Pacific Graphics 2023
R. Chaine, Z. Deng, and M. H. Kim
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 42 (2023), Number 7

Reconstructing 3D Human Pose from RGB-D Data with Occlusions (Supplementary Material)

Bowen Dang[1] Xi Zhao[†1] Bowen Zhang[1] He Wang[2]

[1]Xi'an Jiaotong University, China
[2]University College London, England

1. Computational Efficiency

Running time for all configurations is listed in Table 1. We can see that our method achieves better performance than PROX-D with less running time. Besides, volume matching outperforms surface matching both in time efficiency and performance. We also calculated the inference time of the free zone network and found it costs about 0.5 seconds, which means that the bottleneck lies in the optimization process.

Table 1: *Running time for all configurations.*

	E_p	E_c	E_{fz}	E_{tsv}	VM	Running time / s
SMPLify-D						36.63
PROX-D	✓	✓				52.13
Ours (w/o FZ)		✓		✓	✓	47.79
Ours (w/o TSV)		✓	✓		✓	48.87
Ours (w/o VM)		✓	✓	✓		51.49
Ours		✓	✓	✓	✓	**50.76**

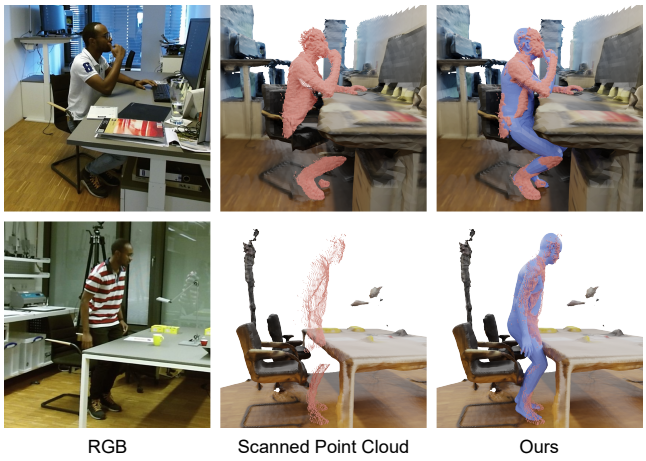

Figure 1: *Additional results on cases occluded in the middle.*

2. Performance on Difficult Data

Cases Occluded in the Middle: In some situations, the human body might be divided into multiple parts due to occlusions. However, we do not perform any special treatments for these cases because we match the body with the confined region holistically. We find it does not affect the performance of our method. Figure 1 presents two representative examples of these cases. In these examples, the scanned body point cloud is divided into the upper part and lower part by the table. Our method can still reconstruct the accurate and plausible pose where the body part matches with the corresponding scanned point cloud regions correctly.

Cases with Loose and Large Clothes: We selected some samples where the human subjects are wearing jackets or coats and observed that our method can still perform well in these cases. Figure 2 presents two representative examples of these cases. When the human is wearing a jacket or a coat, our method can reconstruct the correct pose while aligning the shape with the scanned point cloud. It is worth noting that the body shape might seem a little fatter than the real body shape because SMPL-X model is trained on data of humans wearing tight clothes. To better handle such cases, using a clothed human body model might be a more suitable solution.

3. Failure Cases

Error Brought by Discretization: Discretization might bring some errors, especially in non-convex parts like hands, because the interpolated points can be wrongly located outside the body surface and cause inaccurate matching. In non-convex parts, the two vertices in the same interpolation vertex pair might come from different convex blocks, so the interpolated points might go out of the body, causing the part to lose the original pose and shape. Take the hand as an example, interpolated points of two vertices from different fingers might appear in the finger gaps, making it hard to match the hand with the scanned point cloud accurately. However, these errors can be ignored in our system because the non-convex parts only occupy a small ratio of the whole body.

Cases with Poor Scanned Point Cloud: Our method might fail when the quality of the scanned body point cloud is poor, especially when the body segmentation mask is inaccurate or there ex-

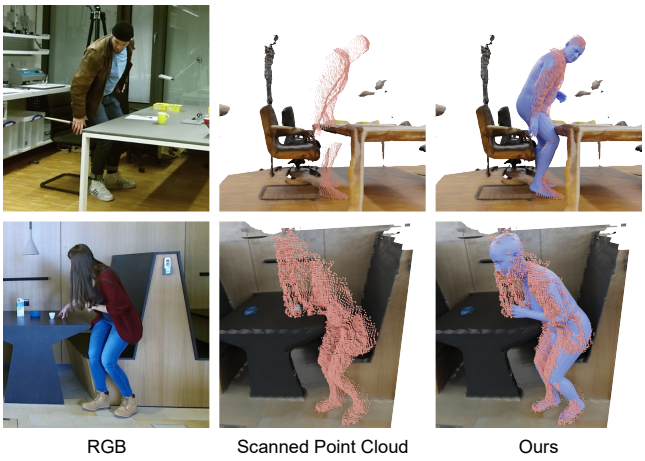

Figure 2: *Additional results on cases with loose and large clothes.*

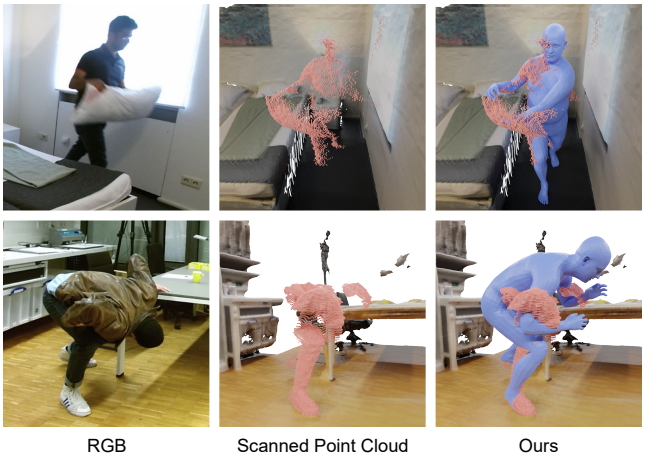

Figure 3: *Additional results of failure cases.*

ists complex self-occlusion. Figure 3 presents two failure cases of our method. In the first example, when the human is holding a pillow, and the pillow is wrongly recognized as part of the body, our method produces an incorrect pose where the left hand penetrates into the pillow. In the second example, although the human is not occluded by the scene, the scanned point cloud is incomplete due to the wrong body segmentation mask and complex self-occlusion. Our method fails to reconstruct the accurate pose that matches with the input.