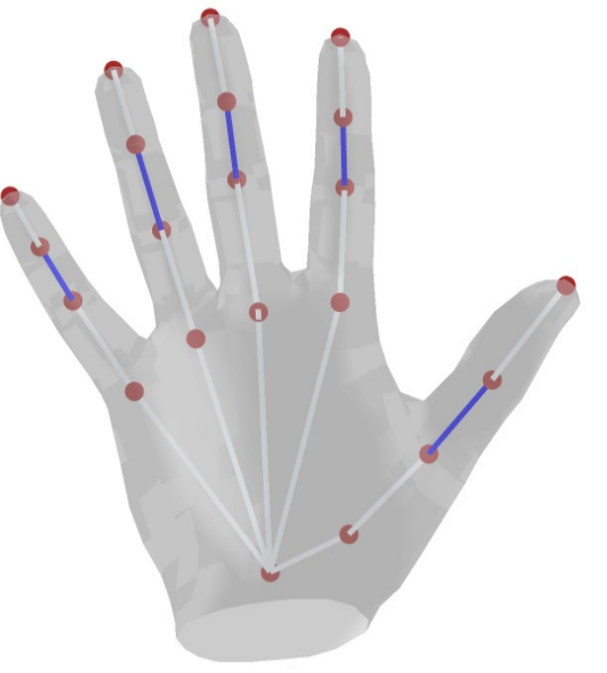


Reconstructing Hand Shape and Appearance for Accurate Tracking from Monocular Video

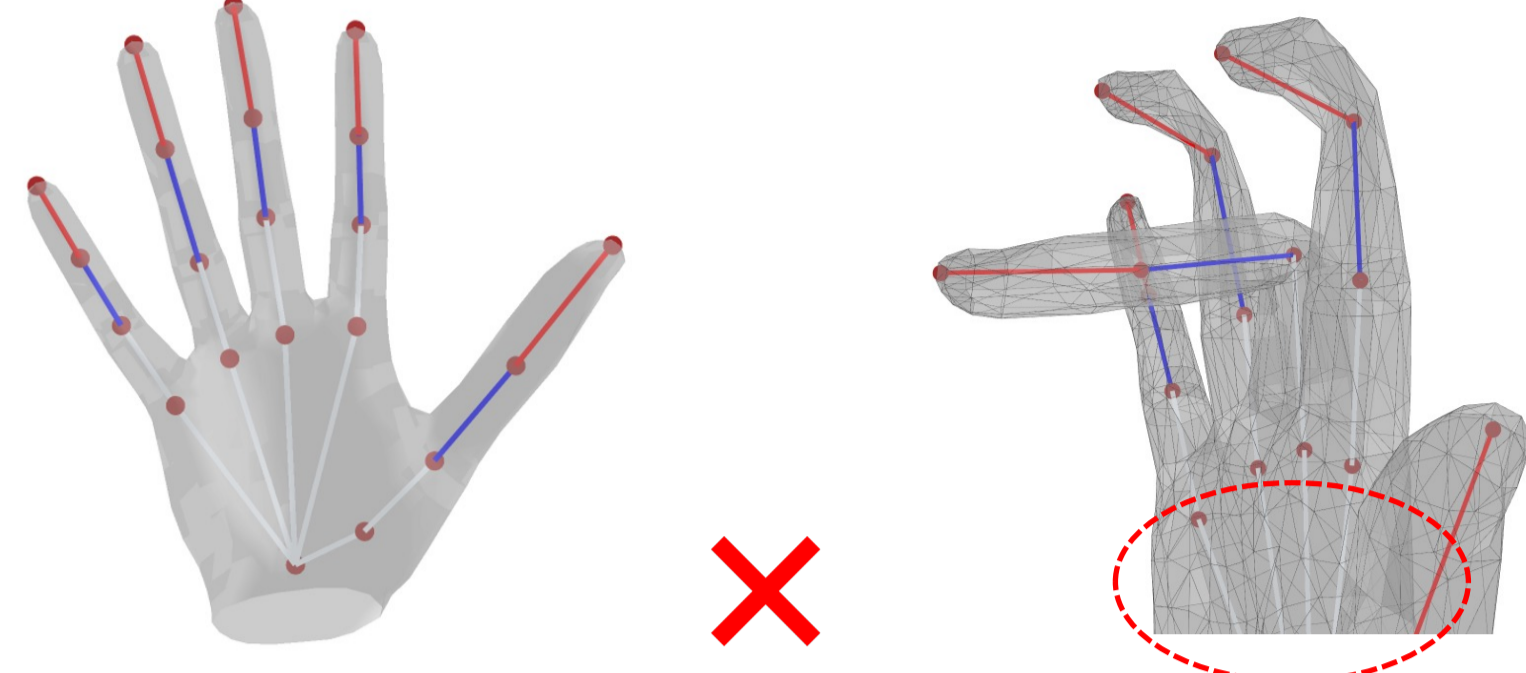
Pratik Kalshetti
Indian Institute of Technology Bombay

