IMPERIAL

CVPR Nashville JUNE 11-15, 2025

Dyn-HaMR: Recovering 4D Interacting Hand Motion from a Dynamic Camera

CVPR 2025 Highlight

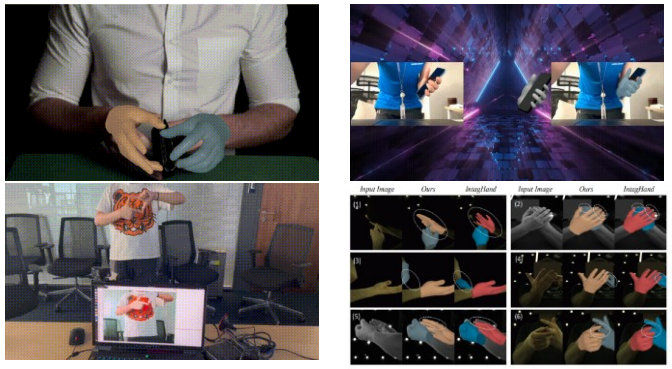
Zhengdi Yu, Stefanos Zafeirou, Tolga Birdal

Imperial College London

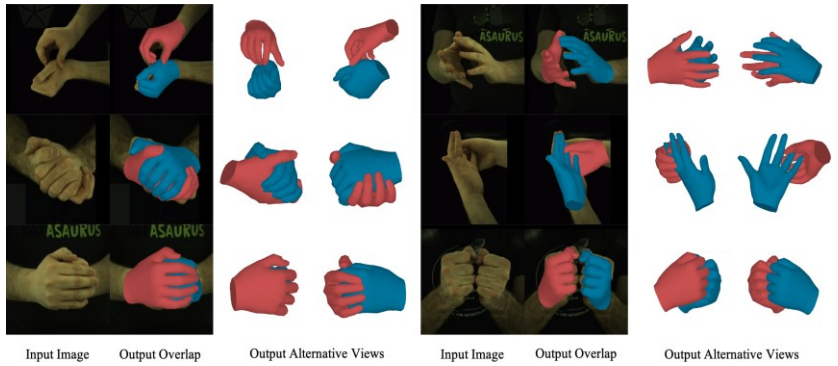
Existing hand reconstruction approaches

HaMeR [1]

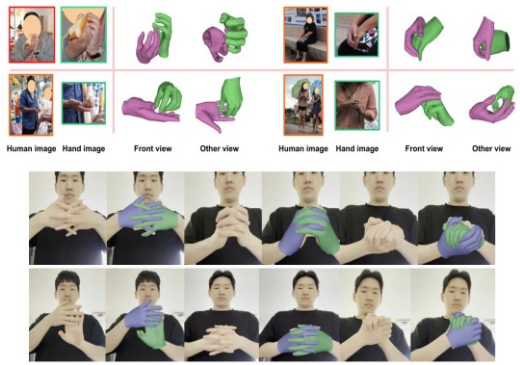
ACR [2]



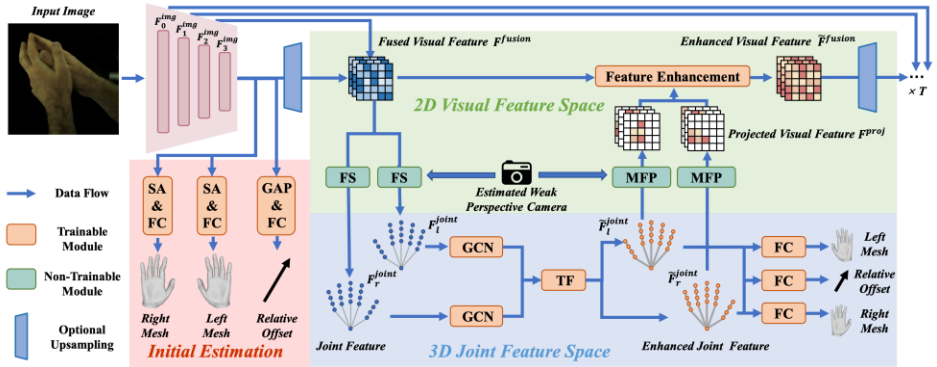
IntagHand [3]



InterWild [4]



DIR [5]

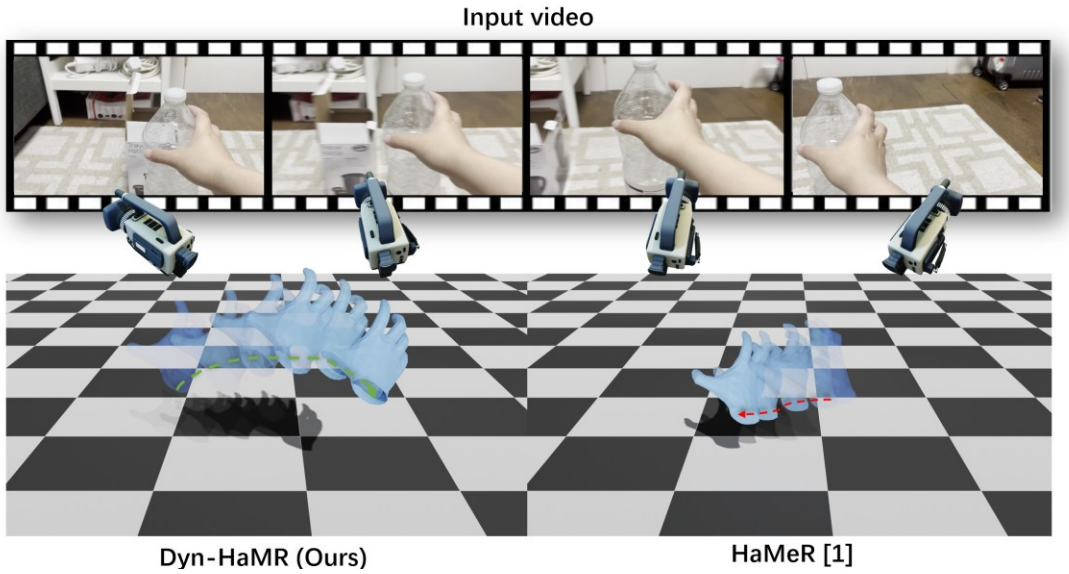


[1] G. Pavlakos et al. Reconstructing Hands in 3D with Transformers. CVPR, 2024
[2] Z. Yu et al. ACR: Attention Collaboration-based Regressor for Arbitrary Two-Hand Reconstruction. CVPR, 2023
[3] M. Li et al. Interacting Attention Graph for Single Image Two-Hand Reconstruction. CVPR, 2022
[4] G. Moon et al. Bringing Inputs to Shared Domains for 3D Interacting Hands Recovery in the Wild. CVPR, 2023
[5] P. Ren et al. Decoupled iterative refinement framework for interacting hands reconstruction from a single rgb image. ICCV 2023

