Three-Dimensional Image Capture and Applications V,
Proc. of SPIE, San Jose, California, 2002, Vol. 4661

Modeling Human Faces with Multi-Image Photogrammetry

Nicola D'Apuzzo*, Institute of Geodesy and Photogrammetry, ETH Zürich, Switzerland

ABSTRACT

Modeling and measurement of the human face have been increasing by importance for various purposes. Laser scanning, coded light range digitizers, image-based approaches and digital stereo photogrammetry are the used methods currently employed in medical applications, computer animation, video surveillance, teleconferencing and virtual reality to produce three dimensional computer models of the human face. Depending on the application, different are the requirements. Ours are primarily high accuracy of the measurement and automation in the process. The method presented in this paper is based on multi-image photogrammetry. The equipment, the method and results achieved with this technique are here depicted. The process is composed of five steps: acquisition of multi-images, calibration of the system, establishment of corresponding points in the images, computation of their 3-D coordinates and generation of a surface model. The images captured by five CCD cameras arranged in front of the subject are digitized by a frame grabber. The complete system is calibrated using a reference object with coded target points, which can be measured fully automatically. To facilitate the establishment of correspondences in the images, texture in the form of random patterns can be projected from two directions onto the face. The multi-image matching process, based on a geometrical constrained least squares matching algorithm, produces a dense set of corresponding points in the five images. Neighbourhood filters are then applied on the matching results to remove the errors. After filtering the data, 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 0.2 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 surface model from the point cloud and the application of smooth filters. Moreover, a color texture image can be draped over the model to achieve a photorealistic visualisation. The advantage of the presented method over laser scanning and coded light range digitizers is the acquisition of the source data in a fraction of a second, allowing the measurement of human faces with higher accuracy and the possibility to measure dynamic events like the speech of a person.

Keywords: surface measurement, face modeling, CCD camera, least squares matching

1. INTRODUCTION

Modeling and measurements of the human face have wide applications ranging from medical purposes [1,7,15,21,23] to computer animation [2,16,17,18,24,27], from video surveillance [5] to lip reading systems [20], from video teleconferencing to virtual reality [3,9,11,26]. How realistic and accurate the obtained shape is, how long it takes to get a result, how simple the equipment is and how much the equipment costs are the issues that must be considered to model the face of a real person.

The different approaches to enable the reconstruction of a human face can be classified depending on the requirements. For animation, virtual reality and teleconferencing purposes, the photorealistic aspect is essential. In contrast, high accuracy is required for medical applications. Two major groups can also be distinguished based on their data source: the first using range digitizers and the second using only images.

To date, the most popular measurement technique is laser scanning [13,18,21,23], for example the head scanner of Cyberware [6]. These scanners are expensive and the data are usually noisy, requiring touchups by hand and sometimes manual registration. Another solution is offered by the structured light range digitizers [25,27,28] which are usually composed of a stripe projector and one or more CCD cameras. These can be used for face reconstruction with relatively inexpensive equipment compared to laser scanners. The accuracy of both systems is satisfactory for static objects, however their acquisition time ranges from a couple of seconds to half of a minute, depending on the size of the surface to measure. Thus, a person must remain stationary during the measurement. Not only does this place a burden on the

---

* nicola@geod.baug.ethz.ch; phone +41-1-6333054; fax +41-1-6331101; http://www.photogrammetry.ethz.ch; Inst. of Geodesy and Photogrammetry, ETH-Hoenggerberg, CH-8093 Zurich, Switzerland.

subject, but it is also difficult to obtain stable measurement results. In fact, even when the acquisition time is short, the person moves slightly unconsciously.

A different approach to face modeling uses images as source data. Various image-based techniques have been developed. They can be distinguished by the type of used image data: a single photograph, two orthogonal photographs, a set of images, video sequences or multi-images acquired simultaneously.

Parametric face modeling techniques [2] start from a single photograph to generate a complete 3-D model of the face. Exploiting the statistics of a large data set of 3-D face scans, the face model is built by applying pattern classification methods. The results are impressively realistic, however the accuracy of the reconstructed shape is low.

