1. Albedo and enviroment lighting estimation

With known normal $\boldsymbol{n}(v)$ of proxy mesh $\mathcal{P}_{proxy}$ at point v, similar to Eq. 14, we can compute the radiance L emitting from point v as:

$$L(v) = \rho(v)S(\boldsymbol{n}(v)) = \rho(v)\sum_{i=1}^{n} l_i Y_i(\boldsymbol{n}(v)), \tag{1}$$

where $\rho(v)$ denotes the surface albedo, Y_i the ith basis of spherical harmonics, l_i the corresponding weight. By representing albedo with *BFM* parameters, we have:

$$L(v) = (\boldsymbol{a}_{alb}^v + \mathbf{E}_{alb}^v \cdot \boldsymbol{\gamma})\sum_{i=1}^{n} l_i Y_i(\boldsymbol{n}(v)), \tag{2}$$

with $\boldsymbol{a}_{alb}^v$ and $\mathbf{E}_{alb}^v$ being the mean and principle component albedo at vertex v. We use the first nine harmonic basis and rewrite in matrix form:

$$L(v) = (\boldsymbol{a}_{alb}^v + \mathbf{E}_{alb}^v \cdot \boldsymbol{\gamma})\mathbf{H}_v \cdot \boldsymbol{l} \tag{3}$$

where $\mathbf{H}_v = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} \otimes \begin{bmatrix}Y_1(\boldsymbol{n}(v)) & \cdots & Y_9(\boldsymbol{n}(v))\end{bmatrix}$ and $\boldsymbol{l} = \begin{bmatrix}l_1^1, & \cdots & l_9^1, & l_1^2, & \cdots & l_9^2, & l_1^3, & \cdots & l_9^3\end{bmatrix}^T$. Accordingly, a reconstructed face image $\mathcal{I}_{recon}$ can be represented by

$$\mathcal{I}_{recon} = (\boldsymbol{a}_{alb} + \mathbf{E}_{alb} \cdot \boldsymbol{\gamma}) \odot (\mathbf{H} \cdot \boldsymbol{l}) \tag{4}$$

where $\mathbf{H} = \begin{bmatrix}\mathbf{H}_{v_1}^T, & \cdots, & \mathbf{H}_{v_n}^T\end{bmatrix}^T$, and $\mathbf{H} \in \mathbb{R}^{3n\times 27}$.

We estimate lighting and albedo by minimizing the following energy function on the illumination coefficients $\boldsymbol{l}$ and the albedo parameters $\boldsymbol{\gamma}$.

$$E(\boldsymbol{l}, \boldsymbol{\gamma}) = \|\mathcal{I}_{input} - \mathcal{I}_{recon}\|_2^2 \tag{5}$$

where $\mathcal{I}_{input}$ is the intensity value at pixels where vertices re-project to input image. In order to achieve a reliable estimation, in our implementation, we first use a self-adaptive mask to select vertices that have reliable normals with which to apply the optimization. We adopt an iterative optimization scheme similar to [5]. The complete algorithm is shown in Algorithm 1 where M, ξ_1, ξ_2 are termination threshold. They are set as 50, 0.05 and 50 in our experiments.

Algorithm 1 lighting and albedo estimation

Require: $\mathcal{I}_{input}, \mathbf{H}, \boldsymbol{a}_{alb}, \mathbf{E}_{alb}, M, \xi_1, \xi_2, i = 0$
Ensure: $\boldsymbol{l}, \boldsymbol{\gamma} = \arg\min_{\boldsymbol{l},\boldsymbol{\gamma}} E(\boldsymbol{l}, \boldsymbol{\gamma})$

