Automatic Face Verification from 3D and Grey Level Clues

Charles Beumier and Marc Acheroy

Signal & Image Centre,
Royal Military Academy,
Avenue de la Renaissance 30,
B1000 Brussels, Belgium
e-mail: beumier@elec.rma.ac.be

Abstract: *We address in this paper automatic face verification from 3D (3-dimensional) facial surface and grey level analysis. 3D acquisition is performed by a structured light system, adapted to face capture and allowing grey level acquisition in alignment. The 3D facial shapes are compared and the residual error after 3D matching is used as a first similarity measure. A second similarity measure is derived from grey level comparison. As expected, fusing 3D and grey level information increases verification performances. The acquisition system, the 3D and grey level comparison algorithms were designed to be integrated in security applications in which individuals cooperate.*

Keywords: *Face Verification, 3-D acquisition, structured light, 3-D matching.*

1. Introduction

Although PIN codes and cards have been the only viable solution for access control till recently, biometric verification appears more and more attractive considering loss or theft of cards. However, existing prototypes based on user accepted biometric modalities such as speech or face still suffer from limited performances in practical situations. One way to override these limitations with little additional cost is to combine modalities [2, 7].

Our approach consists in combining 3D and grey level clues. On the one hand, 3D facial descriptions bring much information, with little dependence from pose, lighting conditions or makeup. On the other hand, grey level clues complement well 3D information. They are localized in hair, eyebrows, eyes, nose, mouth, and facial hairs, precisely where 3D capture is difficult and not accurate. It also allows for the integration of skin and hair colours.

The adequacy of geometrical analysis was supported by our own experience in building a practical prototype for real-time profile recognition [3] and by the success of many profile works [10]. A facial 3D description brings much more information, possibly rotation and scale independent. Although 3D facial modelling for compression and synthesis as in videoconferencing [13] or medical applications is not a new field of interest, 3D facial recognition activities are still weakly addressed [1, 8, 9, 12] in the literature.

A huge research effort has been devoted to grey level analysis of frontal face images [10]. Most of the methods are sensitive to pose and lighting conditions. Our approach is to use 3D information to enhance the grey level analysis. First, the orientation of the projected light and the normal to the surface allow to estimate the albedo, a surface characteristic independent from viewpoint or illumination. Secondly, the facial 3D coordinates ease the registration of grey values. Thirdly, face detection is much easier from range data.

3D capture is usually expensive and slow. From the design of a structured light system adapted to facial surface acquisition, we collected 3D descriptions with sufficient quality and facial covering to perform recognition experiments. The aligned 3D and grey level data have been captured by the same equipment, switching the projector on and off.

The next section presents the 3D acquisition system. A deeper presentation is to be found in [4]. The acquisition system was tested during the collection of a database, matter of section 3, and used for recognition experiments. Section 4 introduces several approaches considered to compare 3D facial representations and focuses on the most promising one. Section 5 explains how grey level values have been integrated in the recognition process. Recognition results are presented in section 6. Section 7 concludes the paper.

2. 3D Acquisition

2.1. Motivations for structured light

Structured light acquisition systems use the projection of a known pattern of light (in our case, parallel 'stripes') to recover 3D coordinates. Compared to other 3D acquisition techniques [14], structured light benefits from some interesting properties.

First, 3D acquisition can be very fast. A single image with stripes contains 3D information so that subject's motion is not a problem. Secondly, the additional cost only concerns the projector and its slide. Thirdly, the projector illumination reduces the influence of ambient light. Fourthly, switching the projector on and off allows to acquire volume and texture information in correspondence.

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. Set-up

The prototype consists of a normal grey level camera and a projector whose slide was drawn by an accurate microelectronics process. With an appropriate selection of lenses, the field of view covers 30x40 cm at a distance of 1m40, and with 40 cm as depth of focus, what fulfils the needs for sitting attitudes in cooperative situations.

As seen in Fig. 2 and 5, the prototype is rotated so that the face is patterned diagonally by the stripes. Horizontal striping on the face indeed creates interferences with horizontal dark features (mouth, eyebrows, eyes, nostrils). On the other hand, vertical striping implies a different density of stripes (and thus resolution) between the left and right regions.

