Supplementary Material of "RobustFusion: Human Volumetric Capture with Data-driven Visual Cues using a RGBD Camera"

Zhuo Su[1*], Lan Xu[1,2*], Zerong Zheng[1], Tao Yu[1], Yebin Liu[1], and Lu Fang[1†]

[1] Tsinghua University
[2] ShanghaiTech University

1 More Experiment

In this section, we provide more experiment to demonstrate the effectiveness of our RobustFusion system, including the comparisons to other RGB or RGBD based off-line methods as well as the evaluation using a sequence from another public dataset.

Comparison with RGB-based off-line methods. Recall that in our model completion stage, a data-driven implicit occupancy network is utilized to generate the water-tight initial geometry model. For further evaluation, we compare our RobustFusion against the state-of-the-art per-frame pixel-aligned implicit approaches PIFu [2] and PIFuHD [3]. Fig. 1 illustrates the qualitative results on various frames with challenging fast motions. Note that these RGB-based approaches suffer from complicated poses and depth ambiguity. In contrast, our approach achieves significantly better and temporally-coherent geometry reconstruction results especially for challenging human motions, with the inherent advantage of performance capture and the aid of various visual priors.

Comparison with RGBD-based off-line methods. For the comparison to RGBD-based methods, we utilize the Poisson reconstruction [1] to blend the tracked initial model and current depth input for each live frame. Note that for fair comparison, we use the tracking pose parameters from our method in this per-frame Poisson reconstruction baseline. As shown in Fig. 2, such per-frame Poisson reconstruction suffers from severe noise from the raw depth input and cannot recover fine geometry details. In contrast, the proposed RobustFusion based on an effective fusion strategy can handle challenging motions and recover temporally-coherent geometry details. Furthermore, we compare to the variation using Poisson reconstruction to blend the partial TSDF and the learned complete mesh in our model completion stage. For fair comparison, the Poisson reconstruction is applied after the same aligning the partial geometry and the learned complete mesh. As illustrated in Fig. 3, without inherent seamless

* Equal Contribution. † Corresponding Author.

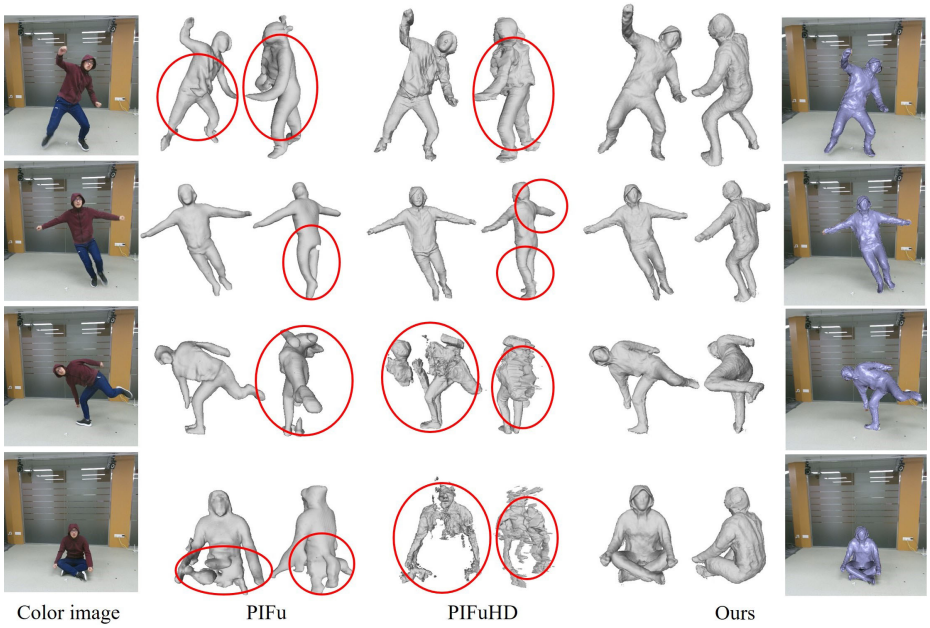

Fig. 1. Qualitative comparison against per-frame RGB based methods including PIFu [2] and PIFuHD [3].

blending advantage of our TSDF-based strategy, Poisson reconstruction suffers from coarse surface without any clothes wrinkles and facial details.