1: $i \leftarrow 0$
2: $\boldsymbol{\gamma} \leftarrow \mathbf{0}$
3: **while** $i \leq M$ **do**
4: $\quad \boldsymbol{l} \leftarrow \arg\min_{\boldsymbol{l}} \|\mathcal{I}_{input} - (\boldsymbol{a}_{alb} + \mathbf{E}_{alb} \cdot \boldsymbol{\gamma}) \odot (\mathbf{H} \cdot \boldsymbol{l})\|_2^2$
5: $\quad \delta\mathcal{I} \leftarrow \mathcal{I}_{input} - (\boldsymbol{a}_{alb} + \mathbf{E}_{alb} \cdot \boldsymbol{\gamma}) \odot (\mathbf{H} \cdot \boldsymbol{l})$
6: $\quad \delta\boldsymbol{\gamma} \leftarrow \arg\min_{\delta\boldsymbol{\gamma}} \|\delta\mathcal{I} - (\mathbf{E}_{alb} \cdot \delta\boldsymbol{\gamma}) \odot (\mathbf{H} \cdot \boldsymbol{l})\|_2^2$
7: $\quad \boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma} + \delta\boldsymbol{\gamma}$
8: $\quad i \leftarrow i + 1$
9: $\quad$ **if** $\|\delta\boldsymbol{\gamma}\|_2^2 < \xi_1$ **or** $\|\delta\mathcal{I}\|_2^2 < \xi_2$ **then return** $\boldsymbol{l}, \boldsymbol{\gamma}$
10: **return** $\boldsymbol{l}, \boldsymbol{\gamma}$

2. Additional Support Figures

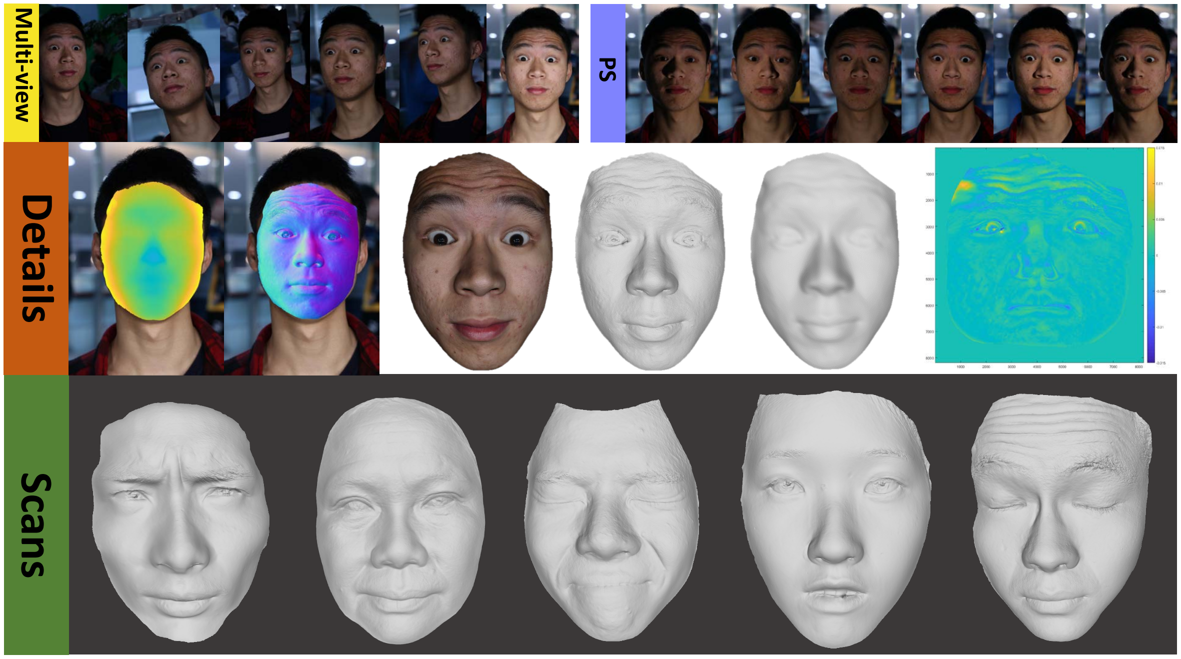

Figure 1. A preview of our dataset. Top row: Multi-view and photometric stereo images. Middle row: facial scan reconstruction and detail extraction. Third row: a subset of our facial scans.

3. Additional Results

The following test images are selected from related papers and *AffectNet* dataset [2], which we select based on less occlusion and high image resolution (width/height > 1500px). Our detail-synthesized models exhibit realistic details that outperform state-of-the-art methods.

References

[1] Yue Li, Liqian Ma, Haoqiang Fan, and Kenny Mitchell. Feature-preserving detailed 3d face reconstruction from a single image. In *Proc. of the 15th ACM SIGGRAPH European Conference on Visual Media Production*. ACM, 2018.

[2] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *arXiv preprint arXiv:1708.03985*, 2017.

