Automatic Face Authentication from 3D surface

Charles Beumier and Marc Acheroy
Royal Military Academy, Signal & Image Centre (c/o ELEC)
Avenue de la Renaissance, 30 B1000 Brussels, Belgium
{beumier, Marc.Acheroy}@elec.rma.ac.be
http://www.sic.rma.ac.be/index.html

Abstract

This paper presents automatic face authentication based on facial surface analysis. The success of a previous profile-based approach, exclusively relying on geometrical features of the external contour, led us to consider the full facial surface. This motivation was further supported by the independence of viewpoint and lighting conditions of 3D information. The geometry also carries information which is complementary to grey-level based approaches, supporting the combination with those techniques. The facial surface is captured by a system based on structured light and adapted to face to deliver a cheap, fast and sufficiently precise solution. Typical applications concern security in cooperative situations.

1 Introduction

More and more developments in the field of security concentrate on biometric solutions in order to get rid of PIN codes and cards which can be stolen or lost. Among the possible clues, speech and face modalities receive the largest acceptance from the users, but they still lack reliability in real situations. In order to bring robustness with limited development efforts, several modalities (speech, profile, face and 3D) can be combined [1, 2].

A previous profile analysis [3] 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 [4].

More information than the single profile is to be found in the whole facial surface. Particularly, the chin, nose, forehead and cheek regions will bring important clues, precisely where grey-level features lack. Real 3D measures will help solving scale and rotation dependence typically encountered in 2D analysis. Depth segmentation is a trivial way to highlight the face out of 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 [5] or medical applications is not a new field of interest, 3D facial identification activities are still poorly addressed [6, 7] in the literature in comparison with frontal or profile developments.

The success of the 3D approach largely depends on the quality and cost of the 3D data. We designed an active 3D acquisition prototype based on structured light which is adapted to facial surface acquisition. Its resolution, high speed and sufficient facial cover

for a low price make it appropriate for practical implementations. The selected solution also allows discretion thanks to infrared lighting and texture capture in alignment with 3D by switching the projector on and off.

The next section describes the structured light acquisition system. The hardware choices are motivated and the calibration and 3D extraction procedures are briefly explained as they are out of the scope of this article. Section 3 reviews four different approaches we considered to compare 3D facial representations: a direct use of striped images, a feature extraction approach and two facial surface matching algorithms, one globally matching the facial surfaces and the second using the symmetry of the face. Section 4 concludes the paper.

2 3D Acquisition

2.1 Motivations for structured light

Among the possible range acquisition systems [8], passive stereo techniques were rejected due to their slowness and problems encountered in the non-textured zones of the face. On the other hand, structured light techniques, actively projecting a given pattern (in our case parallel 'stripes'), capture depth information from the deformation of the light pattern. Four reasons supported the structured light solution.

First, the additional cost is limited to a projector and a slide. Common cameras are precise enough to get most of the geometrical information of faces.

Secondly, a standard camera benefits from the low price and high speed of video hardware. With an appropriate slide, a single image with stripes suffices to recover 3D information. This enables 3D sequence analysis and time integration.

Thirdly, switching the projector on and off is a simple method to acquire the geometry and texture in correspondence.

Fourthly, the projector illumination reduces the influence of ambient light and allows dark situations. In particular, near infra-red light is more discreet and does not dazzle the individual.

The bulkiness of the structured light system is not a drawback compared to stereo techniques which use a second camera. Other range techniques such as depth from motion or from shading, although using one camera, are more complex and slow. The limited field of depth of structured light systems due to the camera and projector lenses constrains object or subject positioning, but automatically hides out-of-focus background.

2.2 Hardware choices

To keep investments low, we opted for off-the-shelf components.

A standard CCD black and white camera is plugged into a 768x576 pixels image digitiser. A 24x36mm projector is used as light source. The slide is made of glass, for good mechanical stability. The pattern is composed of parallel stripes of different thickness (either thick or thin) to code the identity of each stripe in the thickness distribution of neighbouring stripes. The solution is quicker, simpler and cheaper than color or sequencial pattern encoding.

