# Face verification from 3D and grey level clues

C. Beumier, M. Acheroy

*Signal & Image Centre, Royal Military Academy, Belgium*[1]

**Abstract**

We address in this paper automatic face verification from 3D 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 intensity 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.

*Key words:* Face Verification, 3-D acquisition, 3-D matching, intensity comparison, fusion.

## 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.

[1] E-mail: beumier@elec.rma.ac.be

A huge research effort has been devoted to intensity 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.

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 or medical applications is not a new field of interest, 3D facial recognition activities are still weakly addressed (1; 8; 9; 11) in the literature.

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 the most promising of the approaches considered to compare 3D facial representations. Section 5 explains how grey level values have been integrated in the recognition process. Recognition results are presented in section 7. Section 8 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 (12), structured light limits the additional cost to a projector and its slide, reduces the influence of ambient light and allows the acquisition of volume and texture information in correspondence. Moreover, a single striped image contains volume information, enabling fast capture and little sensitivity to motion. The drawbacks of a structured light system are its relative bulkiness, the need for light power and the limited field

Fig. 1. Corresponding a) striped and b) grey images from the database

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, which fulfils the needs for sitting attitudes in cooperative situations. As seen in Fig. 1, the faces are patterned diagonally by the stripes to have the best compromise between the minimum of interference with facial features and the maximal similarity of stripe density (and thus resolution) between the left and right facial parts.

### *2.3 3D extraction*

Automatic 3D extraction is performed by stripe detection and indexing. Stripes are detected by horizontal gradient and followed vertically. Their relative thickness (Fig. 2) 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 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.

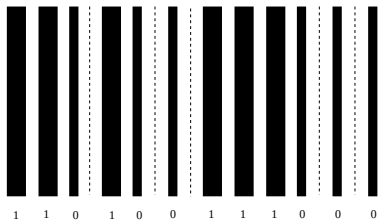


Fig. 2. Slide for projection

## 3 3D database

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

Three sessions (end 97, beginning 98, mid 99) have been completed with more than 100 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. 1 shows a striped image and the corresponding grey image (without projection). Fig. 3 represents 3D reconstructions obtained from the left image of Fig. 1. 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 concerns the glasses (one third of people wearing glasses had to be rejected) and beard, hair and eyes, globally accounting for less than 10 % of rejection.

## 4 3D face comparison

The different approaches considered by the author for the comparison of facial volume information were presented in (6). The most attractive technique relatively to simplicity, performance, memory and speed considerations is based on the extraction and comparison of central and lateral profiles (see Fig. 4).

Matching scaled 3D surfaces is a 6-dimensional optimization consisting of three rotation and three translation parameters. Assuming a vertical symmetry of the faces, three parameters (one translation and two rotations) can be optimized prior to any face comparison by looking for the maximal protrusion of the central profile and the maximal symmetry of left and right lateral profiles. The three remaining parameters are solved when matching the extracted profiles.

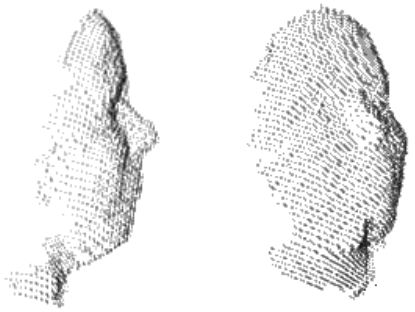

Fig. 3. 3D reconstructions from the left image of Fig. 1

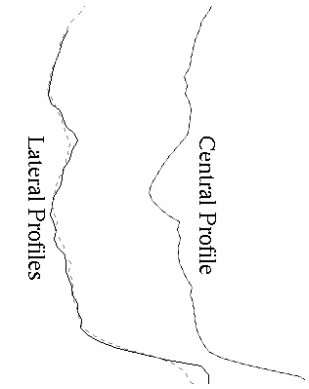


Fig. 4. Central and lateral profiles after intrinsic normalization

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.

First, 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.

Secondly, 3D 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 intensity clues during matching could speed up comparison and reduce local minima by giving more accurate initial conditions.

Finally, facial 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 intensity 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 (13) 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. 1). 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 depend on pose, we used the 3D coordinates of the automatically extracted central profile (see section 4) 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.

