

Short Communication

Head Motion Invariance in Three-Dimensional Video Analysis System of Facial Movement

Barbara Koneczny^{1*}, Nicole M. Artner¹, Eva Gyoeri², Igor Pona², Chieh-Han J. Tzou² and Walter G. Kropatsch¹

¹Department of Pattern Recognition and Image Processing Group, Austria

²Department of Surgery, Medical University of Vienna, Austria

*Corresponding author

Barbara Koneczny, Department of Pattern Recognition and Image Processing Group, TU Wien, Austria, Email: koneczny@rip.tuwien.ac.at

Submitted: 19 November 2016

Accepted: 10 February 2017

Published: 13 February 2017

Copyright

© 2017 Koneczny et al.

OPEN ACCESS

Keywords

- Facial movement analysis
- Multi-view
- Head motion invariance
- Three-dimensional (3D) video analysis
- Facial palsy

Abstract

In the analysis of facial movement an objective, quantitative analysis of the expression is desirable. This can be achieved with the help of 3D systems, like the proposed mirror setup of this paper. These systems analyze the 3D motion of interest points of the face (e.g. left and right corner of the mouth) and compare their trajectories to measure progress. Unfortunately, the motion of the interest points is overlaid by the global motion of the head of the probands. This paper proposes an accurate approach to make the facial movement analysis invariant to head motion. Three static markers are used to model the head motion. These static markers are placed above bony structures so that only motion arising from the head influences the position of the markers. The presented methodology can be used to introduce an object centered coordinate system, in order to distinguish between motion that arises from the head and motion that arises from the facial movements.

ABBREVIATIONS

MUW: Medical University of Vienna; 3D: Three-Dimensional

INTRODUCTION

In the quantitative analysis of facial movement the need arises to objectively and quantitatively evaluate the quality of the facial movement in terms of symmetry and functionality. The main issue with widely used facial grading systems e.g. House-Brackman scale [1] is, that they are based on subjective measurements and observations [2,3], which are subject to investigator rating and provide rough estimates of facial function only. Inter subject and interobserver variability with in subjective grading systems has been reported [4]. The approach presented in this paper uses mirror-reflected, multi-view videos using the hardware setup (3D video analysis system) proposed by Frey et al. [5]. The main advantage of this mirror setup is that it introduces an inexpensive virtual multi camera setup. By using the data gained with this system 3D facial movement analysis system can be introduced.

The 3D video analysis system introduced by Frey et al. [5]. is currently used at the Medical University of Vienna (MUW) to objectively and quantitatively evaluate and document the preoperative status and postoperative recover course of

facial palsy patients after facial reconstructive surgery [6-8]. With reconstructive microsurgeries the paralyzed face can be reanimated to restore facial function and achieve near-symmetrical facial expressions. Template-based particle filter, color-based particle filter and optical strain cannot provide a sufficient accuracy in order to evaluate the quality of facial gestures [9,10]. There are also 3D surface imaging systems, which are able to record 4D data. However, tracking and analyzing of the facial movement is not possible using these systems [11]. Many other applications that use techniques of facial movement analysis track facial features and estimate their 3D position in order to create facial animations [12,13]. In order to create facial animations less accuracy is needed compared to the medical evaluation of facial palsy patients.

The approach presented in this paper is part of a new system that is designed to increase time efficiency and accuracy of the 3D video analysis system. The aim of the proposed methods is to separate the 3D head motion from facial movement. Knowing the head motion assumptions about the physical strain needed to reach the climax of a facial expression can be made.

This paper is organized as follows, in section "Materials and Methods" the hardware setup and methodology are described in more detail. The preliminary results are presented and discussed

in section “Results and Discussion”. In section “Conclusion” a conclusion and outline of future work are presented.