Hand Shape Model

Consider
Localized scaling of
middle phalange (blue)

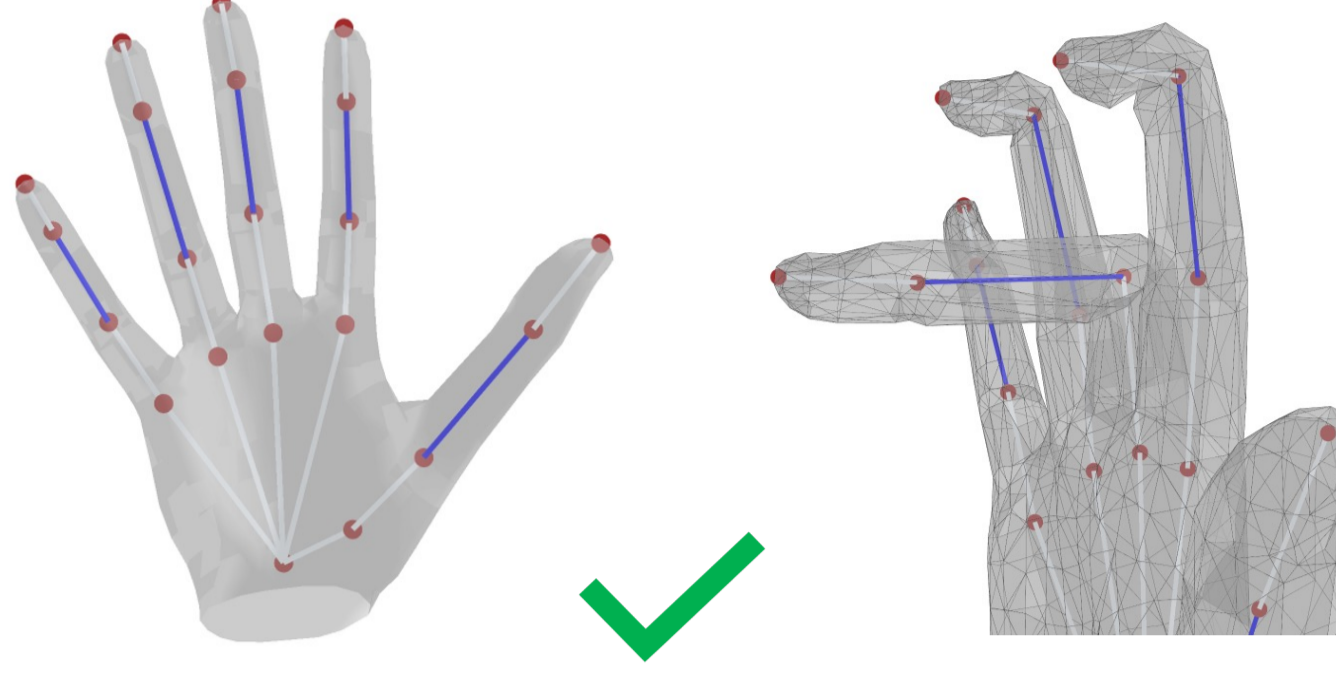


Direct scaling in LBS


$$v'_i = \sum_{j=1}^{n_b} W_{bij} \{a'_j + R_j (\phi_j (-a_j + v_i))\}$$


Shape-aware local scaling in LBS


$$v'_i = \sum_{j=1}^{n_b} W_{bij} \{a'_j + R_j (W_{eij} s_j + (-a_j + v_i))\}$$

$$s_j = (\phi_j - 1)(b_j - a_j)$$



Input depth



MANO [1]
(SIGGRAPH Asia 2017)