2.3. 3D extraction

Automatic 3D extraction is carried out by stripe detection and indexing. Stripes are detected by horizontal gradient and followed vertically. Their relative thickness (Fig. 1) is estimated thanks to horizontal grey profiles by comparing the width of the dark and following bright zone of the stripe. The binary thickness (thin/thick) of a few neighbouring stripes suffices to recover their stripe index. The stripe indices and the x, y image position of points along the stripes are converted into X, Y, Z by triangulation. Refer to [4] for details.

This implementation is very fast (less than 1 second on a Pentium 200) while offering sufficient resolution (more than 2 points per cm in each direction) for recognition purposes. Background objects are not captured as stripes get out of focus on them.

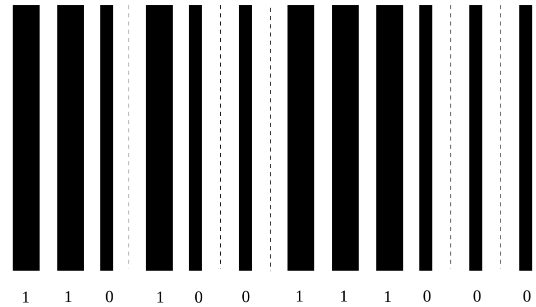

Figure 1. Slide for projection

Figure 2. Samples from the database

3. 3D DATABASE

We created a database to validate the acquisition system and perform recognition tests (see [5]).

A sufficient number of individuals (120) were called from the available people likely to stay in our reach to get long term statistics. Typical complications present in a cooperative scenario were represented (spectacles, smiles, head rotation).

Two sessions separated by two months have been completed with the same 120 people. For each session, the individuals were asked to sit and adopt three different poses (corresponding to central, limited left or right and up or down). A grey level image without projection and corresponding to the central pose was also captured.

Fig. 2 shows four images of the same person, from the two sessions. Fig. 3 represents 3D reconstructions from different points of view corresponding to the left image of Fig. 2.

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

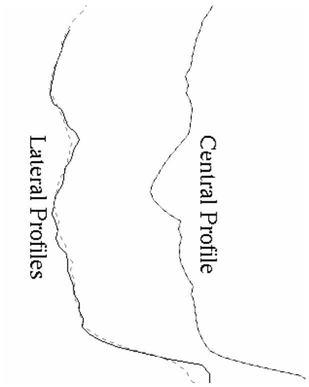

Figure 4. Central and lateral profiles after intrinsic normalization

Running the 3D reconstruction algorithm (see 2.3) on the whole database made us confident in the overall quality of stripe following, indexing and background independence. The important source of acquisition problems concern the glasses (one third of people wearing glasses had to be rejected) and beard, hair and eyes which account for a ten of people rejected out of 120.

4. 3D FACE COMPARISON

4.1. Introduction

When we first addressed 3D face comparison, in 1995, the related litterature was poor. Several studies [12, 17] concentrated on curvature analysis, but the noise present in our 3D data prevented us to work in this direction.

We preferred to analyse the striped images in order to avoid the explicit 3D extraction. Although considered by some researchers [11, 16], this approach appeared difficult due to the dependence of the stripe shapes with the viewpoint.

We thus came back to range processing, trying to reduce the 3D data to a set of geometrical features. Although the prominence and the length of the nose validated the approach, finding normalized features in other facial regions proved difficult. Moreover, we were looking for a way to assess the quality of the acquisition system. We then considered the global matching of the facial surface.

The global surface matching was carried out by comparing corresponding profiles extracted from the 3D representations by parallel planes. The matching process consists in tuning the 3 rotation and 3 translation parameters so that the sum of the distances between corresponding profiles is minimized. Although this optimization was accelerated by an Iterative Conditional Mode procedure and by appropriate initial parameter values, the approach remained slow (about one second to compare two facial surfaces).

4.2. Central and lateral profile matching

In order to speed up the facial surface comparison, the 6-dimensional optimization procedure was divided into a 3-dimensional normalization based on facial symmetry followed by a 3-dimensional profile comparison.

Assuming a vertical symmetry of the faces, three parameters (one translation and two rotations) can be optimized by looking for the maximal protrusion of the central profile and the maximal symmetry of left and right lateral profiles.

The automatically extracted profiles (central and mean lateral, average of the left and right lateral profiles at +3 and -3 cm from the nose) were transformed into local angular values by computing the slope of the curve between two points sliding along the curve with fixed distance (4 cm). The angular curves were compared by the standard deviation of the angle difference between corresponding points. Corresponding points are obtained relatively to the nose reference (the center of the most protrusive points), allowing for one shift parameter to account for imprecise nose localization. The 3-parameter search to match two profiles is thus replaced by a 1-dimensional minimization.

