OSTeC: One-Shot Texture Completion: Supplementary Material

Baris Gecer, Jiankang Deng, and Stefanos Zafeiriou
Imperial College London, Huawei CBG
{b.gecer, j.deng16, s.zafeiriou}@imperial.ac.uk
{baris.gecer, jiankangdeng, stefanos.zafeiriou1}@huawei.com

1. Algorithm

We summarize our method in Algorithm 1 where the notations are described in the original paper.

Algorithm 1: One-Shot Texture Completion

Input : RGB Face Image: $\mathbf{I}_0$
Input : 3DMM Fitting: $(\mathbf{S}, \mathbf{c})$
Input : Novel Camera Views: $\{\mathbf{c}_i\}, i \in \{0 \dots n\}$
Output: Completed UV Texture Map: $\overline{\mathbf{T}}_n$

$\mathbf{T}_0 \leftarrow \mathcal{R}'(t_{coord}, \mathbf{I}_0, \mathbf{S}')$
for $i \leftarrow 0$ **to** n **do**
    $\mathbf{S}'_i \leftarrow \mathcal{P}(\mathbf{S}, \mathbf{c}_i)$
    $\mathbf{V}_i \leftarrow (\frac{[\mathbf{S}'_i, \mathbf{h}]}{||[\mathbf{S}'_i, \mathbf{h}]||_2} \cdot \mathcal{N}(\mathbf{S}_i)^T)$
end
for $i \leftarrow 0$ **to** n **do**
    $\overline{\mathbf{V}}_i \leftarrow \bigcap_{i \neq j} (\mathbf{V}_i > \mathbf{V}_j)$
    $\mathbf{M}_i^{UV} \leftarrow \big((\mathbf{V}_0 > t_1) \cap (2\mathbf{V}_0 > \mathbf{V}_i)\big) \cup \bigcup_{i > j} \overline{\mathbf{V}}_j$
    $\mathbf{M}_i \leftarrow \mathcal{R}(\mathbf{S}'_i, \mathbf{M}_i^{UV}, t_{coord})$
    $\mathbf{I}_i \leftarrow \mathcal{R}(\mathbf{S}'_i, \overline{\mathbf{T}}_{i-1}, t_{coord})$
    $\mathbf{W}_i \leftarrow \mathcal{E}(\mathbf{I}_i)$
    $\mathbf{W}_i \leftarrow \arg\min_{\mathbf{W}_i} \mathcal{L}_{total}(\mathbf{I}_i, \mathbf{M}_i, \mathbf{W}_i)$
    $\mathbf{G}_i \leftarrow \mathcal{G}(\mathbf{W}_i)$
    $\mathbf{T}_i \leftarrow \mathcal{R}'(t_{coord}, \mathbf{G}_i, \mathbf{S}'_i)$
    $\overline{\mathbf{T}}_i \leftarrow \overline{\mathbf{V}}_i \odot \mathbf{T}_i + (1 - \overline{\mathbf{V}}_i) \odot \overline{\mathbf{T}}_{i-1}$
end

2. Pose-Invariant Face Matching: MultiPIE dataset

For the evaluation in **under-controlled** scenario, we compare our method with recent state-of-the-art studies, *e.g.* CPF [12], DR-GAN [9], FF-GAN [13], TP-GAN [5], CAPG-GAN [4], PIM [14], HF-PIM [1] and Rotate & Render [15], on the Multi-PIE dataset [2]. The performances are reported following the protocol of the setting 2 [12, 1] provided by the Multi-PIE dataset. Each testing identity has one gallery image from the first appearance. Hence, there are 72,000 and 137 images in the probe and gallery sets, respectively. In Tab. 1, results are reported across different poses. We employ the strategy of "recognition via generation" and faces at any pose are first frontalized by our model. After the face frontalization, the pre-trained ArcFace model trained on MS1M is employed as the feature extractor. Here, we refer to [15] to train ResNet-18, which is slightly smaller than LightCNN-29 [10] used by [1]. For those poses less than 60°, the performances of most methods are quite good whereas our method almost achieves zero failure rate. However, the profiles with extreme poses (> 60°) on are very challenging. For those extreme poses, our method obviously outperforms other methods, surpassing the "Rotate & Render" method [15] by 0.84% under the pose of 90°. This impressive recognition performance undoubtedly confirms the effectiveness of the proposed identity-preserved UV texture completion.

