AUTOMATED PHOTOGRAMMETRIC MEASUREMENT OF HUMAN FACES

Nicola D'Apuzzo
Institute of Geodesy and Photogrammetry
Swiss Federal Institute of Technology
Zurich, Switzerland
E-mail: nicola@geod.ethz.ch

Commission V, Working Group 4

KEY WORDS: Surface measurement, human face, random texture, least squares matching

ABSTRACT

In recent years, modeling and measurement of the human face has gained importance both for medical and computer animation purposes. While for the latter case, a qualitative measurement is sufficient, high accuracy is required for the former. There are several methods currently employed to produce three dimensional computer models of the human face with high accuracy. Among these, the extensively used ones are digital stereo photogrammetry, laser scanning and coded light based triangulation approaches. This paper presents an automated procedure to measure the human face with multi-image photogrammetry. The system consists of five CCD cameras arranged in front of the subject. The images are captured by a single channel frame grabber switching through the five cameras. The system is calibrated using a reference object with coded target points, which can be measured fully automatically. To facilitate image correlation, texture in form of random pattern is projected from two directions onto the face. The stereo matching process which is based on a least squares algorithm, produces a dense set of more than 15000 corresponding points in the five images. The three-dimensional coordinates of the matched points are computed by forward intersection using the results of the calibration process. The achieved mean accuracy is about 0.3 mm in the sagittal direction and about 0.1 mm in the lateral direction. The last step of data processing is the generation of a triangulated surface from the 3D point cloud. The output data can be transferred to miscellaneous visualisation tools or be used for further quantitative analysis in facial surgery.

1. INTRODUCTION

In order to evaluate the anatomical changes occurring with maxillo-facial and plastic surgery, knowledge of the shape and size of the patient face prior to the operation is required. For this purpose we are developing a robust photogrammetric method to measure the surface of human faces with high accuracy. The project involves three groups: the Computer Graphics Research Group of the Department of Computer Science, the Chair of Photogrammetry and Remote Sensing of the Department of Geodetic Sciences, both of ETH Zurich and the Department of Cranio-Maxillofacial Surgery of the University Hospital of Zurich. The aim of the project, is the development of a system for surgical planing and prediction of human facial shape after craniofacial and maxillofacial surgery.

Two different approaches for modeling the human face can be distinguished. One aims high precision and the other aims a good modeling of the facial features without the requirement of high precision. The latter one is mostly devoted to computer animation purposes (Fua, 1997). The former approach is usually used for medical purposes such as measurement for planing a surgical intervention (Thomas, 1996) or for measurement of changes after surgical intervention (Gäbel, 1996). An interesting and attractive use is the forecast of the result of a facial surgical intervention (Koch, 1996). At this stage of the research high precision measurements of the face before and after the surgical intervention are required to prove and test the performance of the system.

Several methods are currently employed to produce three dimensional computer models of the human face. The most commonly used ones include laser scanning, coded light based triangulation approaches and digital photogrammetry. Laser scanning is extensively used for animation purposes. Several products are commercially available and some of them have been specifically developed for the modeling of human faces, e.g. from Cyberware. This systems normally scans the human face in about 30 seconds. The subject has to remain still while the scan platform moves a digitizing unit around the head. The digitizer is composed of a light beam and video cameras to capture all details of the object, colours included. With triangulation or interferometry methods, 3-D coordinates of the scanned points can be quickly computed. These systems give a dense cloud of measured points and are easy to use. The achieved accuracy is limited to 0.5 mm and smooth filters have to be applied on the modeled surface because of its roughness. The long scanning period contributes to the low precision as the subject cannot remain absolute immobile for so long.

The second most frequently used method is the coded light based triangulation approach (Wolf, 1996). Such a system is usually composed of a camera and a programmable projector. Defined sequences of stripe (black/white) images are projected onto the object during the acquisition. A time-space coding of a sequence of n stripe images allows the differentiation of 2^n different projection directions. Given calibrated projector and camera, the depth information can be computed through triangulation using the acquired images. The system is simple to use and is practical to install (only one camera and a projector). For these reasons it has gained importance in the industrial sector. The method is optimal for static objects. For complex objects such as the human face, multiple acquisitions from different directions and with different projection directions are required.

