Weakly-Supervised 3D Hand Reconstruction with Knowledge Prior and Uncertainty Guidance **Supplementary Material**

Yufei Zhang[1], Jeffrey O. Kephart[2], and Qiang Ji[1]

[1] Rensselaer Polytechnic Institute, [2] IBM Research
{zhangy76, jiq}@rpi.edu, kephart@us.ibm.com

In this supplementary material, we provide additional experiment results:

- Section 1: Evaluation of Incorporating the Hand Physics Knowledge with Comparison to Existing Techniques
- Section 2: Qualitative Evaluation and Comparison with State-of-the-Arts

1 Evaluation of Incorporating the Hand Physics Knowledge with Comparison to Existing Techniques

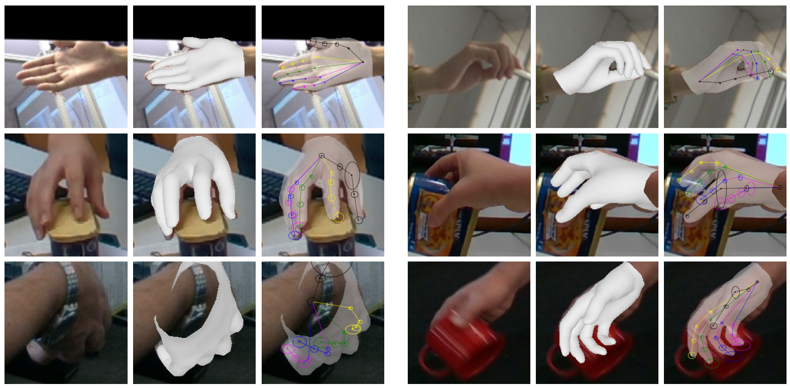
In the main paper, we demonstrate the effectiveness of incorporating the hand physics knowledge to reduce the reconstruction error and improve the reconstruction physical plausibility. Here, we highlight the advantages of our method compared to existing techniques aimed at preventing invalid penetration in 3D reconstructions.

We compare with the method proposed by Tzionas et al. [2] and present the results in Table 1. As shown, the method proposed by Tzionas et al. [2] can reduce the penetration rate but their reconstruction accuracy can suffer. The degraded reconstruction performance stems from their formulation of the non-penetration loss. They only model the colliding mesh triangles located on the surface and impose the loss to separate these colliding triangles. In contrast, our proposed non-penetration loss models the inside mesh vertices and formulates the non-penetration loss based on their distance to the mesh surface, leading to a more effective way of preventing invalid penetrations.

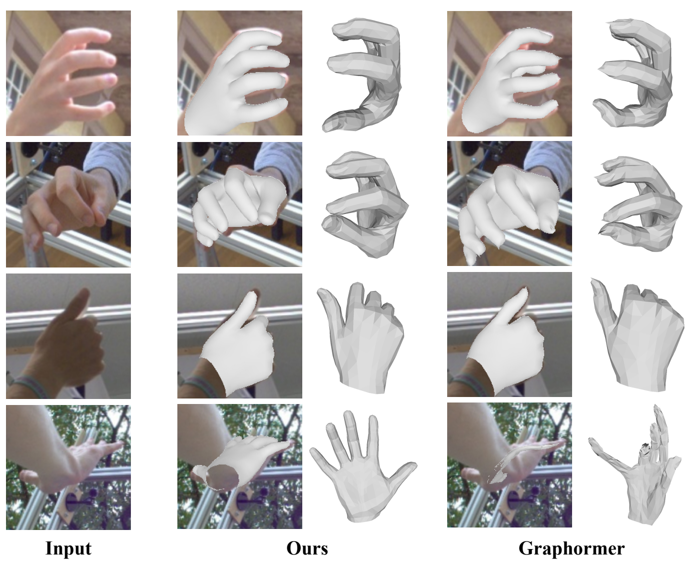
Table 1: Evaluation of Incorporating the Hand Physics Knowledge with Comparison to Existing Techniques. "Baseline" refers to the model trained with 2D hand landmark annotation and incorporating all the hand knowledge except for the hand physics knowledge. The units of E_J and E_J are in mm, while PR is in percentage.

Method	E_J↓	E_V↓	PR↓
Baseline	9.6	10.0	11.4
Tzionas et al. [2]	10.3	10.6	1.0
Ours	9.4	9.8	1.9

2 Qualitative Evaluation and Comparison with State-of-the-Arts



(a) Qualitative Evaluation with Uncertainty Visualization



(b) Qualitative Evaluation with Comparison to the SOTA Method

Fig. 1: Qualitative Evaluation and Comparison with the State-of-the-Arts. (a) The testing images are from FreiHAND (top), HO3D (middle), and DexYCB (bottom). For the uncertainty visualization figures, the colors indicate finger identity: thumb (black), index (yellow), middle (green), ring (blue), and pinky (magenta). The size of the ellipse at each joint represent the magnitude of the estimated variance. (b) The testing images are from FreiHAND.

In Figure 1, we present additional qualitative evaluation and comparison with the State-of-the-Art method to further showcase the improvements of the proposed method. In Figure 1(a), we demonstrate that our method achieves favorable 3D hand reconstruction results on various images, even in cases with significant occlusions or extreme camera views. Meanwhile, the ambiguous image regions, such as image truncations and occlusions, are effectively captured by our method through large variance estimates, offering valuable insights into the observation uncertainty and the prediction confidence. Moreover, in Figure 1(b), we compare our method with Graphormer [1], a SOTA fully-supervised 3D hand reconstruction model. As illustrated, our method demonstrates performance comparable to that of the fully-supervised model. However, it is worth noting that our approach achieves this without leveraging any 3D data; instead, it relies on cheap 2D hand landmark annotation and well-established generic hand knowledge. In challenging scenarios, such as the one with an extreme camera view at the bottom of Figure 1(b), the SOTA fully-supervised model may fail and produce very unrealistic reconstruction results. In contrast, our method exhibits greater robustness under such conditions, showcasing its resilience in challenging real-world situations.

References

1. Lin, K., Wang, L., Liu, Z.: Mesh graphormer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 12939–12948 (2021)
2. Tzionas, D., Ballan, L., Srikantha, A., Aponte, P., Pollefeys, M., Gall, J.: Capturing hands in action using discriminative salient points and physics simulation. International Journal of Computer Vision (IJCV) **118**(2), 172–193 (Jun 2016), `https://doi.org/10.1007/s11263-016-0895-4`