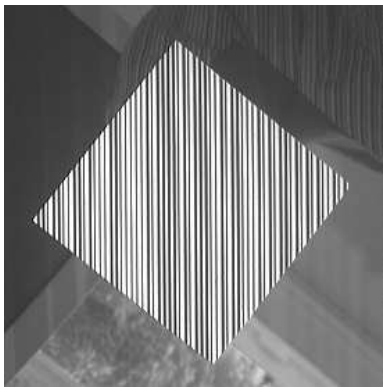

Figure 1: A typical image used for calibration

2.3 Set-up

The camera and the projector have been fixed on a rail. They can be both rotated around one axis, but their optical axes are kept co-planar. This reduces the number of calibration parameters.

Both optical systems have a limited span and depth of focus. We chose lenses to work at 1m40 from the camera/projector head; the field of view covers about 30x40 cm and the depth of focus is about 40 cm. This is sufficient for sitting attitudes in cooperative situations.

2.4 Calibration

The first calibration step consists in roughly measuring the distance and relative angle of the camera and projector. 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 a square object of known size (see Fig. 1) is presented 5 to 10 times in different positions, anywhere in the field of view. The four corners are extracted and corresponding 3D vertices are derived. An automatic procedure refines calibration parameters to bring the four vertices from each image in relative 3D positions coherent with known interdistances and planarity.This calibration procedure has to be done once, as long as the camera and projector settings are not modified.

2.5 3D extraction

Automatic 3D extraction from striped image is done by stripe detection and labelling. Each point of a stripe gives two coordinates which are converted, thanks to the stripe label, into X, Y and Z estimations by triangulation, using the calibration parameters.

Stripe detection is carried out by line following helped by the linear nature of the slide. Grey-level profiles across the stripes allows to classify each stripe as thick or thin. Thickness distribution of neighbouring stripes gives initial stripe labelling thanks to the known thickness distribution of the slide. This initial labeling is checked against normal

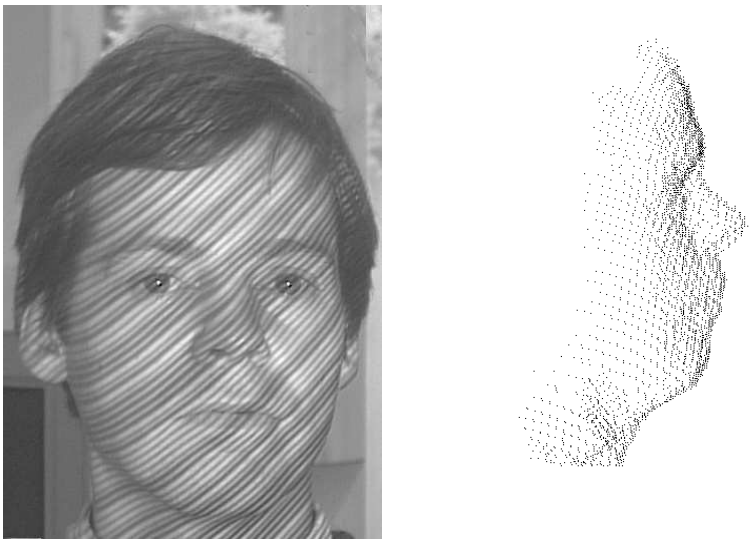

Figure 2: A striped image of a face and its 3D reconstruction from profile

ordering and spacing of stripes to 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 output is a set of ordered points along the stripes from which a mesh is easily derived.

This implementation is very fast (less than 1 second on a Pentium 200) while offering sufficient resolution for recognition purposes as we will see. For a mainly frontal posture, a comfortable cover of the face is acquired, nearly from ear to ear and including the throat. Background objects do not confuse face extraction, as projected stripes are normally out of focus on such objects. Typical problems concern noses and eyes which often disturb the visibility of stripes. Bushy or dark beards and glasses with thick frames impair stripe detection in the concerned regions but grey-level support will be helpful in those areas.

2.6 Database

In order to test the acquisition system and to later estimate the performance of the 3D analysis, a database of 120 persons was recorded. Each individual was asked to sit on a chair and to look in the direction of the camera. Three shots were taken with little posture changes (about 10^o up/down or left/right orientation changes).

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 the problems encountered in bushy beards, glasses, nose and eyes, by order of importance. The quality of the 3D capture was later supported by recognition experiments.

3 3D Face Comparison

3.1 Analysis from Striped Images

We first analysed the 3D information directly from stripe deformation present in the 2D images, without performing the explicit 3D conversion step which consumes time and requests calibration. Similar approaches are part of the studies carried out by some researchers [9, 10], but the specificity of the face makes it inappropriate for a complete analysis, mainly due to the influence of the viewpoint on the shape of the stripes. Only the prominence of the nose and the chin led to their own localisation.

