Multiview Image-based Hand Geometry Refinement using Differentiable Monte Carlo Ray Tracing

Giorgos Karvounas[12]
gkarv@ics.forth.gr
Nikolaos Kyriazis[1]
kyriazis@ics.forth.gr
Iason Oikonomidis[1]
oikonom@ics.forth.gr
Aggeliki Tsoli[1]
aggeliki@ics.forth.gr
Antonis A. Argyros[12]
argyros@ics.forth.gr

[1] Computational Vision and Robotics Laboratory
Institute of Computer Science
Foundation for Research and Technology - Hellas (FORTH)
Heraklion, Crete
Greece
[2] Computer Science Department
University of Crete
Heraklion, Crete
Greece

1 Evaluation on Real Data

In this work we address the problem of hand pose and shape refinement in an unsupervised manner. Our method [1] is based on differentiable rendering using the Monte Carlo ray tracer by Li et al.[2]. To evaluate our method on real data, we relied on the InterHand2.6M [3] dataset. InterHand2.6M is a large-scale dataset with real multi-view images depicting hands of various subjects performing sign language and everyday life gestures. The annotations within InterHand2.6M are partly manual (joint locations) and partly automatic (joint locations and MANO [4] poses) with reported annotation errors in the order of a few millimeters. However, when it comes to fine prediction of hand geometry, the provided annotations can mainly serve as references for the true location and geometry of the hand rather than ground truth. This is evident in the first row of the figures provided below.

As a remedy to imperfect ground truth annotations in InterHand2.6M, we resorted to experiments with synthetic data that closely simulate the capture environment in InterHand2.6M and come with perfect and known ground truth. As described in section 4.2.1 of the main paper, we evidenced the correlation between the improvement in 3D hand geometry estimation and the backprojection error. Specifically, we found a monotonic relation where the error on the estimated 3D geometry decreases as the back-projection error decreases. To further show how the per-camera backprojection improvement actually yields a true amelioration in 3D hand geometry, we conducted extensive qualitative experiments that demonstrate visually consistent and significant improvement in both metrics. Indicative results, in enough detail to highlight the improvements, are presented below.

In addition, we show the impact of our method on the estimated variable c (color map), which implements the hard constraint of color constancy during optimization. The sharpness and definition of the visible parts of the color maps can only be attained through an accurate estimation of the 3D geometry and a proper method (see our discussion on ray tracing in the introduction of the main text [1]) to correctly connect the 2D backprojection error with 3D information (surface color). If the geometry had not been estimated with enough accuracy, the color maps would have been less sharp, due to the positional (on the 3D surface) inconsistencies of the color updates during backpropagation.

In the following figures, we show representative results of hand geometry refinement on InterHand2.6M. In each figure, the first row corresponds to the MANO annotation provided in the dataset that we use as an initialization to our method. The second row shows the result of our method. The third and fourth rows depict the estimated hand appearance before and after refinement. The improvement on the estimated hand geometry as well as the estimated appearance of the hand are evident.

Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan V GPU used for the execution of this research. The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I) under the "1st Call for H.F.R.I Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment" Project I.C.Humans, Number: 91.

