Supplemental Material for Face Alignment Across Large Poses: A 3D Solution

Xiangyu Zhu[1] Zhen Lei[1*] Xiaoming Liu[2] Hailin Shi[1] Stan Z. Li[1]
[1] Center for Biometrics and Security Research & National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences,
[2]Department of Computer Science and Engineering, Michigan State University
{xiangyu.zhu,zlei,hailin.shi,szli}@nlpr.ia.ac.cn liuxm@msu.edu

1. Results Demo

In this section, we demonstrate some alignment results of 3DDFA in AFLW in Fig. 1. Since the landmark visibility can be easily computed from the fitted dense 3D model [1], we also demonstrate the landmark visibility.

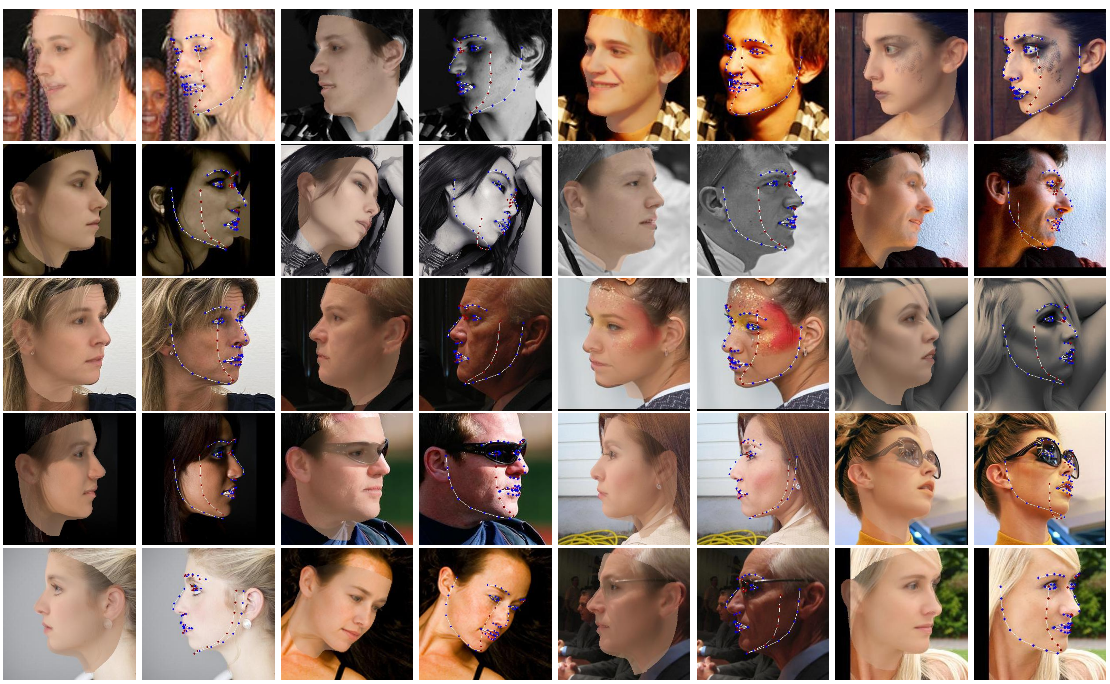


Figure 1. The results of 3DDFA in AFLW. For each pair, the left one renders the fitted 3D shape with the mean texture, which is made transparent to demonstrate the fitting accuracy. The right one shows the landmarks overlayed on the 3D face model. The blue/red ones indicate visible/invisible landmarks.

2. Database

2.1. 300W-3D

This database contains the images in 300W [4] and their ground truth 3D faces. It is constructed by fitting the 3D Morphable Model with the Multi-Features Framework [3]. Different from the original algorithm, the 3D landmarks are

*Corresponding author.