[3] Matan Sela, Elad Richardson, and Ron Kimmel. Unrestricted facial geometry reconstruction using image-to-image translation. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 1585–1594. IEEE, 2017.

[4] Anh Tuân Tran, Tal Hassner, Iacopo Masi, Eran Paz, Yuval Nirkin, and Gérard Medioni. Extreme 3d face reconstruction: Seeing through occlusions. In *Proc. CVPR*, 2018.

[5] Yang Wang, Lei Zhang, Zicheng Liu, Gang Hua, Zhen Wen, Zhengyou Zhang, and Dimitris Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1968–1984, 2009.

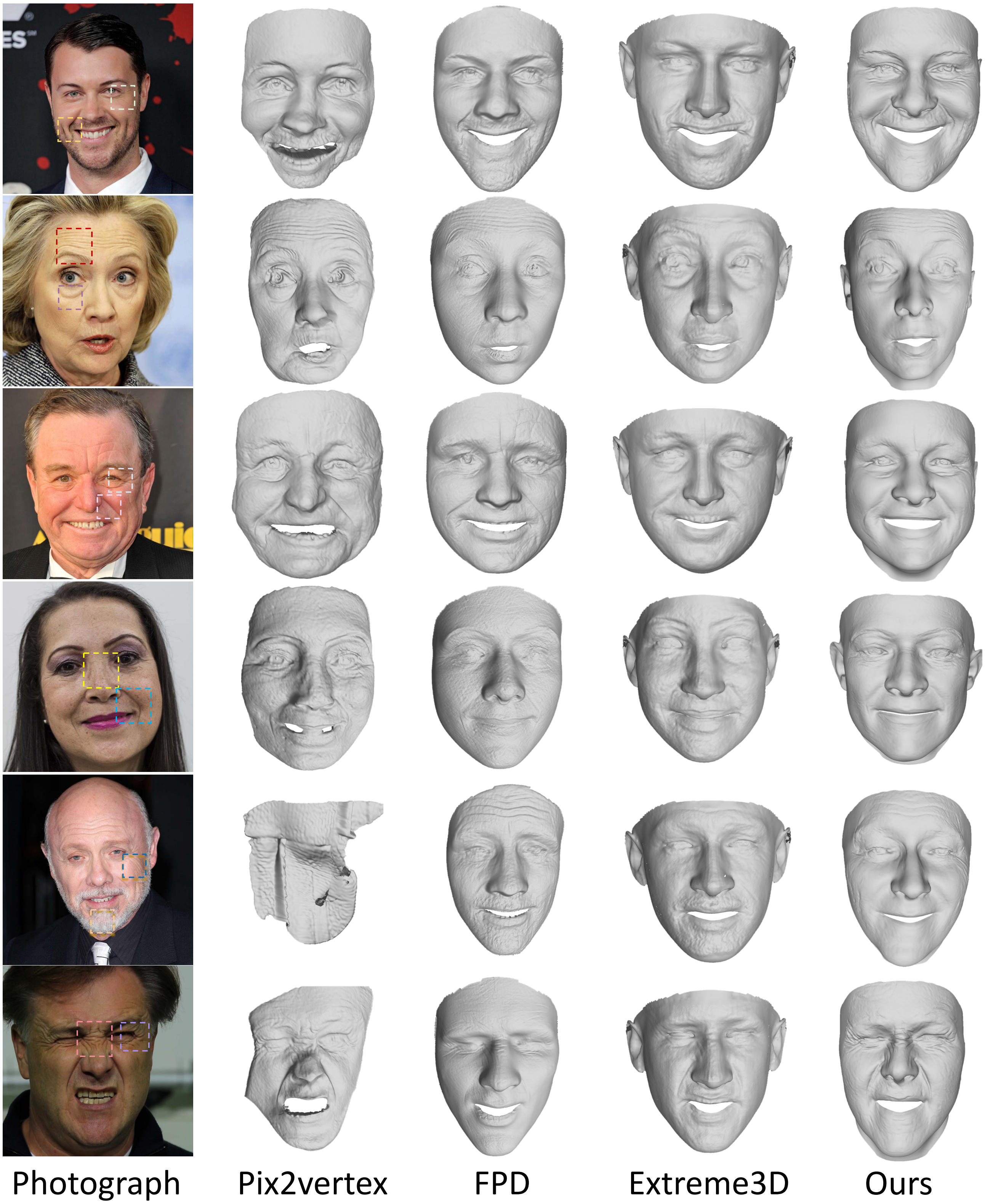

Figure 2. Comparisons of Pix2vertex [3], FPD [1], Extreme3D [4] and ours.

