HTML: A Parametric Hand Texture Model for 3D Hand Reconstruction and Personalization –Supplemental Document–

Neng Qian[1,2], Jiayi Wang[1], Franziska Mueller[1], Florian Bernard[1,3], Vladislav Golyanik[1], and Christian Theobalt[1]
{nqian,jwang,frmueller,fbernard,golyanik,theobalt}@mpi-inf.mpg.de

[1] Max Planck Institute for Informatics, Saarland Informatics Campus
[2] RWTH Aachen University
[3] Technical University of Munich

1 Experiment Details for Hand Model Layer

The provided training dataset contains a total of 130,240 images: 32,560 unique images of hands with foreground masks, times four methods of background composition. However, three of the composition methods attempt to blend the hand into the background, which introduces severe artifacts in the hand appearance (see Fig. 1).

We only use the unaltered 32,560 unique images for our training data to avoid learning these artifacts for texture estimation. The provided foreground masks were used to perform background augmentation without additional image harmonization or image coloration. We train the ResNet-34 [3] using Adam, with a learning rate of 0.001, and for 200 epochs in all of our experiments.

A full comparison of the pose and shape performance is provided in Table 1. Following the evaluation procedure of [5], the meshes were aligned using Procrustes alignment as a rigid body transformation. Errors are measured in Euclidean distance (cm) between corresponding vertex points **(Mesh)** or keypoints **(KP)**. Area under the percentage-of-correct-keypoints curve **(AUC)** and F-scores at two different thresholds **(F@5mm and F@15mm)** are additionally provided. Our method achieves slightly higher pose and shape performance with the photometric loss **(Proposed)** than without **(w/o Photometric)**, and it achieves similar accuracy to the current state-of-the-art methods.

Fig. 1. The provided composed FreiHand data contains noticeable texture artifacts. These images were not used for training.

	KP Error	KP AUC	Mesh Error	Mesh AUC	F@5mm	F@15mm
Zimmerman *et al.* [5]	1.10	0.783	1.09	0.783	0.516	0.934
Boukhayma *et al.* [1]	3.50	0.351	1.32	0.738	0.427	0.895
Hasson *et al.* [2]	1.33	0.737	1.33	0.736	0.429	0.907
w/o Photometric	1.14	0.774	1.14	0.774	0.499	0.925
Proposed	1.11	0.781	1.10	0.781	0.508	0.930

Table 1. Evaluation of our method on the FreiHand dataset [5]. All numbers are from the online leader board. Keypoint (KP) and Mesh errors are measured in cm.

2 Principal Components Without Shading Removal

It is desirable for the parametric texture model to not include lighting effects. Although we aimed to have uniform lighting while acquiring the scans, there are still smooth shading effects, especially at the boundary to the flat background surface. Without the shading removal step in our pipeline, the lighting effects contribute a large portion of the variation in the dataset. Hence, lighting variations are present in some of the first principal components of the PCA space. Refer to Fig. 2 and the supplementary video for visualizations.

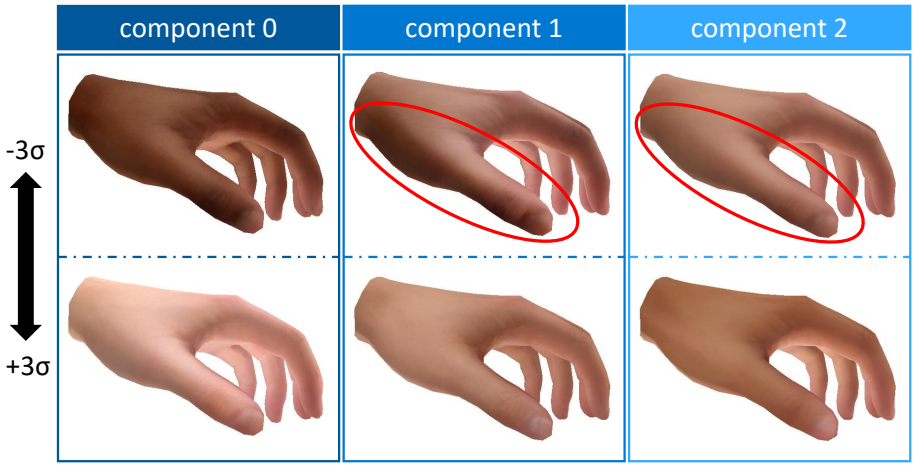

Fig. 2. When we build the texture PCA model without shading removal, the principal components contain a significant amount of lighting variation.

3 Shading Removal on Original MANO Scans

The original scans from which the MANO shape and pose model [4] was built do also include vertex colors. However, they contain strong lighting effects like

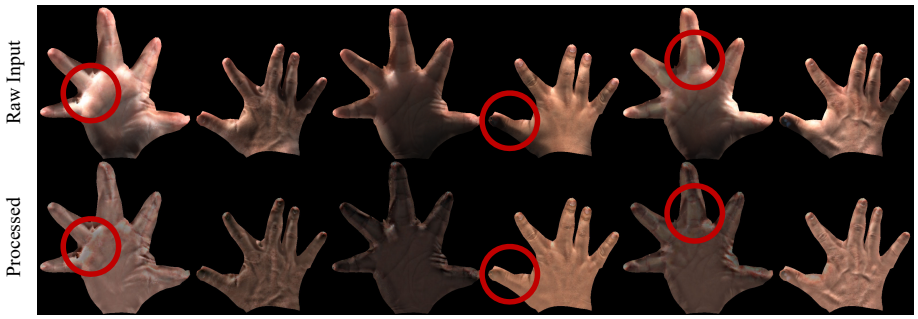

Fig. 3. Results of our shading removal technique when applied to textures extracted from the original MANO scans. The appearance information cannot be properly recovered due to strong lighting effects (left, middle). The strong shadows also contain high-frequency components that are difficult to remove (right).

shadows and over-saturated areas. Fig. 3 shows how our shading removal procedure fails when applied to textures extracted from the original MANO scans. Note that even if more sophisticated approaches are applied, it is inherently not possible to recover accurate appearance information from completely dark or over-saturated areas. The strong shadows also contain high-frequency components that are difficult to remove while preserving texture details.

References

1. Boukhayma, A., Bem, R.d., Torr, P.H.: 3d hand shape and pose from images in the wild. In: Computer Vision and Pattern Recognition (CVPR) (2019)
2. Hasson, Y., Varol, G., Tzionas, D., Kalevatykh, I., Black, M.J., Laptev, I., Schmid, C.: Learning joint reconstruction of hands and manipulated objects. In: Computer Vision and Pattern Recognition (CVPR) (2019)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
4. Romero, J., Tzionas, D., Black, M.J.: Embodied hands: Modeling and capturing hands and bodies together. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) **36**(6) (2017)
5. Zimmermann, C., Ceylan, D., Yang, J., Russell, B., Argus, M., Brox, T.: Freihand: A dataset for markerless capture of hand pose and shape from single rgb images. In: International Conference on Computer Vision (ICCV) (2019)