Method	±15°	±30°	±45°	±60°	±75°	±90°
CPF [12]	95.0	88.5	79.9	61.9	-	-
DR-GAN [9]	94.9	91.1	87.2	84.6	-	-
FF-GAN [13]	94.6	92.5	89.7	85.2	77.2	61.2
TP-GAN [5]	98.7	98.1	95.4	87.7	77.4	64.6
CAPG-GAN [4]	99.8	99.6	97.3	90.3	83.1	66.1
PIM [14]	99.3	99.0	98.5	98.1	95.0	86.5
HF-PIM [1]	99.99	99.98	99.88	99.14	96.40	92.32
R&R [15]	-	**100**	**100**	**99.7**	99.3	94.4
Baseline	99.98	99.86	99.80	98.50	96.19	92.06
Ours	**100**	**100**	99.88	99.62	**99.35**	**95.24**

Table 1: Rank-1 recognition rates (%) across views on the Multi-PIE dataset [2]. The baseline model is ResNet-18 trained on MS1M with the ArcFace loss. Our method further employs face finalization to improve the accuracy.

3. Performance on 'in-the-wild' Scenario

Following 'Pose-Invariant Face Matching' experiment in the original paper, we visualize some of the frontal-profile pairs from CFP dataset [7] to evaluate and verify quantitative experiments qualitatively. Figures 1,2,3a show many pairs of frontal and profile images of the same identity, completed texture UV maps by our method, its rendering, frontalization by our method and cosine similarity scores. The scores are obtained by a ResNet-18 networks [3] on CASIA-WebFace [11] for:

- **'Org.':** the pairs of original images
- **'UV.':** original frontal image and rendered geometry with a completed UV map by our method
- **'Frontalized':** original frontal image and frontalized image by our method

As can be seen in the figures both qualitatively and quantitatively, our approach can generate excellent quality frontal images and UV texture maps with preserved identity, even under low resolution, extreme pose, occlusion, lighting and expression variations. The cosine similarity scores are mostly improved by the generations of our method compared to the original profile images which verifies the qualitative results.

4. Manipulating Frontalized Faces

Frontalization by our approach is achieved by rendering the geometry that is textured by the completed UV map and reconstructing it in StyleGAN [6] latent space. Therefore the frontalized images can be manipulated by common StyleGAN manipulation techniques such as interpolation between different identities and changing/adding some facial attributes.

Fig. 4 illustrates some interpolations performed between the original and the frontalized projections that slowly shift from various poses to the frontal pose. Fig. 5 shows interpolation between different identities, both in the frontalized and the original projections. Please note that, the frontalized interpolation maintain smoother transition between the identities, whereas the original image projections generates artefacts at the intermediate generations due to exhausted latent parameters. Lastly, Fig. 6 illustrates attribute manipulation by extracting some attribute directions with [8] such as age, gender and expression.

5. Limitations and Failure Cases

The biggest strength and the biggest weakness of our approach is being an optimization-based method. Usually, the running time takes around 5-10 minutes depending on the convergence speed. This is mainly due to CPU intensive visibility mask and 3D mesh rendering over the iterations. We believe that the code might be optimized to run under 1 minute which is a reasonable running time for an optimization-based method.

Another limitation of an optimization-based method is the danger of local minima. We observed in some cases, optimization gets stuck at local minima, failing to find a good texture completion and frontalization. This is partially addressed by the encoder network $\mathcal{E}$, but emprically we can still observe this behaviour as can be seen in Fig. 3b.

Another drawback of our approach is that it heavily relies on 3D face reconstruction. Therefore, our method is limited by the accuracy and performance of the 3D reconstruction. That is to say, if some part of the identity cannot be captured by the reconstruction, our method might struggle to compensate. Some of such failure cases are illustrated in Fig. 3b.

Figure 1: Qualitative verification of the *Pose-Invariant Face Matching* experiment. Each block respectively consists of: (1) Original frontal image, (2) Original profile image, (3) Rendered geometry with our completed texture map, (4) Our completed texture map, (5) Our frontalized image.

Figure 2: Qualitative verification of the *Pose-Invariant Face Matching* experiment. Each block respectively consists of: (1) Original frontal image, (2) Original profile image, (3) Rendered geometry with our completed texture map, (4) Our completed texture map, (5) Our frontalized image.

(a) Successful cases (b) Failure cases

Figure 3: Qualitative verification of the *Pose-Invariant Face Matching* experiment. Each block respectively consists of: (1) Original frontal image, (2) Original profile image, (3) Rendered geometry with our completed texture map, (4) Our completed texture map, (5) Our frontalized image.