3.2 3D Feature Extraction

Concluding from the previous section that the 3D extraction was necessary to benefit from independence of volume information relative to rotation and scale, we looked for discriminant (different among people) and reproducible (stable for a given person) features. The description in terms of features synthesizes the 3D data into a more compact representation, leading to an easier and quicker comparison. Also, feature by feature analysis eases development and control during recognition.

The prominence of the nose was estimated relative to points of the cheeks located at a given distance from the nose tip. Among 10 individuals, the values were stable for each person (variations less than 1 mm) although spreading with a span larger than 5 mm.

The nose length was measured by localising the nose tip and the nose saddle (between the eyes). Although this measure was less precise, it brought information thanks to the large variability of the nose length among individuals.

However, the nose seems to be the only facial part providing robust geometrical features for limited effort. Mouthes and eyes may involve disturbances. Foreheads and chins, interesting rigid parts, don't clearly exhibit reference points for normalisation.

We abandoned feature extraction and considered the global matching of the facial surface. The first motivation was to completely normalise the two 3D representations to be compared. A second objective was to study the surface similarity after normalisation as a possible acceptance or rejection criterium.

3.3 Global Surface Matching

The global 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) so that the distance measure is minimal.

The problem of the global approach is its large computational load. Since the face surfaces are captured from different points of view, we must consider the five degrees of freedom (3 rotations and 2 translations). Also, a point to point correspondence must be established between the 2 surfaces to be compared.

To solve the correspondence problem, parallel planes, with an interdistance of 1 cm, are used to extract at most 15 (-7cm .. +7cm) profiles (see Fig. 3). To match two facial surfaces, the corresponding profiles are compared two by two to issue a profile distance based on the area separating the profile pairs. The global distance, which is computed as the sum of the profile pair distances, is minimised by tuning the 5 parameters. To reduce the number of comparisons, the minimisation is made iterative, tuning one parameter

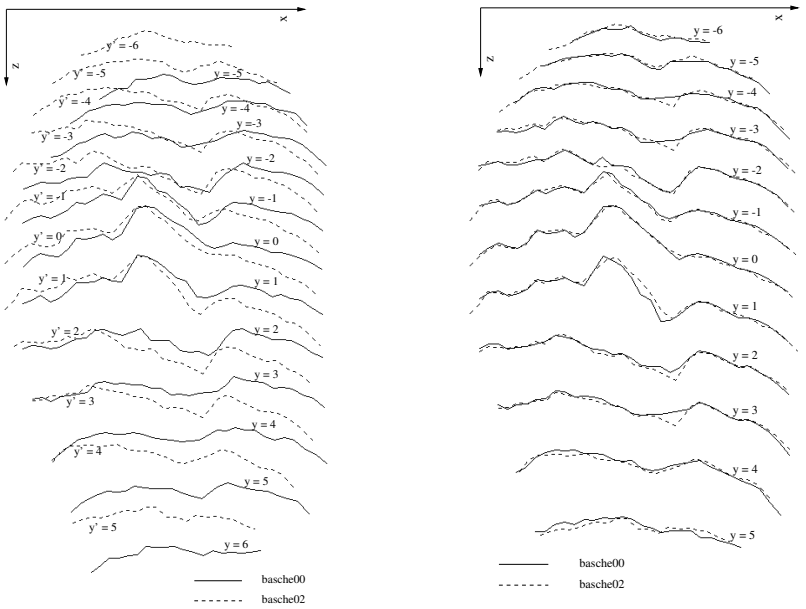

Figure 3: a) Profiles from two 3D representations with noses already in correspondence. b) Profiles of the representations after surface matching

at a time, and organising cycles of minimisations with decreasing search space. On the average, 10 cycles of successive parameter optimisations were necessary, what took about 5 seconds on a Pentium 200. See the results in Fig. 3b.

Although the approach was validated by a large number of successful experiments, the optimisation sometimes falls in a local minimum, either due to bad initial parameter values or noise in the 3D data (typically due to beard, glasses and nose disturbances).