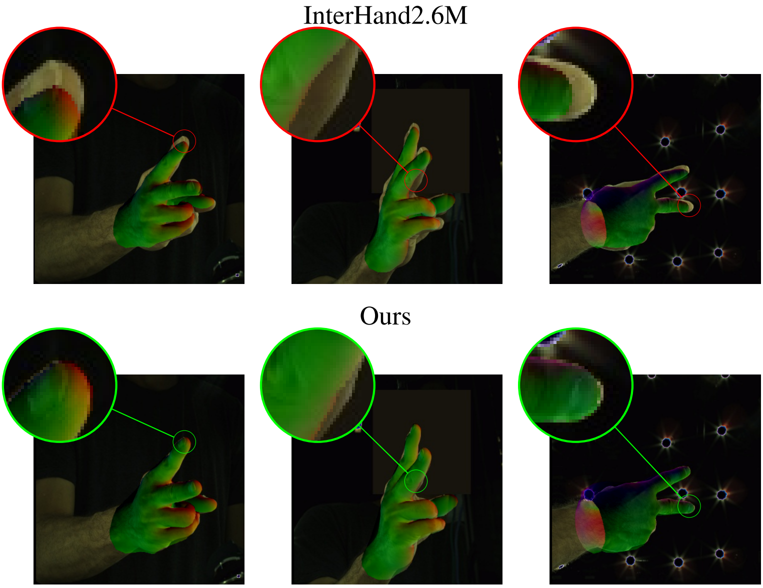
References

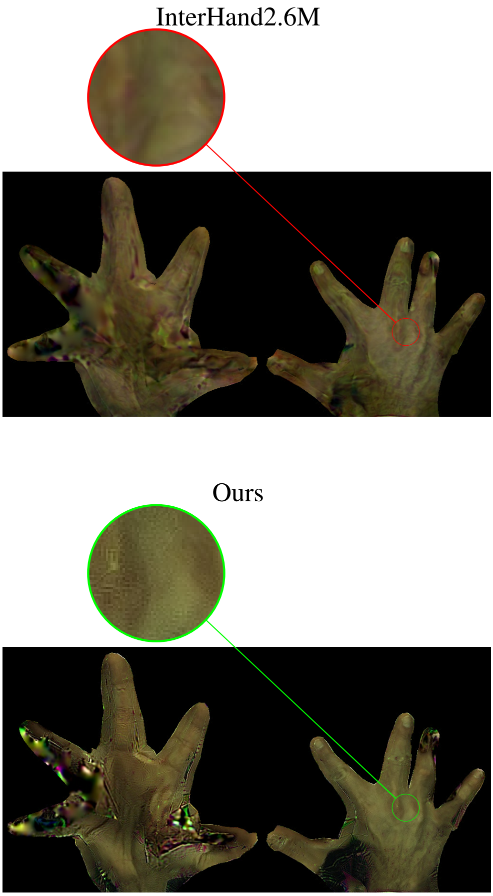
[1] Giorgos Karvounas, Nikolaos Kyriazis, Iason Oikonomidis, Aggeliki Tsoli, and Antonis A Argyros. Multi-view image-based hand geometry refinement using differentiable monte carlo ray tracing. In *British Machine Vision Conference (BMVC 2021)*, Virtual, UK, November 2021. BMVA.

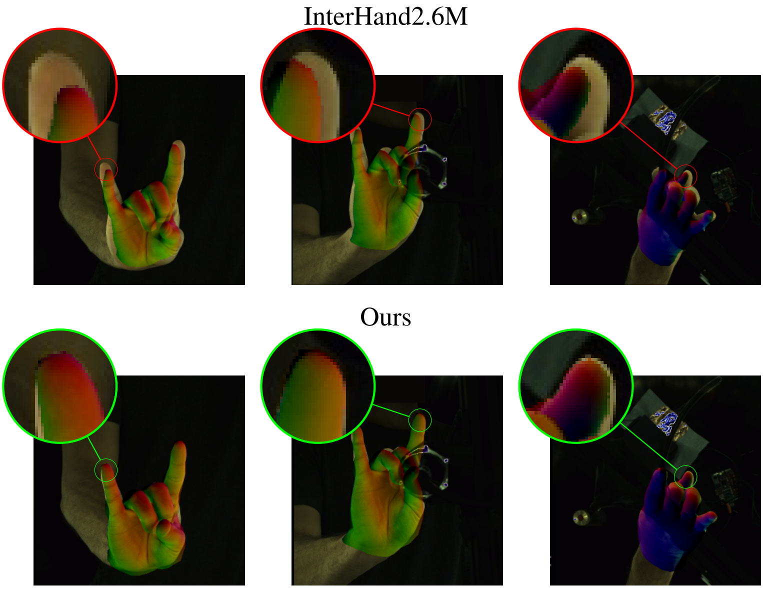
[2] Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. Differentiable monte carlo ray tracing through edge sampling. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 2018.

