3D Facial Surface Acquisition by Structured Light

Charles Beumier, Marc Acheroy

Royal Military Academy, Signal & Image Centre (c/o ELEC)
Avenue de la Renaissance, 30 B1000 Brussels, Belgium
Email: beumier@elec.rma.ac.be

ABSTRACT

This paper presents a 3D acquisition system based on structured light and adapted to 3D facial capture. It was developed in the framework of automatic face recognition with practical constraints so that the design followed speed, low cost and resolution requirements. A single shot with correct posture is sufficient to capture nearly the whole facial surface in half a second. The system has been tested on a database containing more than one hundred persons and has been used to validate person verification from 3D facial analysis.

1. INTRODUCTION

Security applications consider more and more to replace or complement passwords or access cards with biometric solutions. From the available biometric systems, facial analysis remains one of the solutions well accepted by the user.

Although the traditional frontal face analysis showed results highly dependent on illumination conditions, posture and expression, person verification from profile [3] has proven to be efficient in terms of speed, performance and simplicity. We attribute these qualities to the geometrical nature of the profile information, independent of illumination and involving rather rigid parts.

More geometrical clues are available from a full 3D facial representation [2]. This allows for translation, rotation and scale compensation of the point of view. Dealing with range information also simplifies face detection.

To achieve a sufficient level of security, several modalities like speech, profile, frontal can be combined (see [1]). The facial surface is a good candidate for combination as it complements texture images. Accurate geometrical information is obtained where there is little texture such as in the forehead, cheeks or chin without beard. Grey-level or color information is localized more in regions with geometrical disturbancies like the eyebrows, the nose and the mouth. 3D sensing allows for posture and illumination normalization of texture images obtained in correspondence with the facial surface.

This paper is organized as follows. The next section motivates the structured light approach for 3D acquisition. Section 3 presents hardware choices for the retained solution and section 4 explains its calibration procedure. Section 5 lists the steps necessary to recover 3D information from a striped image and section 6 summarizes results of practical tests with the system. Section 7 concludes the paper.

2. MOTIVATIONS FOR STRUCTURED LIGHT

Structured light acquisition systems use the projection of a known light pattern (in our case, parallel 'stripes') to recover 3D coordinates. Among 3D acquisition solutions [4], the choice for structured light has been motivated by the following considerations.

First, compared to a classical camera, the additional cost is limited to a projector and its slide. Secondly, a single image with the projected light pattern suffices to recover absolute 3D coordinates. This allows for quick 3D acquisition enabling sequence analysis. Thirdly, we are convinced that most of 3D facial information lie in the resolution offered by normal cameras and projectors. Fourthly, the proper illumination of the projector reduces the influence of ambient light conditions. Finally, texture (without stripes) and 3-D data can be acquired in near registration by switching the projector on and off.

The main disadvantages of a structured light system with slide projection are the relative bulkiness of the camera-projector head and the limited depth of focus due to optical systems.

This work has been partially supported by the European ACTS programme (AC102 'M2VTS')

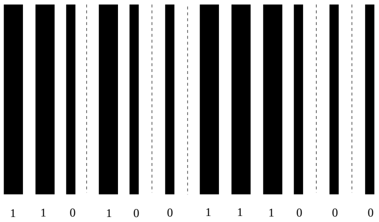

Figure 1: Part of the slide for projection

Figure 2: Image with projected pattern

3. HARDWARE

A normal monochrome CCD camera and a 24x36 mm slide projector have been used. A 1-D stripe pattern consisting of parallel thick and thin stripes has been carefully designed. The index of each stripe is encoded in the binary thickness (thick or thin) of its neighbours.

Optical lenses were chosen to achieve a sufficient field of view (40x30 cm) and depth of focus (30 cm) with limited projected light power, working at a distance of 150 cm. This is compatible with a cooperative scenario for a sitting attitude.

The camera was rotated (portrait) to benefit from the larger horizontal resolution. The 3D acquisition head was rotated of about 45^o (see Fig. 2) so that stripes are nor horizontal (avoiding stripe detection problems in eyebrows or mouth) neither vertical (avoiding a difference in stripe density between left and right).

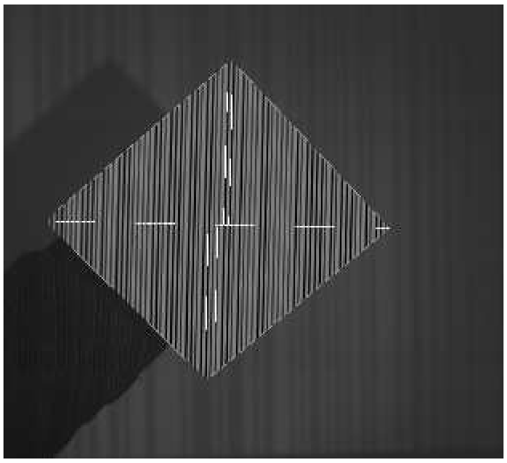

Figure 3: Square object for calibration

4. CALIBRATION

The camera and the projector have been fixed with their optical axis coplanar. In that situation and with the stripe pattern perpendicular to the optical axes plane, the number of parameters is reduced to 7. They concern the camera/projector distance, optical axis angles, pixel to angle and stripe to angle conversion factors.