Our
adaptive MANO




Model near depth data
 Model in front of depth data
 Model behind depth data

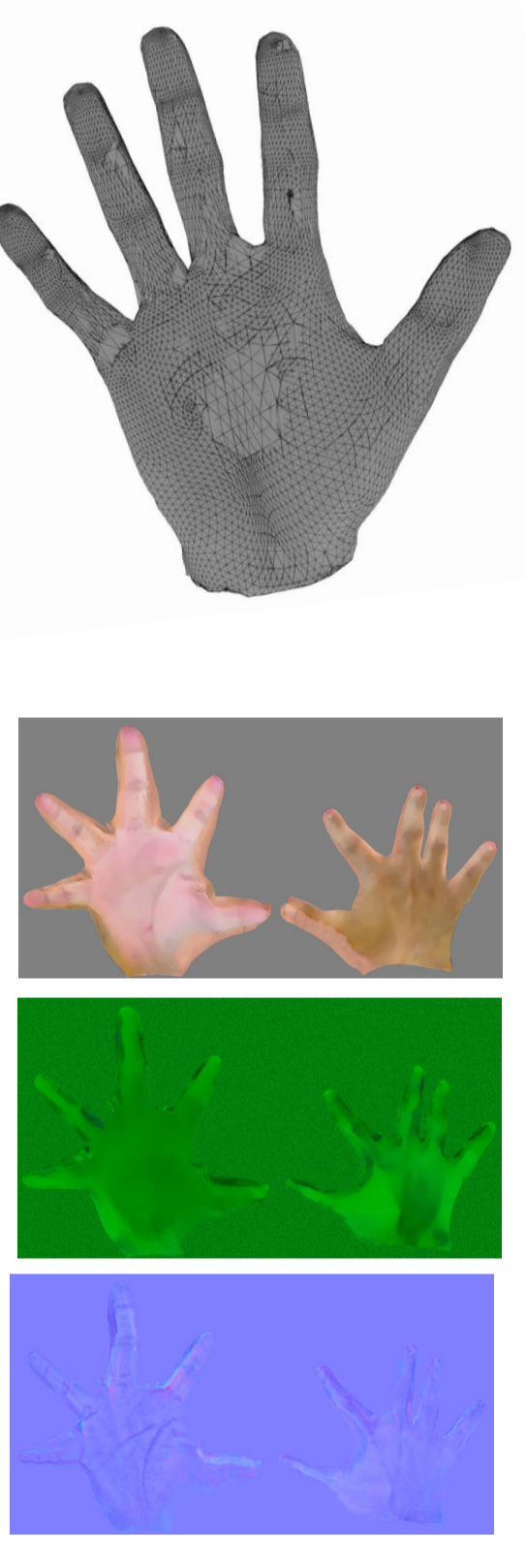
Pratik Kalshetti and Parag Chaudhuri. Local scale adaptation for augmenting hand shape models. In *ACM SIGGRAPH 2022 Posters*, 2022. (2nd place in ACM Student Research Competition)