Since the grey image is registered with the striped image, striped image coordinates of the profile points were used to get intensity values from the grey image. To reduce the influence of noise and to extract more information from the mouth, nostrils, eyes and eyebrows, 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 better describe the face, we later extracted grey values from the lateral profiles obtained during the 3D analysis (see section 4). We then reduced the bandwidth for average of the central profile to about 2.5 cm and we adopted

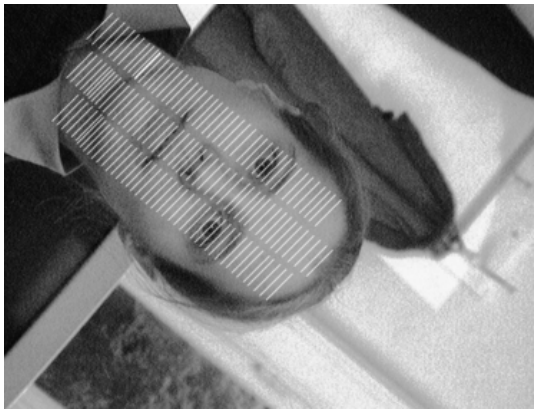

Fig. 5. Areas of grey-level average perpendicularly to the central and lateral profiles

the same width for the lateral profiles (see Fig. 5). 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 intensity values.

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 intensities (across the profile). Fig. 6 shows the compensated grey profiles for the first two sessions of two persons.

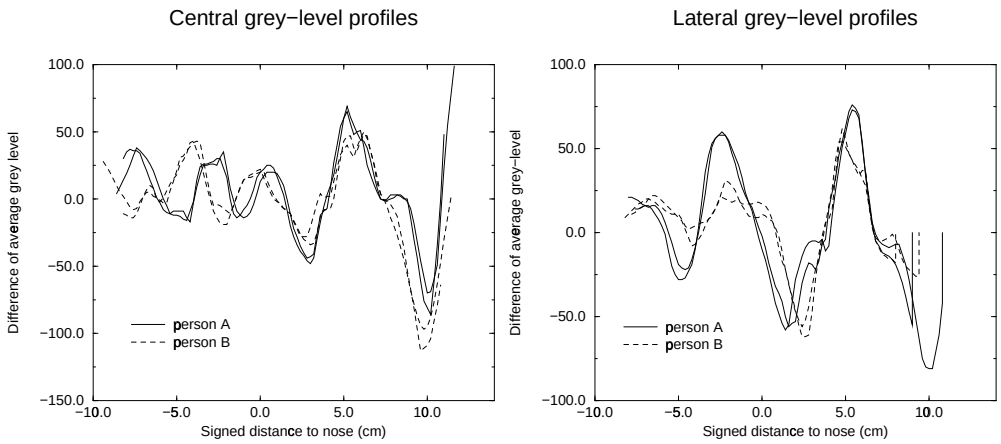


Fig. 6. Central and lateral grey level profiles of 2 persons, sessions 1 and 2

### *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 (−2cm..+2cm) 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 non representative values.

### *5.7 Using striped images*

Thanks to the average perpendicularly to the profiles and diagonally to the stripes, grey level measurements from striped images are not so much influenced by the stripes. Using the striped image simplifies the integration since a single image delivers 3D and intensity data. It also allowed us to conduct more tests thanks to the three shots. The only practical implication was the necessity to increase the bandwith of average (from 2.5 to 3.5 cm) to better smooth out the stripes.

## 6 Fusion

The main objective of the grey level analysis is to access additional information to be combined with the 3D analysis. The fusion of these two experts will probably improve the recognition performance and the robustness of the system. Although the boolean decision of each expert could have been used

("hard fusion"), we preferred to fuse the two experts at their score level ("soft fusion") since each time thresholding is applied, part of the information is lost.

Referring to the study made by (14), we linearly combined the scores of the four experts (namely 3D and grey for the central and lateral profiles) and looked for the minimal error in terms of both false acceptance and false rejection. The coefficients of this linear combination are estimated using the Fisher method which looks for the hyperplane that best separates the client and impostor training scores. A part of client and impostor claims is used to train the fusion engine, while the rest is used to evaluate the operational false acceptance and false rejection rates. The ROC curve drawn from the optimal coefficients represents the false acceptance and false rejection rates for different values of the decision threshold.

## 7 Results

We present here 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).

### *7.1 3D experts*

Tables of Equal Error Rates (EER) for recognition from 3D central and lateral profiles are given in (6). We give in Table 1 the results which are useful for fusion with grey level profiles.

| session2/session1 | Central | Lateral |
|---|---|---|
| 3D Shot 1 | 12.0 % | 8.0 % |
| 3D Shots 1,2,3 | 14.0 % | 12.0 % |
| Grey (grey) Shot 1 | 9.5 % | 16.5 % |
| Grey (striped) Shot 1 | 12.0 % | 17.5 % |
| Grey (striped) Shots 1,2,3 | 16.0 % | 20.5 % |

