

Supplementary Material: Inequality-Constrained and Robust 3D Face Model Fitting

Evangelos Sariyanidi¹, Casey J. Zampella¹, Robert T. Schultz^{1,2}, and
Birkan Tunc^{1,2}

¹ Center for Autism Research, Children’s Hospital of Philadelphia

² University of Pennsylvania

{sariyanide,zampellac,schultzrt,tuncb}@chop.edu

Introduction

This supplementary material provides the mathematical expressions that are necessary for the completeness of the paper (Appendix A) and figures that aid the experimental evaluation of the proposed approach (Appendix B).

Appendix A

This section provides the partial derivatives that are needed to obtain the derivative of the objective function but were omitted from the main text (Section 3.4) due to their lengthy expressions—the partial derivatives $\frac{\partial \hat{\mathbf{g}}_{tx}[k]}{\partial \hat{\mathbf{I}}_t[i]}$, $\frac{\partial \hat{\mathbf{g}}_{ty}[k]}{\partial \hat{\mathbf{I}}_t[i]}$, $\frac{\partial \hat{\mathbf{I}}_t^d}{\partial \mathbf{c}_t}$ and $\frac{\partial \mathbf{x}_t}{\partial \mathbf{c}_t}$. Note that the partial derivative $\frac{\partial \hat{\mathbf{I}}_t^d}{\partial \mathbf{x}_t}$ is not explicitly mentioned in the text but it is a block-diagonal matrix (of 1×2 -sized blocks) very similar to $\frac{\partial \hat{\mathbf{I}}_t^f}{\partial \mathbf{x}_t}$ — the only difference is that the blocks contain the (un-normalized) gradients of the image that corresponds to the diffuse reflection component of the Phong illumination model, $\hat{\mathbf{I}}_t^d$, instead of the estimated (illumination-free) face image $\hat{\mathbf{I}}_t^f$.

The k ith entries of the matrices $\partial \hat{\mathbf{g}}_{tx} / \partial \hat{\mathbf{I}}_t$ and $\partial \hat{\mathbf{g}}_{ty} / \partial \hat{\mathbf{I}}_t$ are:

$$\frac{\partial \hat{\mathbf{g}}_{tx}[k]}{\partial \hat{\mathbf{I}}_t[i]} = \begin{cases} \frac{(\hat{\mathbf{I}}_t[k_b] - \hat{\mathbf{I}}_t[k_a])^2}{h^{\frac{3}{2}}} & \text{if } i = k_r \\ \frac{-(\hat{\mathbf{I}}_t[k_b] - \hat{\mathbf{I}}_t[k_a])^2}{h^{\frac{3}{2}}} & \text{if } i = k_l \\ \frac{-(\hat{\mathbf{I}}_t[k_r] - \hat{\mathbf{I}}_t[k_l])}{h^{\frac{3}{2}}} & \text{if } i = k_b \\ \frac{(\hat{\mathbf{I}}_t[k_b] - \hat{\mathbf{I}}_t[k_a])^{-1} h^{\frac{3}{2}}}{\hat{\mathbf{I}}_t[k_r] - \hat{\mathbf{I}}_t[k_l]} & \text{if } i = k_a \\ \frac{(\hat{\mathbf{I}}_t[k_b] - \hat{\mathbf{I}}_t[k_a])^{-1} h^{\frac{3}{2}}}{\hat{\mathbf{I}}_t[k_r] - \hat{\mathbf{I}}_t[k_l]} & \text{if } i = k_a \\ 0 & \text{otherwise} \end{cases} \quad \frac{\partial \hat{\mathbf{g}}_{ty}[k]}{\partial \hat{\mathbf{I}}_t[i]} = \begin{cases} \frac{(\hat{\mathbf{I}}_t[k_a] - \hat{\mathbf{I}}_t[k_b])^2}{h^{\frac{3}{2}}} & \text{if } i = k_b \\ \frac{-(\hat{\mathbf{I}}_t[k_a] - \hat{\mathbf{I}}_t[k_b])^2}{h^{\frac{3}{2}}} & \text{if } i = k_a \\ \frac{-(\hat{\mathbf{I}}_t[k_r] - \hat{\mathbf{I}}_t[k_l])}{h^{\frac{3}{2}}} & \text{if } i = k_r \\ \frac{(\hat{\mathbf{I}}_t[k_b] - \hat{\mathbf{I}}_t[k_a])^{-1} h^{\frac{3}{2}}}{\hat{\mathbf{I}}_t[k_r] - \hat{\mathbf{I}}_t[k_l]} & \text{if } i = k_l \\ \frac{(\hat{\mathbf{I}}_t[k_b] - \hat{\mathbf{I}}_t[k_a])^{-1} h^{\frac{3}{2}}}{\hat{\mathbf{I}}_t[k_r] - \hat{\mathbf{I}}_t[k_l]} & \text{if } i = k_l \\ 0 & \text{otherwise} \end{cases} \quad (\text{S.1})$$

