Supplementary Material
Joint Face Alignment and 3D Face Reconstruction

Feng Liu[1], Dan Zeng[1], Qijun Zhao[1] and Xiaoming Liu[2]

[1] Sichuan University
qjzhao@scu.edu.cn
[2] Michigan State University

In this supplementary material, we provide additional evaluation results and details, including

- 3D face reconstruction results on the Basel Face Model (BFM) database [1] with comparison to the state-of-the-art methods: Liu et al. [2], SSF [3], and MFF [4].
- Face alignment results on the BU4DFE database [5] with comparison to the state-of-the-art method in [6].
- Detail of evaluation experiment on the LFW [7] and AFW [8] databases.
- Discussion of convergence issue.

1 Evaluation Results on BFM

The BFM database provides a 3D Morphable Model (3DMM) and 10 testing face scans. Each face scan has nine face images of neutral expression and different poses, including one frontal view and eight yaw poses (± 15, ± 30, ± 50 and ± 70 degrees). We use the 3DMM to construct 120 synthetic 3D faces (50% female). From each of them, 2D face images of neutral expression are generated at 21 different poses of (0, ± 10, ± 15, ± 20, ± 30, ± 40, ± 50, ± 60, ± 70, ± 80 and ± 90 degrees) with facial landmarks recorded. The resolution of the rendered 2D face images is 875×656 pixels. These data are used as training data for the proposed method.

The proposed method is compared with several state-of-the-art methods, including Liu et al. [2], SSF [3], and MFF [4], on the 10 testing face scans in terms of 3D face reconstruction accuracy. Figure 1 shows the MAE of different methods on the BFM database with respect to different poses of face images. As can be seen, the MAE of the proposed method is consistently better than that of SSF and MFF, and comparable to that of Liu et al.. It is worth mentioning that Liu et al. need manually marked landmarks as input, whereas the proposed method is fully automatic and can simultaneously reconstruct 3D faces and detect facial landmarks.

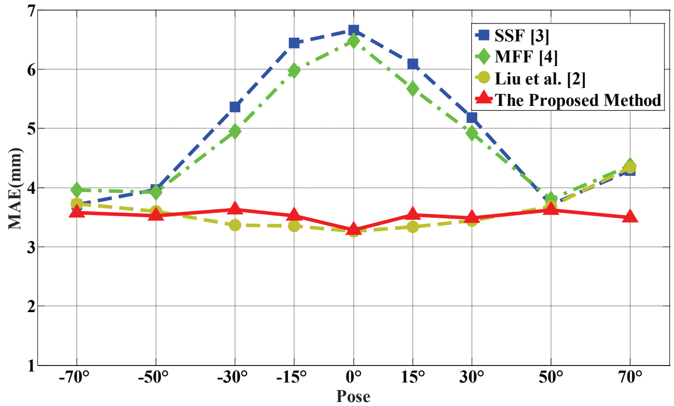

Fig. 1. MAE of the proposed method and three state-of-the-art methods (Liu et al., SSF and MFF) at different poses on the BFM database.

2 Evaluation Results on BU4DFE

The BU4DFE database consists of 2D and 3D videos for six posed prototypical facial expressions (anger, disgust, fear, happy, sad, surprise) for 101 ethnically diverse subjects (58 female and 43 male). In [6], Tulyakov and Sebe sample 3000 face images with the corresponding 3D meshes from the BU4DFE, and generate images with different poses ($[-50, 50]$ degrees for yaw and pitch). They apply a 6-fold cross-validation to split these data into training and testing subsets, and evaluate the localization accuracy of both 2D and 3D landmarks by their method. In order to fairly compare the proposed method with the method of Tulyakov and Sebe, we follow the experimental setup of Tulyakov and Sebe. Note that we use the 3D faces of neutral expression as the ground truth for the proposed method to assess the localization accuracy of 3D landmarks because the proposed method generates pose and expression normalized 3D faces. Besides, we use the method in [9] to establish dense correspondence of the 3D faces as required by the proposed method.

Table 1 reports the NME of the proposed method and Tulyakov and Sebe's method for 2D and 3D landmark localization on the BU4DFE database. Note that here the error is normalized by the inter-ocular distance as Tulyakov and Sebe did in [6]. As can be seen, the proposed method achieves higher accuracy for both 2D and 3D landmarks. Figure 2 shows the reconstruction and alignment results of six subjects by the proposed method. Compared with Tulkyakov and Sebe's method, the proposed method not only locates the landmarks, but also generates dense 3D face shapes for the input images.

Method	2D	3D
Tulyakov and Sebe's method [6]	5.15%	5.92%
The proposed method	3.86%	4.90%

Table 1. NME of 2D and 3D landmarks detected by the proposed method and the method in [6].