A number of researchers have proposed creating models from two orthogonal views [14]. Manual intervention is required for the modeling process by selecting feature points in the images. It is basically a simplified method to produce realistic models of human faces. The obtained shape does however not reproduce the real face precisely. To solve this problem, some solutions [16] work in combination with range data acquired by laser scanners.

Another image-based method consists of automatically extracting the contour of the head from a set of images acquired around the person [19,29]. The obtained data are combined to form a volumetric model of the head. The set of images can be generated moving a single camera around the head or having the camera fixed and the face turning. The systems are fast and completely automatic, however the accuracy of the method is low.

Video sequences based methods [11,17,24,26] uses photogrammetric techniques to recover stereo data from the images. A generic 3-D face model is then deformed to fit the recovered (usually noisy) data. These techniques are full automatic but may perform poorly on face with unusual features or other significant deviations from the normal.

High accuracy measurement of real human faces can be achieved by photogrammetric solutions which combine a thorough calibration process with the use of synchronized CCD cameras to acquire simultaneously multi-images [1,3,7,8,20]. To increase the reliability and robustness of the results some techniques use the projection of an artificial texture on the face [1,7]. The high accuracy potential of this approach results however in a time expensive processing.

For our purposes, we are interested in an automatic system to measure the human face relatively fast and with high accuracy. We have therefore chosen a photogrammetric solution. Five synchronized CCD cameras are used to acquire simultaneously multi-images of a human face and artificial random texture is projected onto the face to increase the robustness of the measurement. The processing consists of five steps: acquisition of images of the face from different directions, determination of the camera positions and internal parameters, establishment of dense set of corresponding points in the images, computation of their 3-D coordinates and generation of a surface model. Due to the simultaneous acquisition of all the required data, the proposed method offers the additional opportunity to measure dynamic events.

In this paper, we present the equipment used, the method and the achieved results.

2. METHOD

In this section, are described the system for data acquisition and the method used for its calibration and depicted the methods for the measurement and modeling of the human face from the acquired multi-images.

An advantage of our method is the acquisition of the source data in fractions of a second, allowing the measurement of human faces with high accuracy and the possibility of measuring dynamic events such as speech. Another advantage of our method is that the developed software can be run on a normal home PC reducing the costs of the hardware. We are developing a portable, inexpensive and accurate system for the measurement and modeling of the human face.

2.1 Data acquisition and calibration

Figure 1 shows the setup of the used image acquisition system. It consists of five CCD cameras arranged convergently in front of the subject. When high accuracy is required, texture in form of random pattern can be projected from two directions onto the face. The cameras are connected to a frame grabber which digitizes the images acquired by the five cameras at the resolution of 768x576 pixels with 8 bits quantisation. A color image of the face without random pattern projection is acquired by an additional color video camera placed in front of the subject. It is used for the realization of a photorealistic visualisation.

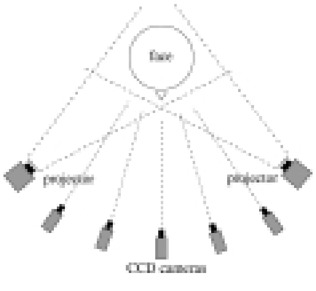

Fig. 1 Setup of cameras and projectors.

The system is calibrated using a 3-D reference frame with coded target points whose coordinates in space are known (see figure 2). These are fully automatically recognized and measured in the images [22]. The results of the calibration process are the exterior orientation of the cameras (position and rotations: 6 parameters), parameters of the interior orientation of the cameras (camera constant, principle point, sensor size, pixel size: 7 parameters), parameters for the radial and decentring distortion of the lenses and optic systems (5 parameters) and two additional parameters modeling differential scaling and shearing effects [4]. A thorough determination of these parameters modeling distortions and other effects is required to achieve high accuracy in the measurement.

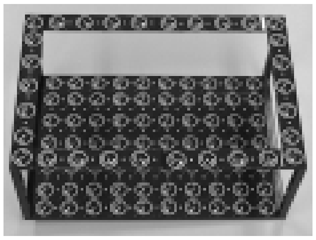

Fig. 2 Calibration frame
with coded targets.