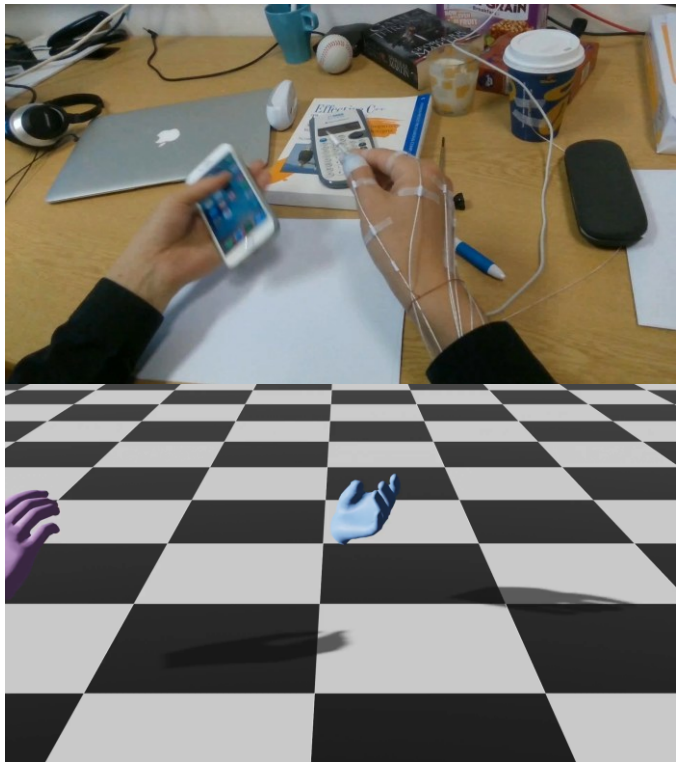
Open problems

- Existing HPE methods typically adopt the weak-perspective camera model

(i) Unable to recover global trajectory in world coordinate system

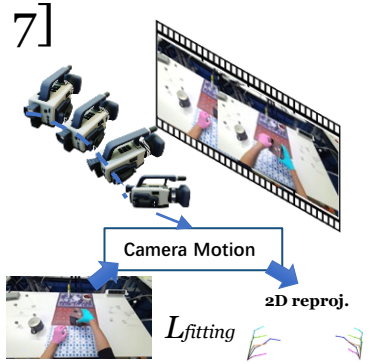


(ii) Weak depth simulation & Interaction



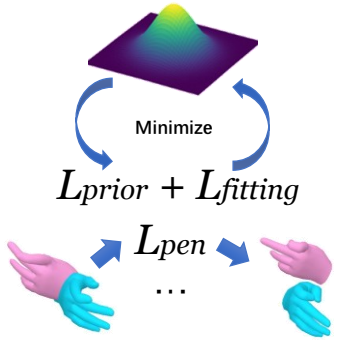
Key idea

- Hierarchical initialization
  - Generative infilling
  - Confidence-based integration for 2D observation [1, 6, 7]
  - Hallucination Handling



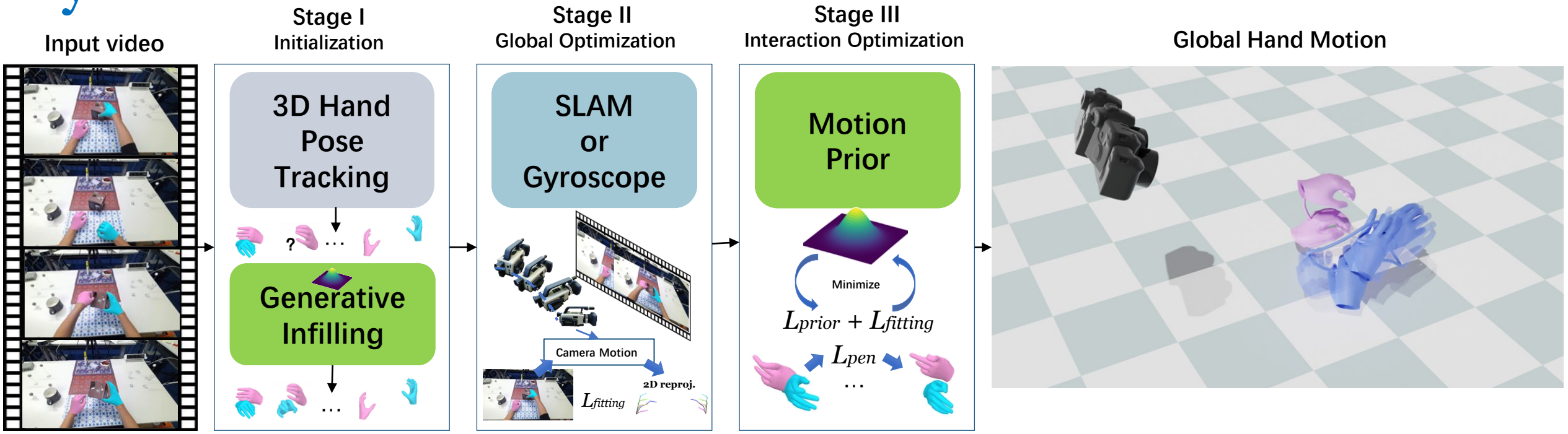
- Global motion optimization
  - Camera motion disentanglement

- Interacting motion prior optimization



[6] Y. Xu et al. ViTPose: Simple vision transformer baselines for human pose estimation. NeurIPS 2022
[7] C. Lugaresi. Mediapipe: A framework for building perception pipelines. 2019

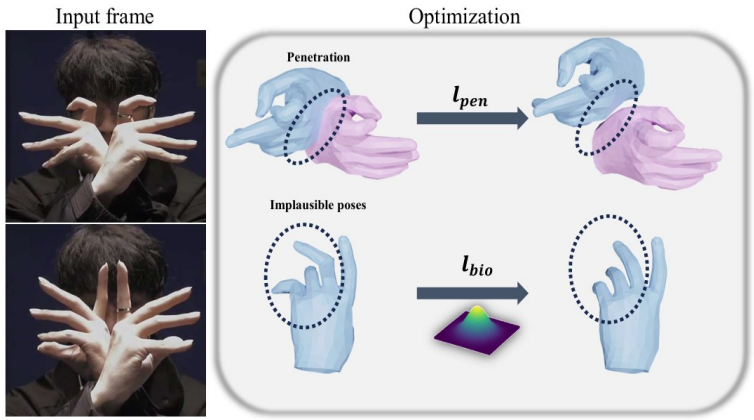
Dyn-HaMR



Objective

$$E_I({}^{\mathrm{w}}\boldsymbol{q}^h, \omega, \mathbf{R}_t, \boldsymbol{\tau}^{\mathbf{c}}{}_t) = \lambda_{2\mathrm{d}}\mathcal{L}_{2\mathrm{d}} + \lambda_{\mathrm{s}}\mathcal{L}_{\mathrm{smooth}} + \lambda_{\mathrm{cam}}\mathcal{L}_{\mathrm{cam}} + \lambda_{\boldsymbol{J}}\mathcal{L}_{\boldsymbol{J}} + \lambda_{\boldsymbol{\beta}}\mathcal{L}_{\boldsymbol{\beta}}.$$

