EXPLOITING 3D FACES IN BIOMETRIC FORENSIC RECOGNITION

Marinella Cadoni, Andrea Lagorio, Enrico Grosso and Massimo Tistarelli

University of Sassari
Computer Vision Laboratory
Porto Conte Ricerche, Tramariglio, Alghero, Italy
email:maricadoni, lagorio, grosso, tista@uniss.it

ABSTRACT

The analysis of 3D face data is very promising in improving the identification performances of individuals. In this paper two algorithms will be presented to match 3D face shapes (3D to 3D) and also to compare a synthetically generated 2D view with a given 2D face image (3D to 2D).

The algorithm to match 3D shapes is based on the extraction and comparison of invariant features from the 3D face template. In order to perform the 3D to 2D matching an algorithm is used to generate a 2D view of the face which is as close as possible to the given face image. Salient features are extracted from the two face images and are used to determine the similarity of the faces.

To simulate a real forensic environment, a small database of 3D faces has been acquired. The dataset will be used to test the identification performances in both cases.

1. INTRODUCTION

Over the last decade, there has been increasing interest in the development of 3D face recognition algorithms leading to great improvements in performance. 3D acquisition systems are also becoming affordable, user friendly and easy to install in different environments. For these reasons it is envisaged that, in the near future, 3D face acquisition and matching can be successfully employed in forensic applications as well.

There are different scenarios where three-dimensional data can be acquired and used to identify criminals. We hypothesize two possible cases:

- (3D to 3D) The three conventional face mug shots taken from arrested criminals (frontal view, left and right profiles) are substituted with a full 3D representation of the face. This can be obtained quickly with different imaging devices at a relatively low cost. A 3D face is obtained from a sequence acquired by a surveillance camera in the crime scene. The two 3D faces are matched to identify a suspect.
- (3D to 2D) As in the previous case, a 3D face representation is available from a list of suspects, but only a 2D face image is available from the crime scene. The 3D face representation of the suspect is used to generate a synthetic 2D view of the face and perform the matching with the face image taken from the crime scene.

In this paper two algorithms are described to match 3D face shapes (3D to 3D) and also to compare a synthetically generated 2D view with a given 2D face image (3D to 2D).

The proposed algorithm to match 3D shapes is based on the extraction and comparison of invariant features from the 3D face template. In order to perform the 3D to 2D matching an algorithm is used to generate a 2D view of the face which is as close as possible to the given face image. A number of salient features are extracted from the two face images. The similarity of the two face images is determined by comparing the two feature-based representations.

In order to reproduce the real operational environment for forensic applications, a small database of 3D faces including both geometrical and texture information has been acquired using a high resolution 3D face scanner. The dataset is used to test the identification performances in both cases.

2. 3D FACE DATA ACQUISITION

To simulate a real scenario two different data sets were acquired:

- a 3D database of ten subjects consisting of one scan per subject, acquired in a controlled environment, which we will call the *training set G*
- a 3D database of the same subjects consisting of one scan per subject, acquired at a subsequent time which we will call the *test set* T_3.

The 3D data sets have been acquired using the Vectra System, which was specifically designed to acquire faces. It is a structured light system (see figure 1) that consists of two high resolution cameras to acquire texture, two structured light projectors, and four calibrated cameras spaced so that in one acquisition they are able to capture the whole face from ear to ear and forehead to neck. The acquisition speed is about 0.2 milliseconds. Illumination does not need to be controlled having the Vectra a build in automatic flashing light system. The output is a 3D virtual model of the face that consists of a mesh of about $8 \cdot 10^4$ points and texture. Figure 2 (left) shows a wireframe 3D scan of a subject, whereas figure 3 (right) shows a partial scan with texture.

Both for the training and the test set, we acquired one scan of each of ten subjects, five males and five females. The time interval between the training set acquisition and the test set acquisition was about six months for six of the subjects, and a few weeks for the others. Despite the subjects where asked to pose with a neutral expression, there are notable differences between the scans taken six months apart, mainly caused by changes in weight.

To simulate a real scenario, a second test set was constructed. Scans of the 3D test set where manually altered to simulate partial acquisition of subjects. The nose area was always preserved whereas in two scans part of one eye was cut, as it would happen in a 3/4 view. Generally, the number of points of the scans was reduced by half, and in the case of figure 3 up to 1/3. We will call this set the 3D *modified test set* T_{3m}.

