathbf{v}_\mathbf{k} \qquad k = 1,\ldots, M \tag{1}$$

where $\mathbf{A}_\mathbf{i}^\mathbf{T}$ is the transpose of the texture (depth) image representing individual $\boldsymbol{i}$ and $\boldsymbol{v_k}$ are the M optimal

projection vectors that maximize the energy of the projected vectors $\boldsymbol{r_k}$ 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. Clearly, the optimization problem for finding the projection vectors has to be reformulated and this will be discussed later. The projection described in Eq.(1) is depicted in Figure 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).

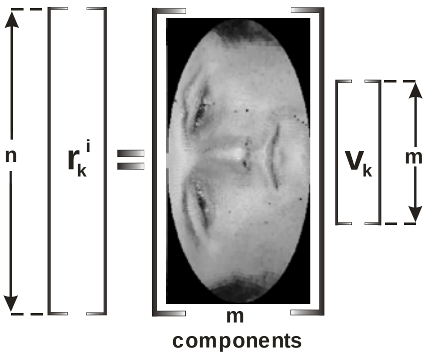

Figure 1. Comparing 2D pictures with a spherical representation of the subject

If complete 3D data of the individual is available the recognition stage is straightforward. In fact, it is only necessary to convert the 3D data to spherical coordinates and compute the resulting M vectors $\boldsymbol{r_k}$. The best match is found for the individual i that minimizes the Euclidean distance:

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

where L represents the number of people in the database.

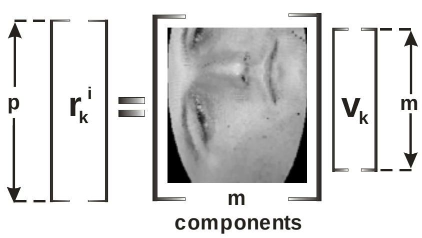

Figure 2. Projection of a partial 2D picture through the vector set $\mathbf{v_k}$ (recognition stage)

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 Figure 2, where it is supposed that only a partial picture of the individual is available.
This partial picture shows a high correlation with the corresponding area of the spherical representation of the 3D image. 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 computed during the training stage. Thus, the measure proposed below can be used to identify the partial available information (p components) through the vectors:

$$\min_{(i,j)} \left\{ \sum_{k=1}^{M} \sum_{l=1}^{p} \left(r_k(l) - r_k^i(l+j) \right)^2 \right\}$$
$$i = 1,\ldots,L \qquad j = 0,\ldots,n-p \qquad (3)$$

The set of vectors which maximize the projection of Eq. (2) may be obtained as the eigenvectors associated to the M largest eigenvalues of the mxm positive definite matrix $\boldsymbol{C_s}$

$$\mathbf{C_s} = \sum_{\mathbf{n}} \mathbf{A_n} \cdot \mathbf{A_n^T}$$

This vector set will be used for feature extraction and recognition from partial information: $\{\mathbf{v_1},\ldots,\mathbf{v_M}\}$.
The procedure for feature extraction from an intensity image $\boldsymbol{A}$ consists of projecting the transposed image through every eigenvector: $\mathbf{r_k} = \mathbf{A^T} \cdot \mathbf{v_k}$, $k = 1,\ldots,M$
Therefore, a total of M feature vectors are available, with n (width) components each, for the image. 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 of Eq. (2).

3. DATABASE DESCRIPTION

Here we describe the data used in order to assess the recognition performance of the P^2CA algorithm.
Since we wanted 3D face models with both profile views, we decided to acquire a new face database with the system described in [4]. The total amount of the individuals corresponds to 17 males and 8 females of a homogeneous human race (Caucasian). There are 12 different models acquired from each person (that means 25×12=300 3D face models), with a certain variability degree over the facial expressions. We have not considered the variations with respect to the presence of glasses, but there exist individuals in the database with beard and moustaches.
Each complete model contains about 30.000 points; these points are used to generate a mesh, a high resolution texture (>1 Mega pixels), and the VRML file of the model (see Figure 3 for an example). We have extracted from the database depth maps and texture images in spherical coordinates of each face as depicted in Figure 4.

Figure 3. Artificially pose variations generated with the acquired 3D model.

