

HOSt3R: Keypoint-free Hand-Object 3D Reconstruction from RGB images

Anilkumar Swamy^{1,2} Vincent Leroy¹ Philippe Weinzaepfel¹
Jean-Sébastien Franco² Grégory Rogez¹

¹NAVER LABS Europe ²Inria centre at the University Grenoble Alpes

HOSt3R: Keypoint-free Hand-Object 3D Reconstruction from RGB images

Supplementary Material



Figure 1. **Example from the Multi-View ObMan synthetic dataset:** Each row represents a single hand-object scene and columns is a different view of the same hand-object. Hand-objects are rendered on a black background, black background is acceptable as pointmap estimation network is trained with only hand-object masked pixels.

1. Multi-view Synthetic Dataset

Multi-View ObMan is an extension of the original ObMan [2] dataset designed to support multi-view hand-object 3D reconstruction tasks. It simulates multiple synchronized camera views of synthetic hand-object interactions, rendered using the MANO [3] hand model and ShapeNet objects [1]. Each interaction instance is captured from several virtual viewpoints with consistent lighting and pose, and includes per-view RGB images, depth maps, segmentation masks, 2D keypoints, and full 3D annotations. Camera intrinsics and extrinsics are provided to enable accurate multi-view geometry reasoning. This extension facilitates research in multi-view pose estimation, shape reconstruction, and occlusion-robust modeling of hand-object interactions. A snapshot of RGB frames are depicted in Figure 1.

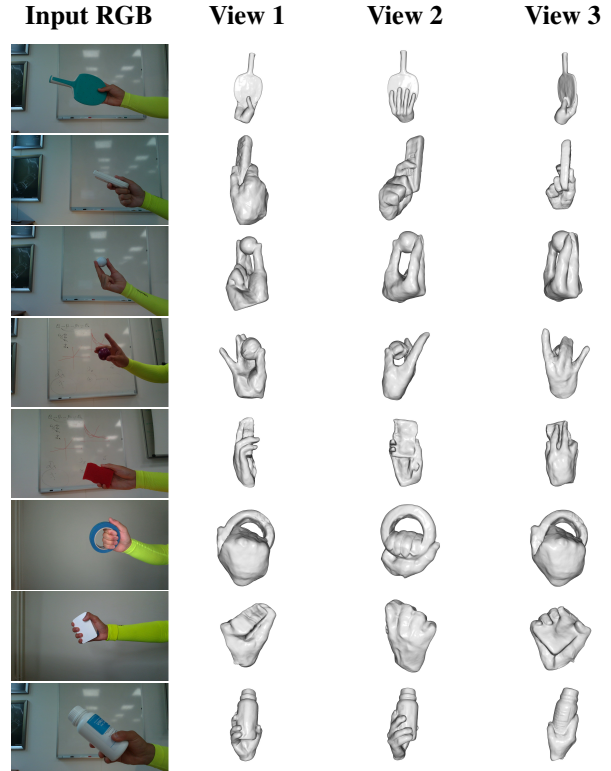


Figure 2. **Qualitative hand-object reconstructions on sequences from the SHOWMe dataset.** Each row shows one sequence: the first image is the RGB input, followed by three views of the reconstructed hand-object shape using our method.

2. Qualitative results

In Fig. 2 we show more qualitative results on SHOWMe dataset.

References

- [1] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 1
- [2] Yana Hasson, Gül Varol, Dimitris Tzionas, Igor Kalevatykh, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning joint reconstruction of hands and manipulated objects. In *CVPR*, 2019. 1
- [3] Javier Romero, Dimitrios Tzionas, and Michael J Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM ToG*, 2017. 1