 ISSN 2076-1465

Figure 1: *Vectra*^{tm} 3D image acquisition system

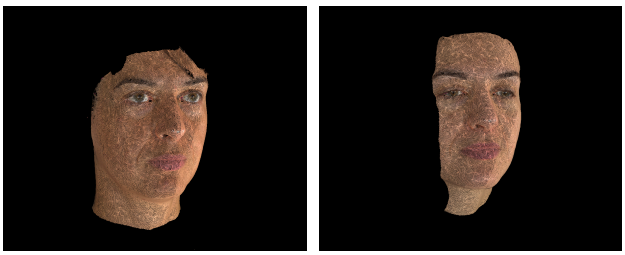

Figure 2: (left) Sample 3D image used for training. (right) 3D image of the same person with a partial occlusion.

3. MATCHING ALGORITHMS

In this section the algorithms we implemented to perform 3D versus 3D recognition and 3D versus 2D recognition are described.

In the first scenario we can assume to have a training database consisting of 3D scans of faces and a 3D model of one or more suspects. We will then compare the models to all subjects in the database to recognise the suspects. Keeping into consideration that the database can be large and consist of highly resolution 3D scans (since many of them will have been taken in a controlled environment), high speed of comparison, minimal manual intervention and reliability are highly desirable if not required. To meet these requirements, we chose to use an algorithm specifically developed for 3D face recognition which is totally automatic and reasonably fast. Its reliability has been tested on a 3D face database available to the scientific community (see [4]). For the second scenario, we choose a well known face image recognition algorithm based on SIFT [12]. In what follows we summarise the algorithms and explain how we adapted them to the forensic scenario.

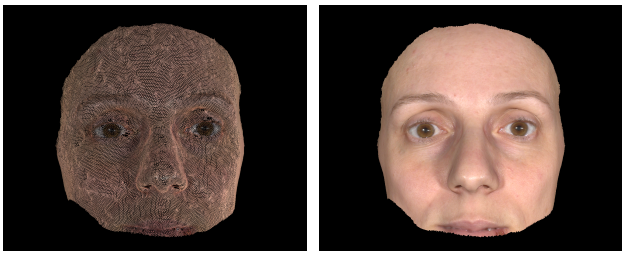

Figure 3: (left) Sample 3D test image with occlusion. (right) Same subject with texture mapped.

3.1 3D to 3D Face recognition

Identification from 3D face data has been extensively studied, particularly based on shape matching. A 3D face recognition system is usually based on four sequential steps: 1) preprocessing of the raw 3D facial data, 2) registration of the two faces, 3) feature extraction, and 4) matching [1]. As for 2D face identification, the 3D face must be first localized in a given 3D image. However, as the available 3D face acquisition devices have a very limited sensing range and the acquired image usually contains only the facial area.

The developed system deploys 3D face data acquired from a Vectra 3D imaging device. Although the acquired 3D scans contain both geometric (shape) and texture information, shape alone is generally sufficient to produce reliable matching scores. For this reason only the geometric information of the scans, given in the form of clouds of points, has been used. The standard approach to point cloud-based 3D face recognition [13] has been applied: given two faces to be compared, we first register them and then use the registration error as a matching score.

The registration algorithm is based on the Moving Frame Theory (for a thorough treatment of the theory see [14]). Given a surface F, the Moving Frame Theory provides us with a framework (or algorithmic procedure) to calculate a set of invariants, say $\{I_1, \dots I_n\}$, where each I_i is a real valued function that depends on one or more points of the surface. By construction, this set contains the minimum number of invariants that are necessary and sufficient to parametrise a "signature" $S(I_1, \dots, I_n)$ that characterizes the surface up to Euclidean motion. The framework offers the possibility of choosing the number of points the invariants depend on, and this determines both the number n of invariants we get and their differential order. The more the points the invariants depend on the lower the differential order. For instance, invariants that are functions of only one point that varies on the surface ($I = I(p)$, $p \in F$) have differential order equals to 2, and they are the classical Gaussian and Mean curvatures. To compromise between computational time and robustness we chose to build invariants that depend on three points at one time. The result is a set of nine invariants, three of differential order zero, and six of order one.

