Tensor-Based Non-Rigid Structure from Motion - Supplementary Material

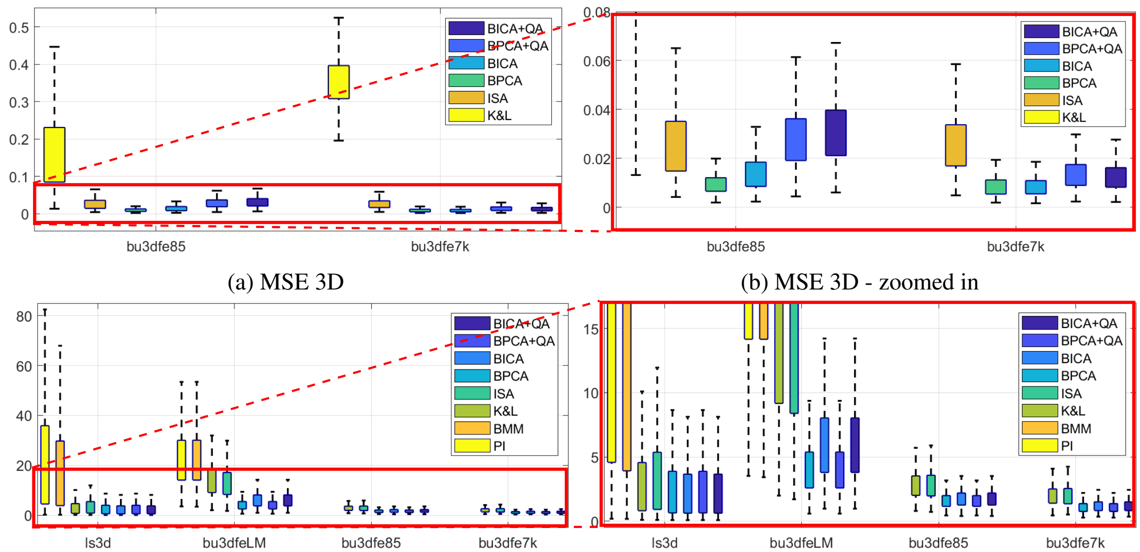
Stella Graßhof and Sami Sebastian Brandt
IT University of Copenhagen, Denmark
stgr@itu.dk, sambr@itu.dk

A. Extended Experiments

In addition to Tab. 1, and Tab. 2, in Sec. 7.3., we present additional boxplots for the 2D and 3D error metrics, evaluated per shape, in Fig. 7. In Fig. 7a, the 3D MSE based on Eq. (19) is presented, for which an additional zoomed in version is shown in Fig. 7b. Accordingly, Fig. 7c shows the corresponding 2D MSE along with the zoomed in version is shown in Fig. 7d. Here it can be seen that the autocalibration only affects the 3D error, i.e., the 2D errors without autocalibration (BPCA, BICA) are the same as with autocalibration (BPCA+QA and BICA+QA). As already stated in the main document (Sec. 7.3), our proposed methods BPCA, BICA, clearly better than all the competing approaches including ISA, K&L, BMM, and PI.

In Fig. 8, we present examples of the 3D reconstruction of samples from the BU3DFE-7k dataset, using the methods referred to as K&L [28], ISA [10], while BPCA and BICA refer to the two variants of the proposed method. These samples illustrate that our methods, BPCA and BICA, lead to qualitatively better results than the method by Kong and Lucey [28], shown in column (b), and by ISA [10] shown in column (c). When inspecting the results, we found that K&L tends to result in rather flat shapes. While the estimated 3D shapes by ISA are better, they are in general not as expressive as by our methods.

We present another example for 3D reconstruction and expression transfer in the wild in Fig. 9, using the proposed combination of tensor model and the BICA factorisation, see Sec. 5. The result shows that the methods yields reasonable results for a sample image that is not part of the training set. Fig. 10 shows an extension of Fig. 7 by illustrating that the rigid part is in fact invariant to the variation of the expression. Finally, Fig. 11 illustrates the performance of the methods with respect to 3D MSE (19) and computational complexity. It can be seen that the method by Kong and Lucey [28] leads to higher errors in 3D and takes longer than the proposed methods BPCA and BICA.