Hand Appearance Reconstruction


Input frames



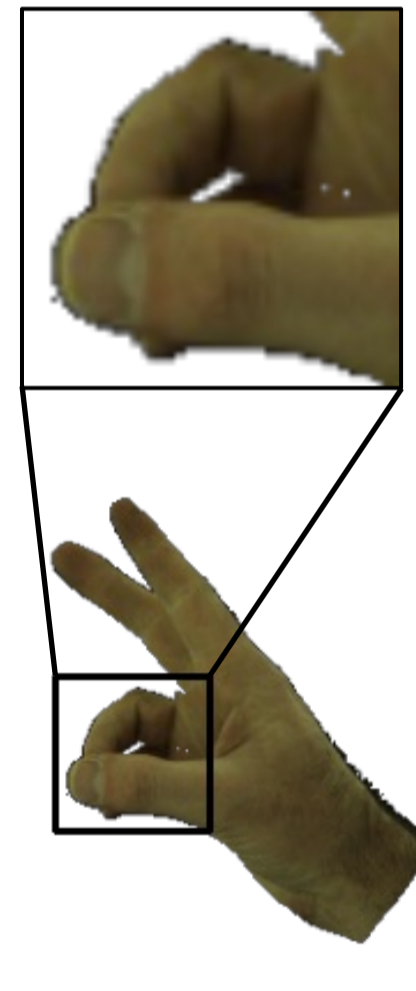
Our output avatar



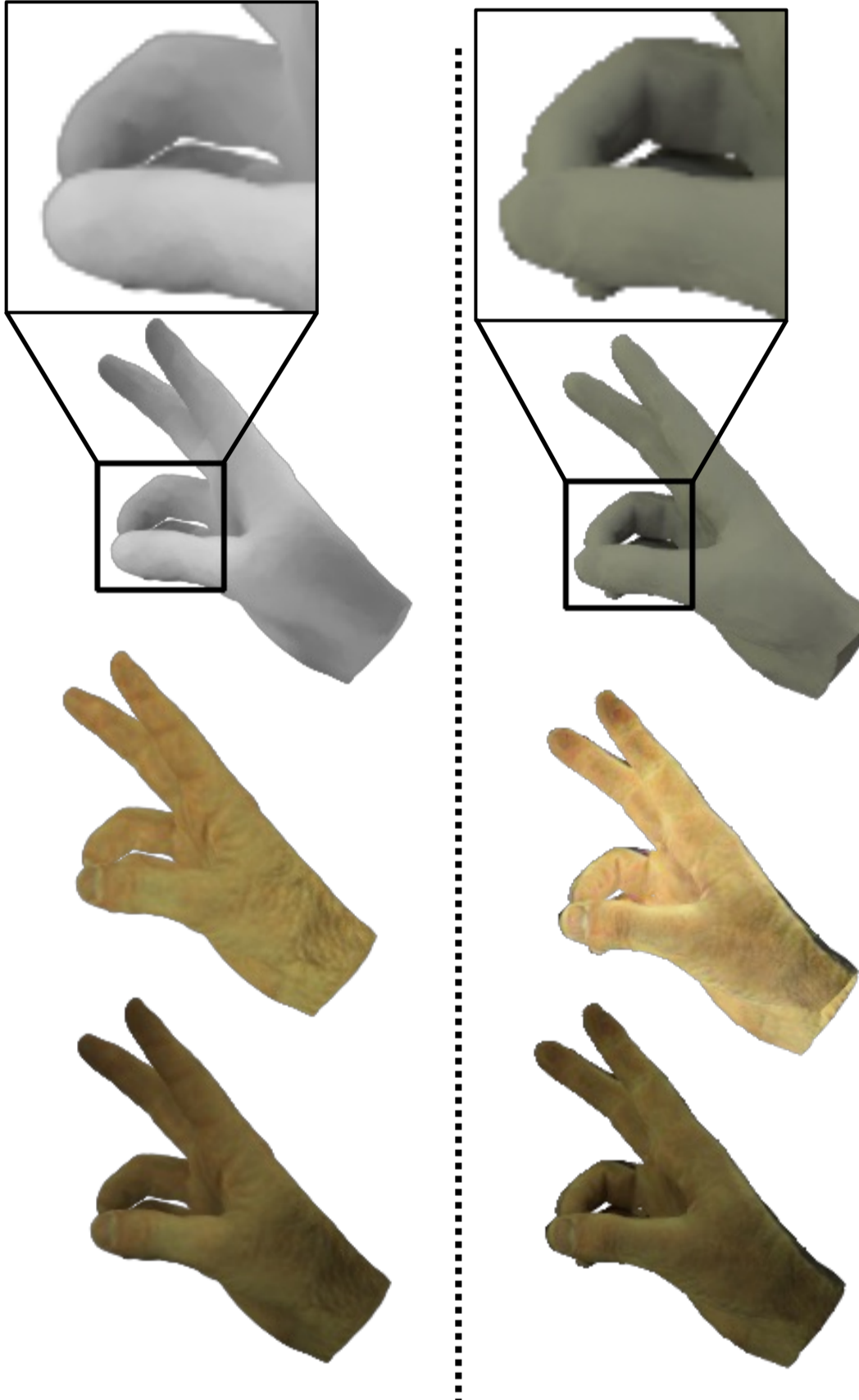
Final rendered Albedo Normal Lighting




Input




HandAvatar [2] (CVPR 2023)



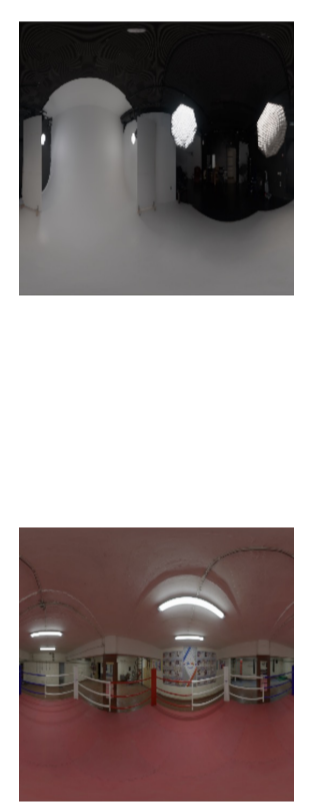
Ours



Novel pose synthesis



Relighting



Method	PSNR↑	SSIM↑	LPIPS↓
SelfRecon [3] (CVPR 2022)	26.38	0.878	0.142
HumanNeRF [4] (CVPR 2022)	27.64	0.883	0.114
Ours	28.66	0.897	0.090