Fig. 2. Reconstruction and alignment results for six BU4DFE subjects under different poses and expressions. Column one are input images. Columns 2 and 4 show the ground truth ('GT') 3D faces from two views. Column 3 and 5 show the reconstructed ('REC') 3D faces. The detected visible/invisible landmarks are shown with green/red color in Column 6.

3 Detail of Experiment on LFW/AFW

3.1 Training Data from LFW

In order to make the obtained regression models robust to pose, illumination and expression (PIE) variations, multiple images of varying PIE are required for each subject in the training dataset. Hence, we choose 150 subjects from the LFW database, each having at least 10 images, to train our proposed regression models. The names and the numbers of images of these chosen subjects are listed in Table 2.

Subjects	#images	Subjects	#images	Subjects	#images
George W Bush	530	Lindsay Davenport	22	Hillary Clinton	14
Colin Powell	236	Naomi Watts	22	James Blake	14
Tony Blair	144	Pete Sampras	22	Kim Clijsters	14
Donald Rumsfeld	121	Amelie Mauresmo	21	Mahathir Mohamad	14
Gerhard Schroeder	109	Carlos Menem	21	Roger Federer	14
Ariel Sharon	77	Jennifer Aniston	21	Yoriko Kawaguchi	14
Hugo Chavez	71	Jennifer Lopez	21	Ari Fleischer	13
Junichiro Koizumi	60	Angelina Jolie	20	Charles Moose	13
Jean Chretien	55	Igor Ivanov	20	Edmund Stoiber	13
John Ashcroft	53	Jiang Zemin	20	George HW Bush	13
Jacques Chirac	52	Michael Bloomberg	20	Gordon Brown	13
Serena Williams	52	Paul Bremer	20	Jackie Chan	13
Vladimir Putin	49	Abdullah Gul	19	Joe Lieberman	13
Luiz Inacio Lula da Silva	48	Carlos Moya	19	Lucio Gutierrez	13
Gloria Macapagal Arroyo	44	John Howard	19	Queen Elizabeth II	13
Arnold Schwarzenegger	42	Joschka Fischer	19	Salma Hayek	13
Jennifer Capriati	42	Julianne Moore	19	Wen Jiabao	13
Laura Bush	41	Nicole Kidman	19	Adrien Brody	12
Lleyton Hewitt	41	Tim Henman	19	Anna Kournikova	12
Alejandro Toledo	39	Fidel Castro	18	Gonzalo Sanchez de Lozada	12
Hans Blix	39	Lance Armstrong	18	Harrison Ford	12
Nestor Kirchner	37	Michael Schumacher	18	Howard Dean	12
Andre Agassi	36	Pervez Musharraf	18	Jeb Bush	12
Alvaro Uribe	35	Richard Myers	18	Jennifer Garner	12
Megawati Sukarnoputri	33	Bill Gates	17	Keanu Reeves	12
Silvio Berlusconi	33	Jean Charest	17	Michael Jackson	12
Tom Ridge	33	John Bolton	17	Rubens Barrichello	12
Kofi Annan	32	John Kerry	17	Ann Veneman	11
Roh Moo-hyun	32	John Snow	17	Catherine Zeta-Jones	11
Vicente Fox	32	Renee Zellweger	17	Condoleezza Rice	11
David Beckham	31	Spencer Abraham	17	James Kelly	11
John Negroponte	31	Venus Williams	17	Jiri Novak	11
Guillermo Coria	30	Halle Berry	16	John Allen Muhammad	11
Recep Tayyip Erdogan	30	Tommy Franks	16	John Paul II	11
Bill Clinton	29	Trent Lott	16	Kim Ryong-sung	11
Mahmoud Abbas	29	Andy Roddick	15	Mark Philippoussis	11
Jack Straw	28	Bill Simon	15	Mike Weir	11
Juan Carlos Ferrero	28	Dominique de Villepin	15	Nicanor Duarte Frutos	11
Ricardo Lagos	27	Hu Jintao	15	Paul Burrell	11
Gray Davis	26	Julie Gerberding	15	Richard Gephardt	11
Rudolph Giuliani	26	Meryl Streep	15	Sergey Lavrov	11
Tom Daschle	25	Mohammed Al-Douri	15	Sergio Vieira De Mello	11
Atal Bihari Vajpayee	24	Nancy Pelosi	15	Tang Jiaxuan	11
Jeremy Greenstock	24	Norah Jones	15	Bill McBride	10
Winona Ryder	24	Pierce Brosnan	15	Ian Thorpe	10
Jose Maria Aznar	23	Taha Yassin Ramadan	15	Jacques Rogge	10
Saddam Hussein	23	Britney Spears	14	Jason Kidd	10
Tiger Woods	23	David Nalbandian	14	Javier Solana	10
George Robertson	22	Dick Cheney	14	Jean-David Levitte	10
Hamid Karzai	22	Eduardo Duhalde	14	Mohammad Khatami	10

Table 2. The names and the numbers of images of the subjects in LFW used to train the proposed regression models.

