Automatic 3D Face Authentication

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. This geometrical approach was motivated by difficulties encountered when considering frontal face recognition. Above the advantages of being independent of viewpoint and lighting conditions, the method exploits information which is complementary to grey-level based approaches, enabling the combination 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 authentication purposes. The acquisition system and the 3D face comparison algorithm were designed to be integrated in security applications with cooperative scenario.

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, 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 [4–6].

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 [7] or medical applications is not a new field of interest, 3D facial identification activities are still poorly addressed [8, 9] 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 cover for a low price 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 briefly explained as they are out of the scope of this article. Section 3 reviews three different approaches considered to compare 3D facial representations: a direct use of striped images, a feature extraction approach and a facial surface matching algorithm. Section 4 presents the results of the surface matching approach. Section 5 describes the strategy and improvements to be adopted to bring the current system to a practical implementation. Section 6 concludes the paper.

2 3D Acquisition

2.1 Motivations for structured light

Among the possible range acquisition systems [10], 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. 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. A single image with stripes suffices to recover 3D information. This enables sequence analysis and time integration.

Thirdly, switching the projected pattern on and off is a cheap method to acquire two complementary modalities 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 of neighbouring stripes.

2.3 Set-up

The camera and the projector have been fixed on a rail, keeping 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. 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 is done by stripe detection and labelling. From each point of a stripe and its label, triangulation allows for X, Y and Z estimations.

Stripe detection is carried out by line following helped by the linear nature of the slide. Stripe thickness is estimated from the grey-level profiles accross 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 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. For an nearly frontal posture, a comfortable cover of the face is acquired, 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 only limited errors except if the beard is bushy and the glasses have thick frames.

3 3D Face Comparison

3.1 Analysis from Striped Images

One interesting 3D comparison approach is to get information directly from the 2D (striped) images in order to postpone the time consuming 3D conversion. Although studies were carried out in that direction by some researchers [11, 12], 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.

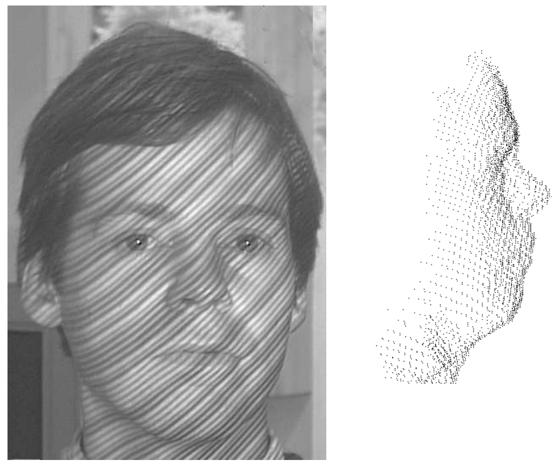

Fig. 1. A striped image of a face and its 3D reconstruction from profile

3.2 3D Feature Extraction

Work has been carried out in looking for discriminant (different among people) and reproducible (stable for a given person) features. The main objective was to reduce the 3D data to a set of features easily and quickly compared.

We estimated the prominence of the nose relative to points of the cheeks located at a given distance from the nose tip. This led to stable values for each person (variations less than 1 mm) with a span of more than 5 mm among 10 individuals.

The nose length was also 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.

3.3 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 geometrical 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. 2). Those planes are initially vertical and centered on the nose tip.

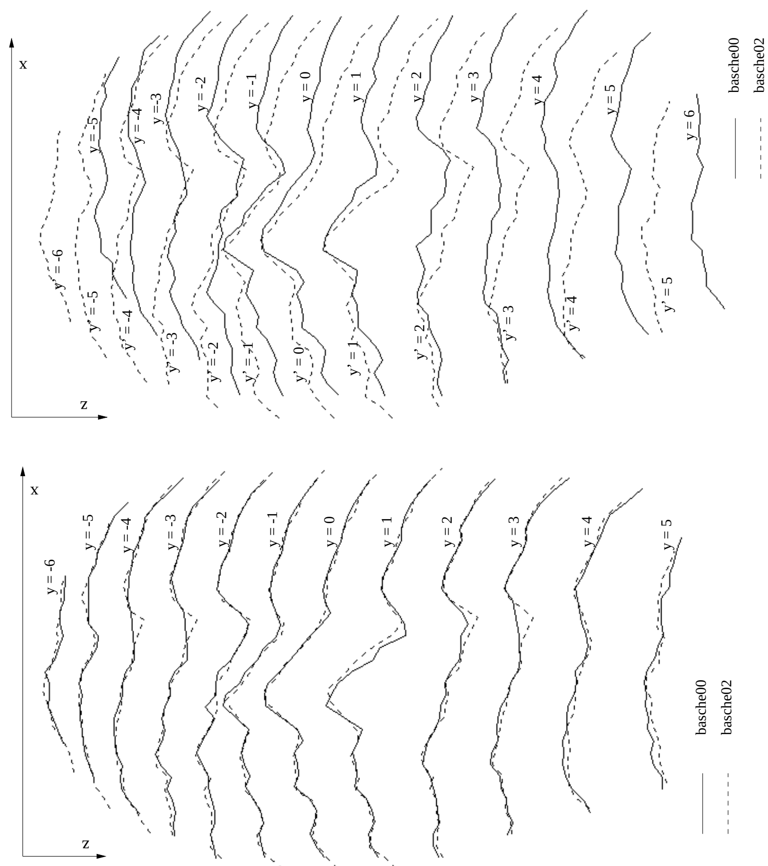

Fig. 2. a) Profiles from two 3D representations with noses already in correspondence. b) Profiles of the representations after surface matching

For different values of the 5 parameters, each profile is compared with the corresponding profile of the other face surface to issue a profile distance. The global distance, which is computed as the sum of the individual profile distances, has to be minimised. To reduce the number of comparisons, we successfully made the optimisation iterative, tuning one parameter at a time, and organising cycles of optimisations with decreasing intervals of search. 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. 2b.

