NeuralHOFusion: Neural Volumetric Rendering under Human-object Interactions Supplementary Materials

Yuheng Jiang[1] Suyi Jiang[1] Guoxing Sun[1] Zhuo Su[2] Kaiwen Guo[3]
Minye Wu[4] Jingyi Yu[1,5] Lan Xu[1,5]

[1]ShanghaiTech University [2]Tencent [3]Meta Reasearch Lab [4]KU leuven
[5]Shanghai Engineering Research Center of Intelligent Vision and Imaging

Table 1. Quantitative comparison of rendering results.

Method	PSNR↑	SSIM↑	MAE ↓
Function4d [1]	29.930	0.981	0.914
Ours	**32.382**	**0.983**	**0.736**

1. Additional Comparison

We further compare our NeuralFusion with the state-of-the-art volumetric capture approach Function4d [1] under our six RGBD system. Please note that we provided the data to the original authors, and the comparison results come from them. Fig. 1 illustrates that our NeuralFusion produces more detailed and complete geometries and more sharp textures, especially under the human-object interaction scenarios. Thanks to the normal refinement and neural blending, NeuralFusion could capture and render fine-grained details of hands in Fig. 1 (a), human body Fig. 1 in (b). When processing human-object interactions Fig. 1 (b,c), our method helps to handle occlusions. Specially, our neural human reconstruction with long term information from key volume features is more robust to severe occlusions comparing to the Function4d with short term information from sliding fusion. Besides, the object geometries captured by our template-aid object fusion is more appealing. The quantitative result in Tab. 1 also demonstrates that our approach is better.

We then compare our NeuralFusion with Function4d [1] on sequences of human-object interactions. As illustrated in Fig. 2, Fig. 3, Fig. 4, Function4d [1] has not taken human-object interactions into consideration while our method adopts a layer-wise neural blending scheme to disentangle human and object for photo-realistic performance rendering.

References

[1] Tao Yu, Zerong Zheng, Kaiwen Guo, Pengpeng Liu, Qionghai Dai, and Yebin Liu. Function4d: Real-time human volumetric capture from very sparse consumer rgbd sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5746–5756, 2021. 1, 2

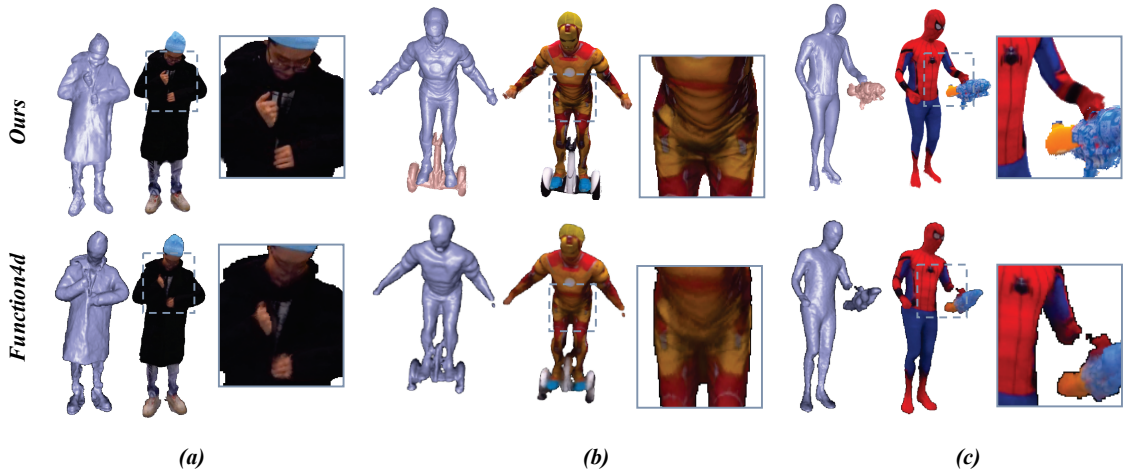

Figure 1. Qualitative comparison of various challenging scenarios. The first line is our result, the second line is the result of Function4d [1].

