XR-MBT: Multi-modal Full Body Tracking for XR through Self-Supervision with Learned Depth Point Cloud Registration - Supplementary Material

Denys Rozumnyi Nadine Bertsch Othman Sbai Filippo Arcadu Yuhua Chen
Artsiom Sanakoyeu Manoj Kumar Catherine Herold Robin Kips

Meta Reality Labs Zurich

1. Extended ablation study results

Table 1 shows an extended version of the ablation study from Table 1 of the main paper. The results on motions with body parts well covered by the depth signal show the benefits of training the MPE network to grasp and leverage semantic information. On the one hand, this is achieved by training an SPC decoder to predict semantics and by adding our novel SPC-loss. However, note that only part of the sequences contains active motions and that our vanilla MPE solution is sufficient for more common and symmetric motion categories, such as bare walking already. The previous state-of-the-art AGRoL is outperformed in all cases.

2. Qualitative results

This section provides qualitative results in Fig. 3 and Fig. 2. Furthermore, we have attached a video in the Supplementary Material.

3. Failure cases

Failure cases occur when lower body motion happens outside the field of view of the sensor, such as in Fig. 1.

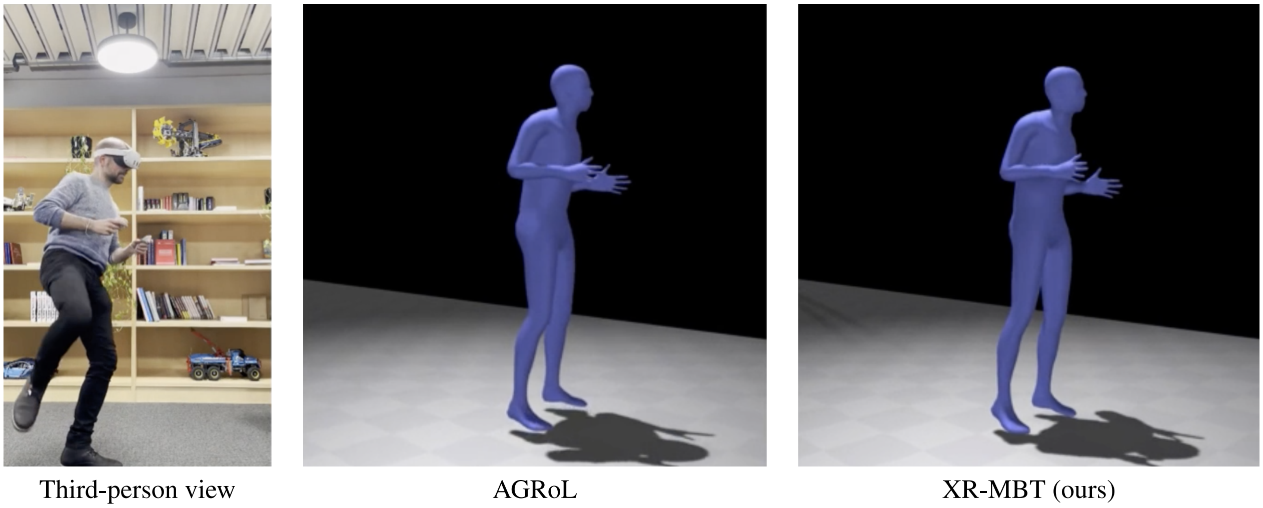

Figure 1. A failure in case of limited field of view, *i.e*. the lifted right leg is not visible to the XR device.

