Supplementary Material: 3D Face Modeling From Diverse Raw Scan Data

Feng Liu, Luan Tran, Xiaoming Liu
Department of Computer Science and Engineering
Michigan State University, East Lansing MI 48824
{liufeng6, tranluan, liuxm}@msu.edu

1. Overview

In this supplementary material, we provide additional experimental results, including

◇ Visualization and analysis of latent space;
◇ Computational efficiency comparison;
◇ Additional dense correspondence comparison and results;
◇ Details of evaluation experiments.

2. Visualization and Analysis of Latent Space

Our 3D face modeling learns two latent spaces, one for identity and the other for expression. To explore the two latent spaces, we conduct two experiments: 1) top 5 elements of identity and expression representations, and 2) identity and expression representations interpolation. A supplementary video is prepared to visualize these two experiments.

Top 5 Elements of Latent Representations. We start with an experiment that explores the learned identity and expression latent spaces by perturbing the top 5 components. We first encode all training scans in the latent space to obtain identity and expression latent vectors via the trained encoder. The mean and standard deviation of each element of the two latent vectors are computed respectively. We then perturb each element of the latent vector with the amount of perturbation equal to the corresponding standard deviation, and use the decoder to transform the perturbed latent vector to a reconstructed shape. By ranking the difference between the reconstruction and the original scans in a descending order, we select the top 5 elements for the identity and expression latent vectors respectively. In other words, varying these elements can have the most significant deformation on the decoded 3D face.

We report visual results of perturbing top 5 elements of identity and expression vectors in the Fig. 1 and 2. As can be seen from this illustration, the first element of the identity latent vector captures the global shape. The next four elements capture deformations of different face regions, e.g., the 2nd element for the nose region and the 5th for the forehead region. On the other hand, the top 5 elements of expression latent vector capture the corresponding face region deformations with expressions. This indicates that our proposed model indeed separates

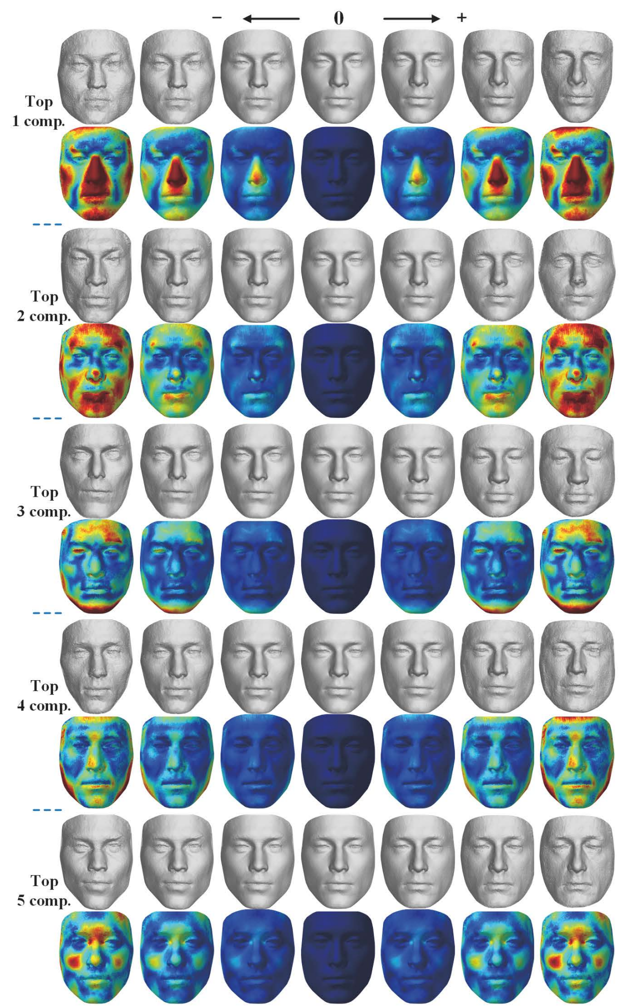

Figure 1: 3D shapes of varying top 5 elements of identity latent vector and the corresponding error maps.

