Automatic 3D Face Authentication

Charles Beumier [1,2] and Marc Acheroy

Signal & Image Centre, Royal Military Academy, Avenue de la Renaissance 30, B1000 Brussels, Belgium

Abstract

This paper presents automatic face authentication from facial surface analysis. This geometrical approach was motivated by difficulties encountered when considering frontal face recognition. Apart from being less sensitive to viewpoint and lighting conditions, the method exploits information which is complementary to grey level based approaches, enabling the fusion with those techniques. A 3D acquisition system based on structured light and adapted to facial surface capture is presented. It is cheap and fast while offering a sufficient resolution for face recognition purposes. The acquisition system and the 3D face comparison algorithm were designed to be integrated in security applications with cooperative scenario.

Key words: Face recognition; structured light; surface comparison

1 Introduction

Biometric measurements receive an increasing interest for security applications where PIN codes and cards are less desired (due to loss or theft). In cooperative environments, speech and face modalities are well accepted by individuals but they still suffer from limited performances. To achieve a sufficient level of reliability, several modalities (speech, profile, face, 3D) may be combined [1,4].

A previous profile analysis [2] has shown the adequacy of geometrical information for automatic person authentication. It takes benefit from the rigidity of the parts involved (forehead, nose, chin) and the little dependence on makeup or lighting conditions. This explains the success of many profile works [6,9].

[1] Corresponding author. E-mail: beumier@elec.rma.ac.be
[2] Supported by the European ACTS program (AC102 "M2VTS")

More geometrical information is taken from a facial 3D description, especially where grey level features lack as in the chin, forehead and cheek regions. The analysis can benefit from real 3D measures (no scale or rotation influence). Depth information also helps segmenting the face from background objects. Those advantages clearly state the 3D geometrical approach as complementary to the grey level analysis. Although 3D facial modelling for compression and synthesis as in videoconferencing [10] or medical applications is not a new field of interest, 3D facial authentication activities are still weakly addressed [8,14] in the literature in comparison with frontal or profile developments.

3D capture is usually expensive and slow. We designed a 3D acquisition prototype based on structured light which is adapted to facial surface acquisition. Its resolution, speed and sufficient facial covering make it appropriate for practical implementations. The emergence on the market of structured light systems for 3D face acquisition supports our choice. However, hair and beard are not properly acquired, and a grey level analysis remains attractive. Switching the projector on and off is a simple way to get geometrical and grey level information in alignment from the same hardware equipment.

The next section describes the structured light acquisition system. The hardware choices are motivated and the calibration and 3D extraction procedures are presented. The acquisition prototype has been tested during the collection of a 3D database, matter of section 3, which was used later for recognition experiments. Section 4 introduces two approaches considered to compare 3D facial representations: a facial surface matching algorithm and a profile matching procedure after intrinsic normalization. Recognition results are presented in section 5. Section 6 concludes the paper.

2 3D acquisition

2.1 Motivations for structured light

Among the possible range acquisition systems [11,12], structured light has emerged as the solution for 3D acquisition in our context. It is based on the projection of a known light pattern (in our case parallel 'stripes'). The light pattern deformation, captured by a camera, contains the depth information of the scene.

Four advantages motivate our choice. First, the additional cost is limited to a projector and a slide. Secondly, a standard camera is precise enough and benefits from the high speed of video hardware. A single image with stripes suffices to recover 3D information. This enables sequence analysis and time integration. Thirdly, switching the projector on and off allows to acquire volume and texture information in correspondence. Fourthly, the projector illumination reduces the influence of ambient light. In particular, near infra-red light is more discreet and does not dazzle the individual.

The drawbacks of a structured light system are its relative bulkiness and its limited field of depth due to the camera and projector lenses.

2.2 Hardware choices

To keep investments low, we opted for off-the-shelf components. We use a standard CCD black and white camera plugged into a 768x576 pixels image digitiser. A 24x36mm projector projects its light through a slide coding the stripe indices in the thickness (thin or thick) of neighbouring stripes.

2.3 Set-up

The camera and the projector have their optical axes co-planar to reduce the number of parameters to be calibrated. Both optical systems have a limited span and depth of focus; the field of view covers about 30x40 cm at 1m40 from the camera/projector head, with a depth of focus of about 40 cm. This is sufficient for sitting attitudes in cooperative situations.