where h , k_l , k_r , k_a and k_b are defined as in Eq. (7) of the main text.

The partial derivative $\frac{\partial \hat{\mathbf{I}}_t^d}{\partial \mathbf{c}_t}$ is an $N \times 6$ matrix, whose each column is the partial derivative w.r.t. one of the camera parameters $\mathbf{c}_t = (q_1, q_2, q_3, \tau_x, \tau_y, \tau_z)$, *i.e.*,

$$\frac{\partial \hat{\mathbf{I}}_t^d}{\partial \mathbf{c}_t} = \begin{bmatrix} \frac{\partial \hat{\mathbf{I}}_t^d}{\partial q_1}, \frac{\partial \hat{\mathbf{I}}_t^d}{\partial q_2}, \frac{\partial \hat{\mathbf{I}}_t^d}{\partial q_3}, \frac{\partial \hat{\mathbf{I}}_t^d}{\partial \tau_x}, \frac{\partial \hat{\mathbf{I}}_t^d}{\partial \tau_y}, \frac{\partial \hat{\mathbf{I}}_t^d}{\partial \tau_z} \end{bmatrix}. \quad (\text{S.2})$$

The i th entries of the vectors on the right-hand-side of (S.2) are computed as

$$\frac{\partial \hat{\mathbf{I}}_t^d[i]}{\partial q_1} = 2\bar{\mathbf{n}}_i^T \begin{bmatrix} q_3\lambda_z + q_2\lambda_y - q_2\tau_y - q_3\tau_z \\ q_2\lambda_x - 2q_1\lambda_y + q_0\lambda_z - q_2\tau_x + 2q_1\tau_y - q_0\tau_z \\ q_3\lambda_x - q_0\lambda_y - 2q_1\lambda_z - q_3\tau_x + q_0\tau_y + q_1\tau_z \end{bmatrix} \quad (\text{S.3})$$

$$\frac{\partial \hat{\mathbf{I}}_t^d[i]}{\partial q_2} = 2\bar{\mathbf{n}}_i^T \begin{bmatrix} -2q_2\lambda_x + q_1\lambda_y - q_0\lambda_z + 2q_2\tau_x - q_1\tau_y + q_0\tau_z \\ q_1\lambda_x + q_3\lambda_z - q_1\tau_x - q_3\tau_z \\ q_0\lambda_x + q_3\lambda_y - 2q_2\lambda_z - q_0\tau_x - q_3\tau_y + 2q_2\tau_z \end{bmatrix} \quad (\text{S.4})$$

$$\frac{\partial \hat{\mathbf{I}}_t^d[i]}{\partial q_3} = 2\bar{\mathbf{n}}_i^T \begin{bmatrix} -2q_3\lambda_x + q_0\lambda_y + q_1\lambda_z + 2q_3\tau_x - q_0\tau_y - q_1\tau_z \\ -q_0\lambda_x - 2q_3\lambda_y + q_2\lambda_z + q_0\tau_x + 2q_3\tau_y - q_2\tau_z \\ q_1\lambda_x + q_2\lambda_y - q_1\tau_x - 2q_2\tau_y \end{bmatrix} \quad (\text{S.5})$$

$$\frac{\partial \hat{\mathbf{I}}_t^d[i]}{\partial \tau_x} = -2\bar{\mathbf{n}}_i^T \begin{bmatrix} 0.5 - q_2^2 - q_3^2 \\ q_1q_2 - q_0q_3 \\ q_0q_2 + q_1q_3 \end{bmatrix}, \quad (\text{S.6})$$