3.1.1 3D Invariants based on point triples

Let $p_1, p_2, p_3 \in F$ and n_i be the normal vector at p_i. We define r to be the direction vector of the line between p_1 and p_2 and n_t the normal to the plane through p_1, p_2, p_3:

$$r = \frac{p_2 - p_1}{\|p_2 - p_1\|} \quad \text{and} \quad n_t = \frac{(p_2 - p_1) \wedge (p_3 - p_1)}{\|(p_2 - p_1) \wedge (p_3 - p_1)\|}.$$

The zero order invariants are the 3 inter-point distances $I_k(p_1, p_2, p_3)$ for $k = 1, 2, 3$:

$$I_1 = \|p_2 - p_1\|, \; I_2 = \|p_3 - p_2\| \quad \text{and} \quad I_3 = \|p_3 - p_1\|$$

whereas the first order invariants are

$$J_k(p_1, p_2, p_3) = \frac{(n_t \wedge r) \cdot n_k}{n_t \cdot n_k} \quad \text{for } k = 1, 2, 3$$

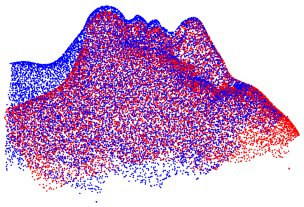

Figure 4: Registration of two complete scans of the same subject

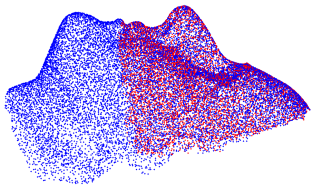

Figure 5: Registration of an occluded scan with a complete one of the same subject

and

$$\tilde{J}_k(p_1,p_2,p_3) = \frac{r \cdot n_k}{n_t \cdot n_k} \quad \text{for } k = 1,2,3.$$

To each triplet (p_1,p_2,p_3) on the surface we can then associate a point of the signature in 9-dimensional space whose coordinates are given by $(I_1,I_2,I_3,J_1,J_2,J_3,\tilde{J}_1,\tilde{J}_2,\tilde{J}_3)$.

3.1.2 *Features extraction*

Having calculated the invariants, we could use them to parametrise a signature S and then compare the signature with that of another scan to establish if the two surfaces are the same up to Euclidean transformation. Unfortunately the size of scans makes it impossible to calculate the invariants for all triplets of points of the cloud. A face scan acquired with the Vectra system can contain up to $8 \cdot 10^4$ points which would lead to $8^3 \cdot 10^{12}$ signatures points. We therefore extracted up to 10 feature points from the scan (i.e. points that are invariant to Euclidean motions). To achieve this, we opted to calculate the Gaussian curvature at each point and find local maxima. Amongst these we selected up to 10 points with highest curvature as feature points. The curvature at each $p_i \in F$ was calculated using PCA analysis on the points of intersection between F and a ball $B(p_i,\rho)$ centered at p_i and with radius $\rho = 10r$, where r is the average resolution of F.

3.1.3 *Registration and matching*

Once the feature points from F are extracted we calculate the invariants described on triples of them and save them into a signature S that characterizes the scan. Given another scan F' (we can think of it a test scan), we apply the same procedure to get a signature S'. At this point, all is left to do is to compare the signatures. If S intersects S' in a subset, then there is a subset of feature points of the two scans that have the same characteristics, i.e. same inter-point distances and normal vectors (up to Euclidean motion). We therefore fix a threshold σ and compare signature points with the Euclidean distance: if $s \in S$, $s' \in S'$ and $|s-s'| \leq \sigma$ then the triplets that generated the signature points are considered to be a good match and so we calculate the rototranslation $(\mathbf{R},\mathbf{t})$ that takes the second into the first. We then apply $(\mathbf{R},\mathbf{t})$ to F' to align it with F, and call the transformed match $F'' = \mathbf{R}F' + \mathbf{t}$. Figures 4 shows the registration of two different scans of the same subject. The neck misalignment reflects the different pose of the subject in the two scans (it was leaning forward when the red scan was acquired). Figure 5 shows a partial scan of a subject registered with a full one.

To measure the registration error, we need to measure the "closeness" of F'' to F. For each point $q_i \in F''$ we find the closest point p_i in F and we save their Euclidean distance $d_i = \|q_i - p_i\|$. We get a set of distances $D = \{d_i\}_{i \in I}$ where I is the cardinality of F''. Since we need to take into account that the test scans will not necessarily be taken under controlled conditions and that the subject will most likely be uncooperative they might fail to contain the whole face. For instance, a hat can be worn or the hair or a scarf might cover part of the face and so on. And this can be true also for training scans obtained from uncontrolled conditions. This means that some of the d_i can be quite big even if the subject is the same and the registration is accurate. In [4], experiments on a large database showed that the median of $D = \{d_i\}_{i \in I}$ was enough to eliminate such outliers and so measure the registration error of scans that overlap only partially. For this reason we adopted it also in this context. In summary, given two scans F, F', they are first registered and then the registration error $e(F,F')$ is measured. It might happen that the scans are so different that the registration step fails. In that case, we will simply consider the result a negative match.