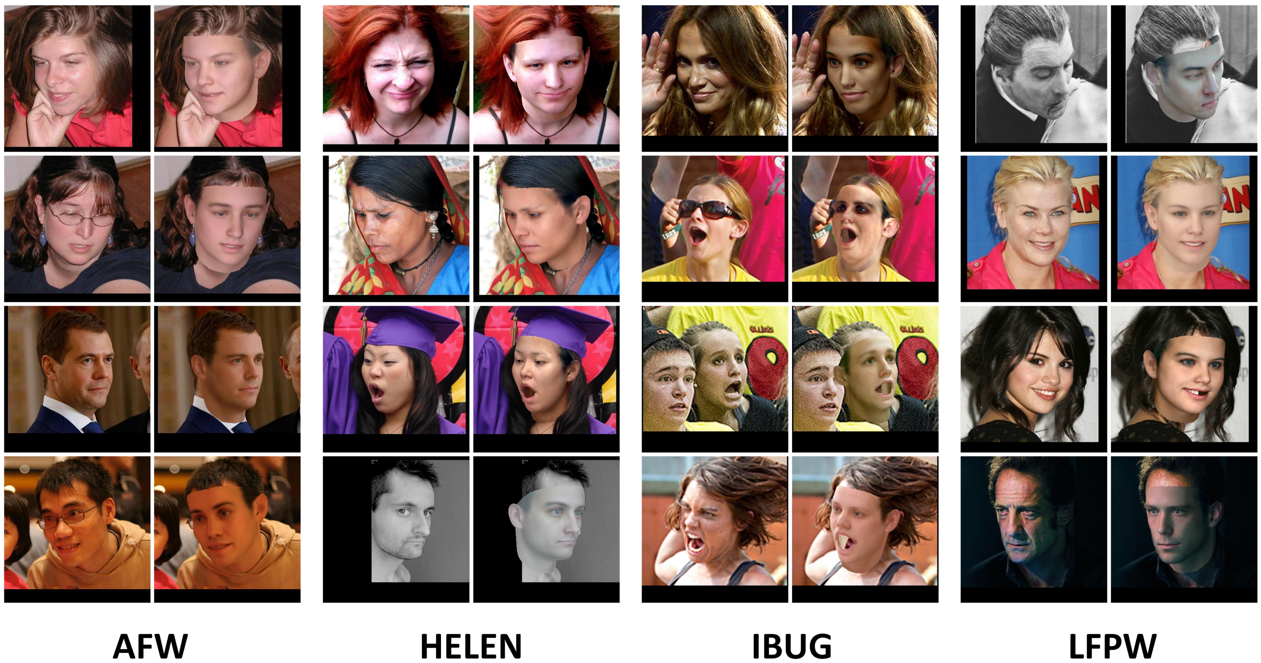
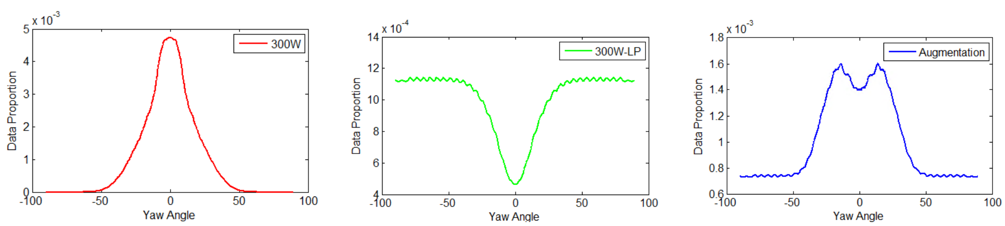


Figure 2. The **300W-3D** database. For each sample, the left one is the original image, the right one is the fitted 3DMM.

adjusted by landmark marching [5] and the 68-landmarks constraint is adopted throughout the fitting process. Each sample is checked, few failed samples are adjusted manually. Fig. 2 demonstrates some samples in the database.

2.2. 300W-LP

The 300W across Large Poses (300W-LP) database contains the synthesized face images from the face profiling algorithm described in Section 4. Some samples are demonstrated in Fig. 4. Besides, Fig. 3(a) and Fig. 3(b) demonstrate the yaw angle distribution before and after face profiling respectively. The distribution is reversed since face profiling only enlarges the yaw angle without shrinking it. Considering that the fidelity of a synthesized sample is negatively related with the Δyaw, we augment each training sample with the times of $\lceil (5 * 0.8^{\Delta yaw/5^{\circ}}) \rceil$. Fig. 3(c) shows the yaw distribution after augmentation.



(a) (b) (c)

Figure 3. The yaw angle distribution of (a) 300W, (b) 300W-LP, (c) After augmentation.