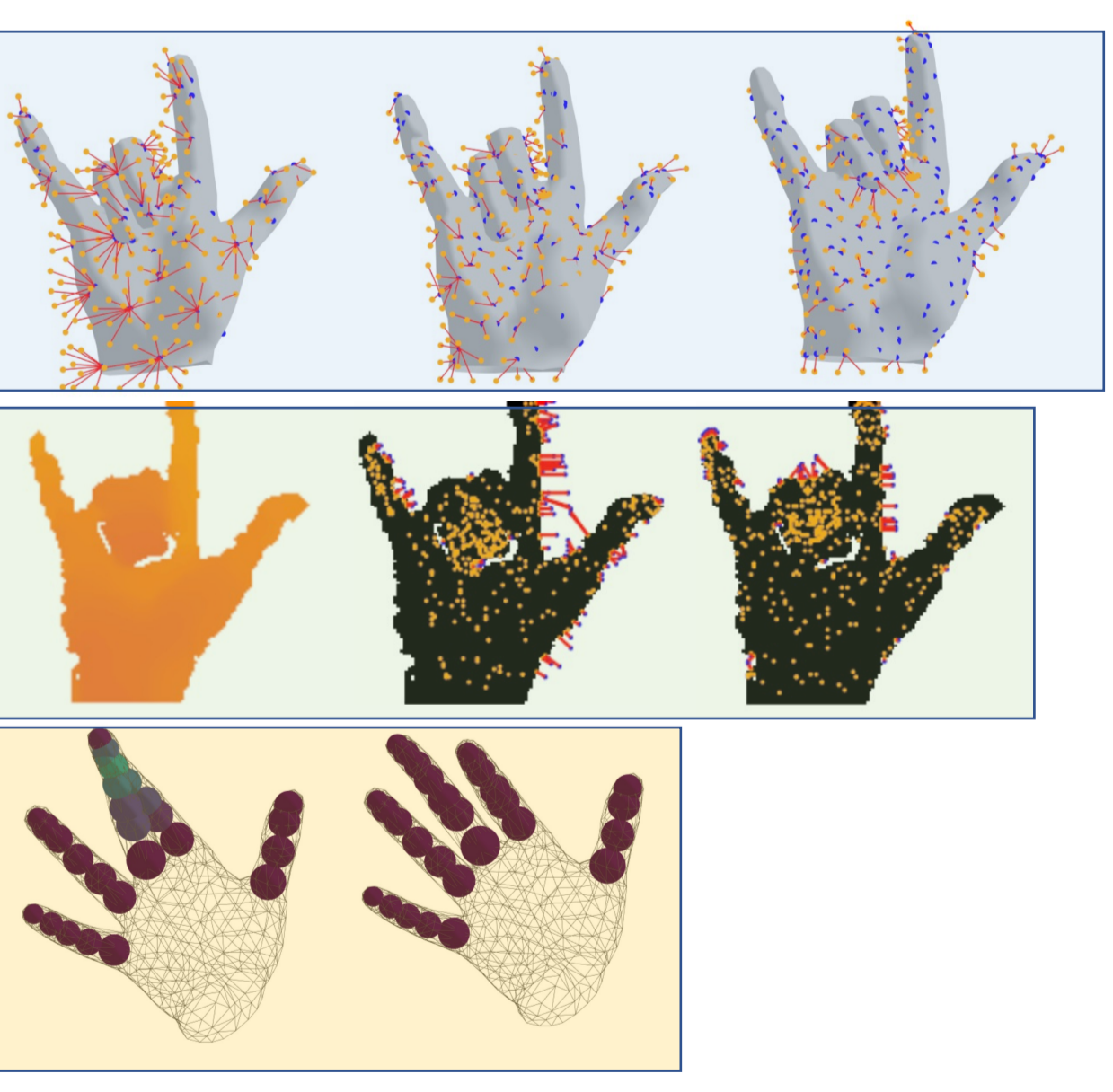
Pratik Kalshetti and Parag Chaudhuri. Intrinsic hand avatar: Illumination-aware hand appearance and shape reconstruction from monocular RGB video. In *WACV*, 2024.