$$\frac{\partial \hat{\mathbf{I}}_t^d[i]}{\partial \tau_y} = -2\bar{\mathbf{n}}_i^T \begin{bmatrix} q_1q_2 + q_0q_3 \\ 0.5 - q_1^2 - q_3^2 \\ -q_0q_1 + q_2q_3 \end{bmatrix}, \quad (\text{S.7})$$

$$\frac{\partial \hat{\mathbf{I}}_t^d[i]}{\partial \tau_z} = -2\bar{\mathbf{n}}_i^T \begin{bmatrix} -q_0q_2 + q_1q_3 \\ q_0q_1 + q_2q_3 \\ 0.5 - q_1^2 - q_2^2 \end{bmatrix} \quad (\text{S.8})$$

where $\bar{\mathbf{n}}_i$ is the surface normal for the i th point of the mesh (w.r.t. canonical pose; *i.e.* without view transformation), and $(\lambda_x, \lambda_y, \lambda_z)$ is the 3D location of the illumination source, as defined in Section 3.1 of the main text.

The other partial derivative term, $\frac{\partial \hat{\mathbf{x}}_t}{\partial \mathbf{c}_t}$, has been calculated in previous studies (see Supplementary material of Booth *et al.* [7]), but we present them here for completeness and coherence with our notation.³ The partial derivative $\frac{\partial \hat{\mathbf{x}}_t}{\partial \mathbf{c}_t}$ is a $2N \times 6$ matrix,

³ We also do a minor correction to [7]. The last derivative in Eq. (15) of Supp. material of [7] has a denominator with an expression in the power of 3. As we show in Eqs. (S.14) and (S.17) of this supplementary material, the expression in the denominator should be in the power of 2.

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{c}_t} = \begin{bmatrix} \frac{\partial x_{t1}}{\partial q_1} & \frac{\partial x_{t1}}{\partial q_2} & \frac{\partial x_{t1}}{\partial q_3} & \frac{\partial x_{t1}}{\partial \tau_x} & \frac{\partial x_{t1}}{\partial \tau_y} & \frac{\partial x_{t1}}{\partial \tau_z} \\ \frac{\partial y_{t1}}{\partial q_1} & \frac{\partial y_{t1}}{\partial q_2} & \frac{\partial y_{t1}}{\partial q_3} & \frac{\partial y_{t1}}{\partial \tau_x} & \frac{\partial y_{t1}}{\partial \tau_y} & \frac{\partial y_{t1}}{\partial \tau_z} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial x_{tN}}{\partial q_1} & \frac{\partial x_{tN}}{\partial q_2} & \frac{\partial x_{tN}}{\partial q_3} & \frac{\partial x_{tN}}{\partial \tau_x} & \frac{\partial x_{tN}}{\partial \tau_y} & \frac{\partial x_{tN}}{\partial \tau_z} \\ \frac{\partial y_{tN}}{\partial q_1} & \frac{\partial y_{tN}}{\partial q_2} & \frac{\partial y_{tN}}{\partial q_3} & \frac{\partial y_{tN}}{\partial \tau_x} & \frac{\partial y_{tN}}{\partial \tau_y} & \frac{\partial y_{tN}}{\partial \tau_z} \end{bmatrix}, \quad (\text{S.9})$$

whose entries are the derivatives of all 2D image points w.r.t. camera parameters. For the i th point, those derivatives are computed as

$$\frac{\partial x_{ti}}{\partial q_j} = \frac{\phi_x}{(v_{ti}^z)^2} \left(\frac{\partial v_{ti}^x}{\partial q_j} v_{ti}^z - \frac{\partial v_{ti}^z}{\partial q_j} v_{ti}^x \right) \quad (\text{S.10})$$

$$\frac{\partial y_{ti}}{\partial q_j} = \frac{\phi_y}{(v_{ti}^z)^2} \left(\frac{\partial v_{ti}^y}{\partial q_j} v_{ti}^z - \frac{\partial v_{ti}^z}{\partial q_j} v_{ti}^y \right) \quad (\text{S.11})$$

$$\frac{\partial x_{ti}}{\partial \tau_x} = \phi_x \frac{1}{v_{ti}^z} \quad (\text{S.12})$$

$$\frac{\partial x_{ti}}{\partial \tau_y} = 0 \quad (\text{S.13})$$

$$\frac{\partial x_{ti}}{\partial \tau_z} = -\phi_x \frac{v_{ti}^x}{(v_{ti}^z)^2} \quad (\text{S.14})$$

$$\frac{\partial y_{ti}}{\partial \tau_x} = 0 \quad (\text{S.15})$$

$$\frac{\partial y_{ti}}{\partial \tau_y} = \phi_y \frac{1}{v_{ti}^z} \quad (\text{S.16})$$

$$\frac{\partial y_{ti}}{\partial \tau_z} = -\phi_y \frac{v_{ti}^y}{(v_{ti}^z)^2}. \quad (\text{S.17})$$