The digital photogrammetry method (Grün, 1989 or Maas, 1992) employs more cameras to acquire stereo images of an object. Matching algorithms can be developed to automatically establish correspondences in the images. The result usually consists of a set of corresponding points. Through an accurate calibration of the interior and exterior parameters of the cameras, the 3-D coordinates of the matched points can be computed with high accuracy. The objects to be measured often show insufficient surface texture, which is necessary for the determination of correspondences. In such cases, an artificial texture is projected onto the surface. It can be for instance a dot pattern, a grid pattern or a random pattern. Since all the information needed to model a surface can be acquired in only one step, digital photogrammetry can be described as a simultaneous method to measure surfaces with high accuracy.

This paper describes a prototype system for an automated measurement and modeling of the human face with digital photogrammetry.

2. METHOD

The method employed is multi image photogrammetry. Five images are taken from different directions. Texture in form of random pattern is projected onto the face from two different directions. The stereo matching process is based on a least squares algorithm which establishes correspondences by minimizing the sum of the square of the differences between the grey levels in patches of the images. The projection of an artificial texture is required to obtain sufficient grey level differences on the surface. The system is calibrated using a reference object with coded target points, which can be measured fully automatically. The computation of the 3-D coordinates is done with forward intersection using the result of the calibration. A meshed surface is generated from the 3-D point cloud and for a photorealistic visualisation, the natural texture is draped over the model of the face.

Early studies to implement the system and to optimize the hardware setup, have been done with a face mask (D'Apuzzo, 1997). In that case three simplifications were present: the stillness of the subject, the use of a single camera to move into the five stations and the possibility of a separate random pattern projection for the left and right side of the face. These differences between the early studies and the presented method make the approach more difficult. The use of 5 cameras requires a finer calibration process than with a single camera. The simultaneous projection from two sides produces poorly focused texture in the region of overlap. During the image acquisition (about 10 seconds), the subject makes natural movements such as respiration or eye movement, those could also reduce the precision of the measurement. The differences to the system with the face mask and a single camera are the causes of the less accurate results of the measurement of real human face compared with the measurement of a face mask.

2.1. Image acquisition

The system consists of five Sony XC-77CE CCD cameras with 35mm lenses (Figure 2) arranged convergently at a distance of 1.5 m from the subject (Figure 1).

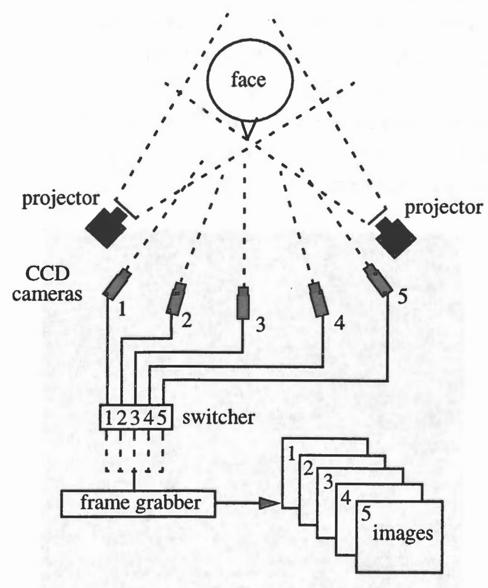

Fig. 1: Setup of cameras and projectors

Fig. 2: CCD camera with 35 mm lens

The cameras are connected to a manual video switcher whose output goes to a DataCell S2200 frame grabber on a Sun Workstation. The size of each image is 768x576 pixels at 8 bits pixel quantisation. The images are taken sequentially switching through the five cameras. This system works manually so that the entire process of image acquisition takes about 10 seconds. During this time the subject has to remain still. For this purposes, the person sits on a dentist chair with head support (Figure 3), which allows a comfortable position.

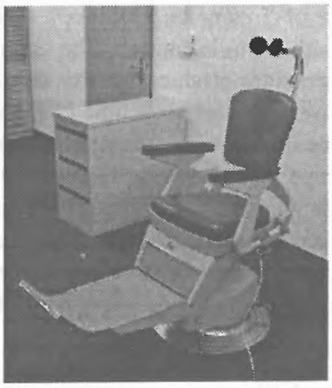

Fig. 3: Dentist chair (courtesy of the University Hospital Zurich)

The natural texture of the human skin presents in wide regions of the face a relative high uniformity. In order to facilitate the estimation of corresponding points in the images of the different views, a random pattern texture is projected onto the face from both left and right direction by two slide projectors. The result is an uniform distribution of the additional texture even on the lateral sides of the face (Figures 4, 5). Such results could not be achieved with only one central projection.