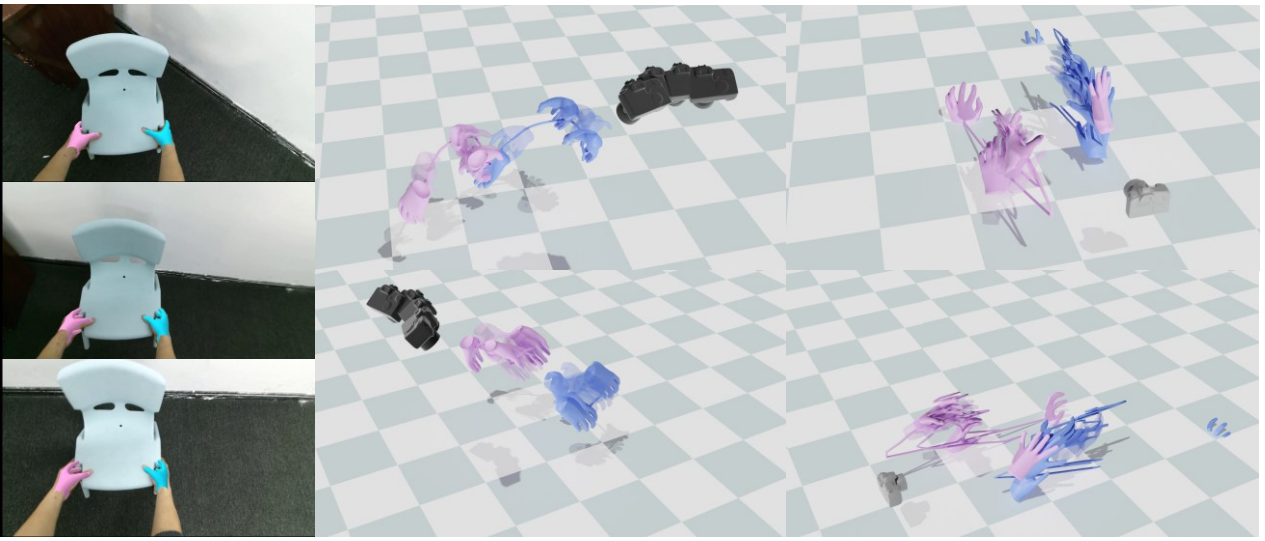
$$E_{II}({}^{\mathrm{w}}\boldsymbol{q}^h, \omega, \mathbf{R}_t, \boldsymbol{\tau}^{\mathbf{c}}{}_t) = \mathcal{L}_{\mathrm{prior}} + \mathcal{L}_{\mathrm{pen}} + \mathcal{L}_{\mathrm{bio}} + \lambda_{2\mathrm{d}}\mathcal{L}_{2\mathrm{d}} + \lambda_{\mathrm{s}}\mathcal{L}_{\mathrm{smooth}} + \lambda_{\mathrm{cam}}\mathcal{L}_{\mathrm{cam}} + \lambda_{\boldsymbol{J}}\mathcal{L}_{\boldsymbol{J}} + \lambda_{\boldsymbol{\beta}}\mathcal{L}_{\boldsymbol{\beta}}.$$



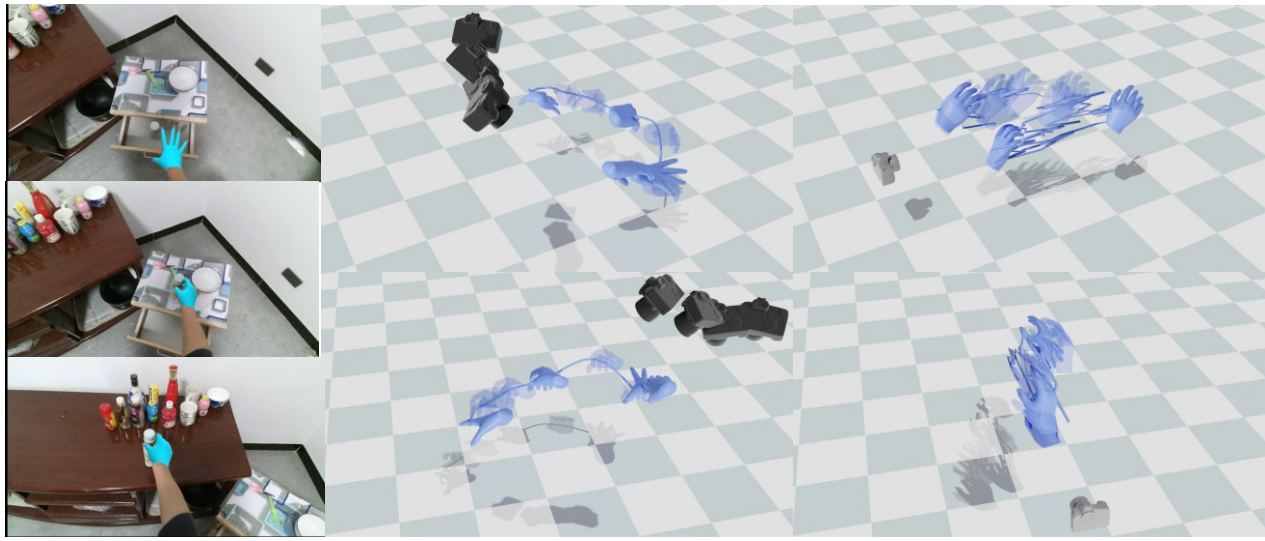
[8] Z. Teed. DPVO: Deep patch visual odometry. 2024
[9] E. Duran. Hmp: Hand motion priors for pose and shape estimation from video. WACV 2024