Some of the formulae above require the partial derivatives of the (view-transformed) 3D points w.r.t. rotation parameters (*i.e.*, quaternions q_1, q_2, q_3), which are com-

puted as

$$\frac{\partial v_{ti}^x}{\partial q_1} = 2(q_2 p_{ti}^y + q_3 p_{ti}^z) \quad (\text{S.18})$$

$$\frac{\partial v_{ti}^x}{\partial q_2} = 2(-2q_2 p_{ti}^x + q_1 p_{ti}^y + q_0 p_{ti}^z) \quad (\text{S.19})$$

$$\frac{\partial v_{ti}^x}{\partial q_3} = 2(-2q_3 p_{ti}^x - q_0 p_{ti}^y + q_1 p_{ti}^z) \quad (\text{S.20})$$

$$\frac{\partial v_{ti}^y}{\partial q_1} = 2(q_2 p_{ti}^x - 2q_1 p_{ti}^y - q_0 p_{ti}^z) \quad (\text{S.21})$$

$$\frac{\partial v_{ti}^y}{\partial q_2} = 2(q_1 p_{ti}^x + q_3 p_{ti}^z) \quad (\text{S.22})$$

$$\frac{\partial v_{ti}^y}{\partial q_3} = 2(q_0 p_{ti}^x - 2q_3 p_{ti}^y + q_2 p_{ti}^z) \quad (\text{S.23})$$

$$\frac{\partial v_{ti}^z}{\partial q_1} = 2(q_3 p_{ti}^x + q_0 p_{ti}^y - 2q_1 p_{ti}^z) \quad (\text{S.24})$$

$$\frac{\partial v_{ti}^z}{\partial q_2} = 2(-q_0 p_{ti}^x + q_3 p_{ti}^y - 2q_2 p_{ti}^z) \quad (\text{S.25})$$

$$\frac{\partial v_{ti}^z}{\partial q_3} = 2(q_1 p_{ti}^x + q_2 p_{ti}^y). \quad (\text{S.26})$$

Appendix B

This section provides visualizations of facial shapes obtained with the Basel 2017 morphable model (Fig. 1); images of the Synthesized dataset (Fig. 6); visual illustrations of the facial shapes generated with the methods compared in the experiments (3); additional illustrations of morphable model fitting results on the AFLW2000-3D dataset (Fig. 4); and the results of our multi-frame method with 3 and 9 frames in addition to 5 frames (Fig. 5).

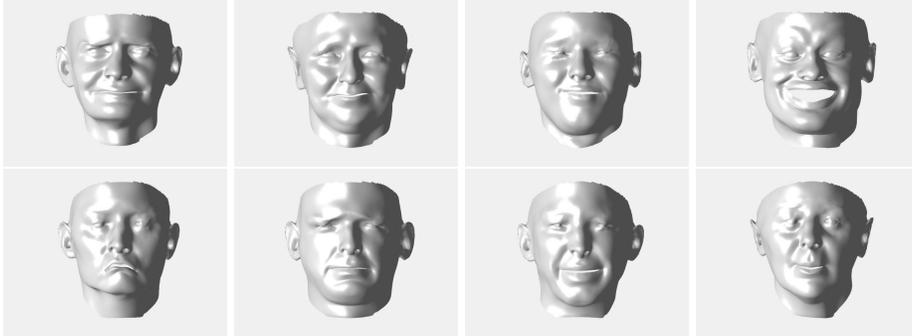


Fig. 1. Random faces generated using the Basel 2017 morphable model. The basis coefficients used to generate the faces were sampled from Normal distributions. Specifically, each basis coefficient was sampled independently from the others, using a zero-mean Normal distribution whose standard deviation, σ_i , is learned while constructing the model; therefore, the coefficients are mostly in the interval $[-3\sigma_i, 3\sigma_i]$. However, the faces from this interval tend to look awkward as can be seen with a close inspection; for example, some faces on the bottom row have unrealistic-looking eyes due to protrusions that are unlikely to exist in real faces. Therefore, in experiments we used a reduced interval of $[-1.5\sigma_i, 1.5\sigma_i]$, which is well capable of representing faces with very diverse characteristics and facial expressions with very large magnitude (*e.g.*, see Fig. 4 below)