To measure the recognition performance of the 3D approach, we applied an automatic version of the surface matching algorithm, using the residual distance after matching as a similarity measure between people. Comparing the first 30 people of the database and rejecting 3 individuals due to clear 3D acquisition problems (beard and glasses), 81 client and 3159 impostor tests were carried out, leading to an EER (Equal Error Rate) of 14.5% (Fig. 4). Then, we manually refined the surface matching for clients and best impostors to reduce the influence of local minima in the optimisation process. The obtained EER of 4.5% (see Fig. 4) is very encouraging, considering that the 3 shots were taken from different angles and that only one shot was used as reference to take the acceptance/rejection decision. Additional client and impostor tests out of the remaining 90 persons and results on a second session of three presentations confirmed this EER.

3.4 Facial symmetry

As presented in the previous section, surface matching suffers from a high computational load. Several possibilities can improve the timing figures.

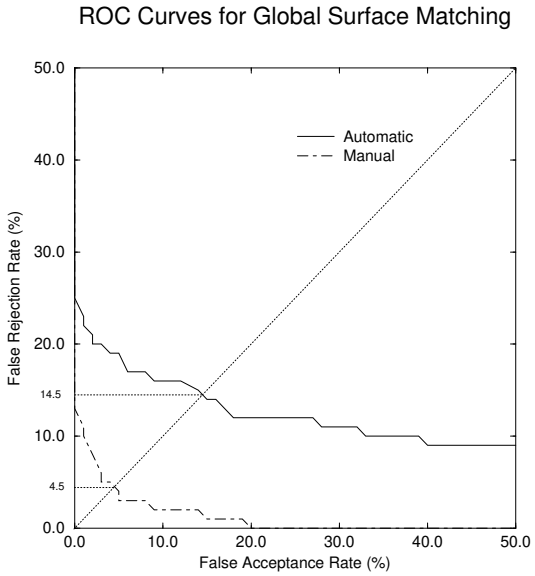

Figure 4: ROC curves of 3D surface matching for part of the database, with (*Manual*) and without (*Automatic*) manual tuning (see text)

First, refined initial estimates of the parameters help to reduce the search space of the related parameters. These initial estimates must be robust enough to avoid local minima.

Secondly, the transitivity of relative translation and orientation values can be used. For instance, knowing the relative positioning between references of the same person by an off-line procedure, the relative positioning of a test representation with one reference gives good initial estimates of relative positioning with other references.

Thirdly, the parameter set may sometimes be divided into subsets optimised separately. The high-dimensional search space is then replaced by two low-dimensional search spaces. In our case, the vertical symmetry of the face allows for a first normalisation, intrinsic to the face, and depending on three parameters (two rotations and one translation). The remaining two parameters are tuned when the comparison with another representation takes place.

Until now, we have concentrated research on this third possibility. More precisely, three parallel, mainly vertical planes are used to extract profiles from the facial 3D representation to be compared. The central profile passes through a rough localisation of the nose tip and the two lateral profiles are 3 cm away from it. The horizontal translation, Left/ Right rotation and tilt rotation parameters are tuned to maximise the similarity of the lateral profiles and the prominence of the central profile.

Although the face is not perfectly symmetrical, this algorithm leads quickly (about 0.5 second on a pentium 200) to a clear optimum. Moreover, this intrinsic optimisation can be done off-line for the database representations and once for the representation to be recognised. Effective 3D comparison then only concerns the remaining two parameters, the vertical translation and Up/Down rotation.

For assessment of this normalisation procedure, automatically extracted central pro-

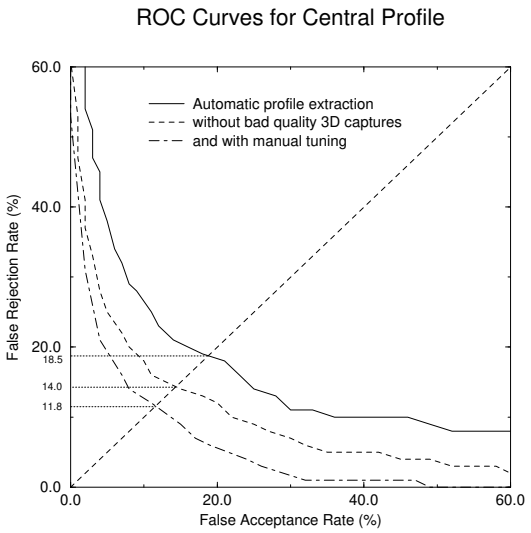

Figure 5: ROC curves for recognition based on central profile automatically extracted, rejecting bad quality 3D acquisitions and with manual refinement

files were compared two by two, measuring the variance of the local angle difference between the profiles for several offsets. This stands for the remaining translation and rotation parameter search.