Figure 3. Sample results of our method.

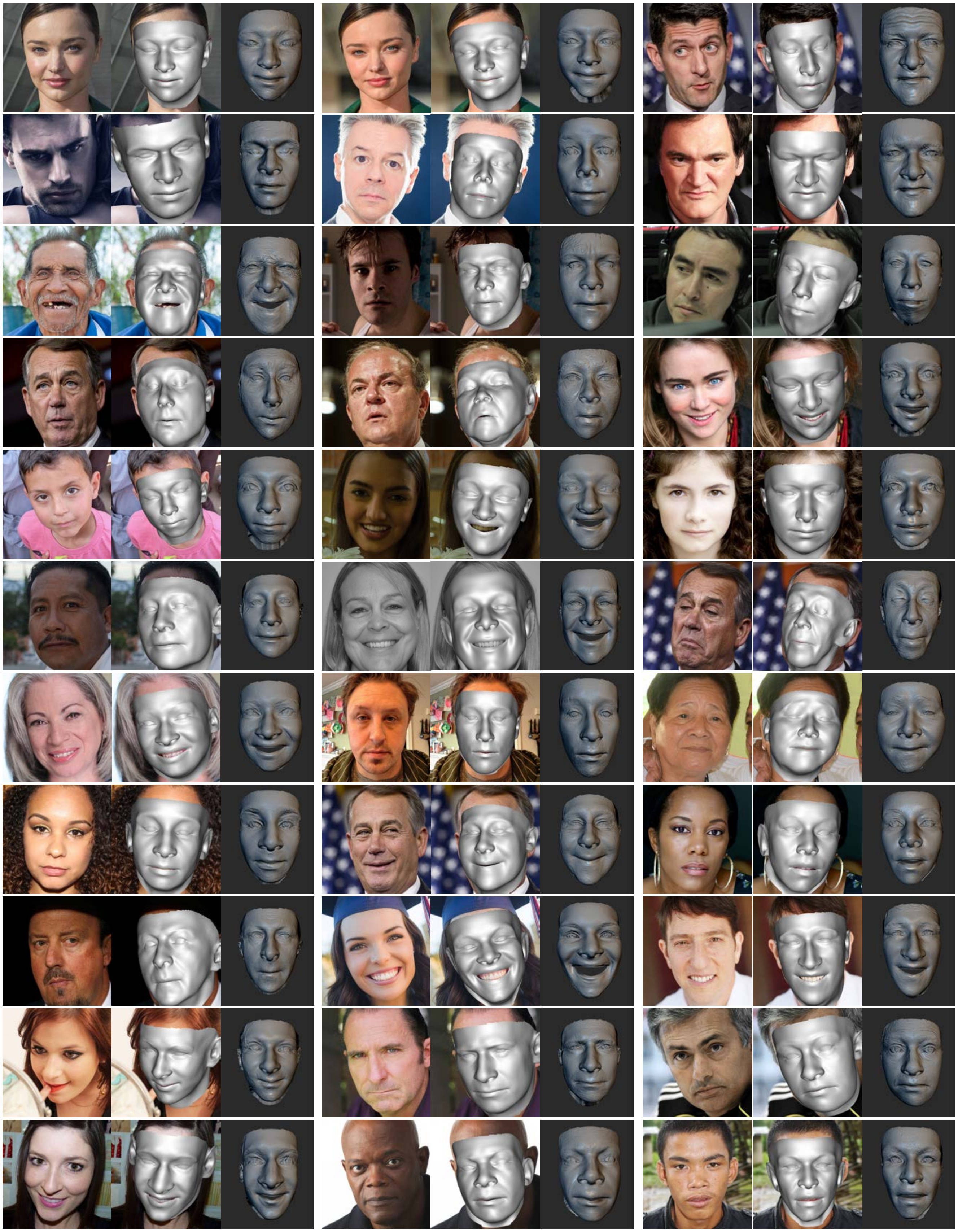

Figure 4. Sample results of our method.

Figure 5. Sample results of our method.