(a) MSE 3D

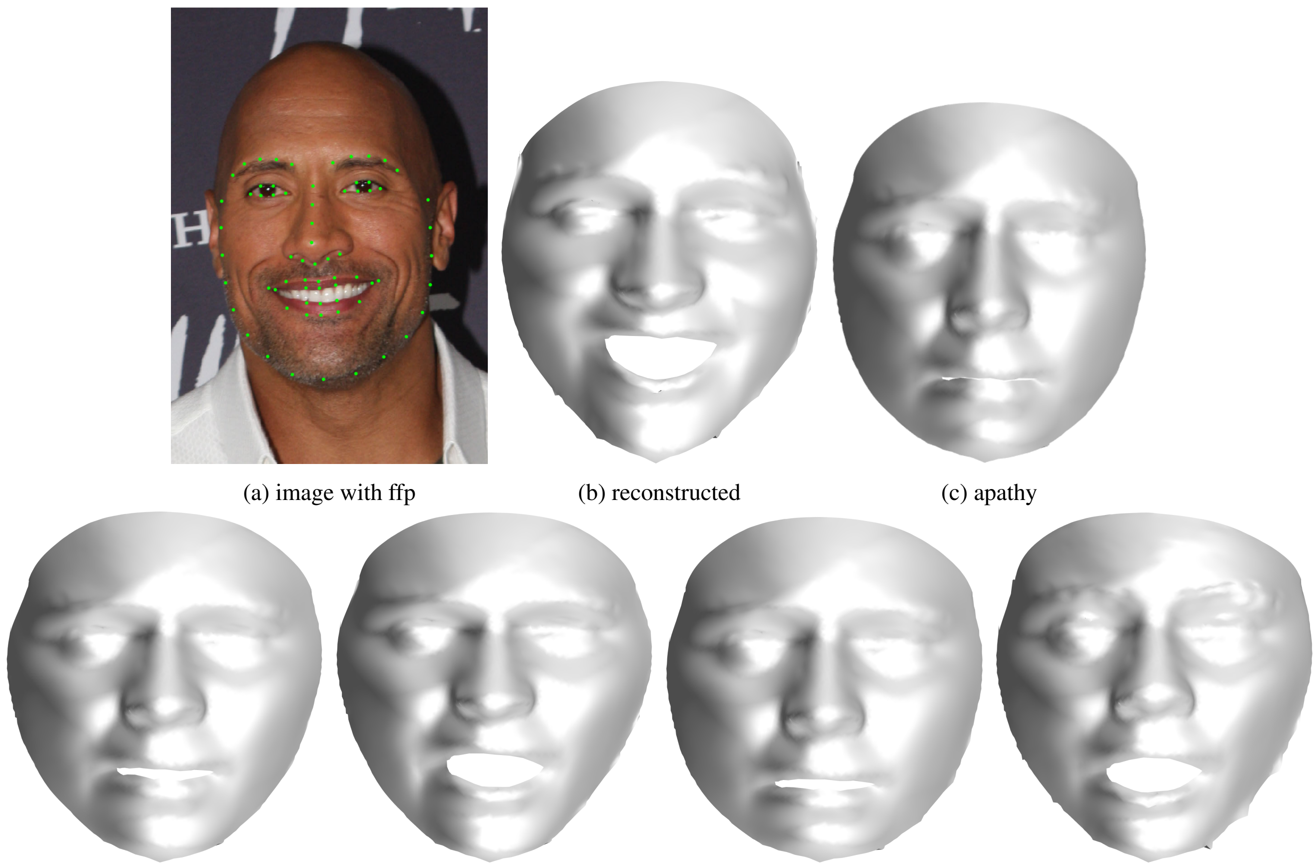
(b) MSE 3D - zoomed in

(c) MSE 2D

(d) MSE 2D - zoomed in

Figure 7: Comparison of NRSfM approaches based on either (a) 3D reconstruction error measured by 3D MSE between estimated to 3D GT shapes, where available, and, (c) 2D MSE between estimated and input 2D shapes. Please note that the 3D GT is not used by any of the approaches.