2.2 Matching process

Our approach is based on multi-image photogrammetry using images acquired simultaneously by synchronized cameras. The multi-image matching process is based on the adaptive least squares method [12] with the additional geometrical constraint of the matched point lying on the epipolar line. Figure 3 shows an example of the result of the least squares matching (LSM) algorithm: the black boxes represent the patches selected in the template image (left) and the affine transformed in the search images (center and right), the epipolar lines are drawn in white.

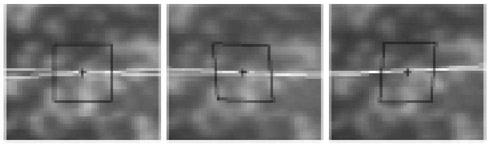

Fig. 3 Geometrical constrained LSM.
Left: template image.
Center and right: two search images.

The automatic matching process produces a dense and robust set of corresponding points, starting from few seed points. The seed points may be manually defined in each image, generated semi-automatically (defining them only in one image) or fully automatically. The manual mode is used for special cases where the automatic modes could fail; the seed points have to be selected manually with an approximation of at least 2 pixels in each image: LSM is then applied to find the exact position. In the semi-automated mode the seed points have to be selected manually only in the template image; the corresponding points in the other images are established automatically by searching for the best matching results along the epipolar line. This mode is the most convenient for normal cases of static surface measurement: it is fast but leave the operator the choice where to set the seed points. The fully automatic mode is useful in cases with dynamic surface measurement from multi-image video sequences, where the number of multi-image sets to be processed could be very large. In this case, Foerstner interest point operator [10] is used to automatically determine in the template image marking points where the matching process may perform robust results; the corresponding points in the other images are then established with the same process as for the semi-automatic mode.

After the definition of the seed points, the template image is divided into polygonal regions according to which of the seed points is closest (Voronoi tessellation). Starting from the seed points, the set of corresponding points grows automatically until the entire polygonal region is covered (see figure 4).

The matcher uses the following strategy: the process starts from the seed point, shifts horizontally in the template and in the search images and applies the least squares matching algorithm in the shifted location. If the quality of the match is good, the shift process continues horizontally until it reaches the region boundaries; if the quality of the match is not satisfactory, the algorithm computes the matching again, changing some parameters (e.g. smaller shifts from the neighbor, bigger sizes of the patches). The covering of the entire polygonal region of a seed point is achieved by sequential horizontal and vertical shifts. The process is repeated for each polygonal region until the whole image is covered.

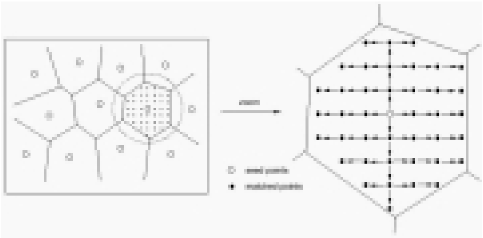

Fig. 4 Search strategy for the matching process.

To evaluate the quality of the result of the matching, different indicators are used: a posteriori standard deviation of the least squares adjustment, standard deviation in x and y directions, displacement from the start position in x and y directions and distance to the epipolar lines. Thresholds for these values can be defined for different cases, according to the texture and the type of the images. Nevertheless, errors are expected in the produced set of corresponding points and filters have therefore to be applied.