Hand Tracking


Articulated Registration

$$E(\beta, \theta) = \sum_{\tau \in \mathcal{T}} \omega_{\tau} E_{\tau}(\beta, \theta)$$

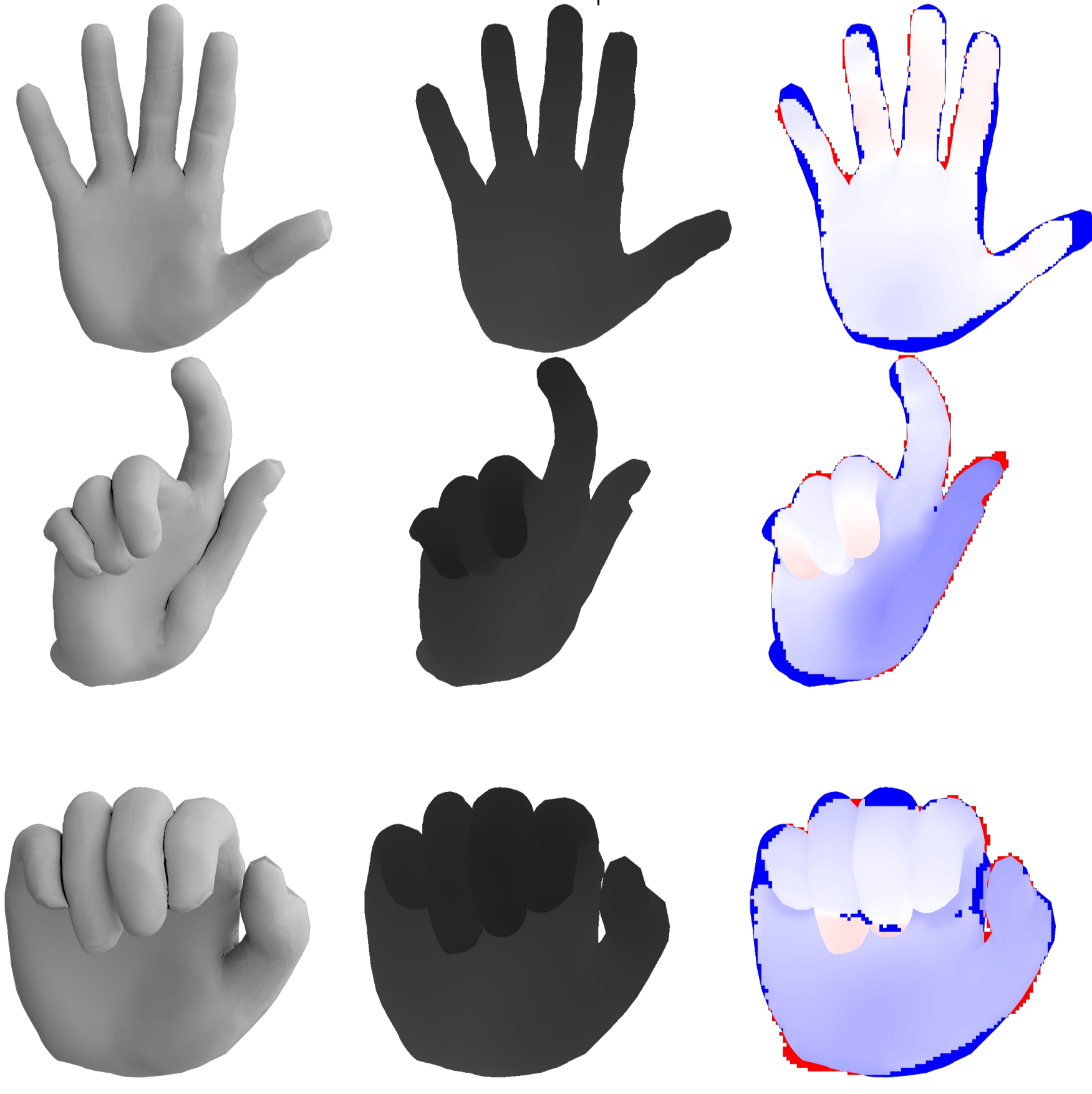
- E_{data3D}
- E_{data2D}
- E_{bound}
- E_{pca}
- E_{int}
- E_{reinit}
- E_{shape}
- E_{temp}