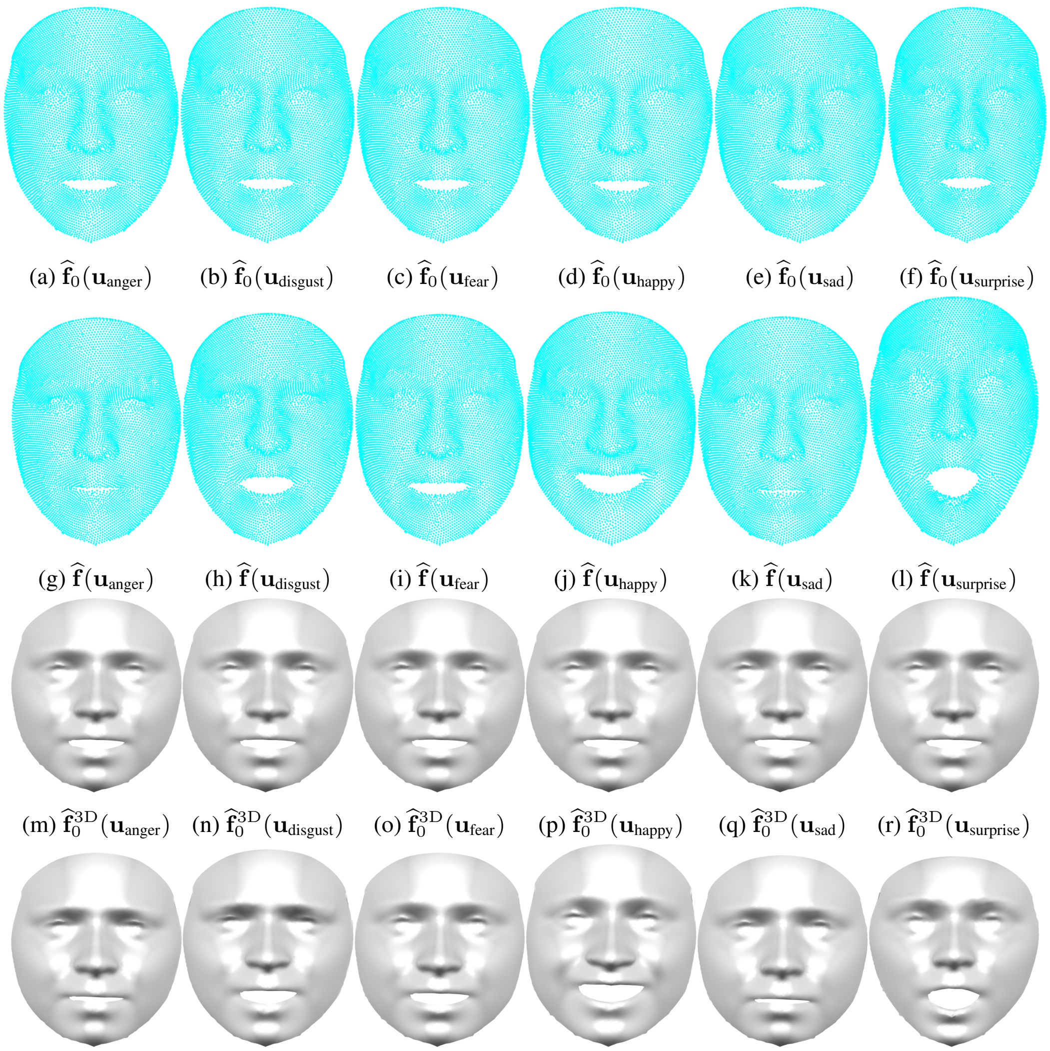
Figure 8: Selected examples of the 3D (affine) reconstruction from the 2D input with 3D ground truth (GT).



(a) image with ffp (b) reconstructed (c) apathy

(d) anger (e) disgust (f) sadness (g) surprise

Figure 9: In the wild 3D reconstruction and expression transfer based on the input image with facial feature points, 3D reconstruction, and synthesized expressions. Please note that the reconstruction is only known up to an unknown affine transform, including flipping, skewing and scaling on each axis, which can be seen when comparing the 3D reconstructions with the ground truth (GT). E.g. when comparing (a) and (b) the slightly asymmetric smile is mirrored. (Image from wikipedia, creative common license, by Eva Rinaldi.)