3.1.4 *Experimental results*

Two 3D face recognition tests were performed. For the first test G was used as training set and T_3 as test set. Each test scan $F'_i \in T_3$ was compared to all scans in G using the 3D face recognition algorithm described in section 3.1 giving a vector of scores $\mathbf{s}_i = (e_1,\ldots,e_{10})$ where $e_j = e(F'_i,F_j)$ is the registration error between F'_i and F_j. Let $e_m = \min_{j=1,\ldots,10}\{e_j\}$. Then F'_i is matched to F_m. Each subject in the test set was matched to the right subject of the train set, giving a matching score of 100%.

The second test was identical to the first one except in the choice of the train set which was T_{3m}. Despite heavy occlusions in the test scans, we still obtained 100% recognition rate.

These results are very promising, especially when using occluded scans, which represent a more realistic scenario in forensic contexts.

3.2 3D to 2D Face recognition

In the second scenario considered, 3D face data is acquired for enrolment while 2D face images are used for identification. This situation emulates the case of convicted criminals whose 3D faces were acquired and stored, while 2D snapshots or a video clip is available from the the crime scene. In this case the police officer should be able to identify the criminal whose face is depicted in the captured image or video. In most cases identification from images taken from a surveillance camera is quite difficult because the face is often rotated with respect to the camera.

3.2.1 2D Database generation

Having 3D face data with texture allows to re-project face images with any orientation and use these images to perform the matching. The training set G_{2D} consists of a series of 2D projections of scans from the 3D training set G. These correspond to 9 different head orientations, spaced 30 degrees along the horizontal and vertical axes. The test set T_{2D} is obtained in an analogous way starting from the 3D test set T_3 but projecting along 5 different directions.

3.2.2 2D Matching

The face matching algorithm is based on the comparison of the Scale Invariant Feature Transform (SIFT) feature sets extracted from the images in the G_{2D} and T_{2D} datasets [11], [12]. One of the interesting features of the SIFT approach is the capability to capture the main gray-level features of an object's view by means of local patterns extracted from a scale-space decomposition of the image.

The matching score between two images is computed, as proposed in [11], by counting the number of similar SIFT features.

As several views are projected for each subject, it is expected that the 2D projection corresponding to the closest head orientation of the test image will produce the greatest matching score.

3.2.3 Experimental results

The training and test images are first aligned and scaled according to the positions of the eyes and the mouth manually located. To ensure the proper scale and position on the 2D image plane a simple planar affine transformation is adopted. Finally each test image is matched, as described above, against all test face images. The training image with the highest number of corresponding features is taken as matching image.

As an example, six out of the total nine 2D projected images of one subject are shown in figure 6. Two sample test and training images from the same subject, with different head orientations, are shown in figure 7. In figure 8 a test and a probe registered images with similar head orientation are shown on the top line and the extracted SIFT features are shown in the bottom line. The genuine and impostor score distributions, obtained by performing a complete matching test on the acquired dataset, are shown in figure 9. The equal error rate computed from the two distributions is equal to 4%. Considering the different projection directions between the training and test images the obtained results are encouraging.

4. CONCLUSION

Face-based identification has been extensively used in forensic applications. Generally 2D face images are captured both from convicted criminals and in the scene of crime. As a consequence standard 2D face matching techniques are deployed.

In this paper we argue that 2D face images do not convey enough information to perform automatically a reliable matching of a test and training pair. Extremely different acquisition conditions between the enrolment set-up and the crime scene make it difficult to compare images from the

Figure 6: Sample 2D images obtained by projecting the 3D texture mapped model.

Figure 7: (left) Projected image used for training. (right) 2D image of the same subject used for testing.

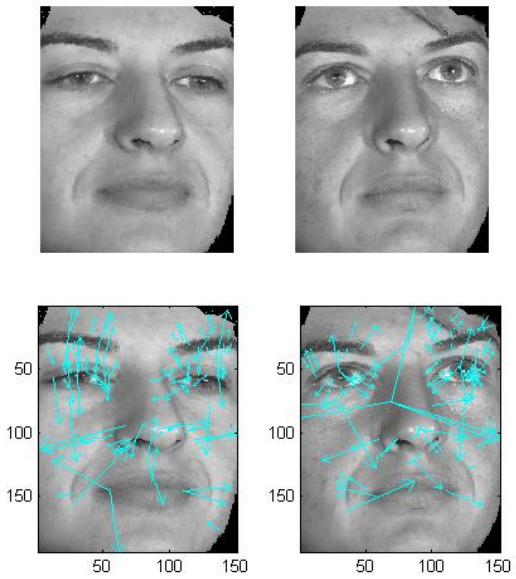

Figure 8: **Upper row**: (left) sample 2D test face image, (right) corresponding pose training face. **Lower row**: SIFT extracted from the two images.