Figure 2: 3D error maps of varying top 5 elements of expression latent vector.

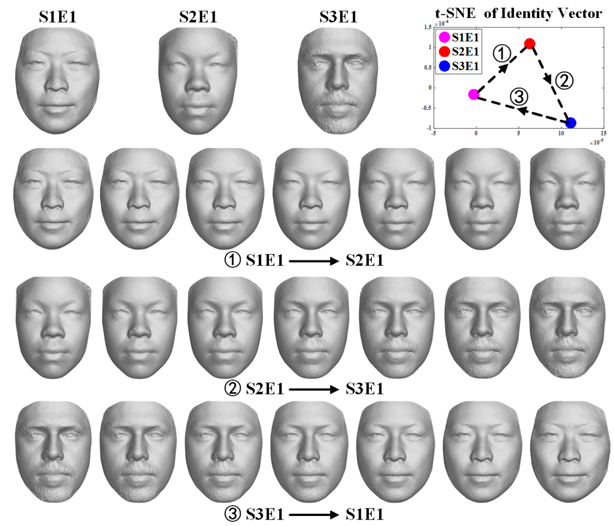

Figure 3: Exploring interpolation results on identity latent space. S1E1, S2E1 and S3E1 denote 3 subjects with neutral expression. We interpolate the latent vector in stride of $1/6$.

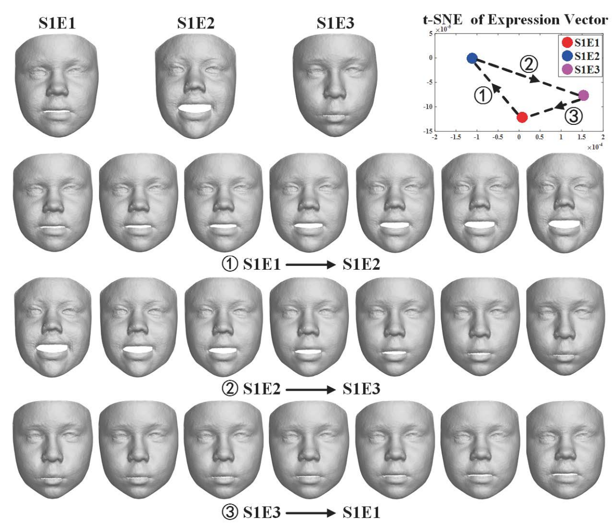

Figure 4: Exploring interpolation results on identity latent space. S1E1, S1E2 and S1E3 denote one subject with 3 different expressions. We interpolate the latent vector in stride of $1/6$.

the geometrical deformation into different semantically meaningful elements of the latent representations.

Latent Representations Interpolation. An alternative way to explore the latent spaces, which is usually employed in deep generative model, is to evaluate the interpolation capability of the model. For the identity interpolation experiment, we select 3 subjects with neutral expression which are denoted as S1E1, S2E1 and S3E1. In the expression setting, 3 different expression scans of one subject are used and denoted as S1E1, S1E2 and S1E3. We apply t-SNE [8] to the latent vectors to visualize the face manifold in a 2D space. As observed in the Fig. 3 and 4, our proposed model generalizes well on different identity and expression deformations. It further demonstrates that our model allows us to easily synthesize 3D faces with new identity and expression, by sampling the latent spaces.

3. Computational Efficiency Comparison

In this experiment, we compare the computational efficiency with two 3D face dense correspondence methods, NICP [1] and [6]. The runtime of [6] is reported in their paper. We run NICP and our method on a PC (with an Inter Core i7-7700K @ 4.20GHz, 16GB RAM and a NVIDIA GeForce GTX 1080Ti) for 700 samples of BU3DFE database [13], and calculate the average runtime. Table 1 gives per scan runtime of various methods. Here, the NICP code [1] is re-implemented with additional landmark constraint for faster convergence. Our proposed method requires only 0.00219 seconds on GPU and 0.26 seconds on CPU, which is at least two order of magnitude faster than the existing methods. This is owing to the data-driven methodology we employ in our implementation, which avoids the slow optimization procedure common in almost all prior dense correspondence methods.