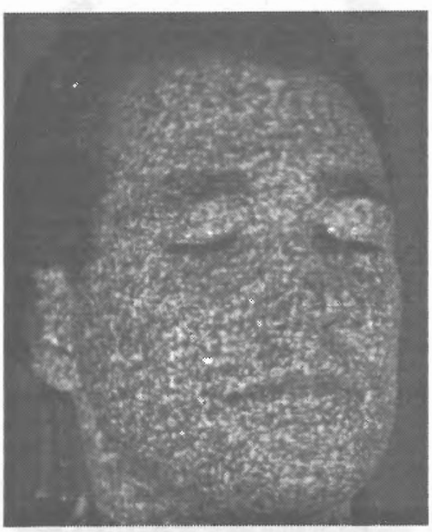

Fig. 4: Face with random pattern projection

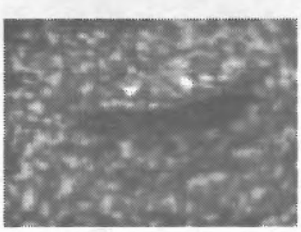

Fig. 5: Zoom around eye

An additional central colour image of the face without random pattern projection is taken by a video camera placed in front of the subject. It will be used for the realisation of a photorealistic visualisation by applying the acquired natural texture to the 3D model of the face.

2.2. Calibration

To increase the automatic functionality of the entire process, the system is calibrated using a 3-D reference frame with coded target points (Figure 6). These are fully automatically recognised and measured in the images (Niederöst, 1996). The external orientation and the internal calibration of the five cameras are determined with the bundle method. The 3D-coordinates of the points have been previously photogrammetrically determined with a mean standard deviation of 20 µm.

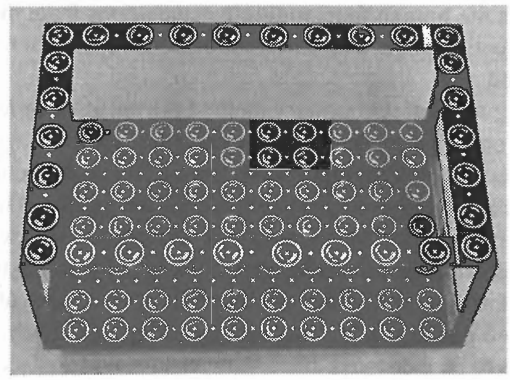

Fig. 6: 3-D calibration frame with coded targets

In order to prevent calibration errors, the five images of the reference object are taken before and after the acquisition of the images of the face.

2.3. Stereo matching

The stereo matcher is based on the adaptive least squares method (Grün, 1985). It considers a patch of area around a selected point. One image is used as template and the others as search images. The patch in the search image is modified by an affine transformation (translation, rotation, sheering and scaling) and the grey levels are varied by multiplicative and additive constants. The algorithm finds the corresponding point in the neighbourhood of the selected point in the search images by minimizing the sum of the square of the differences between the grey levels in these patches. Figure 7 shows the result of the least squares matching with an image patch of 13x13 pixels. The black boxes represent the patches selected and the white box represents the affine transformed patch in the search image.

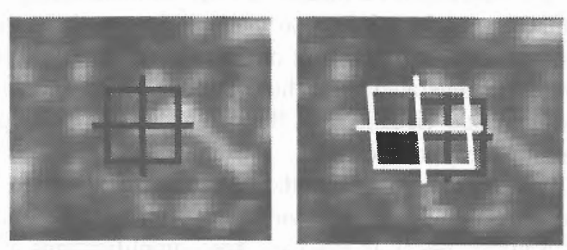

Fig. 7: Result of the least squares algorithm

The stereo matching process works independently from the orientation data. To define the seed points of the proc-

ess, approximations for a few corresponding points (about 10) have to be manually selected in the five images. A least squares algorithm is applied to find their exact location in the pictures. Starting from the seed points, the stereo matcher automatically determines a dense set of corresponding points. The process is done separately for the left and right side of the face. For the left side the images taken by the cameras 1, 2, 3 are used and for the right side the images taken by the cameras 3, 4, 5 (Figure 1). The images 2 and 4 are used as template images. The stereo matcher searches the corresponding points in the two search images (1 and 3 for the left side, resp. 3 and 5 for the right side) independently. At the end of the process, the data sets are merged to become triplets of corresponding points (points matched in the three images 1, 2, 3 for the left side, resp. 3, 4, 5 for the right side). The resulting set of corresponding points is a mixture of triplets and stereo pairs, because of the presence of points matched in two images only (1 and 2, 2 and 3 for the left side, resp. 3 and 4, 4 and 5 for the right side). The process can therefore be defined as quasi-multi-image-matching (but without geometrical constraints).
To define the regions between the different seed points, a Voronoi tessellation is done in the template image. The picture is divided into polyhedral regions according to which of the seed points is the closest (Figure 8). The boundaries are perpendicular to lines joining pairs of neighbouring seed points.