MATERIALS AND METHODS

In this section the hardware setup and the applied methodology to calculate the 3D head position are described in more detail. In the proposed system static markers are positioned above bony structures so that the position of the static markers reflects the global head motion and can be distinguished from the facial muscle motion. The used hardware setup is described in section “Hardware setup”. The position of the static markers in the video sequences is detected and then their 3D position is estimated applying the method described in section “3D position estimation”. The experimental setup is described in more detail in section “Experimental setup”.

Hardware setup

The hardware setup consists of a mirror setup, a calibration grid and a commercial video camera. The multi-view video sequences are recorded using a mirror setup described by Frey et al. [5] (Figure 1). The angle between the two mirror planes is flexible. Depending on the angle the camera records three (frontal and two mirror-reflected views) or more views (frontal, mirror-reflected and multiple mirror-reflected views) simultaneously. The video sequences are recorded at 30 frames per second with a Nikon Cool pix photo camera with HD resolution (1920x1080).

The method to calibrate a camera proposed by Zhang [14] is utilized to calculate the intrinsic parameters of the camera. The intrinsic camera parameters are essential for the calibration of the video sequence and further calculations.

3D position estimation

Lin et al.[15], proposed a robust and inexpensive method to estimate the 3D position of marker points using multi-view video sequences. The 3D positions are used to create a realistic 3D facial animation. In the method proposed by Lin et al. [15], the mirror properties are taken into account, which decreases the computational effort. Furthermore, the angle between the mirrors can be chosen arbitrarily.

The 3D position of marker points are estimated according to Lin et al. [15], The location and orientation of the two mirror planes can be described by a plane equation

$$ax + by + cz = d \tag{1}$$

The unit normal of the plane is $u = (a, b, c)$. According to Lin et al. [15] the 3D position of an arbitrary point $m_i = (x_{mi}, y_{mi}, z_{mi})$ and the corresponding virtual 3D position $m'_i = (x'_{mi}, y'_{mi}, z'_{mi})$ can be presented as follows

$$m_i = \frac{z_{mi}}{f} \cdot p_i \tag{2}$$

$$m'_i = \frac{z'_{mi}}{f} \cdot p'_i \tag{3}$$

with $p_i = (x_{pi}, y_{pi}, f)$ and $p'_i = (x'_{pi}, y'_{pi}, f)$ representing the projections onto the mirror plane. The focal length is given $f \cdot z_{mi}$ and z'_{mi} can be estimated by solving the system of linear

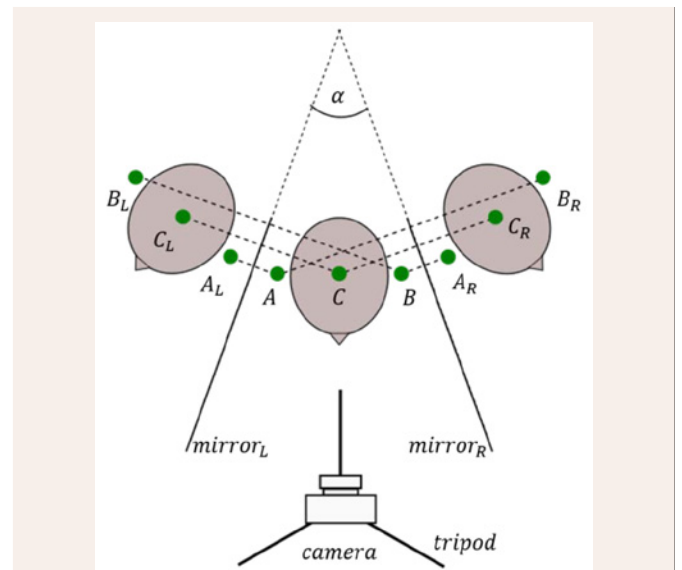


Figure 1 The mirror setup is composed of two mirrors at an arbitrary angle α . The subject is positioned inside this setup. The points A, B and C are the static marker points in the central view. The points A_L, B_L and C_L are the virtual marker points in the left mirror, A_R, B_R and C_R the virtual marker points in the right mirror. The camera recording the scene is positioned on a tripod.