2.3. AFLW2000-3D

The AFLW2000-3D database contains the first 2000 samples in AFLW [2] with their ground truth 3D faces. This database is more challenging to construct than 300W-3D because the AFLW ignores the occluded landmarks (both occluded and self-occluded) and the landmarks do not contain expression information (without lip landmarks). As a result we need a specifical method to deal with the unstandardised landmarks. Firstly, since AFLW provides the ground truth pose information, we train a pose-dependent SDM (a SDM model for each yaw interval) with 68-landmark makeup on 300W-LP, with which the AFLW2000 is coarsely aligned for initialization. Secondly, we run Multi-Features Framework (MFF) 3DMM fitting method initialized by the 68 landmarks and constrained by the provided 21 visible landmarks. Finally, for the failed results, we label

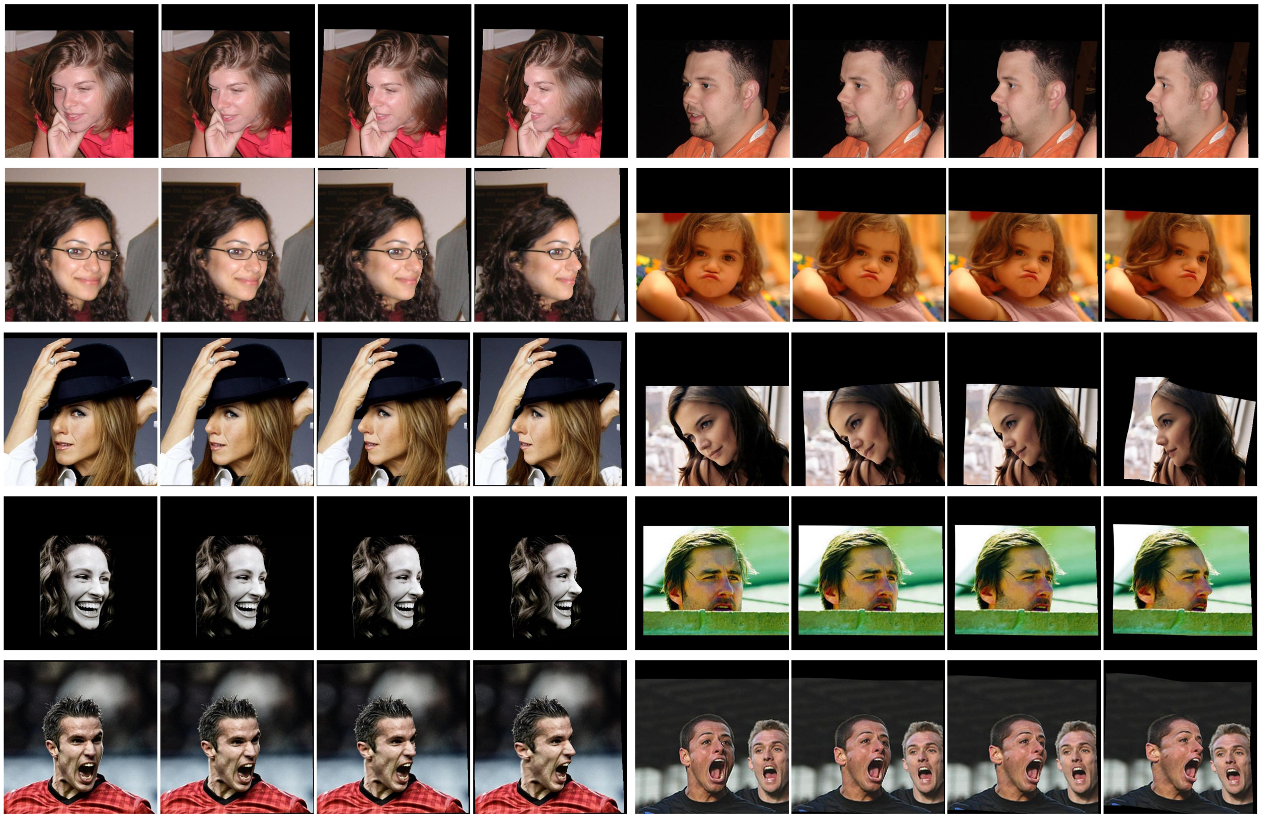


Figure 4. The **300W-LP** database. For each sample, the first is the original image, followed by synthesized larger-pose faces, each with increased 10 degree yaw.

