BNV-Fusion: Dense 3D Reconstruction using Bi-level Neural Volume Fusion

Kejie Li¹, Yansong Tang¹,², Victor Adrian Prisacariu¹, Philip H.S. Torr¹
¹University of Oxford, ²Tsinghua-Berkeley Shenzhen Institute, Tsinghua University

Figure 1. Reconstructions from different methods of two augmented ICL-NUIM sequences. Our method is less affected by noisy depth measurements (e.g. flat walls) and is able to reconstruct fine details (e.g. cables in the office1 and branches in livingroom1) better than other methods.

1. Qualitative Comparisons

Learning-based methods, iMAP [3] and DI-Fusion [2], fail to reconstruct faithfully as shown in Fig. 1. Although TSDF-Fusion [4] is able to capture more details in the scene than these learning-based methods, it is susceptible to noise in depth measurements. As a result, the walls in both scenes are reconstructed roughly. Our method consistently produces reconstructions that are closer to the ground-truth models than other methods.

Fig. 2 visualizes reconstructions of 5 ScanNet sequences. Note that reconstructions in ScanNet are more challenging than the other two datasets due to camera drift in the poses provided. Nevertheless, BNV-Fusion is less affected by the noise in poses and depth measurements than TSDF-Fusion.

2. Video Demo

Readers are encouraged to watch the video that demonstrates a reconstruction of the “lounge” scene in the 3D scene dataset [1]. We refresh the reconstruction at every 10 frames.

References

Figure 2. Reconstructions of 5 sequences in ScanNet. Note that our reconstructions are generally smoother and more complete.