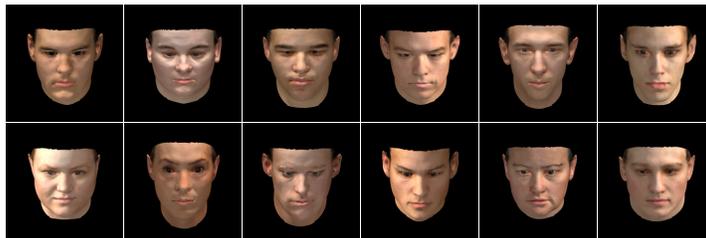


Fig. 2. Illustrations of facial images of the Synthesized dataset, which was generated using the Basel 2009 morphable model and used in the paper for experimental evaluation.

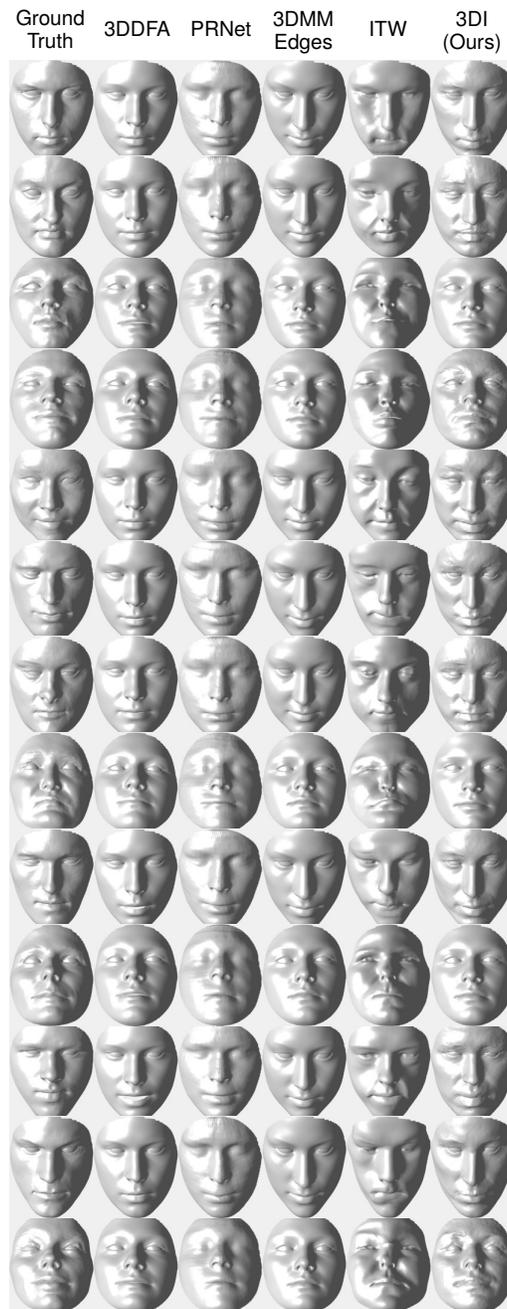
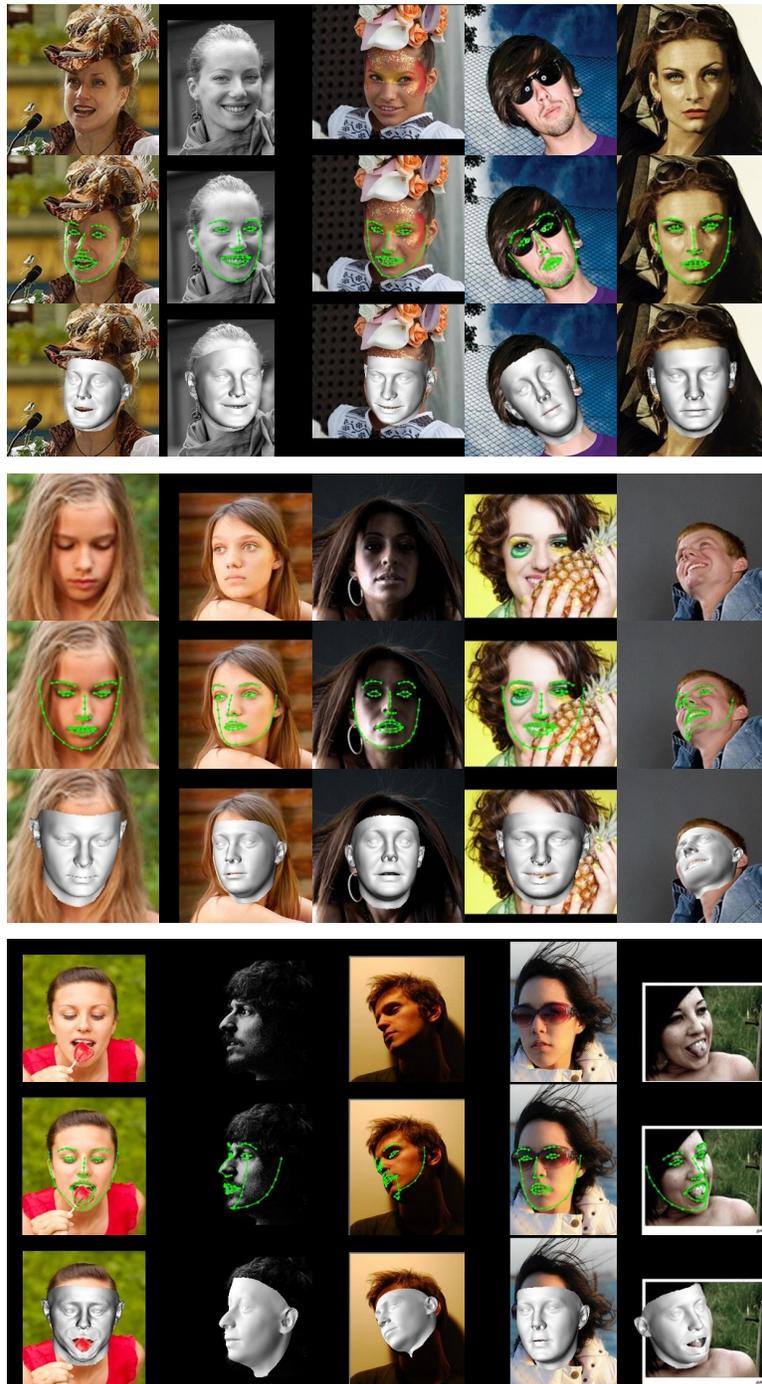