Although the approach was validated by a large number of successful experiments, the optimisation often falls in a local minimum, either due to bad initial parameter values or incorrect input data (noise in the 3D representations). Beard, glasses and nose discontinuities are the most common problems.

4 Results

In order to test the 3D acquisition system and to 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 limited orientation changes (about 10^o) of the head.

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.

To measure the recognition performance of the 3D information, we first applied an automatic version of our surface matching algorithm, using the residual distance after matching as a similarity measure of the people. Comparing 24 people to the 120 people of the database, 72 client and 25848 impostor tests were carried out, leading to an EER (Equal Error Rate) of 16% (Fig. 3). We then rejected 1 person who suffered from a clear 3D acquisition problem (due to beard and glasses). Finally, we manually tuned the automatic surface matching for clients and best impostors to reduce the influence of local minima in the optimisation process. The obtained EER of 4% (see Fig. 3) is very encouraging, considering possible improvements of the acquisition system and the foreseen inclusion of more representations. Additional client tests out of the remaining 96 persons and results on a second session of three presentations confirmed this EER.

5 Strategy

To speed up the surface matching process and reduce the importance of local minima, we propose the following strategy.

First, rough values of the angles and offsets are estimated.

Secondly, the natural vertical symmetry of faces allows for a simpler and quicker normalisation of three parameters. The parameter values, obtained by this intrinsic normalisation, can be saved with the 3D data file.

Thirdly, the remaining 2 parameters are estimated by surface matching. The residual distance is then used as Acceptance/Rejection criterium. Possibly, more discrimination power from the 3D information will be achieved by comparing the normalised 3D data locally or by extracting normalised features.

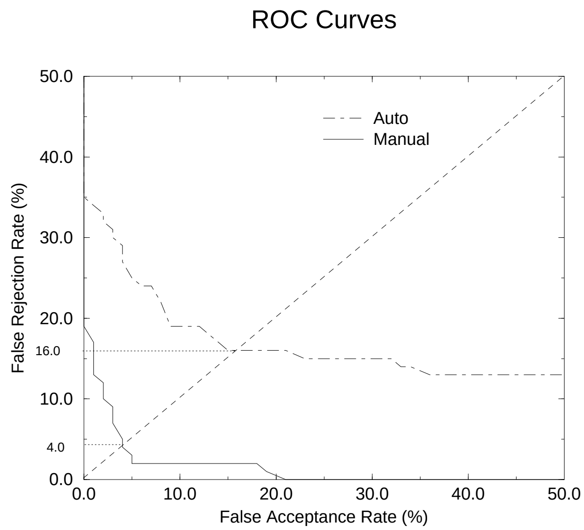

Fig. 3. ROC curves of 3D surface matching for part of the database, with (*Manual*) and without (*Auto*) manual tuning (see text)

Finally, aligned grey-level information, either measured between the stripes or from an acquisition without projection, should increase the performances by its complementarity with 3D information. That grey-level support will be crucial for individuals with facial hair leading to 3D capture difficulties.

6 Conclusions

A complete system for automatic 3D face authentication has been presented.

The 3D acquisition equipment, based on structured light, was adapted to facial surface acquisition to give 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.

We conclude from the current results that surface matching implemented with parallel profiles is a valid way to recognise people from their face surface. The discrimination power seems very high, especially if we foresee the possible improvements of the acquisition system and the further exploitation of normalised data (feature extraction or local comparison).

The proposed strategy for face surface matching also meets the speed and memory requirements of classical security applications. Other potential benefits with small additional efforts such as 3D and grey-level combination or 3D temporal analysis makes the system a challenging face identifier.

7 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, "Multimodal 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, July 1995, pp. 311-323.
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, May 1995, pp. 705-740.
5. L.D. Harmon, S. C. Kuo, P. F. Ramig and U. Raudkivi, "Identification of human face profiles by computer", *Pattern Recognition*, vol 10, pp. 301-312, Pergamon Press Ltd., 1978.
6. S. Pigeon, L. Vandendorpe, "Profile Authentication Using a Chamfer Matching Algorithm", In *Audio- and Video-based Biometric Person Authentication*, Crans-Montana, Switzerland, 12-14 March 1997.
7. 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.
8. G.G. Gordon, "Face recognition based on depth maps and surface curvature", SPIE Geometric methods in Computer Vision, San Diego, vol. 1570, 1991.
9. 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.
10. R.A. Jarvis, "A perspective on range finding techniques for computer vision", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-5, No. 3, March 1983.
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. 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.