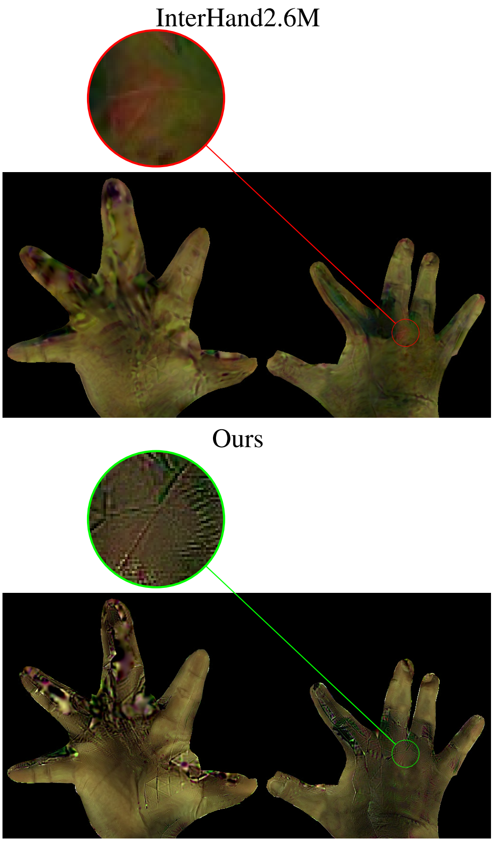
[3] Gyeongsik Moon, Shoou-I Yu, He Wen, Takaaki Shiratori, and Kyoung Mu Lee. Interhand2.6m: A dataset and baseline for 3d interacting hand pose estimation from a single rgb image. In *European Conference on Computer Vision (ECCV)*, 2020.

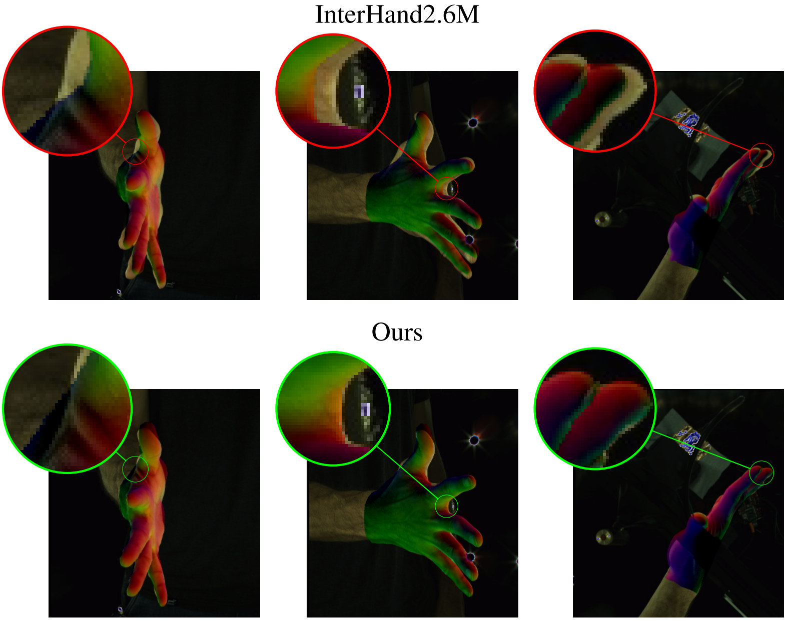
[4] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 2017.