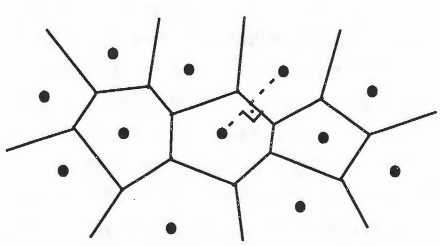

Fig. 8: Seed points Voronoi tessellation of the template image

The search strategy of the stereo matcher is the following: the process starts from one seed point, makes a horizontal shift in the template and in the search image and then the least squares algorithm is applied in the shifted location. If the quality of the match is good, the shift process continues horizontally until the boundaries of the region are reached. The entire polygon region of a seed point is covered with subsequently vertical and horizontal shifts (Figure 9). If the quality of the match is not satisfactory, the algorithm works adaptively by changing parameters (e.g. smaller shift, bigger size of the patch). The normally used value of the shift is 1 pixel but it can be defined as a subpixel value in the cases where the match has not given satisfactory results.

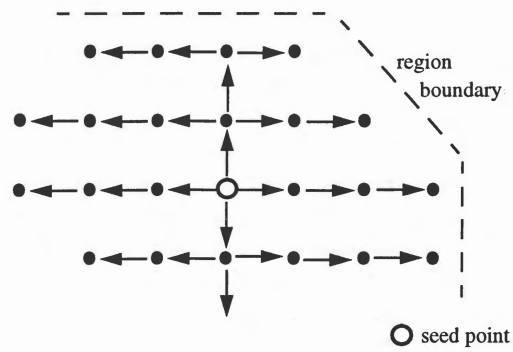

Fig. 9: Search strategy for the establishment of correspondences between images

The search process is repeated for each seed point region until the whole image is covered. At the end of the process it is possible that holes of not analysed areas do appear in the set of matched points. The algorithm tries to close these holes by searching from all directions around.
Different tests have shown consistent results in the matching process: the mean number of matched points on the face is about 15000 and the mean precision of the match is about 0.05 pixel in x- and y-directions in the picture.

2.4. 3-D model of the face

The 3-D coordinates of the matched points are determined by forward intersection using the calibration results. The results show a mean standard deviation of about 0.3 mm in the sagittal direction and about 0.1 mm in the lateral direction, which corresponds to about 0.2 pixel in the image.
With the set of 3-D points a triangulated surface is then generated using the Delauney method (Figure 10).

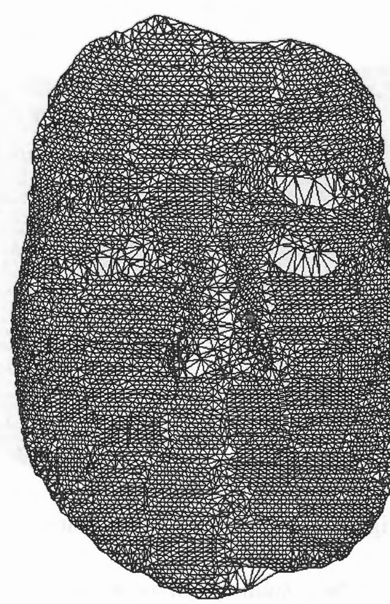

Fig. 10: Triangulation mesh

As can be seen in the figure, the triangulated surface has an irregular mesh structure and contains regions with a low density (eyes, eyebrows). These problems will be discussed in the third paragraph of this paper.

2.5. Visualisation

The results can be visualised as a rendered model of the face (Figure 11). No filters are applied to smooth the surface and no interpolation between points is done, the figure shows raw data. The peaks and discontinuities of the surface represent measuring errors. One aim of the future work is measurements without errors or an automatic removal of the errors.

Fig. 11: Model of the face

A photorealistic visualisation can be achieved by draping the natural texture of the face over the 3D model (Figure 12).

Fig. 12: Photorealistic visualisation

The purpose of the system which we are developing is the measurement of human faces for facial surgical intervention forecasts. The measurement of the face of a patient before and after the surgical intervention is then required. Figure 13 shows an example of the face models of a patient before and after surgery.

