Weight Compensated Motion Estimation for Facial Deformation Analysis

Jürgen Rurainsky

Fraunhofer Heinrich-Hertz-Institute, Einsteinufer 37, 10587 Berlin, Germany
rurainsky@hhi.fraunhofer.de
http://www.hhi.fraunhofer.de

Abstract. Investigation of the motion performed by a person's face while speaking is the target of this paper. Methods and results of the studied facial motions are presented and rigid and non-rigid motion are analyzed. In order to extract only facial deformation independent from head pose, we use a new and simple approach for separating rigid and non-rigid motion called Weight Compensated Motion Estimation (WCME). This approach weights the data points according to their influence to the desired motion model. A synthetic test as well as real data are used to demonstrate the performance of this approach. We also present results in the field of facial deformation analysis and used basis shapes as description form. These results can be used for recognition purposes by adding temporal changes to the overall process or adding natural deformations other than at the given database.

Key words: motion, facial deformation, personalized

1 Introduction

In this paper, we target for the analysis of the dynamic behavior of facial motion and thus the sampling rate, in which the motion states are recorded, is an important issue. Important transitions from one state to another maybe get lost if only a video frame rate of 25 fps is used and then these details are not available for the natural animation of 3D models.
High-end motion capture systems, as used for movie productions, can realistically animate another object, a person, or a creature by mapping an actor's motion to it as described in the publication of Perlman [1]. Rather than only animating faces with the motion information, facial motion and specific facial states are also analyzed for medical purposes, treatment, and diagnosis published by Faraway and Trotman [2, 3]. In this case, the resolution of the analyzed facial motion is mostly limited to the anatomically interesting points and is focused to facial expressions rather than facial motion caused by speech.
Although different approaches for the specification of static expressions are available like the Facial Action Coding System or the MPEG-4 Facial Animation Parameters FAPs, much less has been reported about the dynamic modeling of

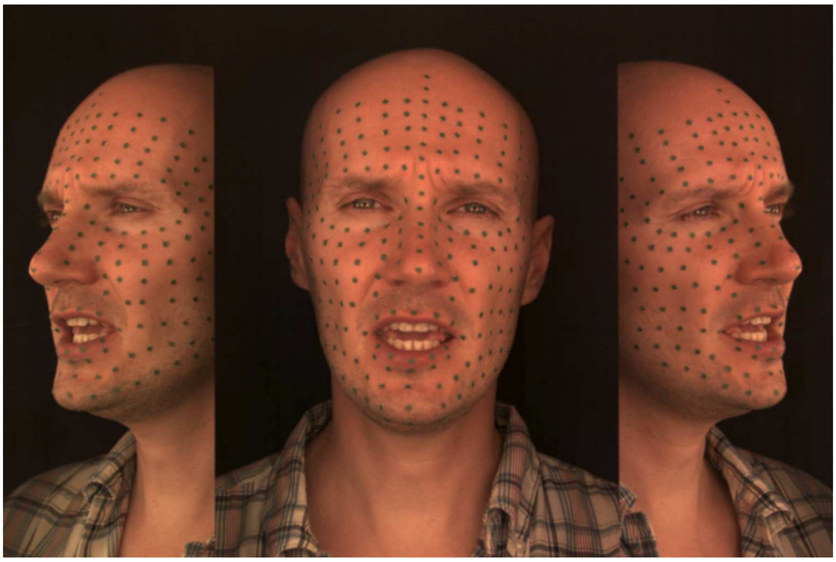


Fig. 1: One frame of the captured sequence of a multiple view recording using two virtual cameras turned $\pm 31^\circ$ to z-axis.

these motions. With the publication of Essa and Pentland [4], a dynamic extension to FACS system is presented. In the conference publication of Kalberer and Van Gool as well as in the journal publication of Odisio *et al.* [5, 6] results of 3D speech movement analysis by using facial deformation states are given and used for animation and tracking purposes.
In this paper, we analyze the dynamics of facial expressions. A 3D model sequence extracted from a multi view capture system with a capture rate of 200 fps is used. Due to the high capture rate important transitions are not lost and available for the analysis and synthesis of the dynamic behavior of facial motion while speaking.
In order to analyze only local facial deformation and not the global head movements, we introduce a new and simple approach for the separation of rigid-body and non-rigid motion named Weight Compensated Motion Estimation (WCME). This approach defines weights for each data point depending of the influence to the desired motion model. Important for such analysis is, that the deformation will be not represented by a set of predefined deformations as used for a model-based approach. The result of this analysis should provide possible deformations for a model-based approach. Additionally, initial results for the analysis of non-rigid deformations in the form of basis shapes are provided.