4. Additional Dense Correspondence Comparison and Results

Figure 5 compares Bolkart *et al.* [5], NICP [1] and the proposed method in terms of 3D face dense correspondence results on one BU3DFE subject under sever

[1]https://github.com/charlienash/nricp

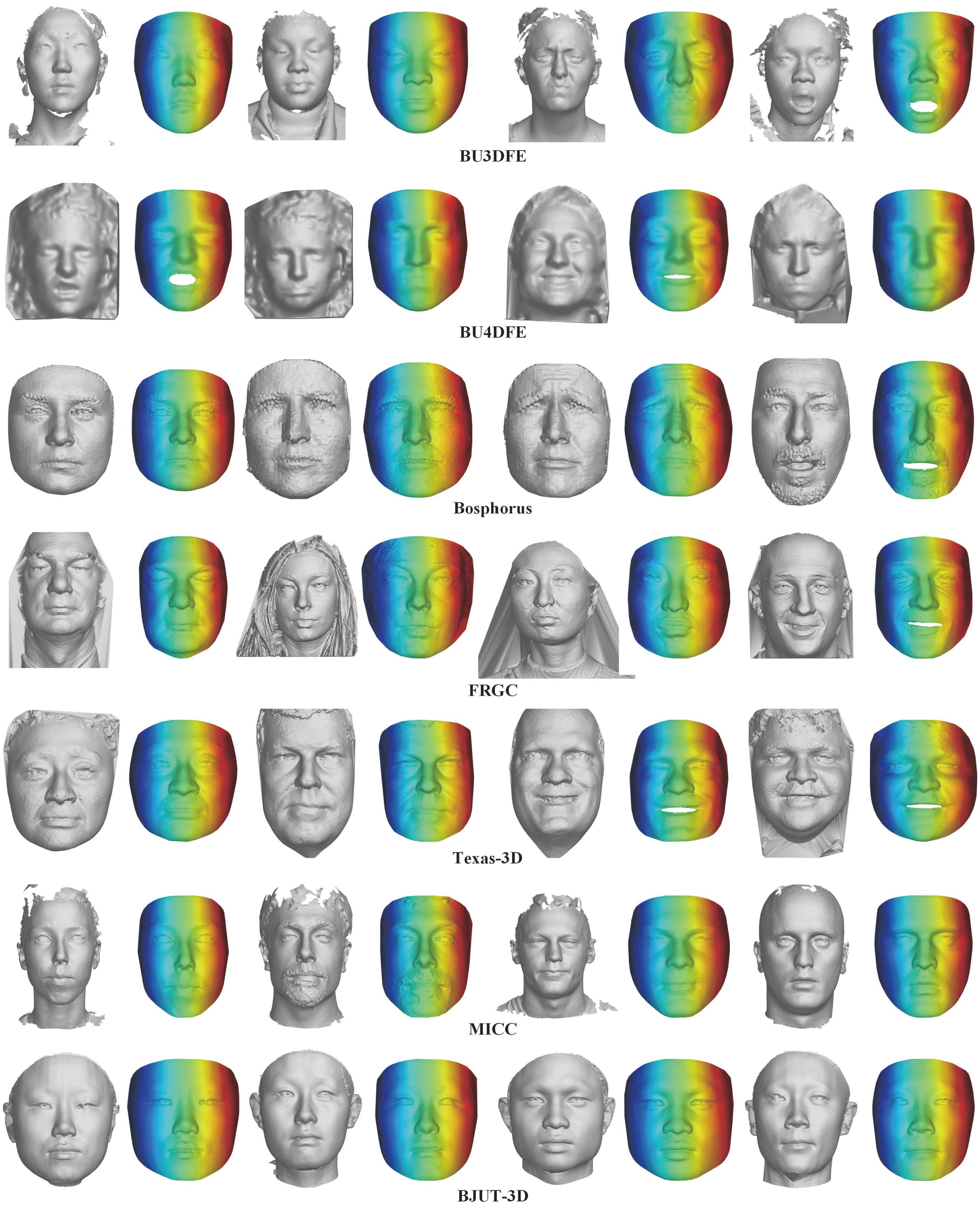

Figure 6: Dense correspondence results by our proposed method on scans from BU3DFE [13], BU4DFE [12], Bosphorus [11], FRGC [9], Texas-3D [7], MICC [3] and BJUT-3D [4]. Note the diversity of scan resolutions and scanners' noises among