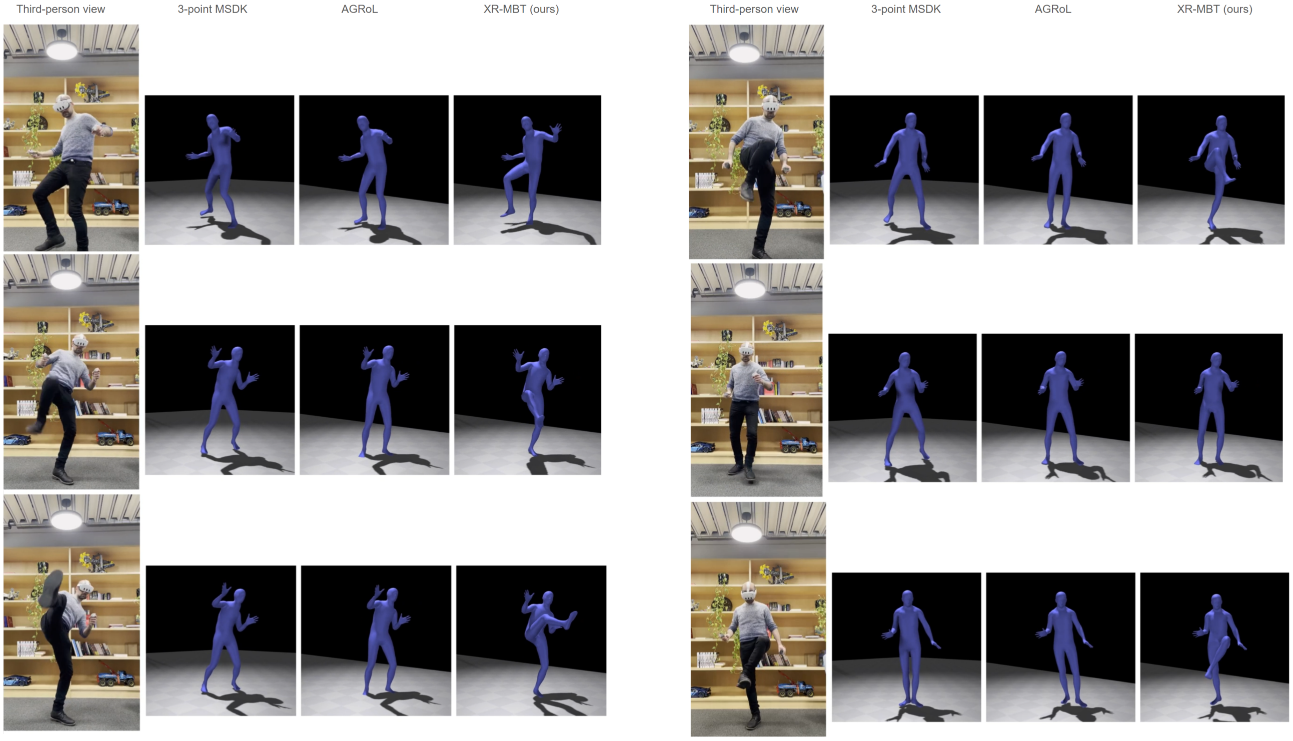

Figure 2. Additional results on real data.

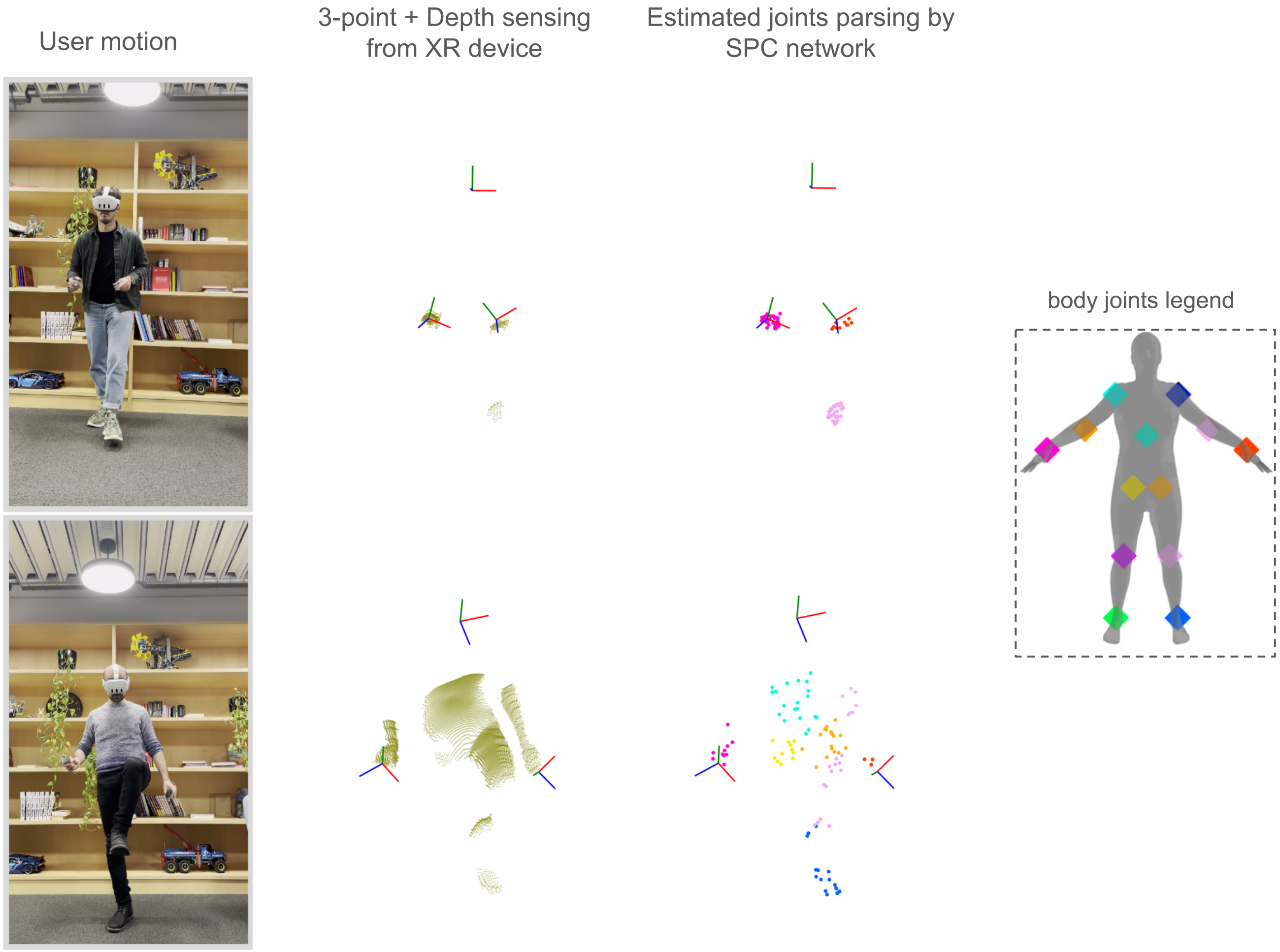
User motion
3-point + Depth sensing from XR device
Estimated joints parsing by SPC network
body joints legend