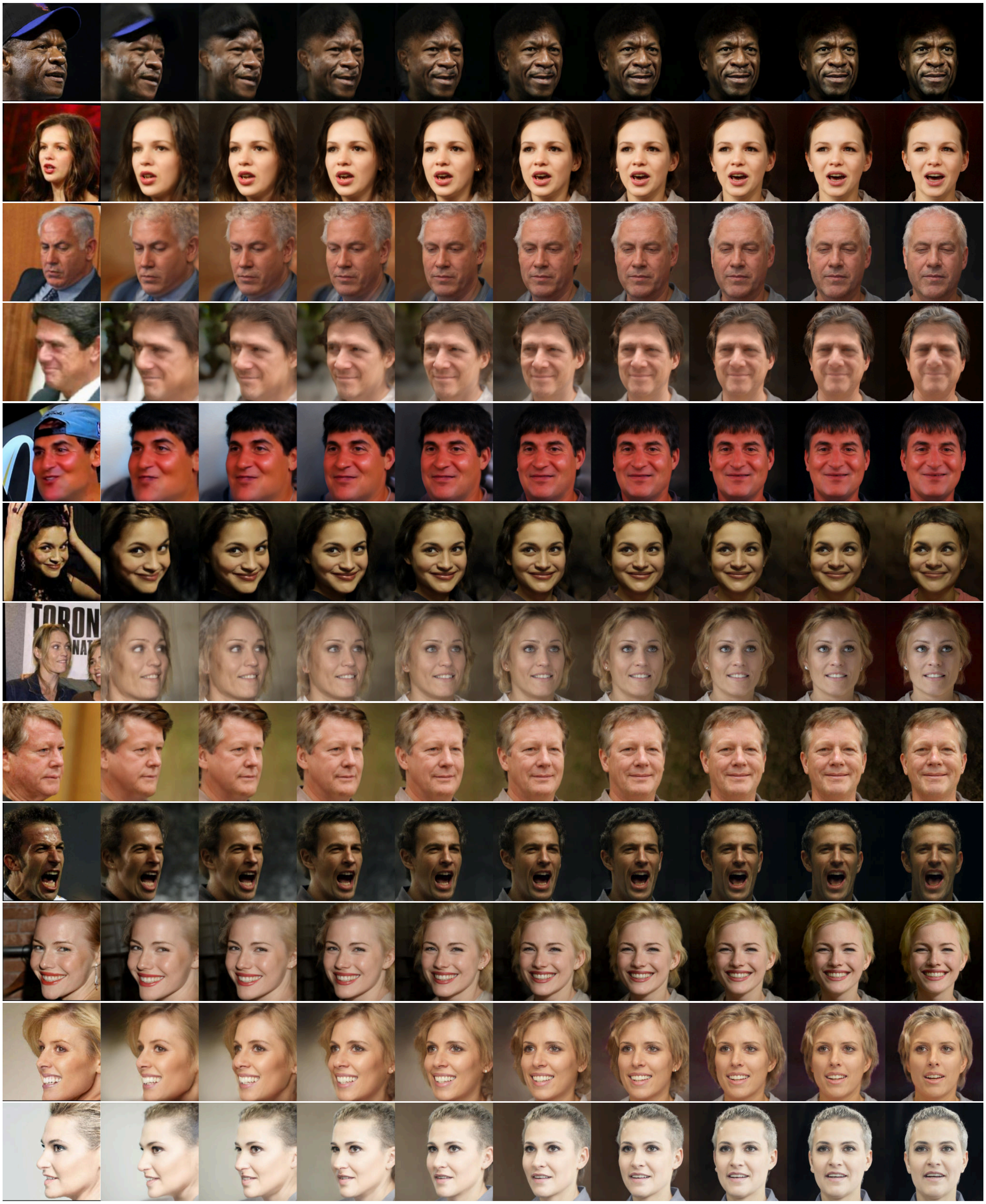

Figure 4: Interpolation between the original poses and the frontalized versions. First column is the original image. Second column is its projection to the StyleGAN space. Last column is the frontalized generated image by our approach. And other columns are the interpolation in-between.

Figure 5: Interpolations between different identities. First and Last colums are the original images and other columns are interpolations. Odd rows are interpolating frontal projections and even rows are interpolating the original image projections. Please note that, the frontalized interpolation maintain smoother transition between the identities, whereas the original image projections generates artefacts at the intermediate generations due to exhausted latent parameters.