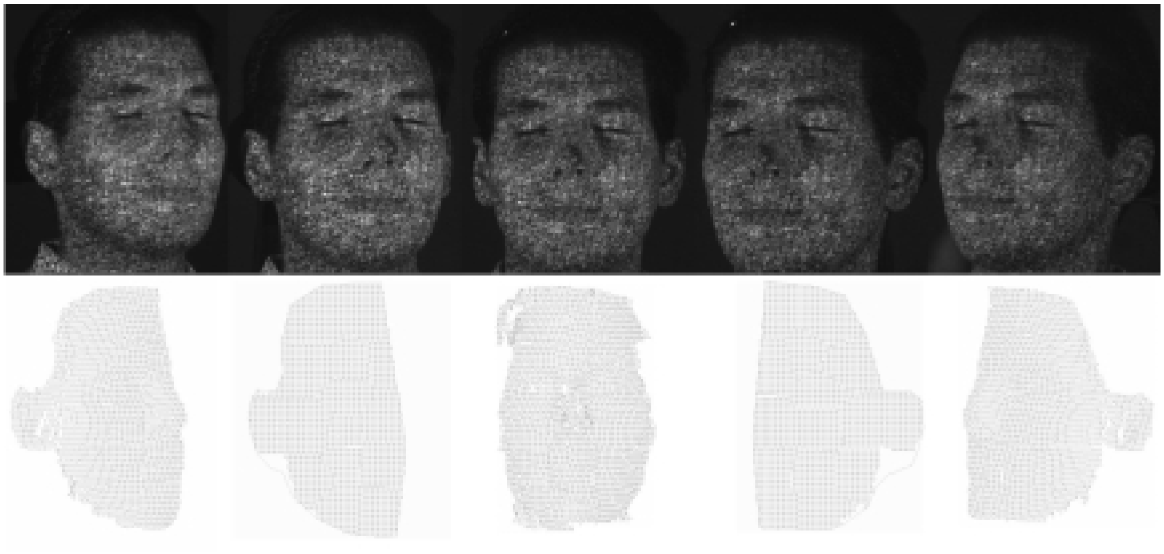

Fig. 5 Five images of a face with random pattern projection and set of corresponding points matched on the face.

Figure 5 shows five images of a face with random texture projections and the matched corresponding points established by the matching process. Since the human face is a steep surface and both sides of the face are not visible to the same camera, the five acquired images are used as two separate set of triplets, one for each side of the face. They are processed separately and at the end, the results are merged into a single data set.

Before beginning the three dimensional processing, filters can be applied to the 2-D matching data to minimize the number of possible errors. The Voronoi tesselation produces an irregular grid (see figure 8, left) of points in the template image, therefore, the set of matched points has first to be uniformed to a regular grid before the application of any filters. This is achieved by matching all the points shifted to the regular grid (see figure 6).

For the definition of the filter, the smoothed characteristic of the surface of the human face is taken in account: as shown in figure 7, the transformed image patches of neighboring points belonging to a common smoothed surface have similar shapes. A neighborhood filter is therefore applied to the set of matched points checking for the local uniformity of the shape of the transformed image patches. Figure 8 shows the results before and after grid regularization and filtering: on the left are displayed the template matched points together with the seed points, the effect of the Voronoi tesselation can be clearly observed; on the right are shown the results after regularization and filtering.

The complete matching process (definition of seed points, automatic matching, filtering) is flexible and can also be performed without orientation and calibration information. This functionality can be useful, for example, if the orientation is not accurate enough or unknown. In these special cases, only the image information is used by the least squares matching algorithm. Obviously, the robustness of the result of the process decreases; however the quality of the set of matched points remains satisfactory.

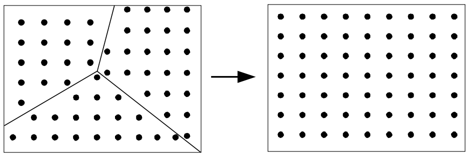

Fig. 6 Regularization of the matched point grid.

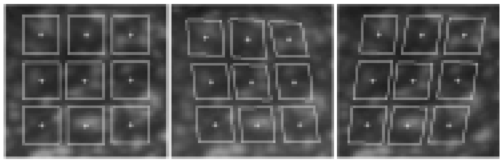

Fig. 7 Points matched in the neighborhood.

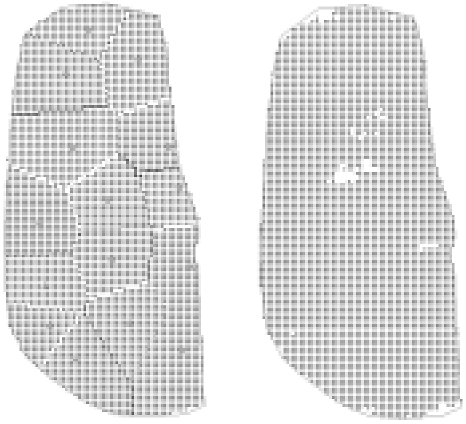

Fig. 8 Regularization and filtering results.
Left: template matched points and seed points.
Right: after regularization and filtering.