In addition, note that all these off-line methods require two to four orders of magnitude more time than our RobustFusion that is available for daily usages.

Reconstruction results on the data from KillingFusion [4]. To demonstrate our robustness on diverse input, we also run our volumetric capture algorithm on a single noisy Kinect depth stream of KillingFusion [4]. Due to the lack of complete human detected in the input image stream, we don't utilize the model completion stage in our approach. Nevertheless, as shown in Fig 4, our approach yields to superior tracking results especially for capturing the temporally smoother geometry details in the face and wrinkles regions.

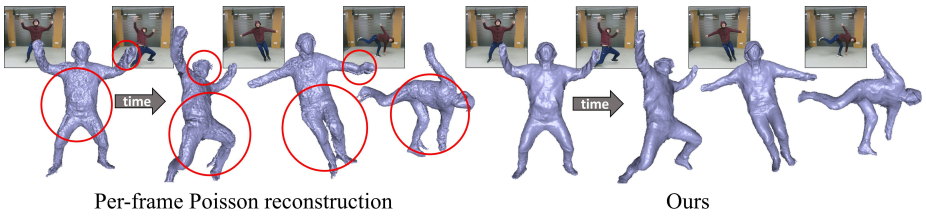

Fig. 2. Qualitative comparison against per-frame Poisson reconstruction [1] for the performance capture stage.

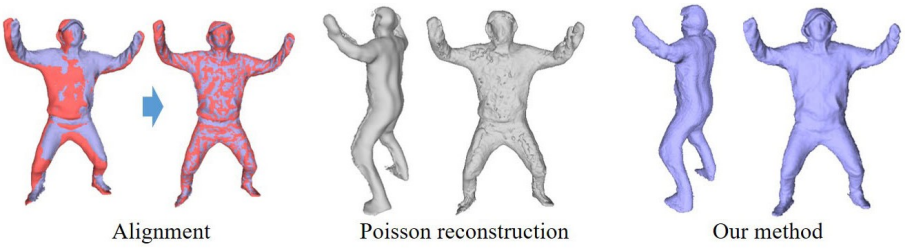

Fig. 3. Qualitative comparison against Poisson reconstruction [1] for the model completion stage.

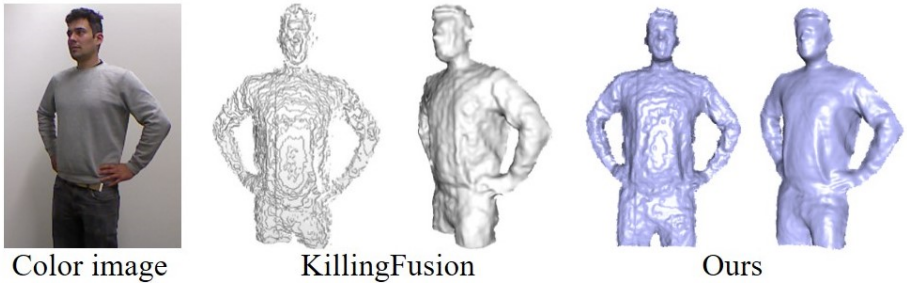

Fig. 4. Reconstruction results on the first and a middle frames from the "Alex" sequence of the public dataset from KillingFusion [4].

References

1. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Proceedings of the Fourth Eurographics Symposium on Geometry Processing. p. 61–70. SGP '06, Eurographics Association, Goslar, DEU (2006)
2. Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., Li, H.: Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In: The IEEE International Conference on Computer Vision (ICCV) (October 2019)
3. Saito, S., Simon, T., Saragih, J., Joo, H.: Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (June 2020)
4. Slavcheva, M., Baust, M., Cremers, D., Ilic, S.: KillingFusion: Non-rigid 3D Reconstruction without Correspondences. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)