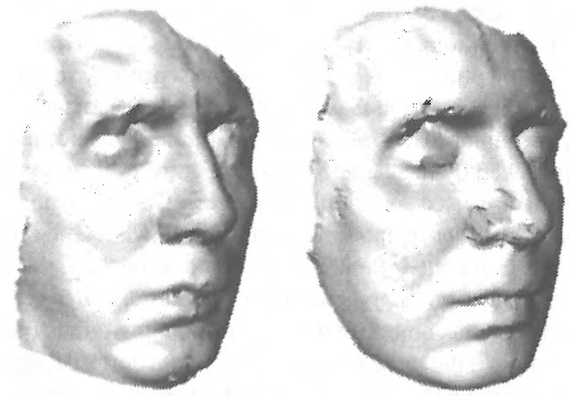

Fig. 13: Left: before, right: after surgical intervention

3. DISCUSSION

As can be seen in figures 11 and 13, measuring errors occur. Three different causes can be distinguished. Firstly, in regions where the texture is insufficient because of the darkness (e.g. eyebrows) or because of the brightness (e.g. regions with strong reflection) the matching process fails so that the meshed surface contains areas without measured points (Figure 10). A second problem appears in regions where the projected random texture is not well focused. In this case the matching process cannot give good results. This problem can be solved by using lenses with aperture for the projectors, that will give a larger depth of field. However, problems will remain in regions where the two projections overlap. The third problem of the matching process occurs in places where big differences between the template and the search images exist (e.g. sides of the nose, the lateral extremes of the face). This is due to the convergent arrangement of the cameras and to topology of the human face. The use of two more cameras or a more accentuated lateral disposition of the five cameras could remove this problem.

4. CONCLUSION AND FUTURE WORK

A photogrammetric method for the measurement of human faces with high accuracy has been described. The project is in the development stage and many improvements of the system can be envisioned. Among these, the most significant could be the implementation of a multi-image geometrically constrained matching algorithm (Grün, 1988), which should reduce the matching errors. A second improvement of the method could be the introduction of an automatic generation of the seed points, thus increasing the automation level of the system. Thirdly, a real 3-D triangulation of the surface has to be implemented or added to the existent method. Until now only a 2.5-D triangulation is computed: the mesh is generated with the projection of the points onto the x,y plane. A definition of an "intelligent" smooth filter to apply to the modeled surface could remove the remaining measurement errors.

A very important improvement on the hardware side would be, instead of the manual switcher, the use of a video multiplexer, which could drastically reduce the time of image acquisition (to less than 1 second). The result would be a better precision of the measurements and the possibility of acquiring sequences of the five images for the reconstruction of facial expression. Finally, the use of other patterns for the projection, such as line pattern has also to be considered.

5. ACKNOWLEDGEMENT

This project is being funded by the Swiss National Science Foundation (SNSF) under no. 2153.049247.96/1.

REFERENCES:

D'Apuzzo N., 1997: Automated measurement of human faces. First step: the face mask. 2nd Workshop on Image Sensing and its Application, Zurich, 1997

Fua P. and Miccio C., 1997: Fitting Sophisticated Facial Animation Models to Image Data. Optical 3D Measurement Techniques IV, pp 3-10

Gäbel H. et al., 1996: Photogrammetric quantification of changes of soft tissue after skeletal treatment of the facial part of the skull. International Archives of Photogrammetry and remote Sensing, 31 (B5), pp 188-193

Grün A., 1985: Adaptive least squares correlation: a powerful image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography, 14 (3), pp 175-187

Grün A. and Baltsavias E. P., 1988: Geometrically Constrained Multiphoto Matching. PERS, vol. 54 (5), pp 633-641

Grün A. and Baltsavias E. P., 1989: Automatic 3-D measurement of human faces with CCD-cameras. Proc. of SPIE, vol. 1030, pp 106-116

Koch R. M. et al., 1996: Simulating Facial Surgery Using Finite Element Models. SIGGGRAPH 96 Conference Proceedings, New Orleans 1996

Maas H. G., 1992: Robust Automatic Surface Reconstruction with Structured Light. International Archives of Photogrammetry and Remote Sensing, 29 (B5), pp 709-713

Niederöst M., 1996: Codierte Zielmarken in der digitalen Nahbereichsphotogrammetrie. Diplomarbeit, Insitut für Geodäsie und Photogrammetrie, ETHZ, Zürich, 1996

Thomas P. R. et al., 1996: Evaluation of a low cost digital photogrammetric system for medical applications. International Archives of Photogrammetry and remote Sensing, 31 (B5), pp 405-410

Wolf H., 1996: Structured Lighting for Upgrading 2D-Vision Systems to 3D. International Symposium on Lasers, Optics and Vision for Productivity and Manufacturing I, pp 10-14, Besançon, 1996