Since we kept an angular resolution of 0.5 degree and the range of θ and ϕ was respectively 180 and 90 degrees, the extracted depth and texture image dimension resulted in 360×180 pixels. The generated depth intensity is proportional to the distance from the origin of the spherical reference frame and it is quantized to 256 levels.

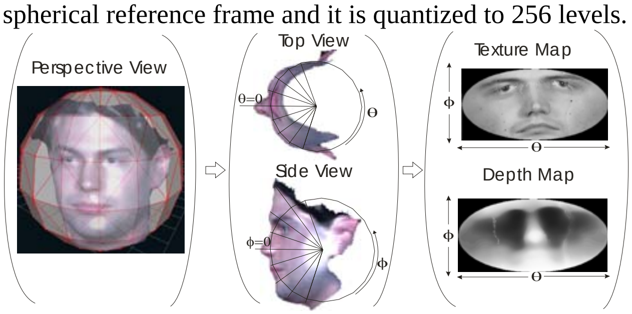

Figure 4. Schema of the generated texture and depth map, an image in spherical coordinates is generated from the 3D model; a symmetry point is taken as origin of the spherical reference frame.

4. EXPERIMENTS DESCRIPTION

Let us now discuss in detail the performance experiments of single-mode over multi-modal operation varying the quantity of information at the input of the P^2CA algorithm. First 6 models for each individual (texture and depth separately from the database described in Section 3) are used to initialize projection vectors, then data is decomposed with P^2CA and calculated vectors are stored. Experiments were conducted by increasing the partial information available (see Figure 5), starting from no information ($\theta=0$) terminating with entire image ($\theta=180°$) (texture, depth or both) available.

The images are manually aligned in a preprocessing step in a way such that eyes and mouth lie on fixed points. Using partial information a first step of pose estimation usually is needed; however in these experiments we already knew which part of the face we used for recognition, therefore no pose estimation was performed. Once the partial images were projected with the P^2CA algorithm, recognition was performed by means of distance comparison between vector representing person to be recognized and those stored in the training step.

As explained in Section 2, the score of the recognition is simply the value of the Eq (3), where only the elements of the vectors in the projection, pertaining to the partial information available are considered.

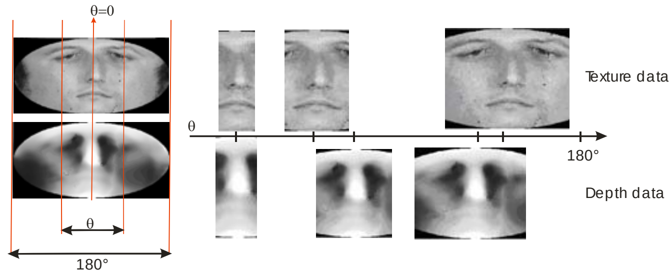

Figure 5. Sketch of the input data for recognition used with the P^2CA algorithm increasing θ implies more information available

Single mode (Single Modality Multiple Sample, *SMMS*) with only texture images and only depth images and Multimodal operation (Multiple Modality Multiple Sample, *MMMS*) [3] were tested.

The first experiment looked at recognition performance from a single modality (SM), either 2D (texture image), or 3D (depth image) using a P^2CA approach. For each modality, data belonging to 3D models, not previously used in the training stage, were considered as a probe and compared with the multiple sample (6 models × 25 individuals) stored in the gallery.

In the second experiment, a person was represented by the combination of one 2D image and one 3D image (Multiple Modality, *MM*). A two-image probe was created by matching the 2D probe image against each of the 2D gallery images and simultaneously, the same was done with the 3D data. The result, shown in the next section, was obtained taking the sum of the two normalized scores.

5. RESULTS

Experimental results for recognition with partial data are shown in Figure 6. The curves refer to SMMS (3D) with depth images only, SMMS (2D) with texture images only and MMSS with multimodal approach. With no information ($\theta=0$) or little information available, recognition is very low for all the experiments conducted. By increasing θ, suddenly recognition reaches higher values. For SMMS with 2D data this happens at lower θ than for SMMS with 3D data. This is probably because much of the information for the recognition task in the texture data is contained on the frontal region of the face, while for the depth data there is still information coming from the side (profile) of the image.