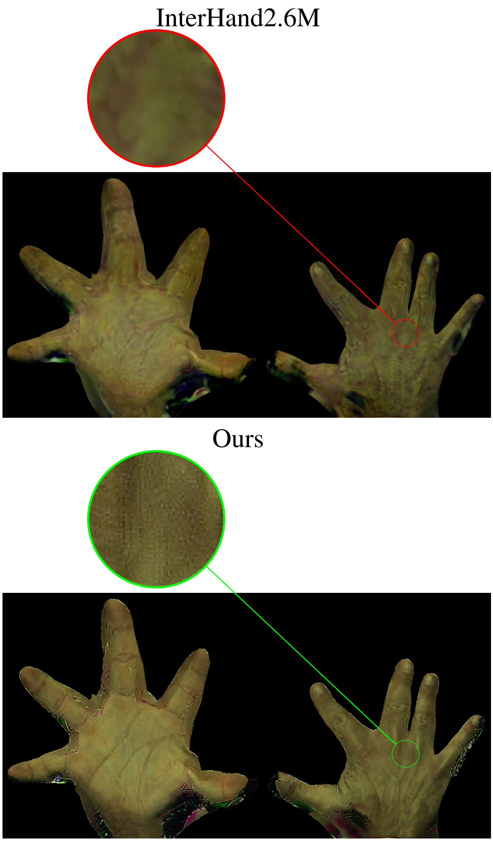

InterHand2.6M
Ours

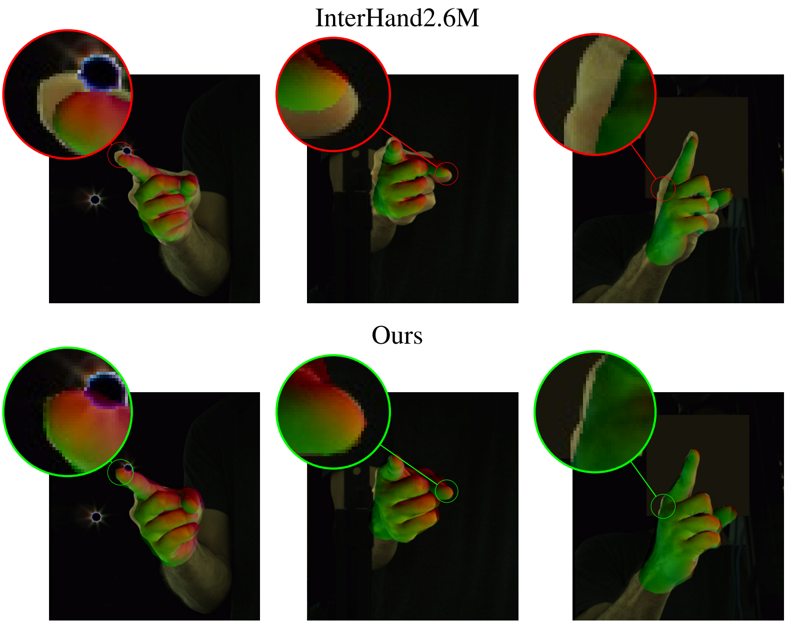

InterHand2.6M
Ours

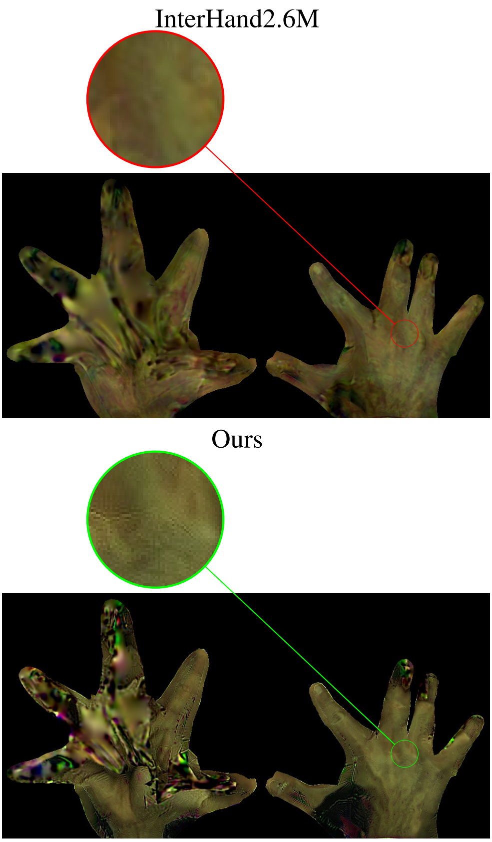

InterHand2.6M
Ours

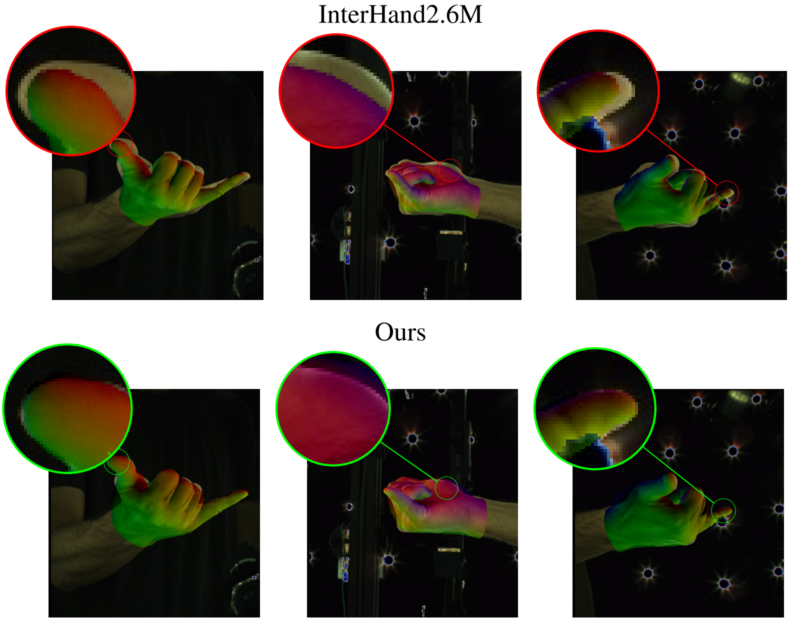

InterHand2.6M
Ours

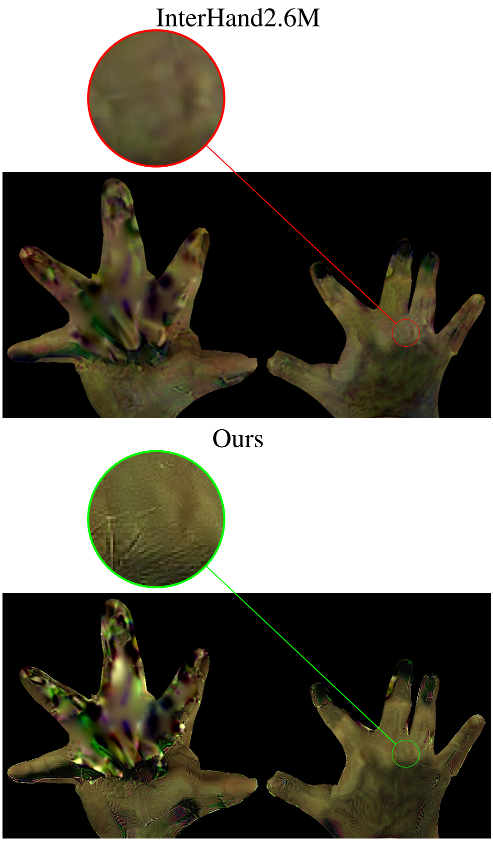