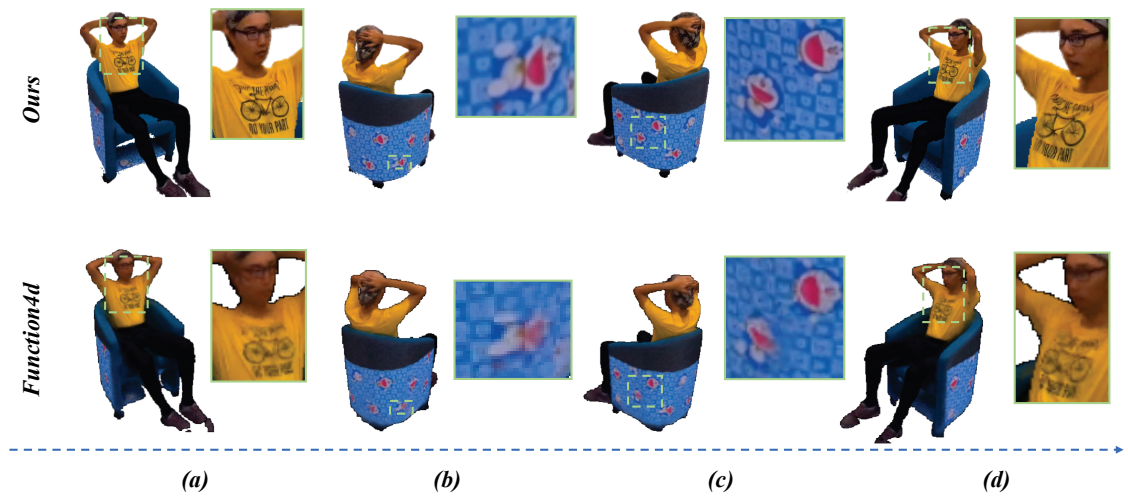

Figure 2. Qualitative comparison on sequence of human-object interaction.The first line is our result, the second line is the result of Function4d [1].

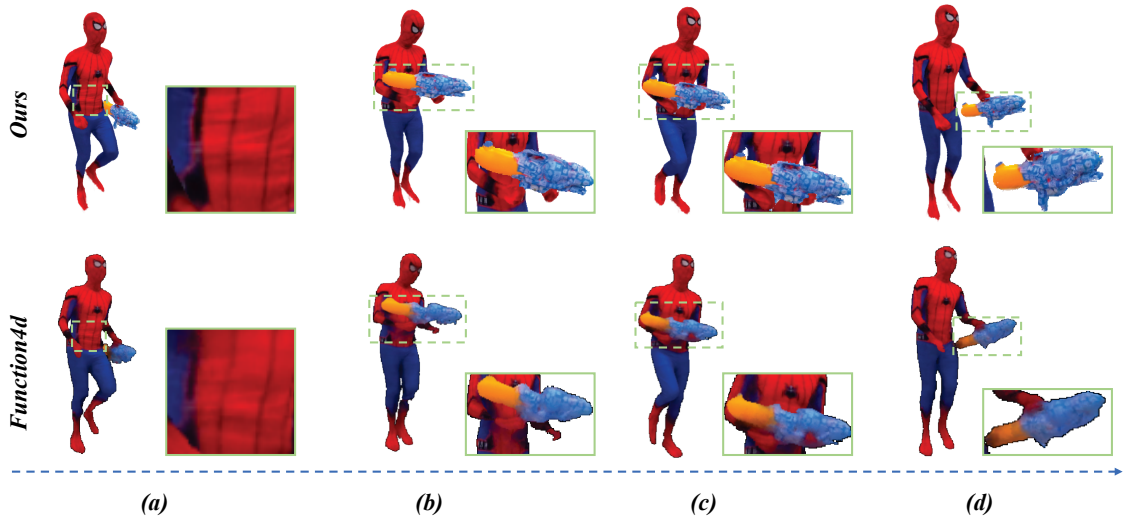

Figure 3. Qualitative comparison on sequence of human-object interaction.

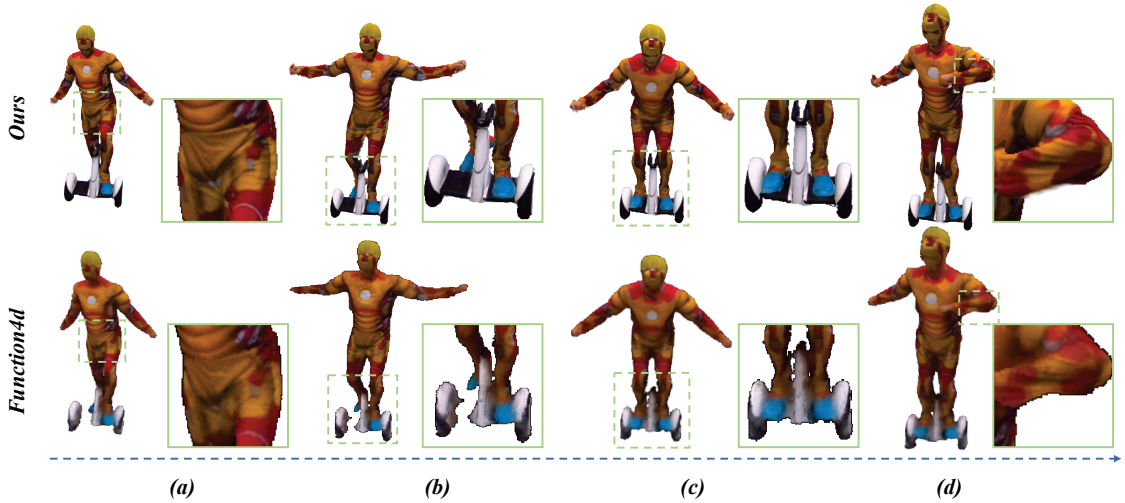

Figure 4. Qualitative comparison on sequence of human-object interaction.