2 Rigid Motion Compensation

For the analysis of facial motion, we separate rigid body motion from deformations using a 3D sequence of facial points. The 3D model sequence is generated by triangulation of markers, which are placed on a human face and captured by a double mirror construction and a high speed camera. Therefore, the requirements of depth data and a high sampling rate (200 fps) are fulfilled for the analysis of facial deformations. An example of the captured sequence is shown in **Fig. 1**. Rotation and translation for all axes (6 DOFs) of the associated 3D model describe the rigid-body motion and all other changes are regarded as deformation and noise.

Rigid-body motion and deformations are very successfully determined by several different approaches. In the journal publication of Eisert and Girod, the journal publication of Li *et al.*as well as the proceedings publication of Yilmaz *et al.* [7–9], methods for motion estimation from a single view using optical flow are described. A neural network was formed to estimate the rigid-body motion in the journal publication of Ploskas *et al.* [10] using multiple views. In the conference publication of Huang *et al.* [11] a simulated annealing approach was introduced to determine the desired motion parameters. The classification of the available 3D model vertices into a rigid and a non-rigid class is described in the publication of Del Bue *et al.* [12].

We present a new and simple approach named Weight Compensated Motion Estimation (WCME) to estimate the rigid-body motion parameters in the presence of non-rigid deformation and noise. This approach is applied to the 3D model and continuously separates the vertices into rigid motion and non-rigid deformation with respect to the reference model.

2.1 Weight Compensated Motion Estimation (WCME)

Our approach to estimate the rigid-body motion is based on the continuous classification of model data into rigid and non-rigid movements. In order to achieve this goal, we have weighted the influence of the vertices used for the rigid-body motion estimation. Weights are also used for a mean filter as described in the article of Oten and Figueiredo [13], in order to represent specific influences. The weights are associated to the Euclidean distance from the rigid-body reference model to the current model. The idea is, that large deviations from the rigid-body constraint is caused by non-rigid deformation. We have used the $cos^2(x)$ function as weight function in the range between 0 and π. The Euclidean distance is scaled such that a weight of 0.5 is associated to the average distance of all vertices classified as rigid-body.

$$\begin{gathered} w(i,n) \cdot \mathbf{v}_0(n) = w(i,n) \cdot (\mathbf{R} \cdot \mathbf{v}_f(n) + \mathbf{t}) \\ f \in \{1, ..., F-1\} \end{gathered}$$

$$\begin{aligned} w(i,n) &= cos^2\left(\frac{e_{3D}(i,n)}{norm(i)} \cdot \pi\right) \\ norm(i) &= \frac{\bar{e}_{3D}(i)}{acos(\sqrt{0.5})} \cdot \pi \end{aligned}$$

Here, $w(i,n)$ represents the weight for each iteration i and for each vertex n and e the Euclidean Distance in 3D space. The rigid-body motion for the frames $\{1, ..., F-1\}$ are estimated with respect to the first frame $f = 0$. The classification rules for the two classes (rigid and non-rigid motion) are shown below, while the non-rigid motion vertices are handled with a zero weight and the rigid motion vertices based on their influence (Euclidean Distance in 3D space).

$$\mathbf{v}(i,n) = \begin{cases} rigid : i = 0 \\ non-rigid : i > 0, \ \frac{e_{3D}(i,n)}{norm(i)} \geq 0.5 \\ rigid : i > 0, \ \frac{e_{3D}(i,n)}{norm(i)} < 0.5 \end{cases}$$