Fig. 3. Illustrations of the (estimated) facial shapes generated by the methods compared in experiments (Section 4) in comparison to the true (*i.e.*, ground truth) facial shapes.



(Continued below)



(Continued below)

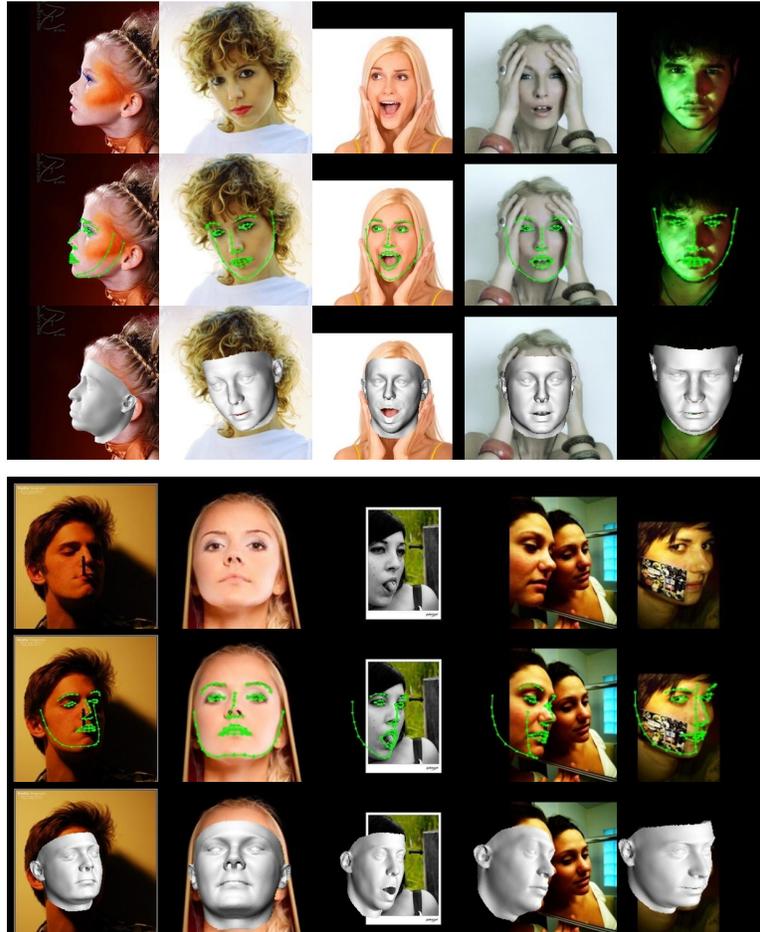


Fig. 4. Additional qualitative illustrations that depict our method’s performance on the AFLW2000-3D dataset. We separately demonstrate each input image, the 2D landmarks estimated in the input image by our method, and the dense 3D shape estimated by our method. It is notable that our method can successfully operate in such uncontrolled conditions, even though we use a 3DMM constructed from controlled data, namely Basel 2017

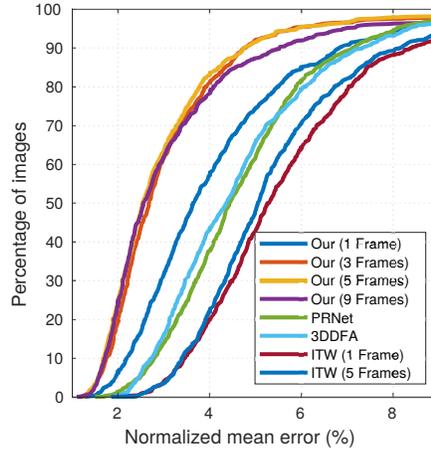


Fig. 5. Cumulative Error Distribution (CED) of compared methods on the BU4DFE dataset. Normalized mean error (NME) was computed by dividing to the distance between the outer eye corners. Our method visibly outperforms other methods, especially when used with multiple frames. The performance of our method with 3 and 5 frames is very similar. Interestingly, performance slightly deteriorates when more (*i.e.*, 9) frames are used; this may be due to the fact that the optimization problem that is solved becomes more difficult as more frame are used, and that using more than, say, 5 frames provides little extra information that cannot be obtained from 5 frames

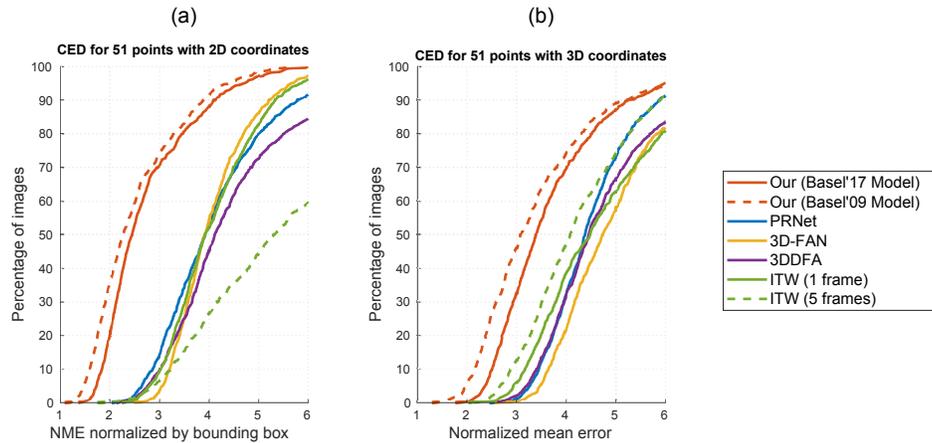


Fig. 6. Cumulative Error Distribution (CED) of compared methods on the Synthesized dataset for $L = 51$ landmark points. Differently from the corresponding figure in the main text, here we provide results with our method obtained with the Basel'17 model in addition to the Basel'09 model (for a single frame). These results prove that our method outperforms other methods even when the model that is used for fitting (Basel'17) is different from the model that is used when generating the dataset (Basel'09).