these databases, and how our reconstructions faithfully preserve those of the inputs. Also, the area inside the mouths can be precisely ignored regardless the amount of expression intensity.

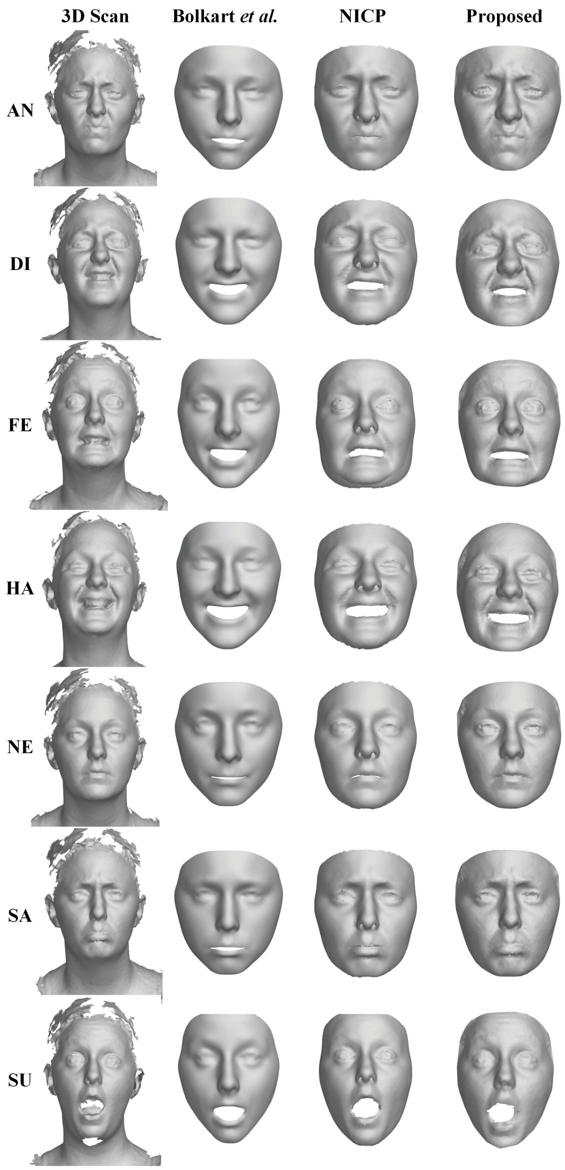

Figure 5: Dense correspondence results for one BU3FE subject under sever different expressions. The first column shows the input scans. Column 2-4 are the results by Bolkart *et al.* [5], NICP [1] and the proposed method.

Table 1: Efficiency comparison of different dense correspondence methods.