2.4 Calibration

The first calibration step consists in rough measurements of the camera and projector distance and relative angle. Rough values are also given to parameters depending on the pixel size of the camera/digitiser pair as well as slide and lenses characteristics. Then automatic refinement is performed by presenting a square object in several orientations and trying to make the 3D corners be at the vertices of a planar square of known size. Typical relative errors on the estimated square side is 2 %.

2.5 3D extraction

Automatic 3D extraction is done by stripe detection and labelling. From each point of a stripe and its label, triangulation allows for X, Y and Z estimations.

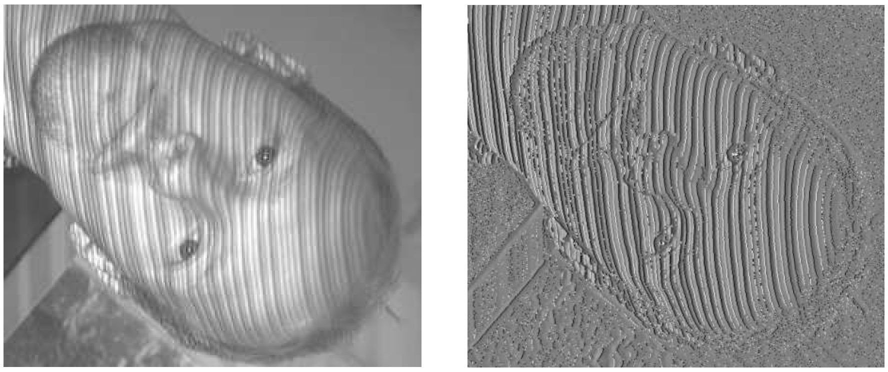

Fig. 1. a) Striped image b) Stripe detection and thickness (white=thin, black=thick)

Stripe detection is carried out by line following helped by the linear nature of the light pattern. Stripe thickness is estimated from the grey level profiles across the stripes. Local thickness distribution of neighbouring stripes helps initiating stripe labeling from the known thickness distribution of the slide. The global coherence of this labeling is checked against normal ordering and spacing of stripes to detect and solve local inconsistencies (commonly found in abrupt transitions of the nose and chin) and propose labels in non-labelled areas (for instance due to grey level troubles in eyes or beard regions). The label of the stripes and the x, y image position of points along the stripes are converted into X, Y, Z thanks to calibrated parameters.

This implementation is very fast (half a second on a Pentium 200) while offering sufficient resolution for recognition purposes. A comfortable covering of the face is possible, nearly from ear to ear and including the throat. Stripes projected on the background are normally out of focus and do not complicate face extraction. Noses and eyes often raise minor problems. Beard and glasses induce large errors if the beard is bushy and the glasses have thick frames.

3 3D database

In order to test the performances of the acquisition system and of the comparison approaches, we prepared a database with the following criteria.

First, a sufficient number of individuals had to be acquired. We started with 120 persons. Secondly, we selected people likely to stay in our reach to get long term statistics from sessions separated by several months. Thirdly, different problems encountered in the cooperative scenario were taken into account (spectacles, smiles, head rotation).

Until now two sessions with the same 120 individuals have been completed. People were asked to sit and to look at the camera. We took three shots per session, corresponding to central, limited left/right and up/down poses.

Fig. 2. Samples from the database

Fig. 3. 3D reconstructions from left image of Fig. 2

From these sessions, we built up two databases. The first one, called the *automatic DB*, used the automatic program described in subsection 2.5 to get the 3D representations of all the individuals. For the second database, called the *manual DB*, the 3D extraction process was performed interactively by clicking initial points on the stripes, assigning them a thin or thick attribute. Only the first 30 individuals were processed manually, delivering sufficient data to analyse the possible enhancement of a better acquisition program.

Fig. 2 shows 4 images of the same person, from the two sessions. Fig. 3 represents 3D reconstructions of the *automatic DB* from different points of views corresponding to the left image of Fig. 2.

Running the 3D reconstruction algorithm (see section 2.5) on the whole database made us confident in the overall quality of stripe following, labelling and background independence. However, it highlighted possible problems encountered in bushy beards, glasses, nose and eyes, by order of importance.