A dedicated software was developed for the face measurement process. Figure 9 shows its user friendly graphical interface. The required intervention of the operator for the matching process is reduced to the semi-automatic definition of about ten seed points and the selection of a contour of the region to measure. The operation can be performed in a couple of minutes, then the process will continue completely automatically. On a Pentium III 600 MHz machine, about 20,000 points are matched on half of the face in approximately 10 minutes.

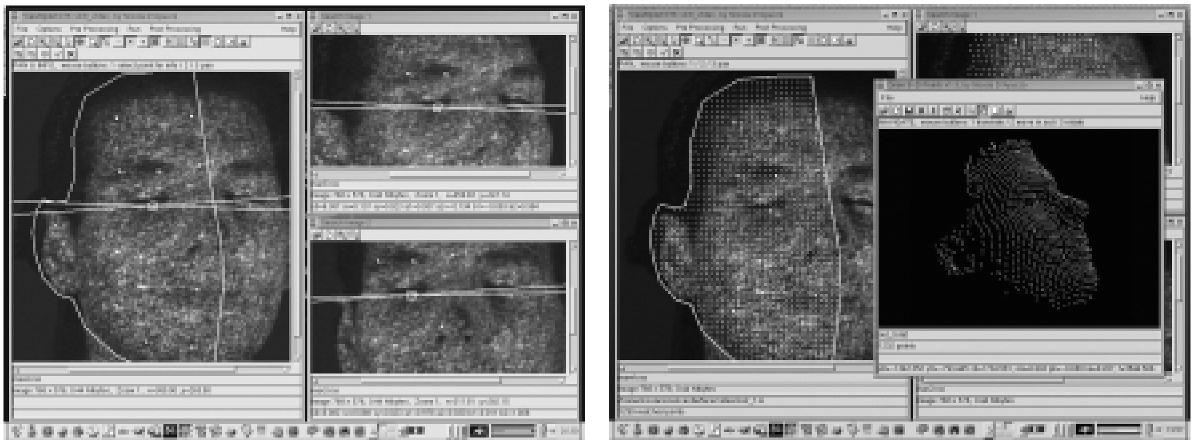

Fig. 9 Graphical user interface of the face measurement software. Left: seed points definition.
Right: matching results and visualisation of the computed 3-D point cloud.

2.3 Modeling and visualisation

The 3-D coordinates of the matched points are computed by forward ray intersection using the orientation and calibration data of the cameras. The achieved accuracy of the 3-D points is about 0.2 mm in the sagittal direction and about 0.1 mm in the lateral direction.

As shown in figure 10 (left), the point cloud is very dense (45,000 points) and the region of overlap of the two joined data set can be observed in the center line of the face. To overcome the redundant data and remove eventual outliers, Gaussian filters [3] are applied to the 3-D point cloud and the data is afterwards thinned (see figure 10 right).

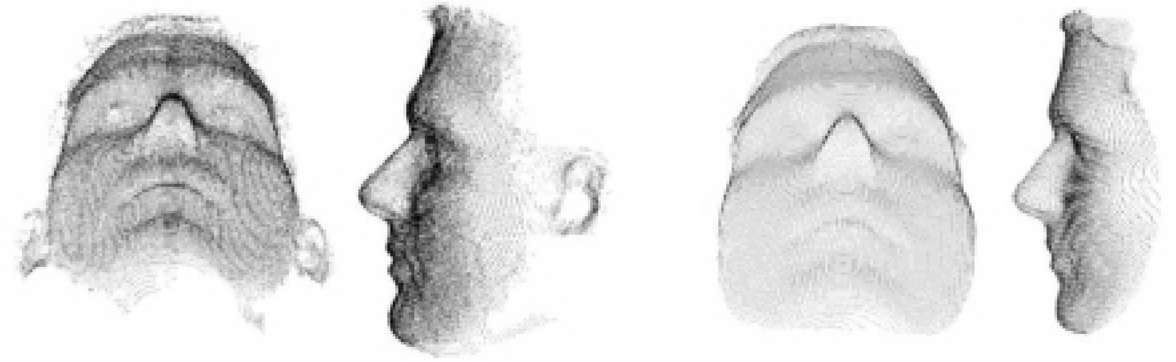

Fig. 10 Left: measured 3-D point cloud (45,000 points). Right: after filtering and thinning (10,000 points).