In a first experiment, all acquisitions of the 120 persons were used to automatically extract and compare central profiles. As Fig. 5 shows, an Equal Error Rate of 18.5% was achieved. We then rejected all persons who had at least one bad representation (either due to 3D acquisition or profile extraction problem) clearly impairing the quality of the central profile. The EER of the profile recognition for the remaining 79 persons was 14%. Finally, we manually tuned central profile extraction. We obtained 11.8% EER.

Then, going on with the same 79 remaining persons and still applying manual refinement, we used the lateral profiles involved in the symmetry optimisation as additional 3D information for recognition. A simple fusion (weighted sum) of the central and lateral profile distances improved the recognition performances from 11.8% and 10.0% respectively, to 6.2% (see Fig. 6). This is similar to the 4.5% obtained by the surface matching approach, considering that 3 profiles (instead of 15) were used. However, the method based on 3 profiles is more sensitive to noise in the data, as shown by the large number of rejected 3D representations. It is also worth noting that lateral and central profiles offer the same level of recognition rate. The central profile, although more discriminant, is more sensitive to acquisition errors in the nose and mouth regions.

Tests performed on a second session composed of three representations per person gave similar results. However, the number of retained persons was 106, much higher than for session 1, mainly because fewer people worn their spectacles for session 2. The comparison of session 2 relative to session 1 gave an EER of 11.5%. This worse result is due to an expression change of some people (rather stressed in session 1 and decontracted or even smiling in session2) and the rarity of people wearing glasses in session 2.

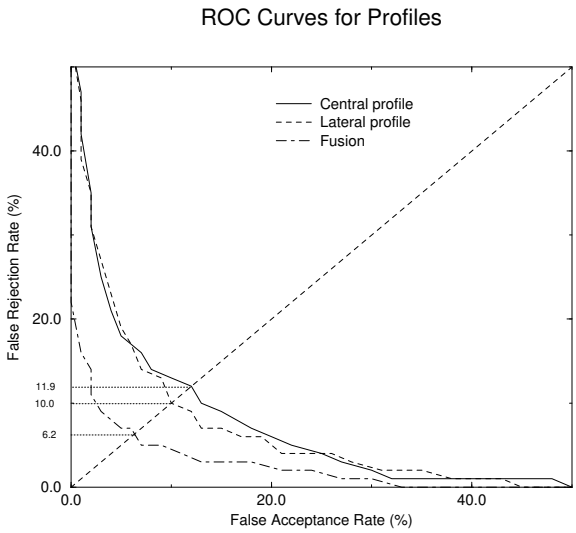

Figure 6: ROC curves for recognition based on central profile, average lateral profile and fusion of both profiles

4 Conclusions

A complete system for automatic person authentication from facial geometry has been presented.

The 3D face acquisition equipment gives appropriate resolution with low cost hardware in cooperative scenarios. Its speed and the adequacy to work with near infra-red projection are additional assets for practical implementations. Further efforts will be devoted to clean typical problems encountered in the nose, eye and ear regions. Beard difficulties are expected to be tackled thanks to grey-level information.

Recognition performances of the surface matching from parallel profiles show its high discrimination power, especially if we consider possible improvements of the acquisition system and the inclusion of additional 3D representations as references. The intrinsic normalisation of the facial surface based on the vertical symmetry assumption is a valid way to speed up facial surface matching. It can also be used to find initial parameter values for the global surface matching which is more robust but which still suffers from slowness in its current implementation.

The proposed solution for facial surface comparison meets the speed and memory requirements of classical security applications. Other potential developments such as facial geometry and grey-level combination or 3D temporal analysis makes the system a challenging face verifier.

5 Acknowledgements

This work has been supported by the European ACTS program (AC102 “M2VTS”).

References

[1] M.P. 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, Louvain-La-Neuve, May 28-30 1996.

[2] R. Brunelli, D. Falavigna, "Person Identification Using Multiple Cues" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-17, No. 10, October 1995.

[3] C. Beumier, M.P. Acheroy, "Automatic Face Identification", In *Applications of Digital Image Processing XVIII*, SPIE, vol. 2564, pp. 311-323, July 1995.

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

[5] Huang Ho-Chao, Ouhyoung Ming and Ja-Ling 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.

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

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

[8] R.A. Jarvis, "A perspective on range finding techniques for computer vision", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-5, No. 2, pp. 122-139, March 1983.

[9] 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.

[10] 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, January 1987.