5. GREY LEVEL COMPARISON

5.1. Motivations for grey level analysis

A grey level analysis complements the 3-D processing steps performed so far.

3D acquisition problems arise from grey level disturbances such as in facial hairs, eyes, nostril and mouth. A possible improvement is to avoid 3D extraction in such regions.

Until now, 3D comparison by matching has been carried out with the sole 3D geometrical information. Noise or local minima may prevent from reaching the optimal solution. The inclusion of grey level based matching could speed up comparison and reduce local minima by giving more accurate initial conditions.

For face recognition, 3D and texture clues are likely to be weakly correlated, so that their combination should enhance performances. 3D matching mainly concerns areas of the forehead, cheeks and chin, where grey information is weak. On the contrary, grey level features are related to parts where 3D sensing is difficult or inaccurate. A grey level analysis can also incorporate facial hairs localization, and skin, eye or facial hairs colour.

In this paper, we consider the inclusion of grey level clues to improve the recognition performances, by an a posteriori combination with 3D comparison scores.

5.2. Grey level measurement

One way to get grey level values in registration with 3D data is to read between the stripes of the striped image. This requires no extra image acquisition or storage and 3D and grey level values are in perfect alignment. Proesmans [15] proposed a nice method to do this, although our projected stripes are too thick to obtain results of a similar quality.

Nevertheless, striped images involve important light reflections due to the directionality of the projected beam. Also, the stripes largely influence grey levels in their neighbourhood. Before trying to compensate for these effects, we preferred to compare grey level images obtained by switching the projector off (see Fig. 5). It seems that ambient light was sufficiently isotropic in grey level images to neglect corrections for reflections.

5.3. Geometry compensation

We first decided to get a 1-D profile of grey level values along the central line (passing through the nose). Because distances between points of a 2-D image highly depend on pose, we used the 3D coordinates of the automatically extracted central profile (see 4.2) to derive a Euclidian indexing of points along the profile. The definition of this profile (intersection with the vertical symmetry plane) also calls for a better reproduction than a straight line on a 2-D image.

Figure 5. Corresponding grey and striped images from the database

Since the grey image is registered with the striped image, striped image coordinates of the profile points were used to get grey level values from the grey image. To reduce the influence of noise and to extract more information from eyebrows and eyes, for each point of the profile, we averaged grey values from a direction perpendicular to the central profile (up to 4 cm each side). To include more information, we later extracted grey profiles from the lateral profiles obtained during the 3D analysis (see 4.2). These left and right grey profiles were summed to increase robustness.

5.4. Grey level compensation

Absolute grey level values are not invariant. They depend on illumination.

A first compensation was achieved by considering the local difference of grey levels along the profile. This reduces the dependence of grey measures on ambient light and reinforces the importance of the position of grey level features relatively to their grey values.

