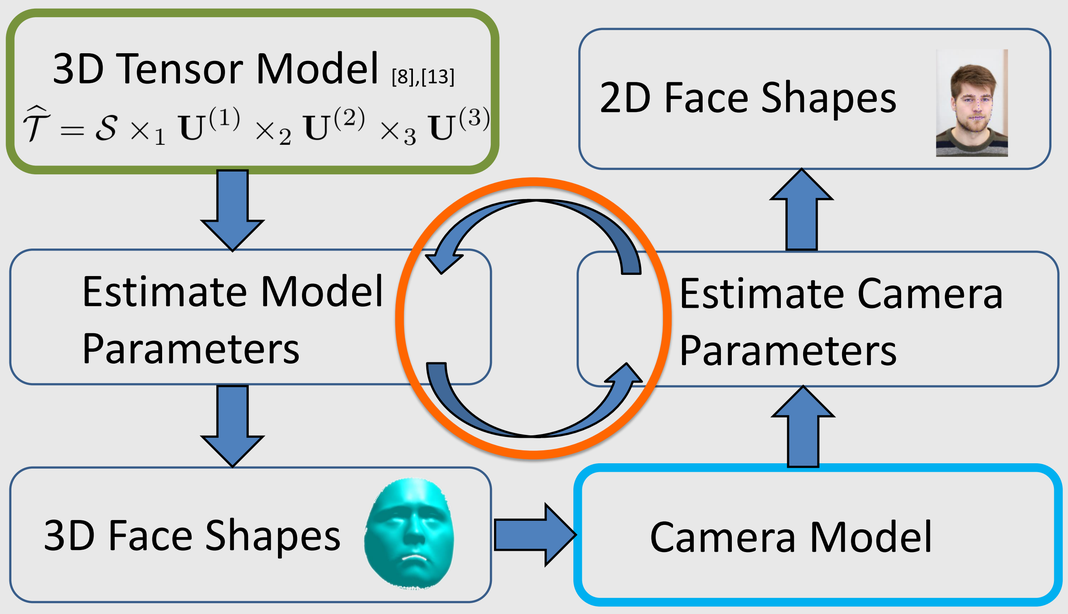
Projective Structure from Facial Motion

Stella Grasshof [1] • Hanno Ackermann [1] • Felix Kuhnke [1] • Jörn Ostermann [1] • Sami S. Brandt [2]

[1] Leibniz Universität Hannover, Germany • [2] University Copenhagen, Denmark

Overview



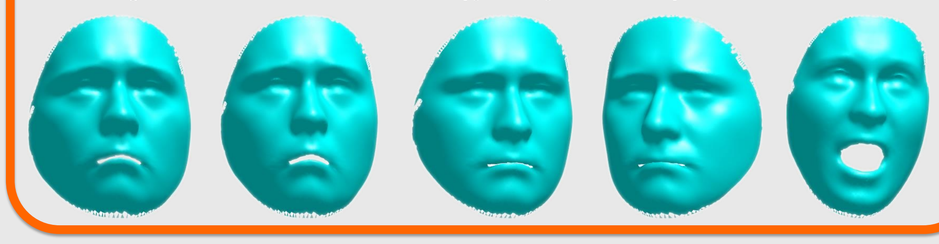
Input

Sparse 2D Landmarks $\mathbf{s}_k^{2D}$

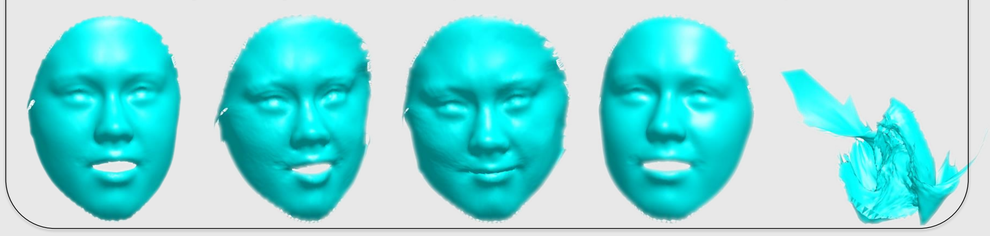
Output

Dense 3D Reconstruction $\widehat{\mathbf{s}}_k^{3D}$

Projective Camera (proposed)



Weak-perspective (State of the Art)



Contributions

- Dense 3D from sparse 2D using a nonlinear projective camera instead of weak-perspective camera
- Linear face model parameter estimation with nonlinear projective camera
- 2-step alternating least squares (ALS) for model parameters (person and expression)
- Reduced run time

References

[8] S. Grasshof, H. Ackermann, S. Brandt and J. Ostermann, “Apathy is the Root of all Expressions”, in *Intern. Conf. on Automatic Face and Gesture Recognition*, (FG) 2017

[13] L. Yin, X. Wei, Y. Sun, J. Wang and M. Rosato, “A 3D facial expression database for facial behavior research”, in 7th *Intern. Conf. on Automatic Face and Gesture Recognition*, (FG) 2006, pp. 1-9.