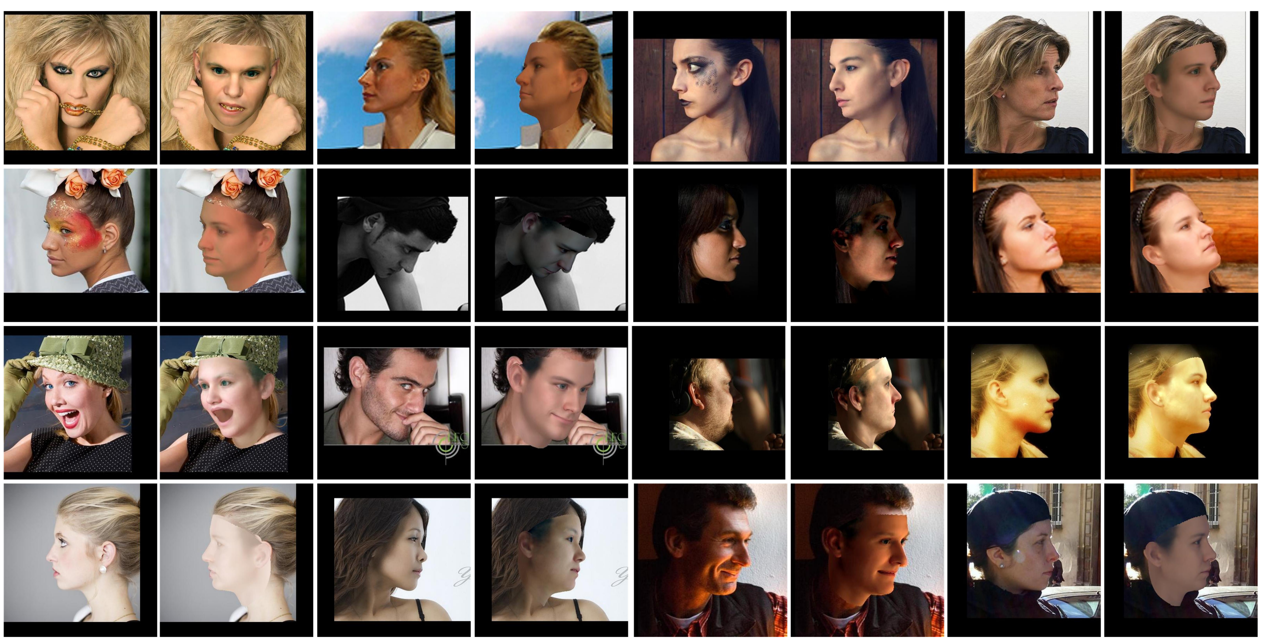


Figure 5. The **AFLW2000-3D** database. For each sample, the left one is the original image, the right one is the fitted 3DMM.

the necessary landmarks (always the upper and lower lip landmarks) and re-run the MFF. Fig. 5 demonstrates some samples in AFLW2000-3D.

3. Derivative of Vertex Distance Cost (VDC)

The Vertex Distance Cost is defined in Section 3.4.2:

$$E_{vdc} = \|V(\mathbf{p}^0 + \Delta\mathbf{p}) - V(\mathbf{p}^g)\|^2 \tag{1}$$

where $\mathbf{p} = [f, pit, yaw, rol, \mathbf{t}_{2d}, \boldsymbol{\alpha}_{id}, \boldsymbol{\alpha}_{exp}]^T$ contains all the parameters including the scale f, rotation angles pit, yaw, rol (which are short for pitch,yaw,roll), translation $\mathbf{t}_{2d}$, shape $\boldsymbol{\alpha}_{id}$ and expression $\boldsymbol{\alpha}_{exp}$. $V(\mathbf{p})$ is the model construction and projection process defined as:

$$V(\mathbf{p}) = f * \mathbf{Pr} * \mathbf{R} * (\overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}) + \mathbf{t}_{2d} \tag{2}$$

$$\mathbf{Pr} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad \mathbf{R} = \mathbf{R}_{pit} * \mathbf{R}_{yaw} * \mathbf{R}_{rol} \tag{3}$$