The weighted motion equation is solved in the least squares sense, similar to the known *Weighted Least Squares* approach shown in the following equation.

$$S = \sum_{i=1}^{n} w_i (y_i - f(x_i))^2$$

The influence of the weights can be visualized by a simple line fitting example, where some points are outliers and incorporated to the closed-form solution via weights. In **Fig. 2** it is clear to see, that the influence of the outliers to the final solution can be controlled by the weights. On the other hand it can also be seen, that the influence can only be reduced to zero if the associated weights are zero as well. Therefore, this approach is an asymptotic approximation with $w \rightarrow 0$ for the non-rigid motion data, where w represents the weights. The weight function

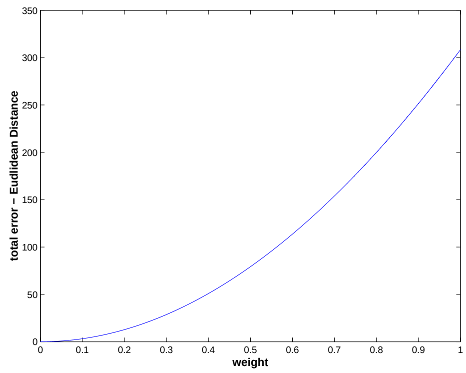


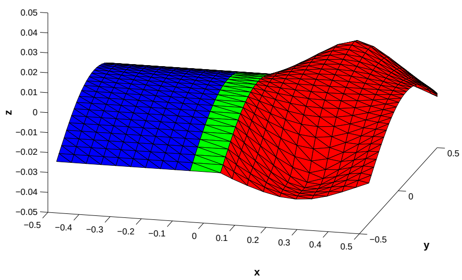
Fig. 2: Line fitting error measured as total sum in the form of a Euclidean Distance between the weighted line fitting and a line fitting with excluded outliers.

describes the relation from the error measurement to the influence (weight) and therefore the level of approximation to the *true* solution. The function $cos^2(x)$ is selected because of the desired mapping from the vertex Euclidean Distance to the associated weight. Mapping functions with a steep deviation ramp are possible with similar behavior ($cos^4(x)$), but if the deviation ramp of the function

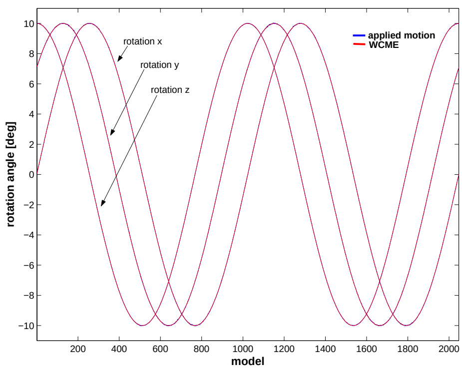
becomes to steep ($cos^8(x)$) a measurable deviation occurs. The same applies for $cos(x)$, where the lowest error is not reached (see **Fig. 2**).
Because of the very small expected relative rotations and translations between successive frames, a linearized version of the rotation matrix is used to determine the rigid-body motion. The use of linearized rotation parameters leads to computational efficient and robust algorithms but requires an iterative process to remove the approximation errors. In this case, it turned out that two iterations are sufficient to converge at an accurate parameter set.
Due to the asymptotic approximation of the WCME approach, an iterative loop is used to minimize the error measurement. With each iterations new weights are determined and applied to the motion estimation equation. Measuring the same Euclidean Distance between two iterations of the associated vertices defines the last iteration and therefore also the final set of weights.
The main advantage of this approach is the unsupervised rigid-body motion estimation by automatic selection of outliers and determination of the weights. It can be shown, that even manually selected vertices can be partially involved in deformation and therefore not used without a compromise. This compromise can be handled by the WCME approach.