4 3D face comparison

When we began this 3D face recognition project, in 1995, the related litterature was poor. Several studies [8,16,13] concentrated on curvature analysis. Our own tests in this direction were unsatisfactory, due to the limited quality of the range data (noise in eyes, nose or mouth).

We preferred to analyse the 2D (striped) images in order to postpone the time consuming 3D conversion. Although studies were carried out in that direction by some researchers [7,15], coping with the influence of the viewpoint on the shape of the stripes seemed too difficult. Only the prominence of the nose led to its localisation.

We thus came back to range processing, looking for characteristics to reduce the 3D data to a set of features easily and quickly compared. We first estimated the prominence of the nose relative to points of the cheeks located at a given distance from the nose tip. As second feature, the nose length was measured by localising the nose tip and the nose saddle (between the eyes).

Although the approach was successful, the nose seems to be the only facial part providing robust geometrical features for limited effort. Mouthes and eyes may involve disturbances. Foreheads, cheeks and chins don't clearly exhibit reference points for normalisation. We abandoned feature extraction and considered the global matching of the facial surface.

4.1 Surface matching

The surface matching approach consists in finding some distance measure which quantifies the difference between two 3D surfaces and in tuning the set of parameters (translations and rotations) to minimize the distance measure.

At most 15 profiles (-7..+7, see Fig. 4) are extracted from each facial surface by intersection with parallel planes spaced with 1 cm. Each pair of corresponding profiles is compared to issue a profile distance: the area between the two profiles divided by the arc length. The average of these profile distances of all corresponding profile pairs is the global error which has to be minimised.

The minimization of the global distance is performed by an Iterative Conditional Mode optimization, which tunes each parameter (3 rotations and 3 translations) on its own, one after the other. The minimization is organized as cycles of ICM, separated by the reduction of each parameter search space. The initial translation parameters are based on the rough localization of the nose tip. The planes are initially vertical and oriented according to the left/right

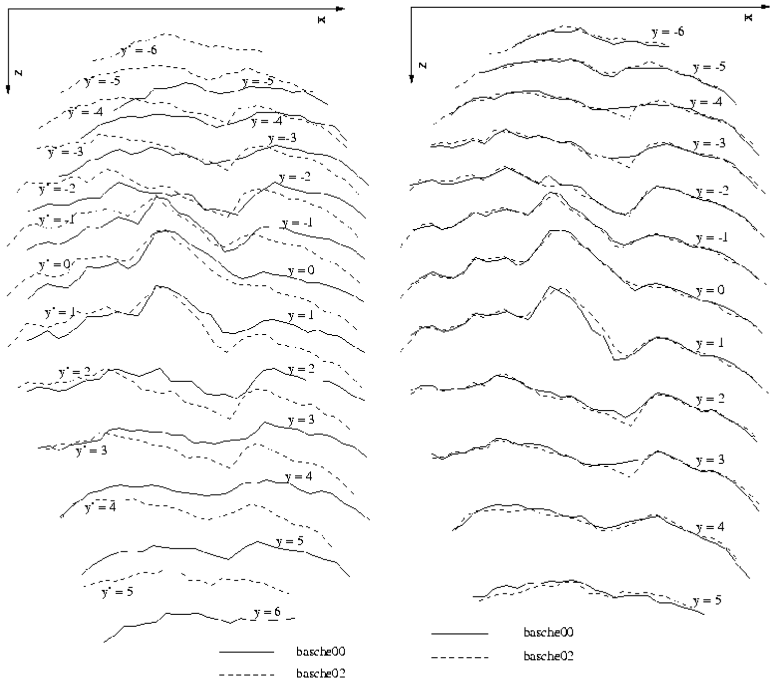

Fig. 4. a) Profiles from two 3D representations with noses already in correspondence. b) Profiles of the representations after surface matching

orientation obtained from the cheeks. The last orientation parameter is initialized from the up/down angle obtained from the forehead and nose. These starting conditions closer to the solution avoid many local minima and speed up the optimization by earlier reduction of the search space.

Some recognition results of this automatic surface matching procedure are shown in Fig. 5. The residual false rejection rates for large false acceptance rates reveals local minima in the automatic optimization either due to bad initial parameter values or important noise in the input data. Beard, glasses and nose discontinuities are the most common problems. Manual refinement from visual profile fitting avoids many local minima.