equations given in (4).

$$\begin{bmatrix} \left(\frac{2a^2-1}{2f}\right)x_{pi} + \left(\frac{ab}{f}\right)y_{pi} + ac & \frac{x'_{pi}}{2f} \\ \left(\frac{ab}{f}\right)x_{pi} + \left(\frac{2b^2-1}{2f}\right)y_{pi} + bc & \frac{y'_{pi}}{2f} \\ \left(\frac{ac}{f}\right)x_{pi} + \left(\frac{bc}{f}\right)y_{pi} + \left(\frac{2c^2-1}{2}\right) & \frac{1}{2} \end{bmatrix} \begin{bmatrix} z_{mi} \\ z'_{mi} \end{bmatrix} = d \begin{bmatrix} a \\ b \\ b \end{bmatrix} \tag{4}$$

The location and orientation of the mirror planes is estimated using a calibration grid (Figure 2). The corners of the checkerboard (which are marked with white stars in (Figure 2) in the frontal view are used as marker points p_i , the corners of the checkerboard in the reflected view are used as marker points p'_i . For the estimation of the position and location of the mirror only one frame is necessary as the position and location of the mirror planes does not change through the video. Knowing the location and position of the mirror planes and the positions p_i and p'_i of an and o arbitrary point in the frontal and reflected view the 3D position of this point can be estimated.

Experimental setup

The videos are recorded using a Nikon Cool pix photo camera. For the camera calibration a planar checkerboard is used. The camera calibration is performed before the estimation of the 3D positions takes place. The same camera settings are used for recording the checkerboard for camera calibration, the calibration grid and the video sequence with the simulated head motion. In order to provide a sufficient ground truth the head motion was simulated using an artificial head model. 3 static markers are attached to the artificial head model (Figure 3). The head is rotated in 5° steps around the z-axis (yaw) counter-

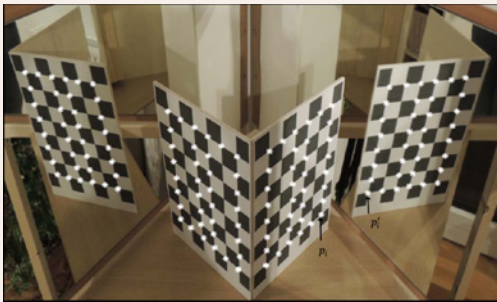


Figure 2 The corners of the calibration grid are used as marker points P_i and P_i (see the white stars) for the estimation of the location and orientation of the mirror planes. The corners marked with white stars are the marker points that are used for the estimation of the location and orientation of the mirror planes.

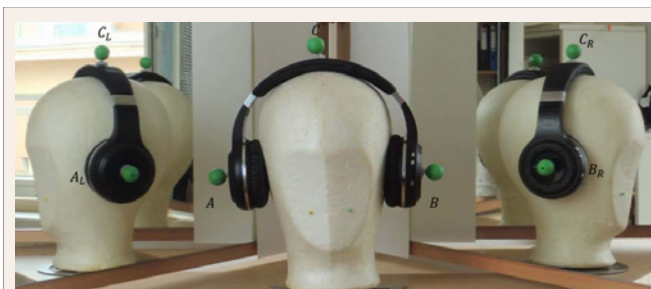


Figure 3 The head motion was simulated using an artificial head model. The static markers A , B and C were attached to the head model.