Qualitative results on HOI4D

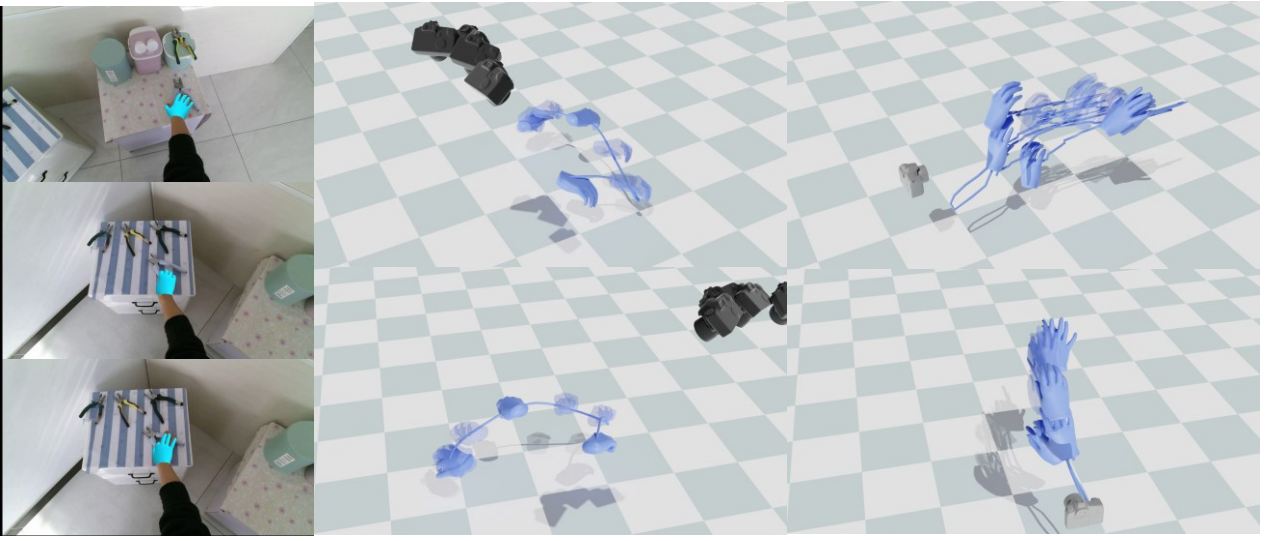
Input video Ours HaMeR



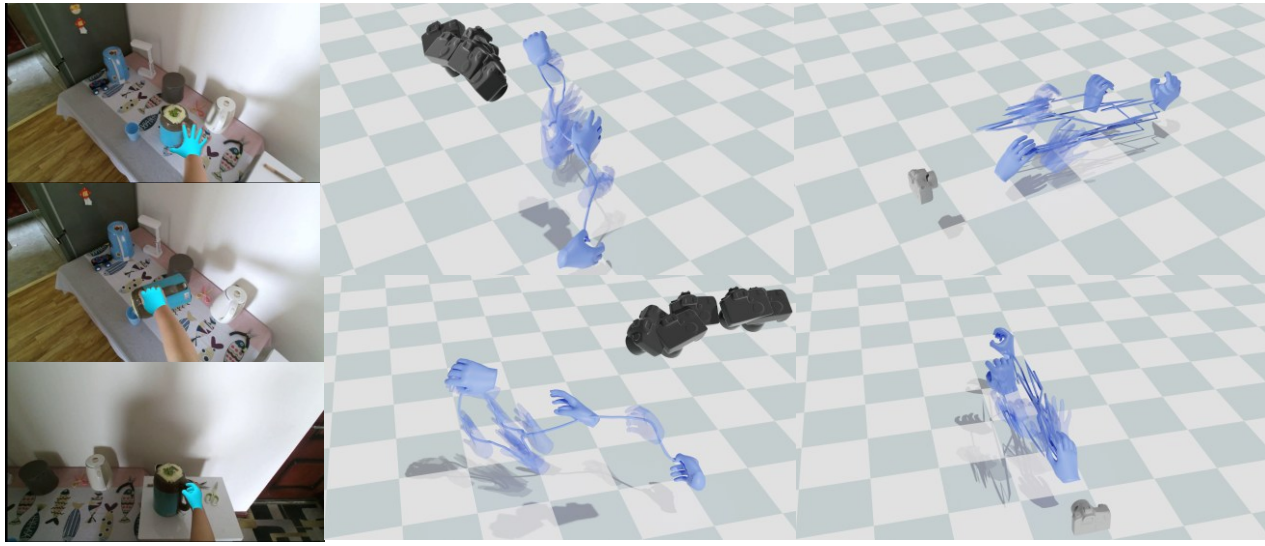
Input video Ours HaMeR



Input video Ours HaMeR

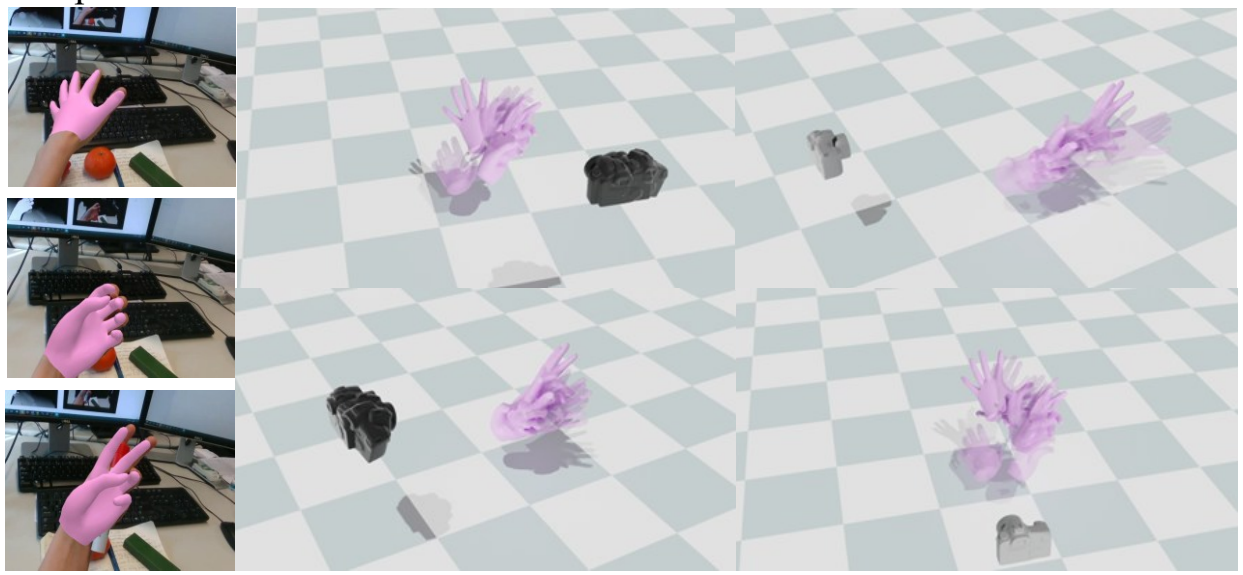


Input video Ours HaMeR

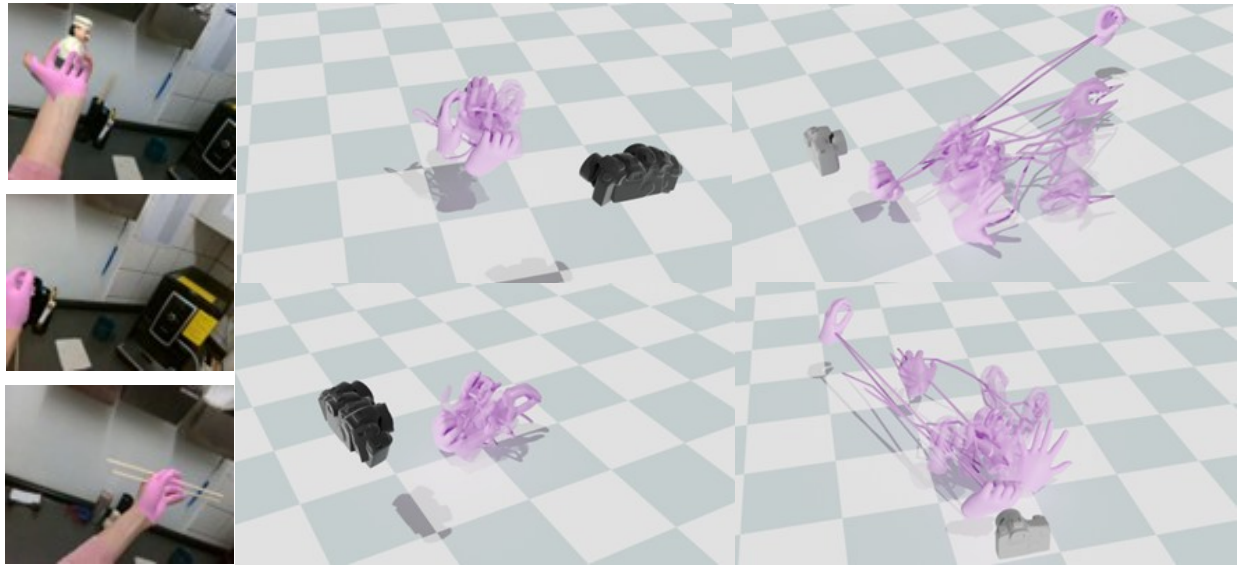


Qualitative results on EgoDexter

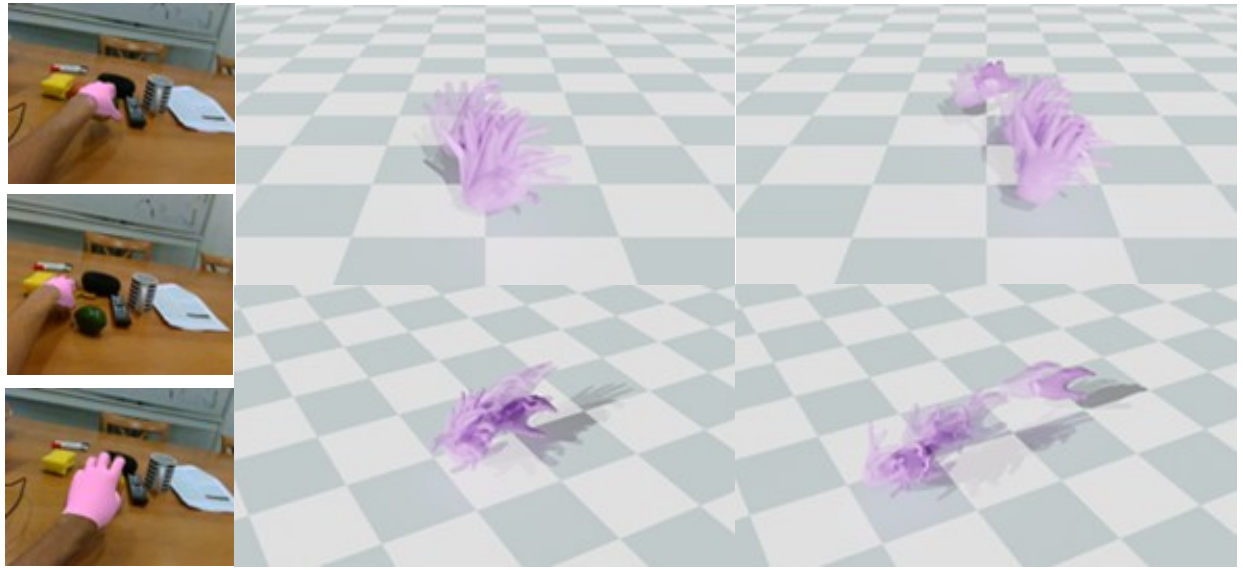
Input video Ours HaMeR



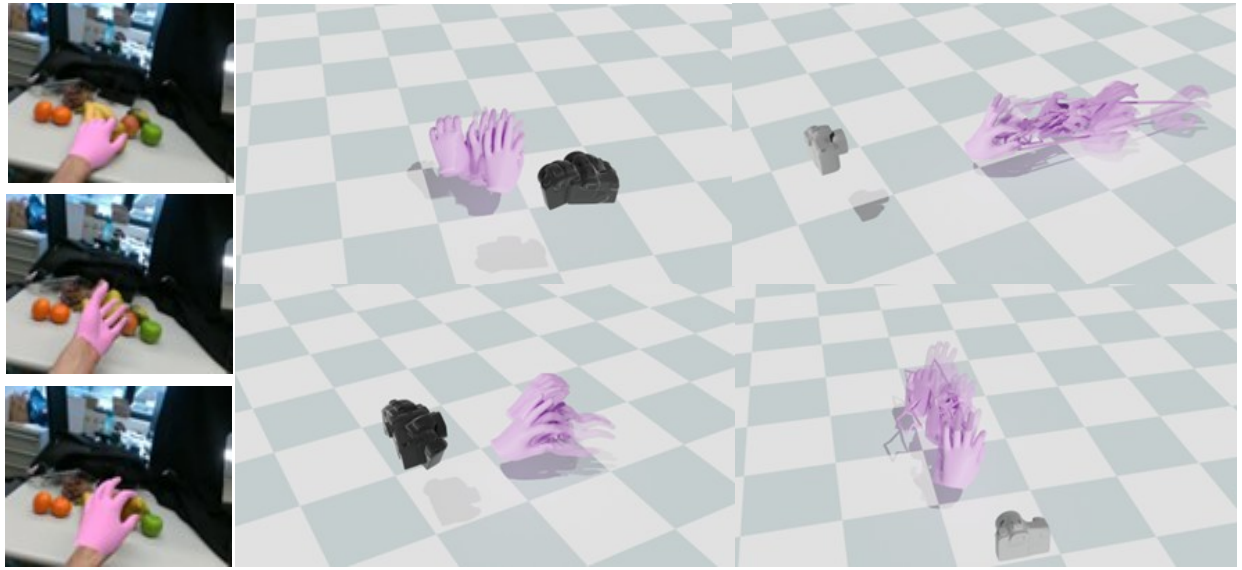
Input video Ours HaMeR



Input video Ours HaMeR