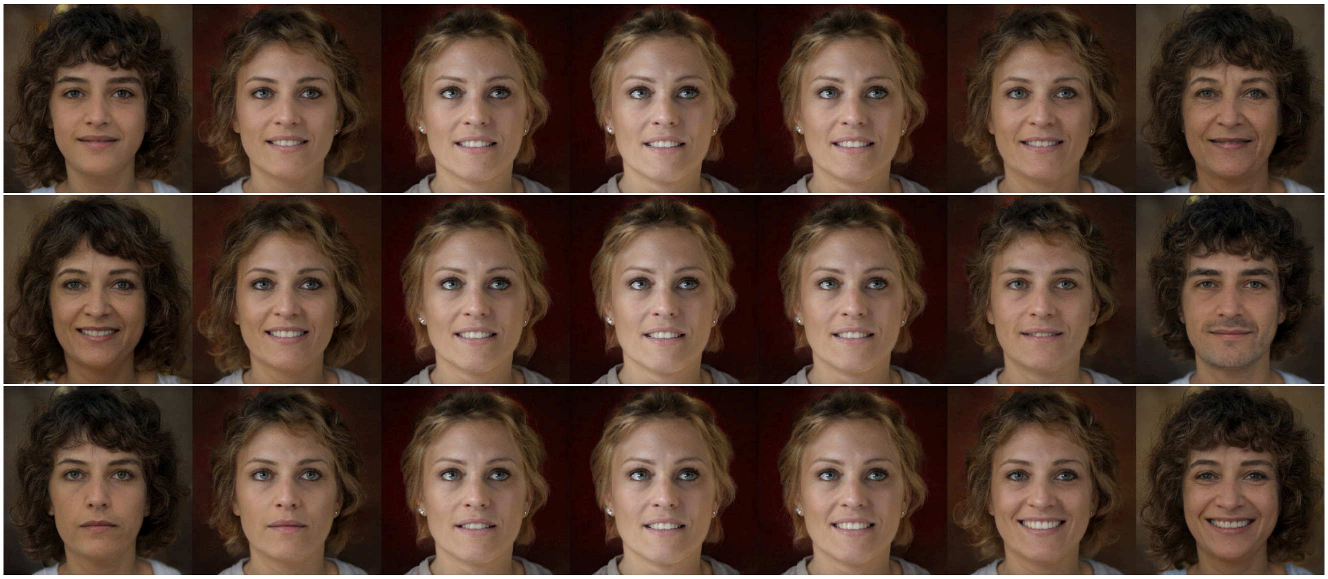

Figure 6: Attribute manipulation by [8] can be performed on the frontalized images.

References

[1] Jie Cao, Yibo Hu, Hongwen Zhang, Ran He, and Zhenan Sun. Learning a high fidelity pose invariant model for high-resolution face frontalization. In *NeurIPS*, 2018. 1

[2] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. *Image and Vision Computing*. 1

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-December, pages 770–778, 2016. 2

[4] Yibo Hu, Xiang Wu, Bing Yu, Ran He, and Zhenan Sun. Pose-guided photorealistic face rotation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 8398–8406, 2018. 1

[5] Rui Huang, Shu Zhang, Tianyu Li, and Ran He. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2439–2448, 2017. 1

[6] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June:4396–4405, 2019. 2

[7] Soumyadip Sengupta, Jun Cheng Chen, Carlos Castillo, Vishal M. Patel, Rama Chellappa, and David W. Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016. 2

[8] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the Latent Space of GANs for Semantic Face Editing. *arXiv*, 2019. 2, 8

[9] Luan Tran, Xi Yin, and Xiaoming Liu. Disentangled representation learning GAN for pose-invariant face recognition. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, volume 2017-January, pages 1283–1292, 2017. 1

[10] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 2018. 1

[11] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z. Li. Learning face representation from scratch. *arXiv:1411.7923*, 2014. 2

[12] Junho Yim, Heechul Jung, ByungIn Yoo, Changkyu Choi, Dusik Park, and Junmo Kim. Rotating your face using multi-task deep neural network. In *CVPR*, 2015. 1

[13] Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu, and Manmohan Chandraker. Towards large-pose face frontalization in the wild. In *ICCV*, 2017. 1

[14] Jian Zhao, Yu Cheng, Yan Xu, Lin Xiong, Jianshu Li, Fang Zhao, Karlekar Jayashree, Sugiri Pranata, Shengmei Shen, Junliang Xing, et al. Towards pose invariant face recognition in the wild. In *CVPR*, 2018. 1

[15] Hang Zhou, Jihao Liu, Ziwei Liu, Yu Liu, and Xiaogang Wang. Rotate-and-render: Unsupervised photorealistic face rotation from single-view images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1