The first calibration step consists in the collection of distances and optical values to derive rough estimates of the parameters. Then parameter refinement is carried out to bring the corners of a reference square object (see Fig. 3) at correct relative positions. For this, two criteria are used: the maximal planarity and the most compatible interdistances between corners. The planarity is measured from the distance of one corner to the plane defined by the three other corners.

Good calibration results are obtained when optimizing concurrently several (typically 5) presentations of the square object. The residual mean errors after calibration amount to about 1% for both criteria.

5. 3D EXTRACTION FROM STRIPED IMAGE

The 3D coordinates of points along each stripe are derived from the position in the image and the index of the stripe. First, the stripes are detected from horizontal gradient and vertical continuity. Secondly, the stripe thickness (thick or thin) is estimated from the horizontal grey level profile. The thickness distribution of consecutive stripes allows for stripe labelling. Thirdly,

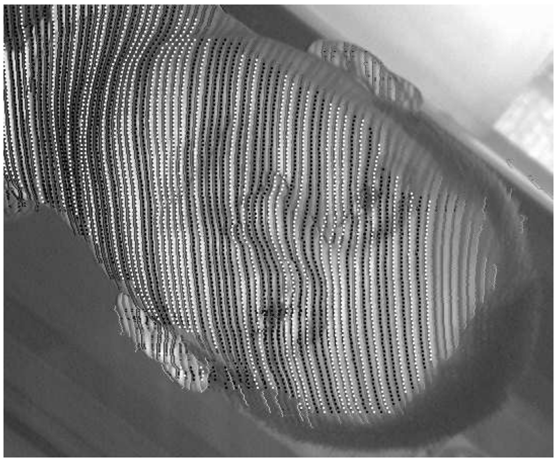

Figure 4: Stripe localization and thickness estimation (thick in black)

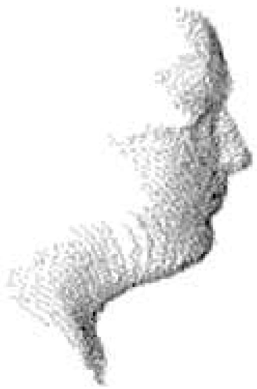

Figure 5: 3D representation from Fig.2

individual labels are globally checked to solve local inconsistencies. Fourthly, median filtering smooths out the surface to clean local bumps incompatible with a 3D facial surface. Finally, the horizontal positions of the stripe points are refined thanks to a subpixel interpolation procedure based on the horizontal grey-level profile. The whole process, from image capture to coordinates delivery takes about 0.5 second on a Pentium 200 MHz.

6. 3D DATABASE AND TESTS

Running the 3D reconstruction algorithm on 120 individuals (2 sessions of 3 shots each, with little orientation changes) 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. Two approaches were developed for facial surface comparison. The first one considers a tenth of parallel planar cuts in the facial surfaces and compares the corresponding planar profiles from the two facial surfaces to be compared. The second approach first localizes the symmetry plane (vertical and passing through the nose) of the head and extracts the central profile and a mean lateral profile. This approach reduces the search in the parameter space for profile matching.

With fully automatic 3D facial acquisition and surface comparison, the EER (Equal Error Rate) of recognition experiments with both methods was about 10%. If manual tuning is applied, the EER falls under 5%. The errors concern acquisition problems and large personal changes due to attitude (e.g. smiles). The sensitivity to attitude is explained by the small representation of the persons: only one image was used as reference in the recognition tests.

7. CONCLUSIONS

The presented system for 3D acquisition by structured light has the advantages of delivering 3D coordinates quickly and automatically, with little background influence. Tests with a large population confirmed those qualities as well as the limitations encountered for bushy beard and spectacles. The system proved adequate to deliver data valuable for 3D face comparison.

8. REFERENCES

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

[2] C. Beumier, M.P. Acheroy, "Automatic Face Authentication from 3D Surface", *British Machine Vision Conference BMVC 98*, University of Southampton, UK, pp. 449-458, 14-17 Sep 1998.

[3] R. Chellappa, C.L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces",

Proceedings of the IEEE, The institute of electrical and electronics engineers, inc., vol. 83, No. 5, pp. 705-740, May 1995.

[4] R. Jarvis, "Range Sensing for Computer Vision", *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.