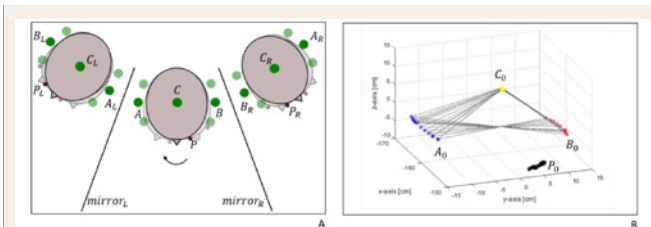


Figure 4 The experimental setup consisted of an artificial head model on a rotary plate, 3 static markers (A , B and C) and the mirror setup. P represents an observation point or facial landmark which is influenced by the rotation. The artificial head model was manually rotated in 5° steps in counter-clockwise direction. (B) The static markers are used to calculate the 3D head motion. The 3D positions of the markers are represented by A_0 , B_0 and C_0 . The 3D position of the observation point is represented by P_0 .

clockwise using a rotary plate (Figure 4A).

RESULTS AND DISCUSSION

By introducing static markers the 3D head motion through the video sequence can be modeled. In Figure (4B) the 3D head rotation of the artificial head model can be seen. As mentioned before an artificial head model was chosen in order to provide a sufficient ground truth. Table (1) shows the rotation angle in degree versus the estimated yaw rotation in degrees. The video sequence was composed of 43 frames. In (Table 1) the mean yaw of each rotation step is presented in (Figure 4B) all 43 measurements are displayed. It also shows the relative and percentual error.

It can be seen in (Table 1) the calculated yaw rotation in counter clockwise direction increases in 5° steps. The slight differences in the corresponding positive and negative angles arise from a systemic error. The systemic error arises from the manual handling of the rotatory plate. In Figure (4B) can be seen that the estimated 3D positions of A and B describe a circular path, which arises from the rotation motion. However, the 3D positions of C are located at one point. Therefore static marker C and is located above the center of rotation and since no translation occurred it is not affected by the rotation motion. The rotation was performed between 30° and -30° since only a small portion of the static markers A and B depending on and (the negative or positive rotation angle) was visible in the frontal view at a higher rotation angle.

Figure (4) also provides an example of an arbitrary observation point or facial landmark represented by P . In this example the observation point is only influenced by the global head motion. By introducing an object centered coordinate system the position of observation point can be given in relation to the static markers. Therefore the position of the observation point is independent from the global head motion.

CONCLUSION

In this paper an approach is proposed that allows estimating the global head motion through the whole video.

The advantage of this system is that the angle between the mirror planes can be arbitrarily chosen. Therefore no time consuming calibration of the mirror setup is needed, in comparison to the currently used system at the MUW. The angle between the two mirror planes should be calibrated for every video recording in order to provide a reliable result. With the static markers an object centered coordinate system can be introduced. This object centered coordinate system can be used to separate the motion that arises from the head from the motion

Table 1: This table shows the simulated rotation angle in degree versus the mean estimated yaw rotation (of each rotation step) in degree. The rotation plate was rotated counter-clockwise direction in 5° steps. The 3D positions of static markers were used to calculate the yaw rotation.

angle [°]	yaw [°]	relative error [°]	percentual error [%]
-30.000	-29.471	-0.018	-1.764
-25.000	-24.673	-0.013	-1.308
-20.000	-19.892	-0.005	-0.539
-15.000	-15.593	-0.040	-3.955
-10.000	-10.377	-0.038	-3.772
-5.000	-5.243	-0.049	-4.868
0.000	0.000	0.000	0.000
5.000	5.131	0.026	2.616
10.000	10.440	0.044	4.403
15.000	15.121	0.008	0.804
20.000	19.309	0.035	3.454
10.000	10.440	0.044	4.403
25.000	24.433	0.023	2.267
30.000	29.178	0.027	2.741

that arises from the facial muscle tonus. This would, for example, allow a more detailed analysis of the progress after a recovery surgery since the physical strain to reach the climax and the quality of performing facial gestures can be evaluated separately. By using this system the arbitrary observation points or facial landmarks can be tracked (manual or by using an automatic or semi-automatic tracking algorithm) without the influence of the global head motion. Therefore facial movement can be analyzed without the need to keep the head motion at a minimum, this facilitates a more natural and intuitive facial expressions. It is also possible to define a custom region of interest that can be tracked through the video sequence. The results gained by this method combined with the tracking of facial landmarks while performing facial expression can be used as input parameters for facial grading systems (e.g. House-Brackman scale or Sunnybrook scale). By using this system the quality of facial movements can be evaluated objectively in comparison to the subjective measurements of the House-Brackman Scale.