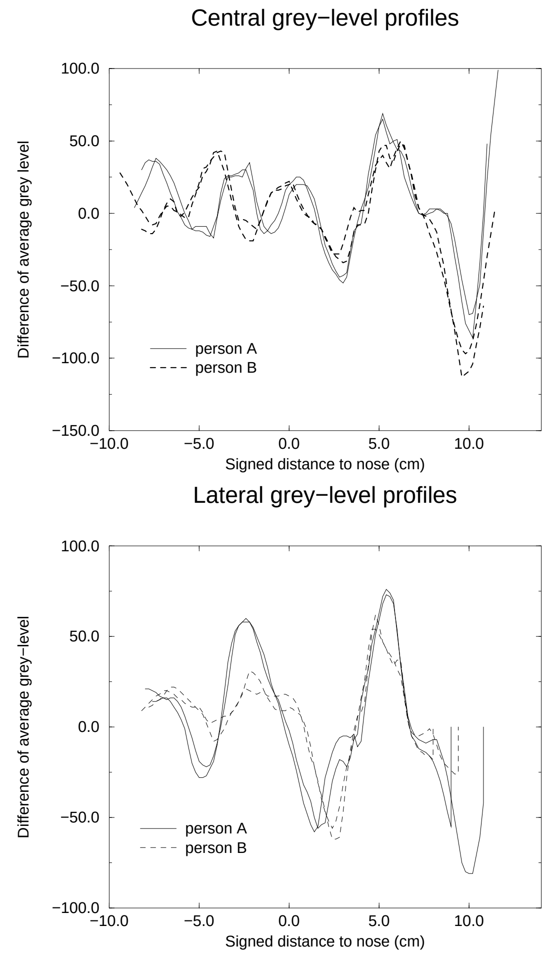

Figure 6. Central and lateral grey level profiles of 2 persons, both sessions

Another compensation should take into account the influence of the local surface orientation on the grey level. From the 3D description and a reflectance model, one is able to derive the albedo, a surface characteristic independent of viewpoint and illumination, and better suited to comparison. However, the grey images had a rather diffused illumination so that grey levels were used without correction.

5.5. Grey profiles

The grey level extraction and compensation process delivers two 2D curves indexed by the signed distance to the nose reference point (negative towards chin, positive towards forehead), giving the local difference (along the profile) of average grey levels (across the profile). See Fig. 6.

One diagram is related to the central profile, considering pixels up to 4 cm away from the vertical symmetry plane. The other diagram corresponds to the mean of the left and right lateral grey level profiles.

5.6. Grey profile comparison

The geometry and grey level compensations allow the direct comparison of extracted grey profiles. However, the nose point used as reference may suffer from imprecision so that several shifts (-1cm..+1cm) between profiles were considered.

For each shift, the histogram of the difference of corresponding profile values is computed. The minimum value of the mean (for all shifts) of the 95 % lowest bins of the histogram is the distance measure. Getting rid of the 5 % highest differences is a way to eliminate spurious or non representative values.

6. RESULTS

Tables of Equal Error Rates (EER) for recognition from 3D central and lateral profiles are given in [6]. We will give here results which are useful for fusion with grey level profiles.

As only one grey level image in alignment with a 3D shot is available for the two sessions of the database, the number of false rejection tests are limited to the number of people. For compatibility with previous experiments, the first 30 alphabetically ordered people were retained. As 3D acquisition errors concerned three persons of session1 and one of session2, 26 clients (and 29 impostors) were used in 3D comparison experiments. Grey level comparisons considered the same population to allow combination of results with 3D.

We summarize below the recognition performances of 3D analysis, grey comparison and their fusion, for comparison of session2 data with session1. All figures concern automatic processing of all steps (3D acquisition, profile extraction and comparison, and grey level measurement and comparison). Although EER values are sensitive due to the limited number of False Rejection tests, we see a clear advantage of combining 3D and grey analysis. The fusion was realized by a weighted sum of the 3D and grey scores. Two fusion possibilities arose (see Table 1): fusing 3D and grey level (2.5 %) or fusing central and lateral profiles (2.0 %).

Three main causes of errors have been identified.

6.1. Error from representation

Each person is currently represented by one grey level image with aligned 3D data per session. From

sess2/sess1	3D Geometry	Grey level	Fusion
Central	12.0 %	10.0 %	5.0 %
Lateral	8.0 %	16.0 %	4.0 %
Fusion	4.0 %	8.0 %	2.5% 2.0%

Table 1. EER for comparison of shot 1 of each session, by geometry and grey level analysis

the two available sessions, one was used as reference and the other for test, limiting to 26 the number of False Rejection tests. Any important change in attitude or physical appearance largely influences the EER. Performances should benefit from a better representation of the people (several shots from different situations).

6.2. Error from acquisition

The 3D acquisition from striped images is not perfect. This obliged us to reject 4 subjects from the tests and affects in general the quality of the representations.

Grey level values also suffer from acquisition errors. On the one hand, image grey values depend on ambient light or reflections. On the other hand, our grey measurement method relies on 3D profile extraction, subject to imprecision. Due to this, the 3D and grey data are not independent.

6.3. Error from matching

3D comparison is carried out by an optimisation procedure which can fail due to noise, local minima or bad initial parameter values. Grey level matching only concerns a shift parameter and is less sensitive to these problems.

6.4. Discussion of the method

The presented grey level analysis was successfully integrated in the existing 3D comparison method from central and lateral profiles. First, grey measurements are obtained directly if one disposes of a grey level image in alignment with the striped image. Under the same assumption, grey normalization based on Euclidian distances and surface normal (for reflection compensation), can be derived from the 3D information. Thirdly, the grey profile comparison procedure is not critical and the one presented here benefits from simplicity and quickness. The current implementation takes 0.5 second to perform central and lateral profile extraction. 0.1 second is needed to compare two faces by the proposed 3D and grey analysis.