Although the approach and the computations were optimized, the algorithm is still slow: about 1 second to compare two facial surfaces.

4.2 Central and lateral profiles

In order to speed up the facial surface comparison, we split the 6-dimensional optimization into a 3-dimensional normalization based on facial symmetry followed by a 3-dimensional profile comparison (realized by a 1-dimensional curvature matching). Refer to [5] for a similar approach which extracts the central profile by looking for the vertical symmetry axis of gaussian curvature

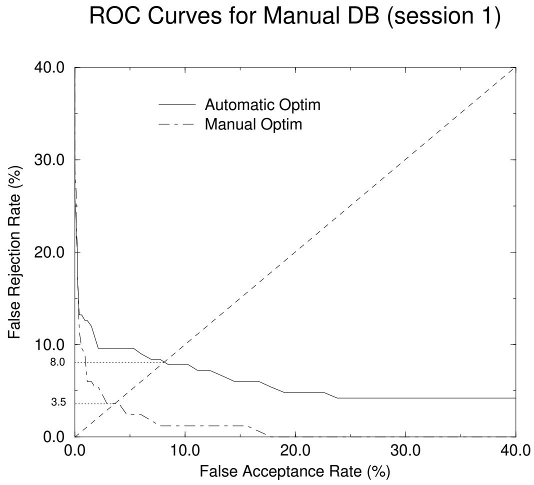

Fig. 5. ROC curves of 3D surface matching for the *manual database*, with (*Manual*) and without (*Automatic*) manual tuning (see text)

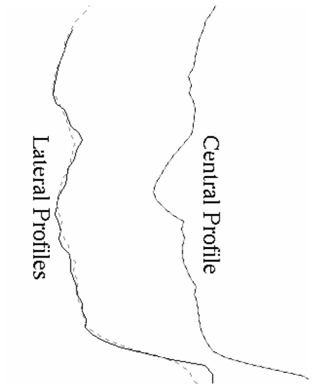

Fig. 6. Central and lateral profiles after intrinsic normalization

values of the facial surface.

Under the assumption of vertical facial symmetry, we automatised central profile extraction by looking for the profile with maximal protrusion (due to the nose) and with maximal symmetry of left and right profiles parallel and 3 cm away from it. This optimum search is quick as it only depends on three parameters (one translation and 2 rotations).

The automatically extracted central profiles were compared in the curvature space, by transforming each two-dimensional profile into the one-dimensional local curvature values along this profile. In the curvature space, values can be compared directly, dealing with only one shift parameter (for instance to bring the nose in correspondence). The 3-parameter search to match 2 profiles is replaced by a 1-dimensional minimization.

To include more 3D information, we analysed the lateral profiles used for central profile extraction. To bring robustness, these left and right curves were

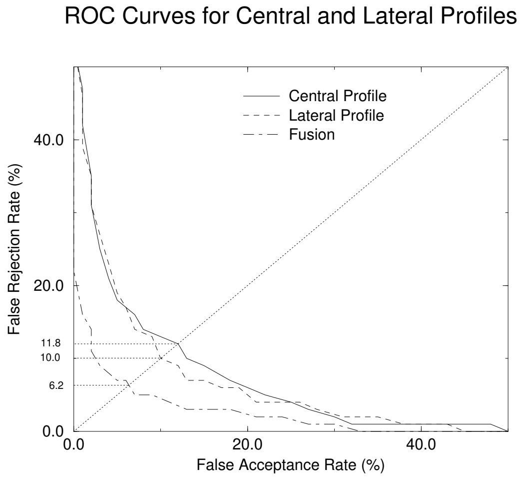

Fig. 7. ROC curves of comparison of central profiles, lateral profiles and the fusion of central and lateral profiles, with manual refinement (*automatic DB, 120 persons*)

averaged to offer a mean lateral profile. Mean lateral profiles were compared in the curvature space as described before. As seen in Fig. 7, the lateral profile discrimination power (EER: 10.0%) is similar to what was obtained with the central profile. Although more specific to individuals, the central profile often suffers from acquisition troubles in the nose and chin regions. The fusion of the scores from central and lateral profile comparisons by simple average brought a clear recognition advantage (EER: 6.2%).