REFERENCES

1. House JW, Brackmann DE. Facial nerve grading system. *Otolaryngol Head Neck Surg.* 1985; 93: 146-147.
2. Coulson SE, Croxson GR, Adams RD, O'Dwyer NJ. Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial grading systems to assess voluntary movement and synkinesis after facial nerve paralysis. *Otolaryngol Head Neck Surg.* 2005; 132: 543-549.
3. Zhai M-Y, Feng G-D, Gao Z-Q. Facial Grading System: Physical and Psychological Impairments to Be Considered. *J Otol.* 2008; 3: 61-67.
4. Linstrom CJ. Objective Facial Motion Analysis in Patients With Facial Nerve Dysfunction. *Laryngoscope.* 2002; 112: 1129-1147.
5. Frey M, Jenny A, Giovanoli P, Stüssi E. Development of a New Documentation System for Facial Movements as a Basis for the International Registry for Neuromuscular Reconstruction in the Face. *Plast Reconstr Surg.* 1994; 93: 1334-1349.
6. Frey M, Giovanoli P, Gerber H, Slameczka M, Stüssi E. Three-Dimensional Video Analysis of Facial Movements: A New Method to Assess the Quantity and Quality of the Smile. *Plast Reconstr Surg.* 1999; 104: 2032-2039.
7. Tzou C-HJ, Pona I, Placheta E, Hold A, Michaelidou M, Artner N, et al. Evolution of the 3-Dimensional Video System for Facial Motion Analysis: Ten Years' Experiences and Recent Developments. *Ann Plast Surg.* 2012; 69: 173-185.
8. Giovanoli P, Tzou C-HJ, Ploner M, Frey M. Three-dimensional video-analysis of facial movements in healthy volunteers. *Br J Plast Surg.* 2003; 56: 644-652.
9. Limbeck P, Kropatsch WG, Haxhimusa Y. Semi-automatic tracking of markers in facial palsy. 2012; 69-72.
10. Shreve M, Jain N, Goldgof D, Sarkar S, Kropatsch W, Tzou C-HJ, et al. Evaluation of Facial Reconstructive Surgery on Patients with Facial Palsy using Optical Strain. In: Real P, Diaz-Pernil D, Molina-Abril H, Berciano A, Kropatsch WG, eds. *The 14th International Conference on Computer Analysis of Images and Patterns (CAIP).* 2011; 6854: 512-519.
11. Tzou C-HJ, Artner NM, Pona I, Hold A, Placheta E, Kropatsch WG, et al. Comparison of three-dimensional surface-imaging systems. *J Plast Reconstr Aesthet Surg.* 2014; 67: 489-497.
12. Guenter B, Grimm C, Wood D, Malvar H, Pighin F. Making Faces. In: *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques.* ACM. 1998: 55-66.
13. Basu S, Oliver N, Pentland A. 3D modeling and tracking of human lip motions. In: *Computer Vision, 1998. Sixth International Conference.* 1998; 337-343.
14. Zhang Z. A flexible new technique for camera calibration. *Pattern Anal Mach Intell IEEE Trans On.* 2000; 22: 1330-1334.
15. Lin I-C, Yeh J-S, Ouhyoung M. Extracting 3D facial animation parameters from multiview video clips. *IEEE Comput Graph Appl.* 2002; 22: 72-80.

Cite this article

Koneczny B, Artner NM, Gyoeri E, Pona I, Tzou CHJ, et al. (2017) Head Motion Invariance in Three-Dimensional Video Analysis System of Facial Movement. *Comput Sci Eng* 2(1): 1007.