As presented in section 5, the simultaneous analysis of grey and geometrical information could be further investigated. Grey level measures can help 3D acquisition and 3D profile extraction. Some gain could be obtained with little effort by comparing 3D and grey profiles simultaneously, using the same shift parameter.

7. Conclusions

Face recognition from 3D facial and grey level clues has been presented. Facial surfaces are acquired by an original, cheap and fast structured light system. 3D comparison is performed by central and lateral profile matching. Grey analysis is made along these profiles, using 3D information for distance normalization.

The recognition rates, improved by the combination of 3D and grey data, supports the approach. The time latency of 2 seconds to get a 3D representation and compare it to the claimed reference, is compatible with a practical application. Background removal, rotation and scale independence are important assets of the method, and infrared projection should further guarantee application success, offering more comfort for the user and being more discrete.

8. Acknowledgements

This work was partially supported by the European ACTS programme (AC102 “M2VTS”).

References

[1] B. Achermann, X. Jiang, H. Bunke, “Face Recognition Using Range Images”, In *Proceedings International Conference on Virtual Systems and MultiMedia ’97 (VSMM ’97)*, Geneva, Switzerland, pp 129-136, September 1997.

[2] M. Acheroy, C. Beumier, J. Bigün, G. Chollet, B. Duc, S. Fischer, D. Genoud, P. Lockwood, G. Maitre, S. Pigeon, I. Pitas, K. Sobatta, L. Vandendorpe, “Multi-modal person verification tools using speech and images”, In *Proceedings of the European Conference on Multimedia Applications, Services and Techniques (ECMAST ’96)*, pp 747-761, May 28-30 1996.

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

[4] C. Beumier, M. Acheroy, “3D Facial Surface Acquisition by Structured Light”, In *International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*, Santorini, Greece, 15-17 Sep, 1999, pp 103-106.

[5] C. Beumier, M. Acheroy, “SIC_DB: Multi-Modal Database for Person Authentication”, In *Proceedings of the 10th International Conference on Image Analysis and Processing*, Venice, Italy, 27-29 Sep, 1999, pp 704-708.

[6] C. Beumier, M. Acheroy, “Automatic 3D Face Authentication”, In *Image and Vision Computing*, Vol. 18, No. 4, pp 315-321, Mar 2000.

[7] R. Brunelli, D. Falavigna, “Person Identification Using

Multiple Cues", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-17, No. 10, Oct 1995.

[8] J.Y. Cartoux, J.T. Lapresté, M. Richetin "FACE AUTHENTIFICATION OR RECOGNITION BY PROFILE EXTRACTION FROM RANGE IMAGES", In *IEEE Computer Society Workshop on Interpretation of 3D scenes (1989)*, pp 194-199.

[9] S. Chang, M. Rioux, J. Domey, "Face recognition with range images and intensity images", Optical Engineering, 36 (4), Apr 1997, 1106-1112.

[10] R. Chellappa, C.L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces", In *Proceedings of the IEEE*, vol. 83, No. 5, pp 705-740, May 1995.

[11] Z. Chen, S.-Y. Ho and D.-C. Tseng, "Polyhedral Face Reconstruction and Modeling from a Single Image with Structured Light", IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, No. 3, May/June 1993.

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

[13] H.-C. Huang, M. Ouhyoung and J.-L. Wu, "Automatic feature point extraction on a human face in model-based image coding", *Optical Engineering*, vol. 32, No 7, pp 1571-1580, July 1993.

[14] R. Jarvis, "Range Sensing for Computer Vision", In *Three-Dimensional Object Recognition Systems*, Advances in Image Communication, Volume 1, A.K. Jain and P.J. Flynn (Editors), Elsevier Science Publisher 1993, pp 17-56.

[15] M. Proesmans, L. Van Gool, "A sensor that extracts both 3D shape and surface texture", International Conference on multisensor fusion for intelligent systems, Washington, USA, December 8-11, 1996, 485-492.

[16] Y.F. Wang, A. Mitiche, and J.K. Aggarwal, "Computation of Surface Orientation and Structure Using Grid Coding", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-9, No. 1, Jan 1987, 129-137.

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