Supplimentary Material
MV2MP: Segmentation Free Performance Capture of Humans in Direct Physical Contact from Sparse Multi-Cam Setups

Sergei Eliseev[1], Leonid Shtanko[2], Rasim Akhunzianov[2], Yaroslav Romanenko[2], and Anatoly Starostin[1]

[1] Yango Israel, Tel-Aviv, Israel
[2] Yandex, Moscow, Russia
{serg-turbo, leonsht, rakhunzy, yromko, starost}@yandex-team.ru

1 Other results

We additionally evaluate our method using several metrics:

1. Mean Per Joint Position Error (MPJPE, mm): Evaluates the accuracy of SMPL joints.
2. Intersection over Union (IoU): Measures the accuracy of person semantic masks generated by our method compared to those produced by Mask2Former on the HI4D dataset.

Table 1 provides a comparison of MPJPE of SMPL joints before and after applying our method across different scenes. Our method generally shows an improvement in MPJPE after optimization, indicating enhanced accuracy in joint position estimation.

We also compared the Intersection over Union (IoU) between the semantic masks generated by Mask2Former and our method. Table 2 details the IoU results across different scenes. Our method consistently achieves high IoU scores, comparable to Mask2Former, demonstrating the effectiveness of our approach in generating accurate semantic masks.

Table 1: Comparison of Mean Per Joint Position Error (MPJPE) before and after applying our method, across different scenes and training views.

Scene	**# Train Views**	**MPJPE, Before (mm)**	**MPJPE, After (mm)**
hug21	7	72.9	73.3
yoga00	7	45.2	39.8
sidehug32	7	54.9	49.5

Table 2: Comparison of Intersection over Union (IoU) between Mask2Former and our method across different scenes and training views.

Scene	# Train Views	IoU, Mask2Former	IoU, Our Method
hug21	7	0.94	0.937
yoga00	7	0.931	0.934
sidehug32	7	0.942	0.928

Table 3: Train/Validation Splits for HI4D and CMU Panoptic Datasets

Dataset	Validation Cameras	Training Cameras
hi4d_pair21_hug21	4	52, 28, 40, 64, 16, 76, 88
hi4d_pair21_hug21	4, 52, 40	28, 64, 16, 76, 88
hi4d_pair21_hug21	4, 52, 40	28, 64, 88
hi4d_pair00_yoga00	16	52, 28, 40, 64, 4, 76, 88
hi4d_pair00_yoga00	16, 52, 64	28, 40, 4, 76, 88
hi4d_pair00_yoga00	16, 52, 64	28, 40, 88
hi4d_pair32_sidehug32	8	32, 56, 72, 96, 12, 18, 90
hi4d_pair32_sidehug32	8, 56, 72	32, 96, 12, 18, 90
hi4d_pair32_sidehug32	8, 56, 72	32, 96, 90

2 Data splits

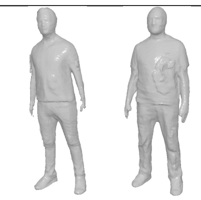
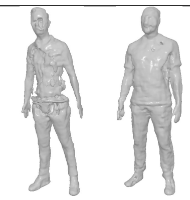
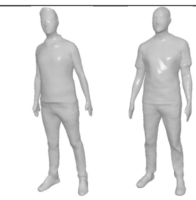
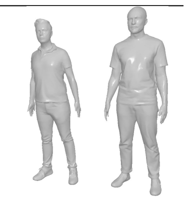
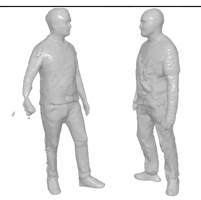
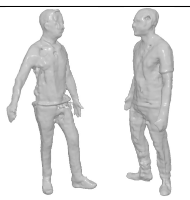
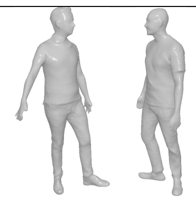
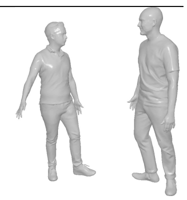
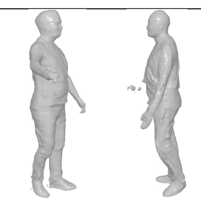
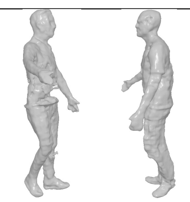
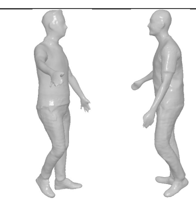
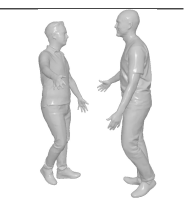
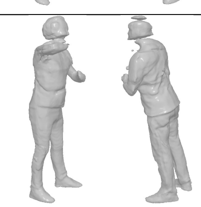
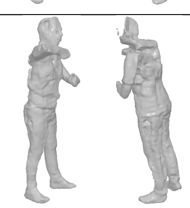
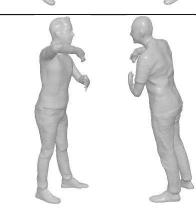
In Table 3 we define specific train and validation splits for the HI4D and CMU Panoptic datasets used to evaluate our method's performance under varying conditions of camera sparsity.

3 MultiPly limitations

We show that MultiPly novel view synthesis is inherently limited by the single training camera and seen sides of the subject as shown in Fig. 1. Our method explicitly targets multi-view reconstruction.

Fig. 1: Novel view synthesis of MultiPly under extreme camera angles. Train view 4, render view 28 and 76.

Table 4: Comparison of Image Outputs for Different Methods

Frame	DMC	MNB	Ours	GT
11				
31				
41				
71				
91