(a) $\widehat{\mathbf{f}}_0(\mathbf{u}_{\text{anger}})$ (b) $\widehat{\mathbf{f}}_0(\mathbf{u}_{\text{disgust}})$ (c) $\widehat{\mathbf{f}}_0(\mathbf{u}_{\text{fear}})$ (d) $\widehat{\mathbf{f}}_0(\mathbf{u}_{\text{happy}})$ (e) $\widehat{\mathbf{f}}_0(\mathbf{u}_{\text{sad}})$ (f) $\widehat{\mathbf{f}}_0(\mathbf{u}_{\text{surprise}})$

(g) $\widehat{\mathbf{f}}(\mathbf{u}_{\text{anger}})$ (h) $\widehat{\mathbf{f}}(\mathbf{u}_{\text{disgust}})$ (i) $\widehat{\mathbf{f}}(\mathbf{u}_{\text{fear}})$ (j) $\widehat{\mathbf{f}}(\mathbf{u}_{\text{happy}})$ (k) $\widehat{\mathbf{f}}(\mathbf{u}_{\text{sad}})$ (l) $\widehat{\mathbf{f}}(\mathbf{u}_{\text{surprise}})$

(m) $\widehat{\mathbf{f}}_0^{\text{3D}}(\mathbf{u}_{\text{anger}})$ (n) $\widehat{\mathbf{f}}_0^{\text{3D}}(\mathbf{u}_{\text{disgust}})$ (o) $\widehat{\mathbf{f}}_0^{\text{3D}}(\mathbf{u}_{\text{fear}})$ (p) $\widehat{\mathbf{f}}_0^{\text{3D}}(\mathbf{u}_{\text{happy}})$ (q) $\widehat{\mathbf{f}}_0^{\text{3D}}(\mathbf{u}_{\text{sad}})$ (r) $\widehat{\mathbf{f}}_0^{\text{3D}}(\mathbf{u}_{\text{surprise}})$

(s) $\widehat{\mathbf{f}}^{\text{3D}}(\mathbf{u}_{\text{anger}})$ (t) $\widehat{\mathbf{f}}^{\text{3D}}(\mathbf{u}_{\text{disgust}})$ (u) $\widehat{\mathbf{f}}^{\text{3D}}(\mathbf{u}_{\text{fear}})$ (v) $\widehat{\mathbf{f}}^{\text{3D}}(\mathbf{u}_{\text{happy}})$ (w) $\widehat{\mathbf{f}}^{\text{3D}}(\mathbf{u}_{\text{sad}})$ (x) $\widehat{\mathbf{f}}^{\text{3D}}(\mathbf{u}_{\text{surprise}})$

Figure 10: The six prototypical emotions, synthesised by the tensor model (14) for the average rotation, and average person, with varying expression $\mathbf{u}_5$. (a)-(f) rigid 2D shapes $\widehat{\mathbf{f}}_0$ based on (16), (g)-(l) 2D shapes including the nonrigid part $\widehat{\mathbf{f}} = \widehat{\mathbf{f}}_0 + \Delta\widehat{\mathbf{f}}$ (17). The corresponding synthesised 3D shapes (see Sec. 5) are shown in (m)-(r) for rigid, and (s)-(x) including the nonrigid part. Please note that the 3D faces do not have corresponding 2D faces in the training data. The depicted figures (g)-(l) correspond to Fig. 7(b)-(g), and (s)-(x) to Fig. 7(i)-(n). See also the additional videos demonstrating the effect of changing the expression parameter.

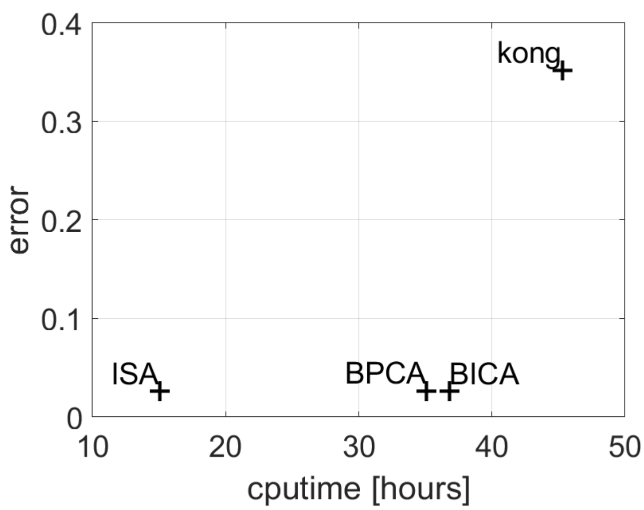


Figure 11: Illustration of the differences between the methods with respect to the 3D error (MSE 3D) (19) and CPU time for the dense dataset BU3DFE-7k.