Model Parameter Estimation

Input: 2D landmarks $\mathbf{s}_k^{2D}$

3D shape represented as sum of mean and mean-free shape

$$\widehat{\mathbf{s}}^{3D} = \mathbf{m} + \widehat{\mathbf{v}}\,(\mathbf{p}_2, \mathbf{p}_3),$$

$$\widehat{\mathbf{v}}\,(\mathbf{p}_2, \mathbf{p}_3) = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{p}_2^{\mathrm{T}}\mathbf{U}^{(2)} \times_3 \mathbf{p}_3^{\mathrm{T}}\mathbf{U}^{(3)}, \quad \mathbf{p}_2 \in \mathbb{R}^P,\ \mathbf{p}_3 \in \mathbb{R}^E.$$

Reformulate tensor-product into matrix-vector product

$$\widehat{\mathbf{v}}\,(\mathbf{p}_2, \mathbf{p}_3) = \mathbf{M}_2\mathbf{p}_2 = \mathbf{M}_3\mathbf{p}_3$$

Linear w.r.t. Model Parameters

project 3D points to 2D

$$\widetilde{\mathbf{u}} = \mathbf{K}\left(\mathbf{R}\mathbf{s}_i^{3D} + \mathbf{t}\right) = \mathbf{K}\left(\mathbf{R}(\widehat{\mathbf{v}}_i + \mathbf{m}_i) + \mathbf{t}\right)$$

$$\widehat{\mathbf{s}}_i^{2D} = \left(s_{i,x}^{2D},\ s_{i,y}^{2D}\right)^{\mathrm{T}} = \left(\widetilde{u}_x/\widetilde{u}_z,\ \widetilde{u}_y/\widetilde{u}_z\right)^{\mathrm{T}}$$

Nonlinear Camera

With $s_{i,x}^{2D} = \widetilde{u}_x/\widetilde{u}_z \ \Leftrightarrow\ \widetilde{u}_z s_{i,x}^{2D} = \widetilde{u}_x$

$$\left[\mathbf{K}\mathbf{R}\mathbf{M}_{2,i}\ \mathbf{p}_2 + \mathbf{K}\mathbf{R}\mathbf{m}_i + \mathbf{K}\mathbf{t}\right]_z s_{i,x}^{2D} = \left[\mathbf{K}\mathbf{R}\mathbf{M}_{2,i}\mathbf{p}_2 + \mathbf{K}\mathbf{R}\mathbf{m}_i + \mathbf{K}\mathbf{t}\right]_x$$

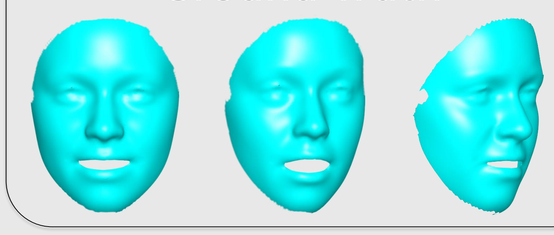
$$\left(\left[\mathbf{K}\mathbf{R}\mathbf{M}_{2,i}\right]_x - s_{i,x}^{2D}\left[\mathbf{K}\mathbf{R}\mathbf{M}_{2,i}\right]_z\right)\mathbf{p}_2 = \left[\mathbf{K}\mathbf{R}\mathbf{m}_i + \mathbf{K}\mathbf{t}\right]_x s_{i,x}^{2D} - \left[\mathbf{K}\mathbf{R}\mathbf{m}_i + \mathbf{K}\mathbf{t}\right]_z$$

Linear Model Parameter Estimation

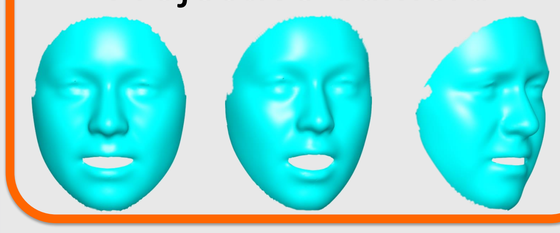
Output: 3D shape $\widehat{\mathbf{s}}^{3D} = \mathbf{M}_2\widehat{\mathbf{p}}_2 + \mathbf{m}$

Results on Synthetic Shapes

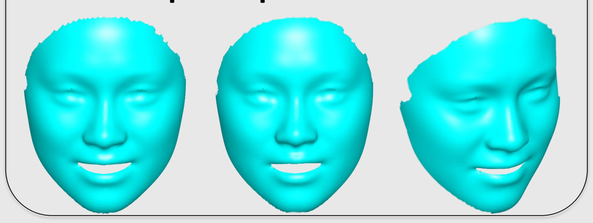
Ground Truth



Projective Camera

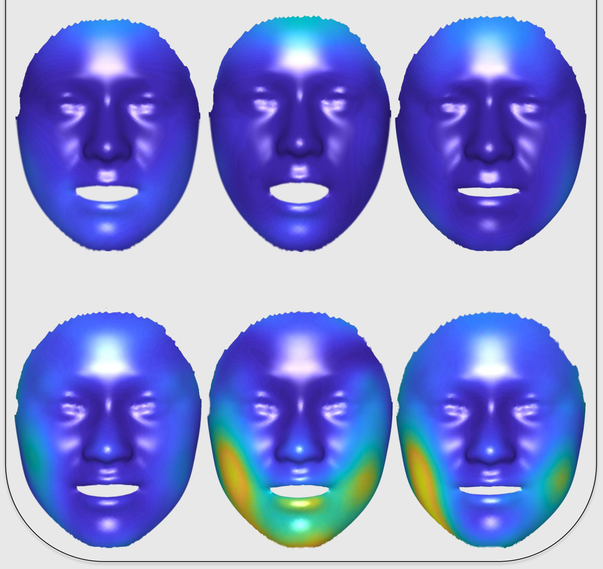


Weak-perspective Camera



Pointwise Shape Error w.r.t. Ground Truth Shape

(blue=small, yellow=high)



Results on Real Images

Institut für Informationsverarbeitung
Stella Graßhof
grasshof@tnt.uni-hannover.de
http://www.tnt.uni-hannover.de/~grasshof

Leibniz Universität Hannover