Input video Ours HaMeR

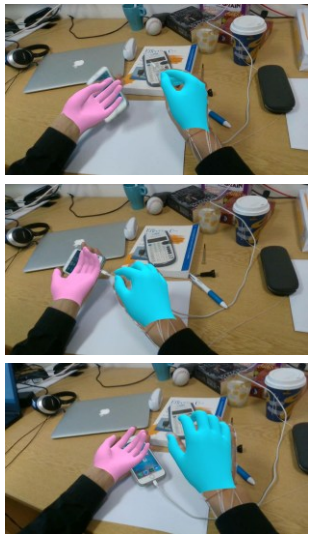


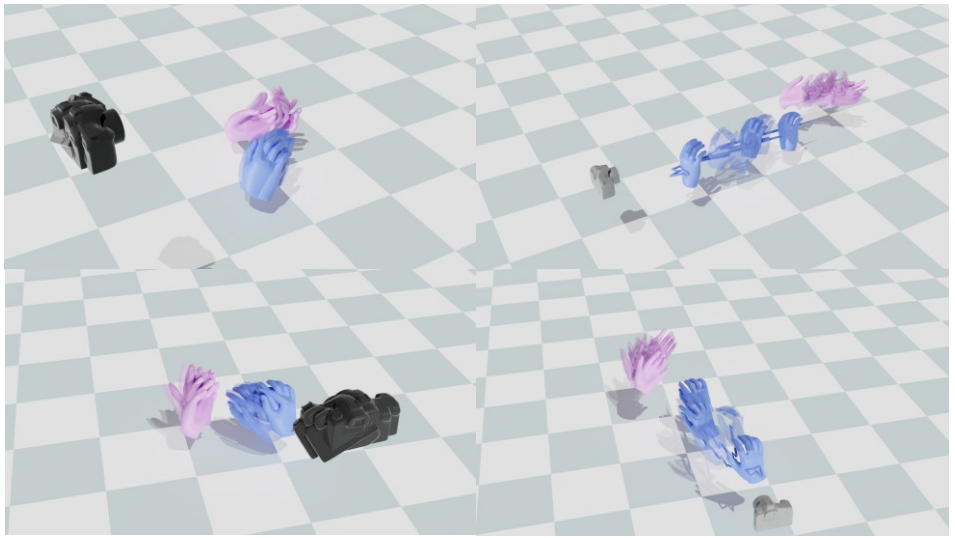
Qualitative results on FHPA

Input video

Ours

HaMeR

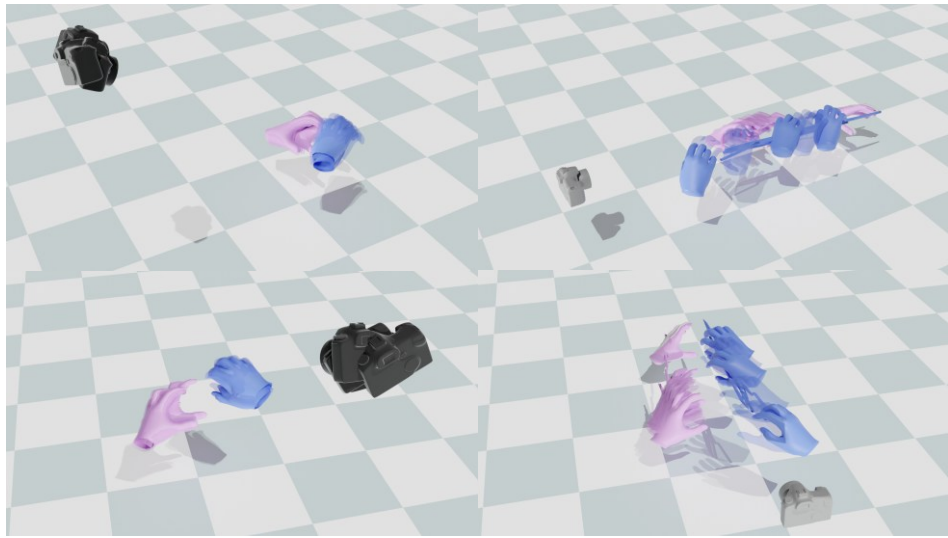




Input video

Ours

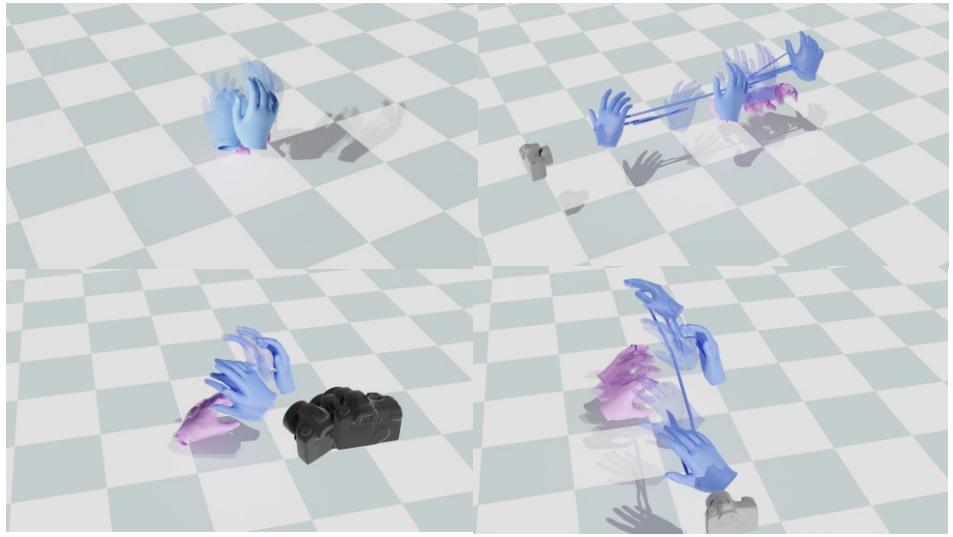
HaMeR



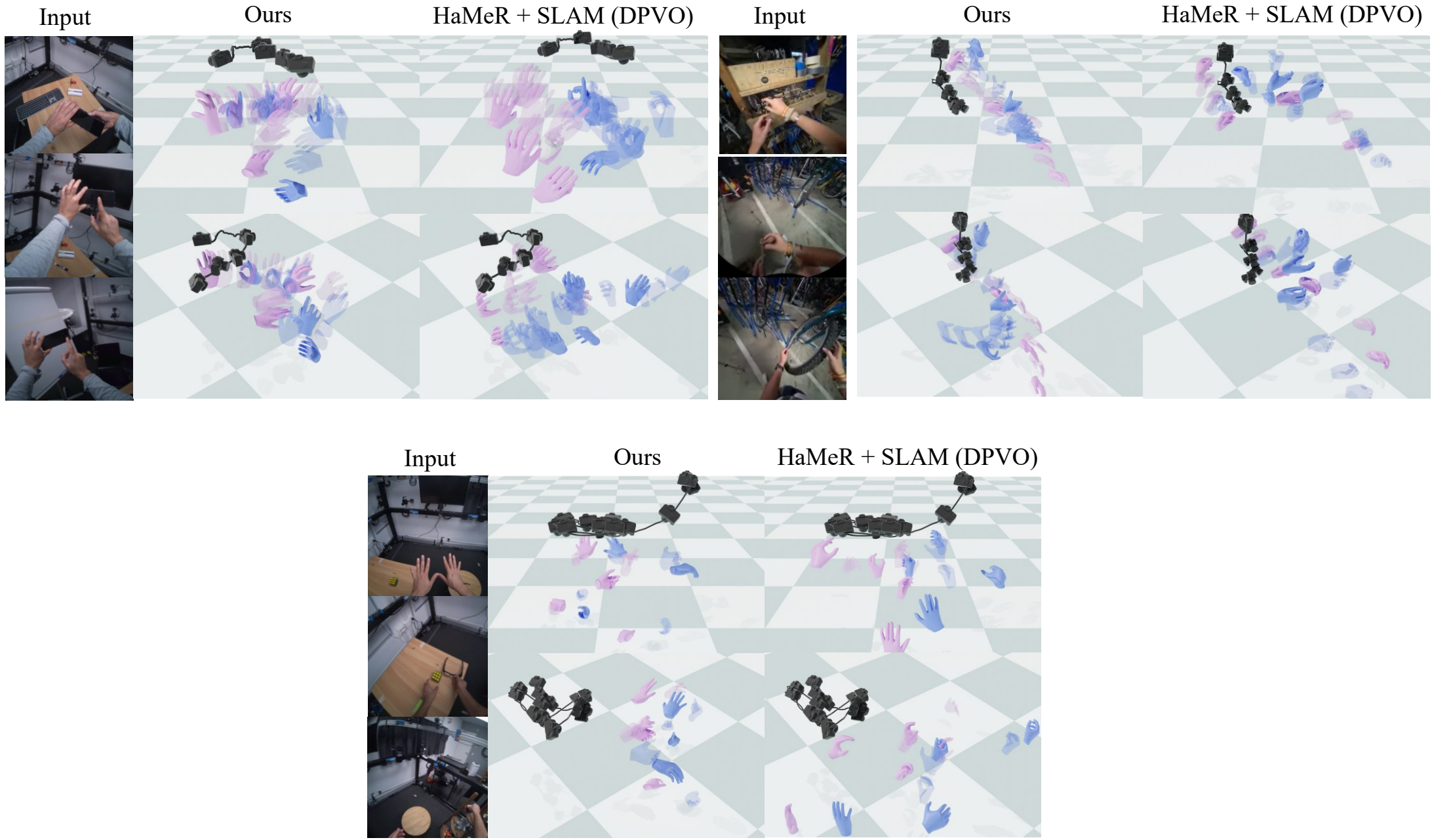
Input video

Ours

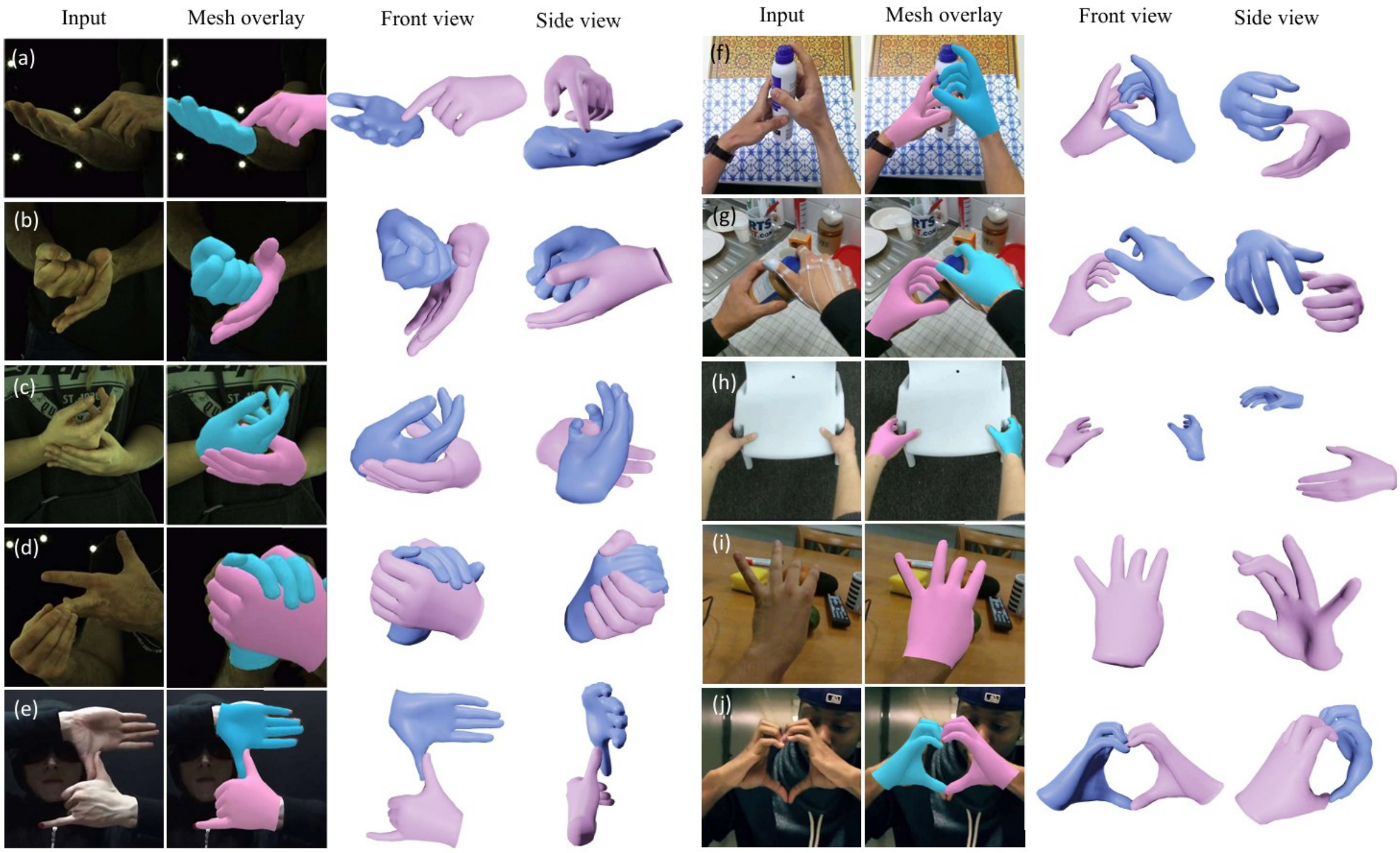
HaMeR

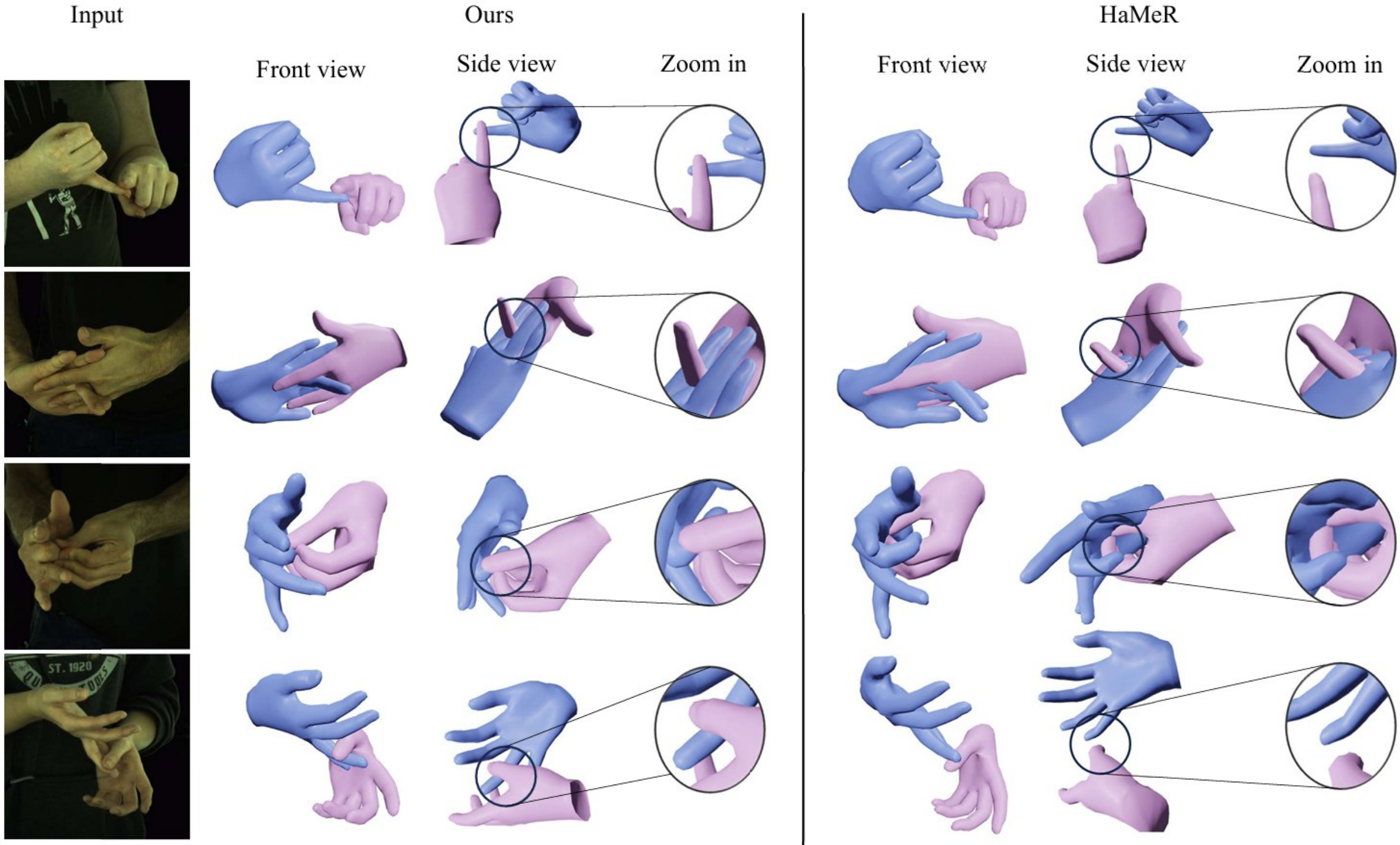


Qualitative results on HOT3D


Input
Ours
HaMeR + SLAM (DPVO)
Input
Ours
HaMeR + SLAM (DPVO)
Input
Ours
HaMeR + SLAM (DPVO)

Interacting hands reconstruction


Input
Mesh overlay
Front view
Side view
Input
Mesh overlay
Front view
Side view
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)


Input
Ours
HaMeR
Front view
Side view
Zoom in
Front view
Side view
Zoom in

Quantitative evaluation

Table 1. **Quantitative evaluation results for InterHand2.6M [35] 30 fps dataset.** We compare our method with the state-of-the-art hand reconstruction methods on local hand poses.

Method	MPJPE ↓	MPVPE ↓	Acc Err ↓
InterWild [32]	12.35	13.45	6.68
DIR [42]	9.09	9.43	8.92
ACR [54]	8.75	9.01	3.99
IntagHand [26]	9.26	9.71	4.41
HaMeR [38]	9.84	10.13	5.13
Ours (w/o III)	8.98	9.25	4.72
Ours (Dyn-HaMR)	**7.94**	**8.15**	**2.76**

Table 2. **Quantitative evaluation results for H2O [22] dataset**. Our method demonstrates significant improvements over state-of-the-art approaches in recovering both local and global 4D hand motion, with additional gains achieved incorporating Stage III.