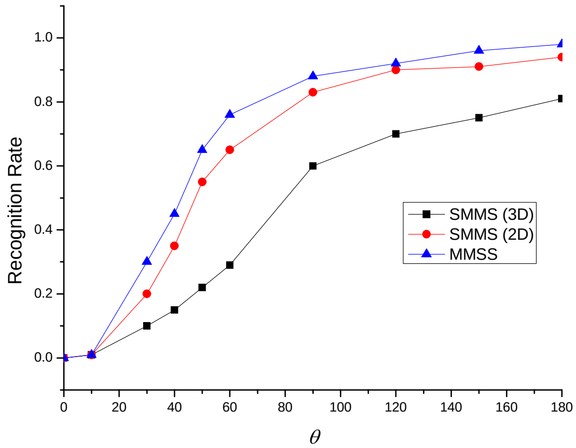

Figure 6. Experimental results for recognition with partial data. Results for SMMS (3D), SMMS (2D) and MMSS are reported.

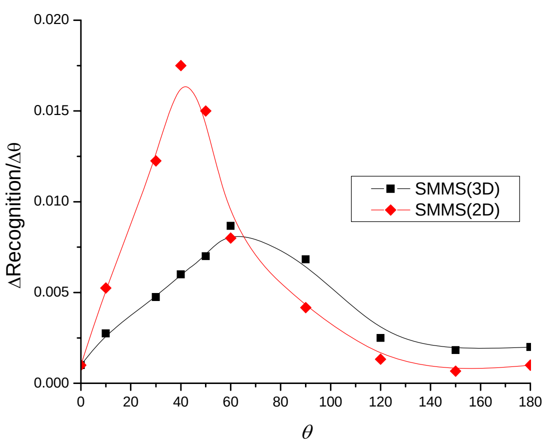

Figure 7. Derivative of the recognition rate with respect to θ shown in Figure 6, it can represent information brought by partial data to the recognition system.

In order to better see this effect we can think that the information provided by partial data to the system is proportionally related to the derivative of the recognition rate with respect to θ. In Figure 7 it is evident that this information has different maxima for the two kinds of data. Looking at the experimental results it appears that MMSS approach outperforms SMMS both 3D and 2D, as reported in other experiments [3]. It seems clear that multimodal 2D+3D face recognition achieves significant improvement over 2D face recognition even when only partial data is available.

6. CONCLUSIONS

In this paper we have presented experimental results of single (2D or 3D) and multimodal (2D+3D) face recognition with the P^2CA technique using partial information. The results support the basic conclusions:

1. Multimodal 2D+3D face recognition performs significantly better than using either 3D or 2D alone.
2. It seems that the useful information for recognition brought by texture and depth data is different. While texture data is recognizable from the frontal region of the face, depth data instead is better recognized from the side of the face, at least for what concerns the P^2CA algorithm.

As future work we intend to compare other recognition methods in a multimodal (2D+3D) scheme with partial data available on faces.

7. REFERENCES

[1] K. Bowyer, K. Chang, and P. Flynn, “A Short Survey of 3D and Multi-Modal 3D+2D Face Recognition,” Proc. Int’l Conf. Pattern Recognition, 2004.

[2] Huang, J., Heisele, B., and Blanz, V., Component-based Face Recognition with 3D Morphable Models, Proc. of the 4th Int. Conf. on Audio- and Video-Based Biometric Person Authenticitation Surrey, UK, 2003

[3] Kyong I. Chang, Kevin W. Bowyer, and Patrick J. Flynn, “An evaluation of multi-modal 2d+3d face biometrics”, IEEE Trans. PAMI, vol. 27, no. 4, pp. 619–624, 2005.

[4] D. Onofrio, S.Tubaro, T. Rama, F. Tarres, “3D Face Reconstruction with A Four Camera Acquisition System”, *VLBV05*, Cagliari (Italy), 2005.

[5] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[6] A.Rama, F.Tarrés, “P^2CA: A new face recognition scheme combining 2D and 3D information”, in IEEE ICIP, Genoa, Italy, September 11th-14th 2005

[7] T. Rama, F. Tarres, D. Onofrio, S. Tubaro, “Using Partial Information for Face Recognition and pose Estimation”, *ICME05*, Amsterdam 2005

[8] F. Tsalakanidou, S. Malassiotis, and M. Strintzis, “Integration of 2D and 3D Images for Enhanced Face Authentication,” Proc. Sixth Int’l Conf. Automated Face and Gesture Recognition, pp. 266-271, May 2004.