Figure 3. Additional point cloud labeling results.

Model	MPJPE (cm)		MPJRE (radians)		MPJVE (m^2/s^3)		jitter (pred/gt)		PC-loss
	up	low	up	low	up	low	up	low	
All Motions									
AGRoL (retrained)	1.87	11.21	1.79	2.30	10.27	34.36	**2.55**	**6.19**	0.61
MPE	1.78	9.62	1.75	2.23	9.94	31.89	3.22	12.66	0.39
MPE + SPC-decoder	1.77	9.30	1.75	**2.21**	9.95	32.41	3.11	12.98	0.34
MPE + SPC-decoder + PC-loss	1.78	10.25	1.76	2.25	9.96	32.97	2.84	9.43	0.48
MPE + SPC-decoder + SPC-loss	**1.76**	**9.27**	**1.74**	**2.21**	**9.78**	**31.40**	2.88	9.47	**0.32**
Kicking									
AGRoL (retrained)	1.58	10.21	1.63	2.05	10.28	39.22	**1.87**	**3.32**	0.80
MPE	1.49	8.27	1.61	2.01	10.11	36.56	2.31	7.6	0.39
MPE + SPC-decoder	**1.48**	7.97	1.61	1.99	10.25	37.77	2.23	7.88	0.33
MPE + SPC-decoder + PC-loss	1.50	8.91	1.62	2.02	10.26	38.98	2.08	5.70	0.51
MPE + SPC-decoder + SPC-loss	**1.48**	**7.78**	**1.59**	**1.98**	**10.11**	**36.42**	2.18	5.93	**0.28**
Knee strikes									
AGRoL (retrained)	1.63	10.67	1.58	2.03	9.97	39.97	**1.91**	**3.34**	0.32
MPE	1.57	**9.36**	1.55	1.95	**9.83**	**38.56**	2.28	7.81	**0.25**
MPE + SPC-decoder	**1.51**	9.38	**1.54**	**1.93**	9.89	38.79	2.28	7.81	0.30
MPE + SPC-decoder + PC-loss	1.53	10.16	1.56	2.00	9.91	40.68	2.21	5.96	0.33
MPE + SPC-decoder + SPC-loss	1.57	**9.36**	1.57	1.95	**9.83**	**38.56**	2.30	7.18	**0.25**
Elbow knee strikes									
AGRoL (retrained)	2.14	11.82	1.87	2.29	2.09	5.54	**1.56**	**2.50**	0.35
MPE	2.09	10.27	1.85	**2.23**	**2.08**	**5.35**	1.75	5.27	**0.24**
MPE + SPC-decoder	**2.06**	10.38	1.82	2.37	2.10	5.54	1.80	5.86	0.26
MPE + SPC-decoder + PC-loss	2.08	11.36	1.83	2.41	2.10	5.64	1.66	4.32	0.35
MPE + SPC-decoder + SPC-loss	2.09	**10.17**	**1.83**	2.34	**2.08**	**5.35**	1.75	5.27	0.26
Walking									
AGRoL (retrained)	1.33	5.77	1.45	1.63	7.41	23.65	**1.79**	**2.41**	0.82
MPE	1.27	**5.42**	1.44	**1.59**	**7.21**	**23.19**	2.09	4.80	**0.62**
MPE + SPC-decoder	**1.22**	5.61	**1.42**	1.64	7.36	24.00	2.19	6.10	0.65
MPE + SPC-decoder + PC-loss	1.23	5.82	1.43	1.66	7.35	24.25	2.00	4.07	0.74
MPE + SPC-decoder + SPC-loss	1.25	5.44	1.45	1.70	7.22	**23.19**	2.09	4.80	**0.62**

Table 1. Ablation study on the Mocap data test set with synthetically generated point clouds. Metrics are reported separately for the upper (up) and lower (low) body.