Table 1
EER from 3D and grey analysis for the central and lateral profiles

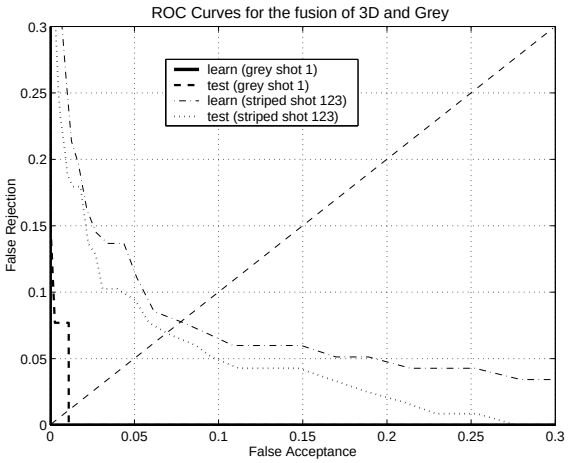


Fig. 7. ROC Curves for the fusion of 3D and Grey, with grey or striped images

### *7.2 Grey experts*

In a first experiment, grey measurements used the grey images (only available for shot 1). For compatibility with previous 3D experiments, the first 30 alphabetically ordered people were tested. Due to acquisition errors, three persons of session1 and one of session2 were rejected, limiting to 26 the number of clients claims.

In a second experiment, grey measurements were obtained from the striped images, enabling the simultaneous capture of grey and 3D data from a single image and allowing a better representation of people (three shots per session, 234 client claims). For comparison purposes, tests with the sole shot 1 were also carried out. See Table 1.

### *7.3 Fusion*

We applied the Fisher method to linearly combine the two grey and two 3D experts. The results are presented in the form of ROC curves for different cases of grey analysis. Grey measurements either come from grey images (shot 1) or striped images (shot 1 or shots 1,2,3). The EER is estimated to respectively 1.2 %, 3.0 % (not in Figure) and 7.8 %. Compared to the individual scores of Table 1, the fusion brings a clear advantage in recognition performances.

The difference in performance levels comes from the variability of the images used for tests. Grey images only concern shot 1, with a diffuse illumination and a rather similar frontal pose for the two sessions. Striped images, shot 1, gives similar results thanks to the similar frontal pose and, consequently, the similar light reflection of the projected light. Striped images, all shots,

introduce pose variations with lighting influences and 3D deficiencies.

EER values have to be compared with caution, knowing that the number of False Rejection tests were small. These values were obtained using a fully automatic version of all processing steps. We are convinced that the figures are underestimated, according to the identification of the three main causes of errors detailed in the next section.

### *7.4 Main sources of error*

#### *7.4.1 Error from representation*

In the tests related to shot 1 only, the number of false rejection tests were limited to 26, implying a high sensitivity to change in attitude or physical appearance. Performances should benefit from a better representation of people, disposing of several shots from different situations. In the tests involving 3 shots, it was possible to take advantage of the three references by combining the matching scores. Using a simple sum, the EER of the fusion of 3D and grey experts fell from 7.8 % to 4.5 %. Fusing the scores of several captures is another way to better represent the people. This last enhancement is more time consuming.

#### *7.4.2 Error from acquisition*

Our 3D acquisition prototype is not perfect. This obliged us to reject four 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.

#### *7.4.3 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. The 3D and grey matching were performed independently. The constraint that the shift parameter should be consistent should bring more discriminative power.

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. We also showed that using the striped images for grey measurement implied a small performance penalty and allows easy 3D and intensity capture from the same image.

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.

## 8 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 infra-red projection should further guarantee application success, offering more comfort for the user and being more discrete.

## 9 Acknowledgements

This work was partially supported by the European ACTS programme (AC102 "M2VTS"). Many thanks to Stéphane Pigeon for his fusion expertise.

## References


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

[2] 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., 1996. 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.

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

[4] Beumier, C., Acheroy, M., 1999. 3D Facial Surface Acquisition by Structured Light. International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, Santorini, Greece, September 1999, 103-106.

[5] Beumier, C., Acheroy, M., 1999. SIC_DB: Multi-Modal Database for Person Authentication. Proceedings of the 10th International Conference on Image Analysis and Processing, Venice, Italy, September 1999, 704-708.

[6] Beumier, C., Acheroy, M., 2000. Automatic 3D Face Authentication. Image and Vision Computing 18 (4), 315-321.

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

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

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

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

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

[12] Jarvis, R., 1993. 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] Proesmans, M., Van Gool L., 1996. A sensor that extracts both 3D shape and surface texture. International Conference on multisensor fusion for intelligent systems, Washington, USA, December 1996, 485-492.

[14] Verlinde, P., Chollet, G., Acheroy, M., 2000. Multi-Modal Identity Verification Using Expert Fusion. Information Fusion, 1(1), 17-33, July 2000.