3.2 Landmarks Considered for Evaluation

Our proposed method detects 68 landmarks, while the baseline methods, CDM [10] and PIFA [11], detect 66 and 21 landmarks, respectively. For a fair comparison among these methods, we evaluate their Normalized Mean Error (NME) on landmarks that are the intersection of the three landmark sets. Figure 3 shows the detail.

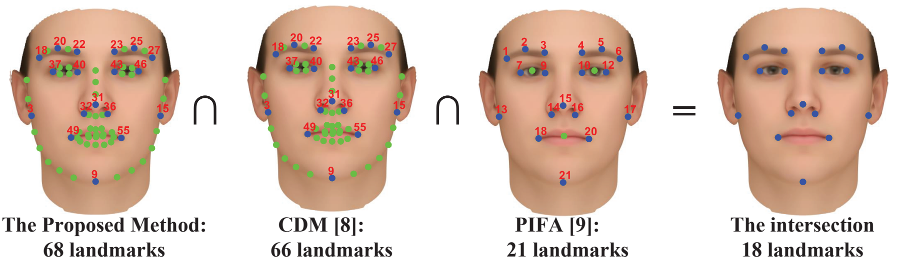

Fig. 3. The evaluated 18 landmarks (shown in blue color) that the intersection of the three landmark sets considered by the proposed method, the CDM and PIFA methods.

4 Discussion of Convergence Issue

The proposed method has two alternate optimization processes, one in 2D space for face alignment and the other in 3D space for 3D shape reconstruction. We experimentally investigate the convergence of these two processes. We conduct ten-fold cross-validation experiments on BU3DFE database, and record down the objective function values through 10 iterations. The average results are shown in Figure 4, Figure 4(a) for face alignment objective function and Figure 4(b) for 3D shape reconstruction objective function. It can be seen that both optimization processes converge in about five iterations.

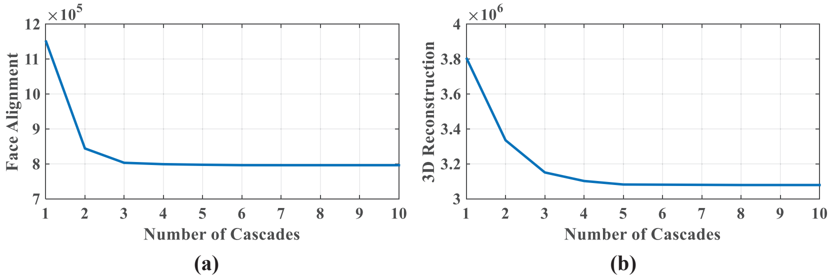

Fig. 4. Face alignment and 3D reconstruction objection function values as a function of cascades.

References

1. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3D face model for pose and illumination invariant face recognition. In: AVSS, IEEE (2009) 296–301
2. Liu, F., Zeng, D., Li, J., Zhao, Q.: Cascaded regressor based 3D face reconstruction from a single arbitrary view image. arXiv preprint arXiv:1509.06161 (2015)
3. Zhu, X., Yi, D., Lei, Z., Li, S.Z., et al.: Robust 3D morphable model fitting by sparse sift flow. In: ICPR. (2014) 4044–4049
4. Romdhani, S., Vetter, T.: Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In: CVPR, IEEE (2005) 986–993
5. Yin, L., Chen, X., Sun, Y., Worm, T., Reale, M.: A high-resolution 3D dynamic facial expression database. In: FG, IEEE (2008) 1–6
6. Tulyakov, S., Sebe, N.: Regressing a 3D face shape from a single image. In: ICCV. (2015) 3748–3755
7. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst (2007)
8. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: CVPR, IEEE (2012) 2879–2886
9. Gong, X., Wang, G.: An automatic approach for pixel-wise correspondence between 3D faces. In: Hybrid Information Technology. Volume 2. (2006) 198–205
10. Yu, X., Huang, J., Zhang, S., Yan, W., Metaxas, D.N.: Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. In: ICCV, IEEE (2013) 1944–1951
11. Jourabloo, A., Liu, X.: Pose-invariant 3D face alignment. In: ICCV, IEEE (2015)