2.2 Synthetic Test

In order to demonstrate the performance of the Weight Compensated Motion Estimation (WCME) approach synthetic rigid motion and non-rigid deformation data are applied to a test model. The test model shown in **Fig. 3a** consists of 651 vertices and 1200 triangles and about half of the vertices are defined as flexible. Rigid motion consists of rotation and translation and is applied to the complete



(a)



(b)

Fig. 3: (a) Test model for the rigid motion estimation. (blue) Only rigid motion. (red) Rigid motion and non-rigid deformation.(b) Rotation estimation test results using WCME. Rotation around axes are shifted.

model. Additionally, non-rigid deformation is applied to the red colored part of **Fig. 3a**. The rotation and translation is sine wave shaped and applied to all

three axes. An example of the applied rotation and translation is given with the blue lines in **Fig. 3b**. Deformation is added by adding an offset to two vertices and using radial basis function for the interpolating of the remaining points. The amplitudes of the deformation wave is shown with the green line in **Fig. 4**. The rotation is limited to $\pm 10°$ and the translation to $\pm 0.01m$. In addition, normally distributed random numbers with a standard deviation of $0.001m$ are applied to each vertex and axes, independently.
The WCME applied to the described test model with rigid motion and deformation leads to the results shown in **Fig. 3b** and **Tab. 1**, where the rotation and translation of the test are estimated. The blue colored line refers to the applied rotation and the red colored line to the detected rotation in **Fig. 3b**. The results in **Tab. 1** are given as absolute distances.
Another parameter for the WCME approach is the number of $3D$ points used

	$1/N \sum_N \lvert d \rvert$	$max \lvert d \rvert$
rot x [deg]	0.012	0.053
rot y [deg]	0.013	0.065
rot z [deg]	0.009	0.042
tran x [m]	0.00006	0.00025
tran y [m]	0.00006	0.00031
tran z [m]	0.00007	0.00032

Table 1: Estimation results for rotation and translation for the application of WCME to the synthetic sequence. The results are specified as absolute difference.

for the rigid motion estimation. In **Fig. 4** this number is given as well as information about the applied deformation for each model in the tested sequence. It is clear to see, that for models with no deformation more points are used for the motion estimation and less during deformation. The minimum number of rigid motion points, which is identical with the number of points colored blue in **Fig. 3a** is not reached.

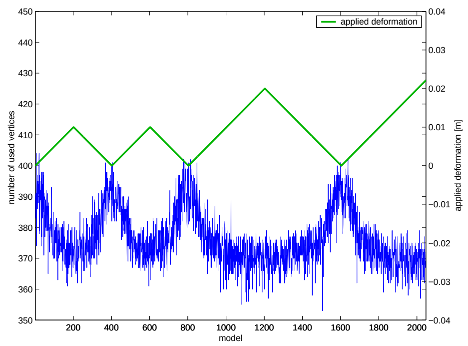
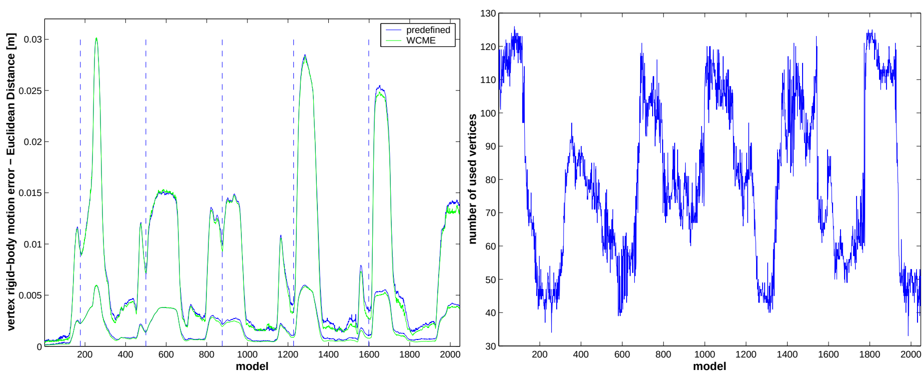


Fig. 4: (blue) Number of $3D$ points used for the estimation approach WCME. (green) Amplitude of the applied non-rigid deformation.