Method	Time (s)
NICP [1]	57.48
Fan *et al.* [6]	164.60
Proposed (CPU)	0.26
Proposed (GPU)	2.19×10^{-3}

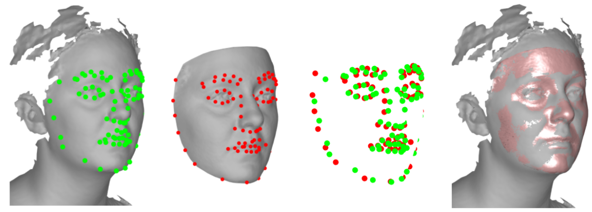

Figure 7: Visualization of the semantic landmark error and per-vertex fitting error on one BU3DFE example.

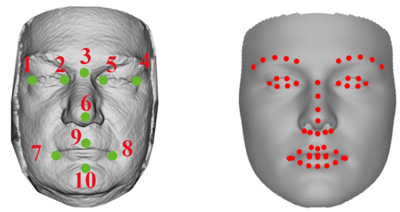

Figure 8: Landmarks used for semantic landmark evaluation of (a) FRGC (10 points in Tab. 4 of the main paper) and (b) BFM (51 points in Tab. 6 of the main paper) databases.

different expressions. From these results, we can clearly see that the proposed method performs well in reconstructing expressive scans and capturing fine details. Additional 3D face dense correspondence results of our proposed method on some scans from the related 7 databases are shown in Fig. 6. One can obviously observe from these results that the reconstructed 3D faces do preserve the high frequency details of the original scans (e.g., wrinkles and expressions), despite the diverse resolutions, and distinct scanner-specific noise among the different 3D face databases.

5. Details of Evaluation Experiments

Dense correspondence accuracy. Figure 7 visualizes the semantic landmark error and per-vertex fitting error of one BU3DFE example. It can be observed that the landmark error is much large than per-vertex fitting error, due to the inconsistent and imprecise annotations.

Shape representation on COMA. COMA database [10] contains 20,466 3D face models. The dataset is captured at 60fps with a multi-camera active stereo system, which consists of 12 classes of expressions from 12 different subjects. These expressions are complex and asymmetric. The shapes in COMA are not specified with expression labels. We manually select 4,000 neutral models from all the 12 subjects for the identity decoder training. We follow the train process as described in Sec. 3.4 of the main paper.

References

[1] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration. In *CVPR*, 2007. 2, 4

[2] Timur Bagautdinov, Chenglei Wu, Jason Saragih, Pascal Fua, and Yaser Sheikh. Modeling facial geometry using compositional VAEs. In *CVPR*, 2018.

[3] Andrew D Bagdanov, Alberto Del Bimbo, and Iacopo Masi. The florence 2D/3D hybrid face dataset. In *J-HGBU workshop*. ACM, 2011. 3

[4] Yin Baocai, Sun Yanfeng, Wang Chengzhang, and Ge Yun. BJUT-3D large scale 3D face database and information processing. *Journal of Computer Research and Development*, 6:020, 2009. 3

[5] Timo Bolkart and Stefanie Wuhrer. A groupwise multilinear correspondence optimization for 3D faces. In *ICCV*, 2015. 2, 4

[6] Zhenfeng Fan, Xiyuan Hu, Chen Chen, and Silong Peng. Dense semantic and topological correspondence of 3D faces without landmarks. In *ECCV*, 2018. 2, 4

[7] Shalini Gupta, Kenneth R Castleman, Mia K Markey, and Alan C Bovik. Texas 3D face recognition database. In *SSIAI*, 2010. 3

[8] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008. 2

[9] P Jonathon Phillips, Patrick J Flynn, Todd Scruggs, Kevin W Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the face recognition grand challenge. In *CVPR*, 2005. 3

[10] Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J Black. Generating 3D faces using convolutional mesh autoencoders. In *ECCV*, 2018. 4

[11] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. Bosphorus database for 3D face analysis. In *Biometrics and Identity Management*, pages 47–56. 2008. 3

[12] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. A high-resolution 3D dynamic facial expression database. In *FG*, 2008. 3

[13] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J Rosato. A 3D facial expression database for facial behavior research. In *FG*, 2006. 2, 3