5 Results

Tables of Equal Error Rates (EER) are given below to show the recognition performance of the two approaches of subsection 4.1 and 4.2.

To see the influence of the acquisition system, results are provided for the *manual* and *automatic DB* limited to 30 people. Tests on the whole *automatic DB* gave similar results. Figures with (*refined*) and without (*auto*) manual refinement allows to estimate the possible improvement of the automatic optimization procedures present in the 3D comparison methods.

The figures only give general comparison clues. In particular, it is dangerous to summarize the recognition ability by the EER.

5.1 Surface Matching

	session1	session2	session1-2
automatic DB, auto	9.0 %	9.0 %	13.0 %
automatic DB, refined	4.5 %	3.25 %	6.0 %
manual DB, auto	8.0 %	7.0 %	9.5 %
manual DB, refined	3.5 %	2.0 %	4.75 %

Table 1
EER of surface matching for the *automatic DB* (30 persons) and *manual DB*, with (*refined*) and without (*auto*) manual refinement

5.2 Central and Lateral Profiles

	session1	session2	session1-2
automatic DB, auto	7.25 %	7.75 %	9.0 %
automatic DB, refined	6.25 %	7.0 %	9.5 %
manual DB, auto	4.75 %	6.75 %	7.25 %
manual DB, refined	2.25 %	3.75 %	6.75 %

Table 2
EER of the central and lateral profiles method for the *automatic DB* (30 persons) and *manual DB*, with (*refined*) and without (*auto*) manual refinement

Many reasons decreased recognition performances. They are classified in the three following subsections.

5.3 Error from representation

The representation of each person is currently limited to 3 facial surfaces per session. For the two existing sessions, people were asked to rotate their head, implying 3D differences (different covering and possible head distorsion). We think from previous experiences that people naturally present themselves more consistently when they are asked to pose. Moreover, the database was acquired completely before being tested. For this, some people are not correctly represented, possibly smiling in some images or wearing glasses. As only one 3D representation was used as reference during the tests, the presented causes for poor representation have a large influence, as depicted by inter-session results ("session1-2").

5.4 Error from acquisition

The extraction of 3D information from striped images is not perfect. Eyes and nose typically introduce local errors due to grey and volume disturbances. Spectacles, especially with thick frames, impair surface acquisition in the eye regions. Bushy or dark moustaches and beards prevent stripe visibility and hence 3D sensing. Since eyes and nose problems are rather local, since the glasses can be taken off and beards or moustaches are easily localized from texture, these problems were not specifically addressed till now, although they explain a large proportion of errors.

5.5 Error from matching

3D face comparison is carried out by an optimisation procedure which can fail due to noise, local minima or bad initial parameter values. The importance of the incurred errors is visible from the results of matching with and without manual refinement. A large part of important matching errors can be explained by acquisition errors.

5.6 Comparison of the methods

The two 3D comparison methods (surface matching and central/lateral profiles) give the same level of recognition performance. The surface matching approach is more sensitive to matching errors, as can be seen from the differences between *auto* and *refined* rows of Table 1. On the contrary, Table 2 shows the importance of acquisition errors on the performance of the central/lateral profiles approach (differences between *automatic DB* and *manual DB* rows).

The main advantages of the central/lateral profiles method are its speed and low storage needs. The intrinsic normalization based on vertical facial symmetry takes 0.5 second and delivers two profiles of a few hundred bytes. These can be matched very rapidly with the reference profiles of the database (extracted off-line). The surface matching algorithm takes a mean time of 0.8 second to compare two 3D representations. This precludes decision methods based on ranking or sequence analysis. 3D representations are about 25 Kbyte large.

For verification applications, both methods satisfy a time constraint of three seconds and only consumes a few hundred Kbyte during execution.

5.7 Comparison with profile performance from 2-D images

The database also contains color images of profile view of the same people. We used a previous approach (see [3]) to extract profile contours from the color images and compare session2 profiles with session1 profiles. The EER of 10.5 % is equivalent to what was obtained with the central profiles extracted from the 3-D facial surfaces, outlining the importance of global information in the profiles (the 2-D profiles are indeed much more precise than the profiles extracted from 3-D).