Input depth




Ours



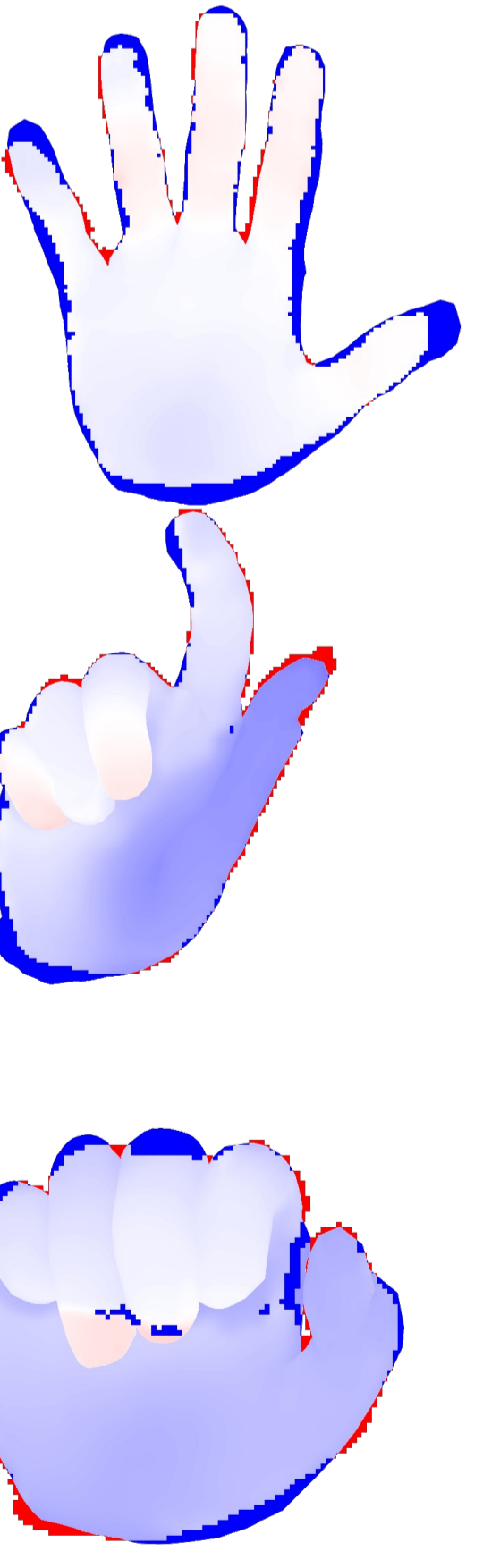
Rendered mesh



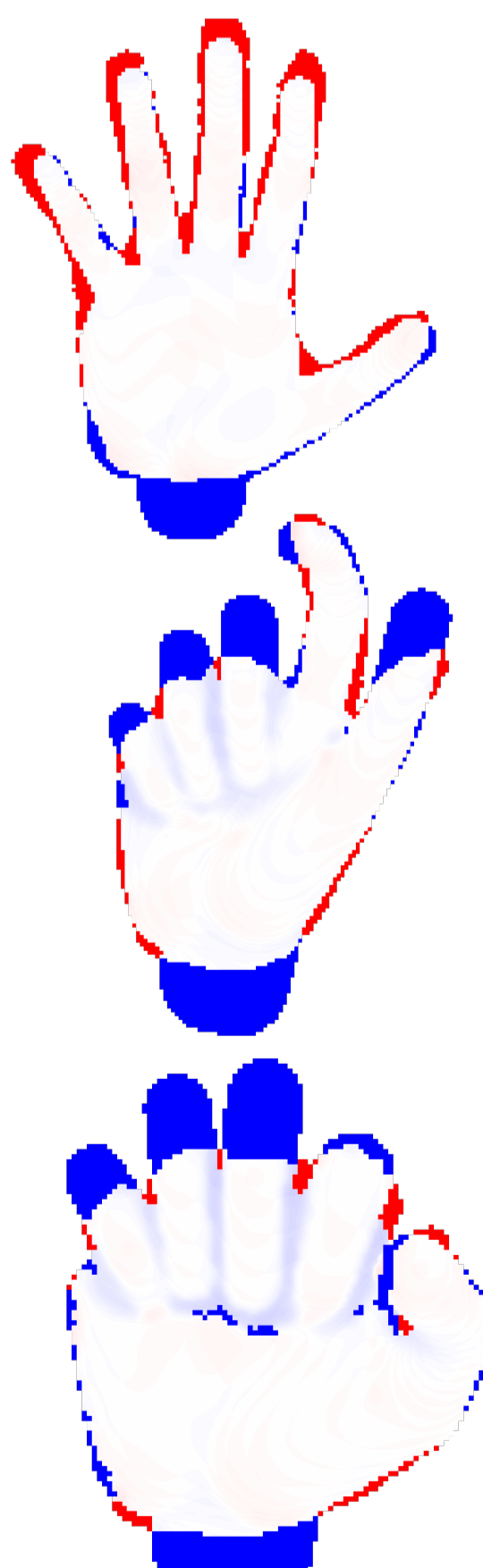
Rendered depth



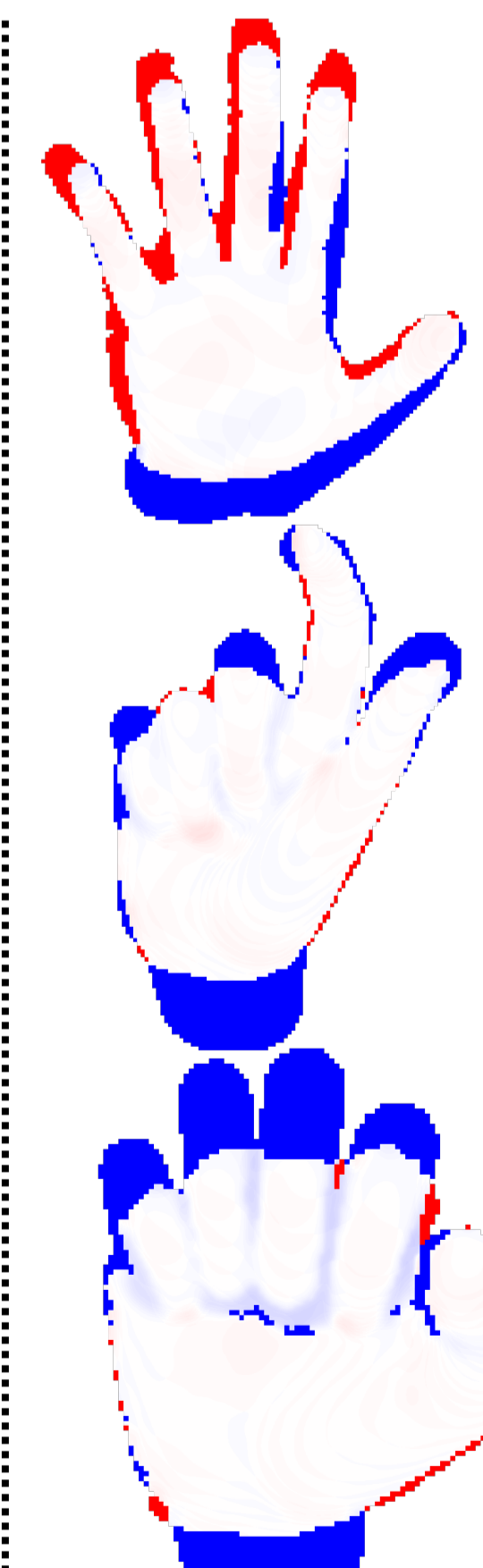
Depth difference



HAdjust [5] (ICCV 2017)



HOnline [6] (SIGGRAPH Asia 2017)



Pratik Kalshetti and Parag Chaudhuri. Local Scale Adaptation to Hand Shape Model for Accurate and Robust Hand Tracking. In *Computer Graphics Forum (SCA)*, 41:219–229, 2022.

References

- Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM TOG*, 36(6):245:1–245:17, 2017.
- Xingyu Chen, Baoyuan Wang, and Heung-Yeung Shum. Hand avatar: Free-pose hand animation and rendering from monocular video. In *CVPR*, 2023.
- Boyi Jiang, Yang Hong, Hujun Bao, and Juyong Zhang. Selfrecon: Self reconstruction your digital avatar from monocular video. In *CVPR*, 2022.
- Chung-Yi Weng, Brian Curless, Pratul P. Srinivasan, Jonathan T. Barron, and Ira Kemelmacher-Shlizerman. HumanNeRF: Free-viewpoint rendering of moving people from monocular video. In *CVPR*, 2022.
- Edoardo Remelli, Anastasia Tkach, Andrea Tagliasacchi, and Mark Pauly. Low-dimensionality calibration through local anisotropic scaling for robust hand model personalization. In *ICCV*, 2017.
- Anastasia Tkach, Andrea Tagliasacchi, Edoardo Remelli, Mark Pauly, and Andrew Fitzgibbon. Online generative model personalization for hand tracking. *ACM TOG*, 36(6):1–11, 2017.

Acknowledgements

I thank my Ph.D. advisor, Parag Chaudhuri, for introducing the problem and supporting me throughout this journey. I am also grateful to be part of the vibrant VIGIL lab at IIT Bombay.