$$\mathbf{R}_{pit} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(pit) & \sin(pit) \\ 0 & -\sin(pit) & \cos(pit) \end{bmatrix} \quad \mathbf{R}_{yaw} = \begin{bmatrix} \cos(yaw) & 0 & -\sin(yaw) \\ 0 & 1 & 0 \\ \sin(yaw) & 0 & \cos(yaw) \end{bmatrix} \quad \mathbf{R}_{rol} = \begin{bmatrix} \cos(rol) & \sin(rol) & 0 \\ -\sin(rol) & \cos(rol) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where $\overline{\mathbf{S}}$ is the mean 3D shape, $\mathbf{A}_{id}$ and $\mathbf{A}_{exp}$ are the shape and expression PCA models respectively and $\mathbf{Pr}$ is the weak perspective projection matrix.

The derivative of VDC is

$$\frac{\partial E_{vdc}}{\partial \Delta\mathbf{p}} = (V(\mathbf{p}^0 + \Delta\mathbf{p}) - V(\mathbf{p}^g))^T \frac{\partial V}{\partial \mathbf{p}}\Big|_{\mathbf{p}=\mathbf{p}^0+\Delta\mathbf{p}} \tag{5}$$

$$\frac{\partial V}{\partial \mathbf{p}} = \left[\frac{\partial V}{\partial f}(:), \frac{\partial V}{\partial pit}(:), \frac{\partial V}{\partial yaw}(:), \frac{\partial V}{\partial rol}(:), \frac{\partial V}{\partial t_{2dx}}(:), \frac{\partial V}{\partial t_{2dy}}(:), \frac{\partial V}{\partial \boldsymbol{\alpha}_{id}}(:), \frac{\partial V}{\partial \boldsymbol{\alpha}_{exp}}(:)\right] \tag{6}$$

$$\frac{\partial V}{\partial f} = \mathbf{Pr} * \mathbf{R} * (\overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}) \tag{7}$$

$$\frac{\partial V}{\partial pit} = f * \mathbf{Pr} * \mathbf{R}'_{pit} * \mathbf{R}_{yaw} * \mathbf{R}_{rol} * (\overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}) \tag{8}$$

$$\frac{\partial V}{\partial yaw} = f * \mathbf{Pr} * \mathbf{R}_{pit} * \mathbf{R}'_{yaw} * \mathbf{R}_{rol} * (\overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}) \tag{9}$$

$$\frac{\partial V}{\partial rol} = f * \mathbf{Pr} * \mathbf{R}_{pit} * \mathbf{R}_{yaw} * \mathbf{R}'_{rol} * (\overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}) \tag{10}$$

$$\frac{\partial V}{\partial t_{2dx}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix} \qquad \frac{\partial V}{\partial t_{2dy}} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{11}$$

$$\frac{\partial V}{\partial \boldsymbol{\alpha}_{id}} = f * \mathbf{Pr} * \mathbf{R} * \mathbf{A}_{id} \qquad \frac{\partial V}{\partial \boldsymbol{\alpha}_{exp}} = f * \mathbf{Pr} * \mathbf{R} * \mathbf{A}_{exp} \tag{12}$$

$$\mathbf{R}'_{pit} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\sin(pit) & \cos(pit) \\ 0 & -\cos(pit) & -\sin(pit) \end{bmatrix} \quad \mathbf{R}'_{yaw} = \begin{bmatrix} -\sin(yaw) & 0 & -\cos(yaw) \\ 0 & 0 & 0 \\ \cos(yaw) & 0 & -\sin(yaw) \end{bmatrix} \quad \mathbf{R}'_{rol} = \begin{bmatrix} -\sin(rol) & \cos(rol) & 0 \\ -\cos(rol) & -\sin(rol) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{13}$$

where $(:)$ is the concatenation operator which is the same as Matlab.

References

[1] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[2] M. Köstinger, P. Wohlhart, P. M. Roth, and H. Bischof. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 2144–2151. IEEE, 2011.

[3] S. Romdhani and T. Vetter. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *Computer Vision and Pattern Recognition (CVPR), 2005 IEEE Conference on*, volume 2, pages 986–993. IEEE, 2005.

[4] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 397–403. IEEE, 2013.

[5] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li. High-fidelity pose and expression normalization for face recognition in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 787–796, 2015.