6 Conclusions

A complete prototype for 3D face recognition has been studied. Facial surfaces are acquired thanks to structured light with the precious advantages of low cost and quickness. The 3D comparison is carried out by profile matching, either globally or more specifically for central and lateral profiles. The person is asked to sit on a chair and center himself in front of the camera. A couple of seconds suffice to get a 3-D representation and compare it to the claimed reference.

The obtained equal error rates clearly support the motivation for an application with a cooperative scenario, especially when considering the possible improvements to be carried out in the acquisition and representation of people. Background removal, rotation, scale and translation independency, quickness and infra-red lighting are additional assets which should guarantee the success of a practical application.

We intend to complement the facial surface information with a grey level analysis to increase the recognition performance. Acquired in correspondence and with the same camera, 3D and grey level data allow for viewpoint normalization of grey levels and grey normalization based on a light source and reflexion models thanks to 3D. Grey level analysis can also assist 3D comparison and vice versa.

References

[1] Acheroy, M., Beumier, C., Bigün, J., Chollet, G., Duc, B., Fischer, S., Genoud, D., Lockwood, P., Maitre, G., Pigeon, S., Pitas, I., Sobatta, K., Vandendorpe, L., Multi-modal person verification tools using speech and images, Proceedings of the European Conference on Multimedia Applications, Services and Techniques (ECMAST '96)(1996), 747-761.

[2] Beumier, C., Acheroy, M., Automatic Face Identification, Applications of Digital Image Processing XVIII, SPIE, vol. 2564, July 1995, 311-323.

[3] Beumier, C., Acheroy, M., Automatic Profile Identification, First International Conference on Audio and Video based Biometric Person Authentication (AVBPA), Crans-Montana, Switzerland, March 12-14 1997, 145-152.

[4] Brunelli, R., Falavigna, D., Person Identification Using Multiple Cues, IEEE Trans. Pattern Anal. Machine Intell. 17 (10), Oct 1995, 955-966.

[5] Cartoux, J.Y., Lapreste, J.T., Richetin, M., Face Authentification or Recognition by Profile Extraction from Range Images, IEEE Computer Society Workshop on Interpretation of 3D scenes (1989), 194-199.

[6] Chellappa, R., Wilson, C.L., Sirohey S., Human and Machine Recognition of Faces, Proceedings of the IEEE 83 (5), May 1995, 705-740.

[7] Chen, Z., Ho, S.-Y., Tseng, D.-C., Polyhedral Face Reconstruction and Modeling from a Single Image with Structured Light, IEEE Transactions on Systems, Man, and Cybernetics 23 (3), May/June 1993.

[8] Gordon G.G., Face recognition based on depth maps and surface curvature, SPIE Geometric methods in Computer Vision, San Diego, vol. 1570, 1991.

[9] Harmon, L.D., Kuo, S.C., Ramig, P.F., Raudkivi, U., Identification of human face profiles by computer, Pattern Recognition 10, 301-312, 1978.

[10] Ho-Chao, H., Ming O. and Wu J-L, Automatic feature point extraction on a human face in model-based image coding, Optical Engineering 32 (7), July 1993, 1571-1580.

[11] Jarvis, R.A., A perspective on range finding techniques for computer vision, IEEE Trans. Pattern Anal. Machine Intell. 5 (3), March 1983.

[12] Jarvis, R., Range Sensing for Computer Vision, in: Three-Dimensional Object Recognition Systems, K. Jain and P. J. Flynn, eds., Advances in Image Communication, vol 1, Elsevier Science Publisher, 1993, 17-56.

[13] Lee, J.C., Milios, E., Matching Range Images of Human Faces, Proc. IEEE Soc. 3rd Int. Conf. on Computer Vision, 1990, 722-726.

[14] Proesmans M., Van Gool L., One-shot 3D-shape and Texture Acquisition of Facial Data, Audio- and Video-based Biometric Person Authentication, Crans-Montana, Switzerland, 12-14 March 1997.

[15] Wang, Y.F., Mitiche, A., Aggarwal, J.K., Computation of Surface Orientation and Structure Using Grid Coding, IEEE Trans. Pattern Anal. Machine Intell. 9 (1), Jan 1987, 129-137.

[16] Yacoob, Y., Davis, L.S., Labeling of Human Face Components from Range Data, CVGIP: Image Understanding 60 (2), Sep 1994, 168-178.