Method	G-MPJPE ↓	GA-MPJPE ↓	MPJPE ↓	Acc Err ↓
ACR [54]	113.6	88.5	46.8	14.3
IntagHand [26]	105.5	81.5	45.6	13.5
HaMeR [38]	96.9	75.7	32.9	9.21
Ours (w/o III)	51.9	41.2	24.9	9.5
Ours (Dyn-HaMR)	**45.6**	**34.2**	**22.5**	**4.2**

Table 3. **Quantitative evaluation results for HOI4D [30] dataset.** We compare our method with the state-of-the-art hand reconstruction methods [26, 38, 54].

Method	G-MPJPE ↓	GA-MPJPE ↓	MPJPE ↓	Acc Err ↓
ACR [54]	251.1	153.5	36.4	12.5
IntagHand [26]	291.3	145.6	40.9	14.1
HaMeR [38]	201.6	129.7	27.6	11.6
Ours (w/o III)	69.2	48.5	23.7	10.9
Ours (Dyn-HaMR)	**58.5**	45.6	**19.5**	**4.1**

Table 4. **Quantitative evaluation Results on Ego-Exo4D [14] dataset.** We only report the positional error metrics as the GT annotation is too discrete to extract a meaningful acc.

Method	Jerk ↓	PA-MPJPE ↓	G-MPJPE ↓
ACR	245.12	19.45	291.34
IntagHand	175.65	24.32	275.89
HaMeR	195.45	15.87	267.75
HaMeR + DPVO	200.45	15.87	213.75
Ours (Dyn-HaMR)	**5.26**	**14.34**	**53.89**

Table 5. **Qualitative results on HOT3D [3] dataset**. We compare our method with the canonical HaMeR + DPVO baseline.

Method	Jerk ↓	PA-MPJPE ↓	G-MPJPE ↓	Acc Err ↓
ACR	153.45	16.41	159.34	16.45
IntagHand	171.24	21.75	165.42	15.12
HaMeR	189.62	10.43	155.98	13.78
HaMeR + DPVO	195.77	10.43	129.45	12.78
Ours (Dyn-HaMR)	**4.18**	**8.87**	**42.36**	**4.95**

Table 6. **Ablation of pipeline components on H2O [22] dataset.** It shows the impact of removing different components from the pipeline on various performance metrics.

Method	G-MPJPE ↓	GA-MPJPE ↓	MPJPE ↓	Acc Err ↓
Stage I	84.5	72.5	25.6	8.8
Stage I+II	51.9	41.2	24.9	9.5
w/o bio. const.	49.6	43.1	24.5	4.3
w/o pen. const.	46.3	34.7	23.6	**4.1**
w/o gen. infill.	48.9	37.8	24.1	5.6
Ours (Dyn-HaMR)	**45.6**	**34.2**	**22.5**	4.2

In-the-wild reconstruction and analysis

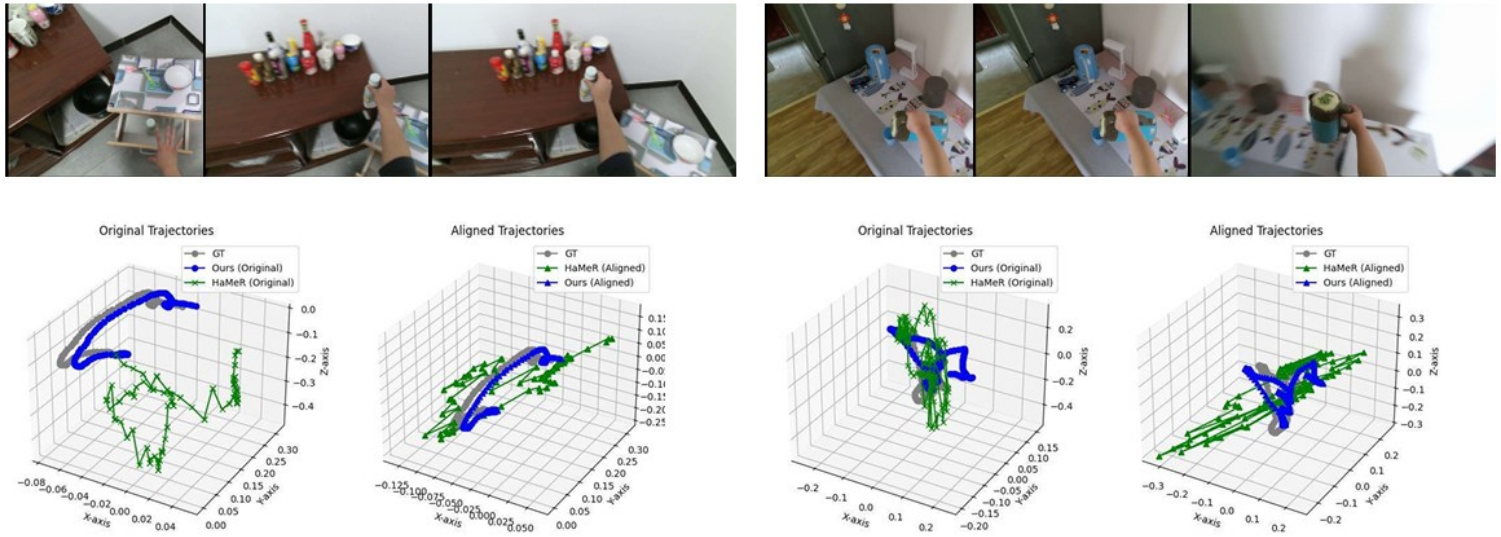


Figure 6. **Comparison of global trajectory on HOI4D** [30].

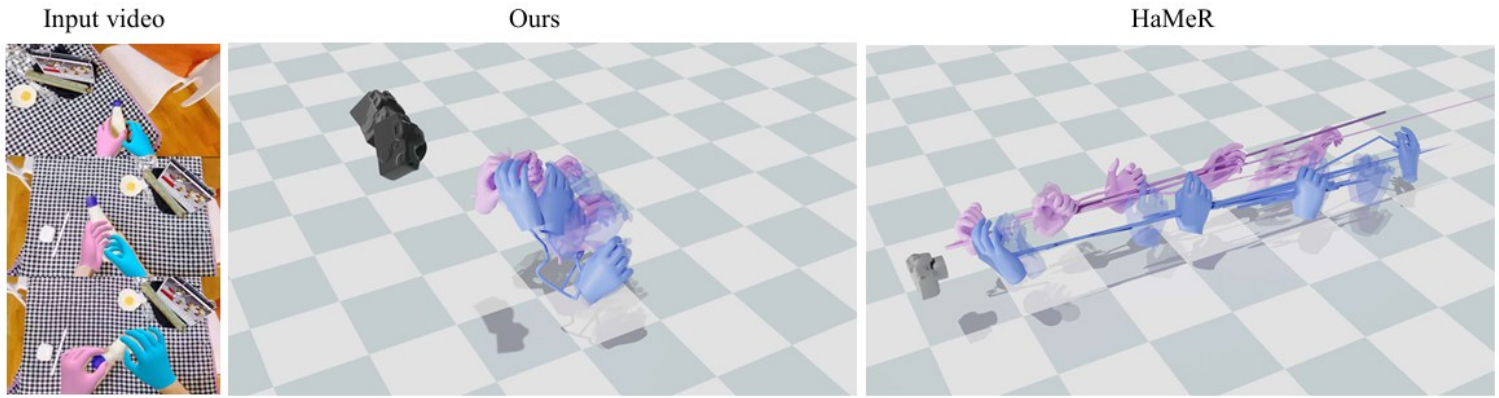


Figure 7. **Qualitative comparison with state-of-the-art method HaMeR [61] on in-the-wild online videos.**