InterHand2.6M
Ours


InterHand2.6M
Ours

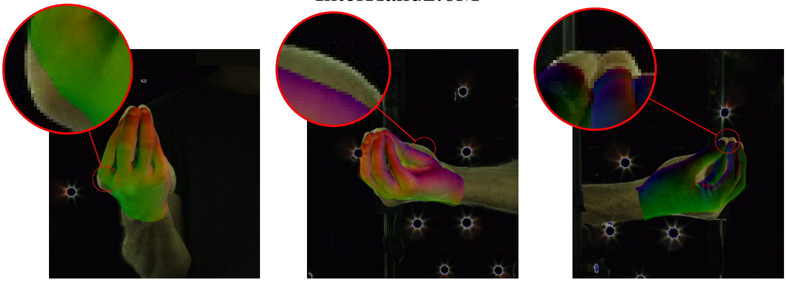

InterHand2.6M
Ours


InterHand2.6M
Ours

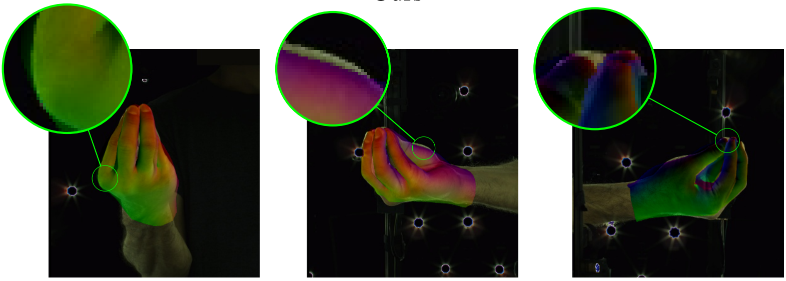

InterHand2.6M
Ours


InterHand2.6M
Ours

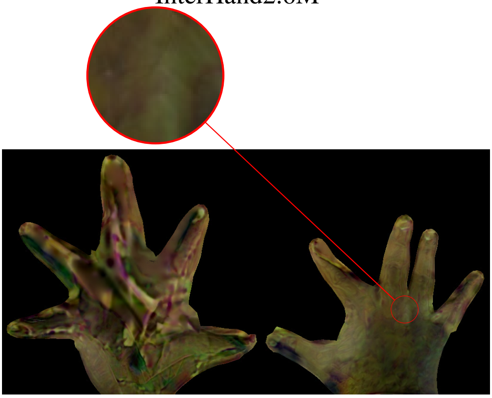
InterHand2.6M



Ours



InterHand2.6M



Ours