2.3 Facial Motion Data

With this and the following sections results are presented, which are achieved by analyzing the facial deformation of a 3D model sequence. The observation consists of a person, who is counting from "one" to "six". The results targeting a compact representation of the observation, which can be used for the analysis and synthesis of facial deformations while speaking.
Before the deformation can be analyzed, the rigid-body motion of the current 3D model with respect to the reference model has to be estimated and afterwards compensated. Subtraction of the rigid-body motion defines the desired model deformation which also contains the measurement noise.
Extracting the model deformation by subtracting the rigid-body motion with the first frame model as reference using our WCME approach provides us with a sequence of model deformations. The results of this extraction is shown in **Fig. 5a**, where the blue colored line shows the Euclidean distance from the reference to the current model using a set of predefined points for the estimation step. The green colored line refers to the WCME approach. The remaining Euclidean distance describes the non-rigid deformation. The vertical dashed lines represent the beginning of the spoken numbers. Between two spoken numbers the person tried to return to the first frame deformation, but an average deformation of around 0.5 mm remained. One result of the rigid-body motion estimation is the almost similar performance of our proposed WCME approach compared to a predefined set of rigid-body motion points.
The WCME approach tries to incorporate as many points as possible for more



(a) (b)

Fig. 5: (a) (upper two) Maximum and (lower two) average vertex error caused by facial deformation by using the first model as reference. The different curve colors refer to different approaches for selecting points for the rigid-body motion estimation. (b) Number of used vertices for the weighted motion estimation. The approach incorporates as many points as possible. Maximum number is reached in silent periods and minimum number in sequence part with a high level of deformation.

reliable results during the rigid-body motion estimation. **Fig. 5b** shows the number of vertex used for estimation over the model sequence. It is clear to see, that the minimum number of points is reached at a high level of deformation. These used vertices are located at the forehead. During silent periodes (non or lowest level of deformation) sometimes the maximum number (all vertices) are used for the rigid-body motion estimation.

2.4 Facial Deformation Map

Another interesting result is the distribution of the deformation in a face, after compensating the face pose for each model. This distribution is a spatial representation of **Fig. 5a**, which shows the deformation for each model and identifies the level of involvement of specific facial parts, but does not allow a distinction between deformations along the y- and z-axis. Such a deformation map is displayed in **Fig. 6**, where the level of deformation is expressed as color (red refers to more and blue less deformation). Please note, that even the forehead vertices show some deformation. Therefore, these vertices cannot exclusively be used for the estimation of the rigid-body motion parameters.

Fig. 6: Non-rigid deformation map: Color encodes the normalized level of deformation, where red refers to a high level of deformation.

3 Conclusions

We presented a new and simple approach for the rigid-body motion estimation named as Weight Compensated Motion Estimation (WCME) based on 3D model sequence. This approach shows almost similar results to manual selected rigid motion points by incorporating as much data as possible. Furthermore, analysis results of a facial deformation sequence captured with 200 fps are presented. This includes the definition of a facial deformation map, which shows that facial deformation can be found in a wide area around the mouth and the cheeks. Such dynamic deformation analysis data can be used to add temporal changes to the recognition system or to enhance the given database by natural deformations. Also the basis shape analysis shows, that the number of Eigen vectors required for a very good reconstruction can be limited to the first eight, because the average reconstruction error is smaller than $0.3mm$ which is shown in **Fig. 7**. This allows to use a hierarchical approach by applying more or less deformations.