For surface measurement purposes, the computed 3-D point cloud is satisfactory. In case of visualisation, a complete model of the face with texture has to be produced. A meshed surface is therefore generated from the 3-D point cloud by 2.5-D Delauney triangulation and to achieve photorealistic visualisation, the natural texture acquired by the color video camera is draped over the model of the face. Figure 11 shows the surface model, the texture image and two views of the resulted face model with texture, figure 12 shows two other examples of face models.

Fig 11 Photorealistic visualisation. Left: shaded surface model, texture image. Right: face model with texture.

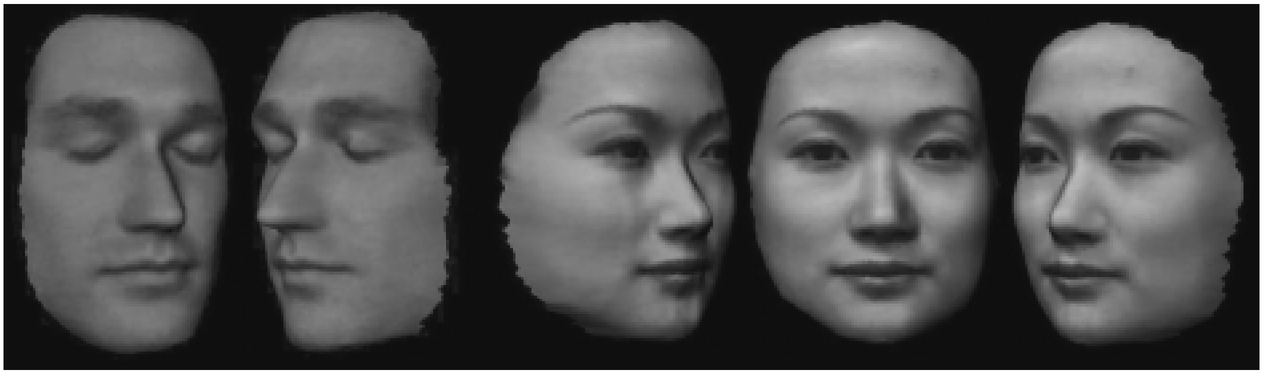

Fig 12 Photorealistic visualisation. Two other examples of face models.

3. CONCLUSIONS

A process for an automated measurement of the human face from multi-images acquired by five synchronized CCD cameras has been presented. The main advantages of this method are its flexibility, the reduced costs of the hardware and the possibility to perform surface measurement of dynamic events.

ACKNOWLEDGEMENT

The work reported here was funded in part by the Swiss National Science Foundation.

REFERENCES

1. Banda F. A. S. et al., "Automatic Generation of Facial DEMs", *Int. Archives of Photogrammetry and Remote Sensing* **29(B5)**, pp. 893-896, 1992
2. Blanz V. and Vetter T., "A Morphable Model for the Synthesis of 3D Faces", *SIGGRAPH'99 Conf. Proc.*, pp. 187-194, 1999
3. Borghese A. and Ferrari S., "A Portable Modular System for Automatic Acquisition of 3-D Objects", *IEEE Trans. on Instrumentation and Measurement* **49(5)**, pp. 1128-1136, 2000
4. Brown D. C., "Close-Range Camera Calibration", *Photogrammetric Engineering and Remote Sensing* **37(8)**, pp. 855-866, 1971
5. CNN, "Facing Up to Airport Security Fear", http://europe.cnn.com/2001/US/09/28/facial.recognition.QandA, September 28, 2001
6. Cyberware, "Head and Face Color 3D Scanner Model 3030", http://www.cyberware.com/products/psInfo.html