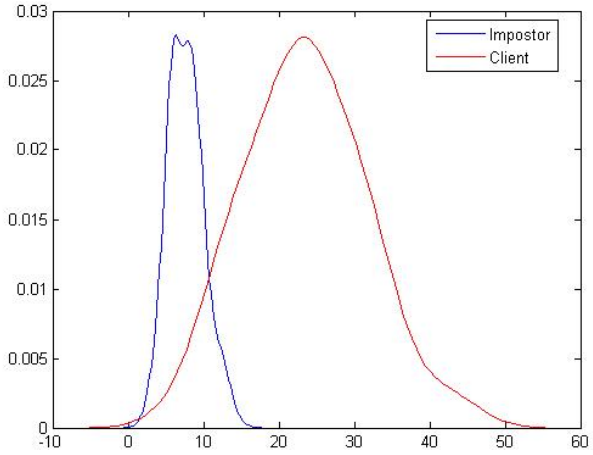

Figure 9: Impostor and client score distribution computed with the 3D to 2D matching using the SIFT features and using the global distance of the features.

same subject. Not only illumination and head pose are different, but even the camera position and view direction are different. In fact, while mug shots are taken from criminals with a camera directly looking at the subject's face, the pictures taken from the crime scene generally originate from surveillance cameras which look at faces from above. For these reason it seems reasonable to exploit the information content of a full 3D face, at least for the enrolment phase.

In this paper two different forensic scenario have been studied. In the former, a full 3D representation of the face is acquired instead of the three conventional face mug shots taken from arrested criminals. A 3D face is obtained from a sequence acquired by a surveillance camera in the crime scene. The two 3D faces are matched to identify a suspect. In the latter scenario, a 3D face representation is available from a list of suspects, but only a 2D face image is available from the crime scene. The 3D face representation of the suspect is used to generate a synthetic 2D view of the face and perform the matching with the face image taken from the crime scene.

A compact 3D data set has been purposely acquired to simulate the two scenarios. Both 3D to 3D and 3D to 2D experiments have been performed producing good results. In particular the best performances are obtained when matching a 3D face model with a full 3D face, as expected. The comparably lower performances obtained in the 3D to 2D matching are presumably due to the small number of back-projected gallery images used for matching. Increasing the resolution in the head pose variations would certainly improve the recognition results.

Even though the experiments have not been performed on real data from a crime scene, both methods can be successfully exploited in scenarios in which the quality of the test images and 3D scans is affected by occlusions or poor resolution.

REFERENCES

[1] A.F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28, pp. 1885–1906, 2007.

[2] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), pp. 239–256, 1992.

[3] A.M. Bronstein, M.M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *Int. Journal of Computer Vision*, 64(1), pp. 5–30, 2005.

[4] M.I. Cadoni, M. Bicego, E. Grosso. "3D Face Recognition Using Joint Differential Invariants" *ICB09-Proceedings*, pp. 11–25, June 2009.

[5] I. Mpiperis, S. Malassiotis, and M.G. Strintzis. 3-D face recognition with the geodesic polar representation. *IEEE Transactions on Information Forensics and Security*, 2(3 Part 2), pp. 537–547, 2007.

[6] F.R. Al-Osaimi, M. Bennamoun, and A. Mian. Integration of local and global geometrical cues for 3D face recognition. *Pattern Recognition*, 41(2), pp. 1030–1040, 2008.

[7] A. Colombo, C. Cusano, and R. Schettini. 3D face detection using curvature analysis. *Pattern Recognition*, 39(3), pp. 444–455, 2006.

[8] C. BenAbdelkader and P.A. Griffin. Comparing and combining depth and texture cues for face recognition. *Image and Vision Computing*, 23(3), pp. 339–352, 2005.

[9] C. Beumier and M. Acheroy. Face verification from 3D and grey level cues. *Pattern Recognition Letters*, 22, pp. 1321–1329, 2001.

[10] K. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding*, 101, pp. 1–15, 2006.

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2), pp. 91–110, 2004.

[12] M. Bicego, A. Lagorio, E. Grosso and M. Tistarelli. On the use of SIFT features for face authentication. *Proc. of Int Workshop on Biometrics, in association with CVPR 2006*, 2006.

[13] Gokberk, B., Akarun, L. Comparative analysis of decision-level fusion algorithms for 3d face recognition. *Proc. of Int. Conf. on Pattern Recognition*, pp. 1018–1021, 2006.

[14] Olver, P.J. Joint Invariants Signatures. *Found. Comput. Math.* Volume 1, pp. 3–67, 2001.