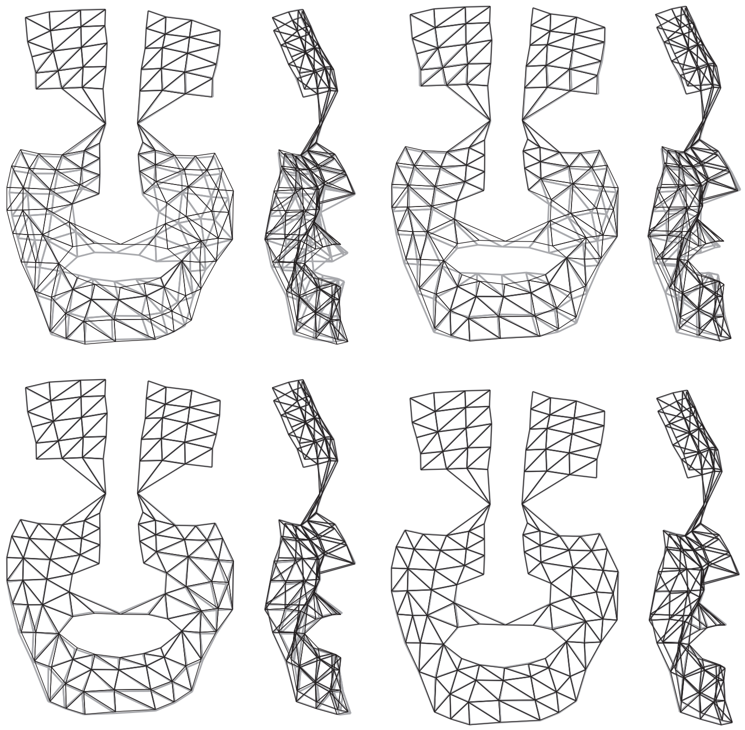


Fig. 7: Reconstruction results (gray belongs to the reconstructed model) at frame model 1975 of the number sequence using different numbers of Eigen vectors: (top left) one, (top right) two, (bottom left) three vectors with an average reconstruction error of $0.42mm$ per vertex, and (bottom right) eight vectors with an average reconstruction error at $0.26mm$ per vertex. All numbers are reflecting the results of the complete sequence.

References

1. Perlman, S.: Contour Reality Capture System Unveiled. In: SIGGRAPH 2006. (July 2006)
2. Faraway, J.: Modeling continuous shape change for facial animation. Statistics and Computing **14**(4) (October 2004) 357–363
3. Trotman, C., Faraway, J.J.: Modeling facial movement: I. A dynamic analysis of differences based on skeletal characteristics. Journal of Oral and Maxillofacial Surgery **62**(11) (November 2004) 1372–1379
4. Essa, I.A., Pentland, A.P.: Facial expression recognition using a dynamic model and motion energy. In: Proc. International Conference on Computer Vision (ICCV), Cambridge, MA, USA (June 1995) 360–367
5. Kalberer, G., Gool, L.V.: Face animation based on observed 3D speech dynamics. In: Proceedings of the 14. Conference on Computer Animation, Nice, France (November 2001) 20–27
6. Odisio, M., Bailly, G., Elisei, F.: Tracking talking faces with shape and appearance models. Speech Communication **44** (October 2004) 63–82
7. Eisert, P., Girod, B.: Analyzing Facial Expressions for Virtual Conferencing. IEEE Computer Graphics Applications: Special Issue: Computer Animation for Virtual Humans **18**(5) (September 1998) 70–78
8. Li, H., Roivainen, P., Forchheimer, R.: 3-D Motion Estimation in Model-Based Facial Image Coding. IEEE Trans. Pattern Anal. Machine Intell. **15**(6) (1993) 545–555
9. Yilmaz, A., Shafique, K., Shah, M.: Estimation of Rigid and Non-Rigid Facial Motion Using Anatomical Face Model. In: ICPR '02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 1, Washington, DC, USA (2002) 10377
10. Ploskas, N., Simitopoulos, D., Tzovaras, D., Triantafyllidis, G., Strintzis, M.: Rigid and non-rigid 3D motion estimation from multiview image sequence. Signal Processing: Image Communication **18**(3) (March 2003) 185–202
11. Huang, W., Zhang, Y., Wang, Y., Cheng, H.: 3D non-rigid motion estimation using the improved simulated annealing algorithm. In: 4th International Conference on Machine Learning and Cybernetics (ICMLC 2005). Volume 9. (August 2005) 5330–5335
12. Bue, A.D., Lladó, X., Agapito, L.: Non-rigid Face Modelling Using Shape Priors. In: Analysis and Modelling of Faces and Gestures (ICCV workshop), Lecture Notes in Computer Science. LNCS 2723, Beijing, China (2005) 96–107
13. Oten, R., de Figueiredo, R.: Adaptive alpha-trimmed mean filters under deviations from assumed noise model. IEEE Transactions on Image Processing **13** (May 2004) 627–639