7. D'Apuzzo N., "Automated Photogrammetric Measurement of Human Faces", *Int. Archives of Photogrammetry and Remote Sensing* **32(B5)**, Hakodate, Japan, pp. 402-407, 1998
8. D'Apuzzo N., "Photogrammetric Measurement and Visualisation of Blood Vessel Branching Casting: A Tool for Quantitative Accuracy Tests of MR-, CT- and DS- Angiography", *Videometrics and Optical Methods for 3D Shape Measurement, Proc. of SPIE* **4309**, San Jose, USA, pp. 204-211, 2001
9. DeCarlo D. et al., "An Anthropometric Face Model Using Variational Techniques", *SIGGGRAPH'98 Conf. Proc.*, pp. 67-74, 1998
10. Foerstner W. and Guelch E., "A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features", *Proc. of the Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, pp. 281-305, 1987
11. Fua P., "Regularized Bundle-Adjustment to Model Heads from Image Sequences without Calibration Data", *Int. Journal of Computer Vision* **38(2)**, pp. 153-171, 2000
12. Gruen A., "Adaptive Least Squares Correlation: A Powerful Image Matching Technique", *South African Journal of Photogrammetry, Remote Sensing and Cartography* **14(3)**, pp. 175-187, 1985
13. Hasegawa K. et al., "A High Speed Face Measurement System", *Proc. of Vision Interface '99*, Trois-Rivières, Canada, pp. 196-202, 1999
14. Ip H. H. S. and Yin L., "Constructing a 3D Individualized Head Model from Two Orthogonal Views", *The Visual Computer* **12**, pp. 254-266, 1996
15. Koch R. M. et al., "Simulating Facial Surgery Using Finite Element Models", *SIGGRAPH96 Conference Proceeding*, New Orleans, USA, 1996
16. Lee W.-S. and Magnenat-Thalmann N., "Fast Head Modeling for Animation", *Image and Vision Computing Journal* **18(4)**, pp. 355-364, 2000
17. Liu Z. et al., "Rapid Modeling of Animated Faces from Video", *Proc. of the 3rd Int. Conf. on Visual Computing (Visual2000)*, Mexico City, pp. 58-67, 2000
18. Marschner S. R. et al., "Modeling and Rendering for Realistic Facial Animation", *Proc. of the 11th Eurographics Workshop on Rendering*, Brno, Czech Replublic, 2000
19. Matsumoto Y. et al., "CyberModeler: A Compact 3D Scanner Based on Monoscopic Camera", *Three-Dimensional Image Capture and Applications II, Proc. of SPIE* **3640**, San Jose, USA, pp. 3-10, 1999
20. Minaku S. et al, "Three-Dimensional Analysis of Lip Movement by 3-D Auto Tracking System", *Int. Archives of Photogrammetry and Remote Sensing* **30(5W1)**, Zurich, Switzerland, 1995
21. Motegi N. et al., "A Facial Growth Analysis Based on FEM Employing Three Dimensional Surface Measurement by a Rapid Laser Device", *Okajimas Folia Anatomica Japonica* **72(6)**, pp. 323-328, 1996
22. Niederoest M., *Codierte Zielmarken in der digitalen Nahbereichsphotogrammetrie*, Diplomarbeit, Institut für Geodaesie und Photogrammetrie, ETHZ, Zurich, (in German), 1996
23. Okada E., "Three-Dimensional Facial Simulations and Measurements: Changes of Facial Contour and Units Associated with Facial Expression", *Journal of Craniofacial Surgery* **12(2)**, pp. 167-74, 2001
24. Pighin F. et al. "Synthesizing Realistic Facial Expressions from Photographs", *SIGGRAPH'98 Conf. Proc.*, Orlando, USA, pp. 75-84, 1998
25. Proesmans M. and Van Gol L., "Reading Between the Lines", *SIGGPRAPH'96 Conf. Proc.*, pp. 55-62, 1996
26. Shan Y. et al., "Model-Based Bundle Adjustment with Application to Face Modeling", *Proc. of the 8th Int. Conf. on Computer Vision (ICCV01) Vol. II*, Vancouver, Canada, pp. 624-651, 2001
27. Sitnik R. and Kujawinska M., "Opto-Numerical Methods of Data Acquisition for Computer Graphics and Animation Systems", *Three-Dimensional Image Capture and Applications III, Proc. of SPIE* **3958**, San Jose, USA, pp. 36-43, 2000
28. Wolf H. G. E., "Structured Lighting for Upgrading 2D-Vision system to 3D", *Proc. of Int. Symposium on Laser, Optics and Vision for Productivity and Manufacturing I*, Besancon, France, pp. 10-14, 1996
29. Zheng J. Y., "Acquiring 3-D Models from Sequences of Contours", *IEEE Trans. Patt. Anal. Machine Intell.* **16(2)**, pp. 163-178, 1994