Robust Reconstruction of Indoor Scenes

Sungjoon Choi[1], Qian-Yi Zhou[2], Vladlen Koltun[2]
[1]Stanford University [2]Intel Labs

Abstract. We present an approach to indoor scene reconstruction from RGB-D video. The key idea is to combine geometric registration of scene fragments with robust global optimization based on line processes. Geometric registration is error-prone due to sensor noise, which leads to aliasing of geometric detail and inability to disambiguate different surfaces in the scene. The presented optimization approach disables erroneous geometric alignments even when they significantly outnumber correct ones. Experimental results demonstrate that the presented approach substantially increases the accuracy of reconstructed scene models.

Introduction. High-fidelity reconstruction of complete indoor scenes is known as a particularly challenging problem. The availability of consumer depth cameras provides an opportunity to develop robust reconstruction systems but does not in itself solve the associated challenges. While 3D models of real-world objects can now be created easily [3, 5], the same combination of quality and reliability has yet to be achieved for complete scenes. Unlike an object, which can be entirely in the field of view of the camera, a large scene must be reconstructed from views acquired along a complex trajectory, each view exposing only a small part of the environment. Camera paths that thoroughly image all surfaces at close range lead to significant odometry drift and the necessity to match and register different views globally.

Prior work on scene reconstruction with consumer depth cameras recognized the importance of global registration. Nevertheless, no prior system appears to be sufficiently reliable to support automatic reconstruction of complete indoor scenes at a quality level appropriate for particularly demanding applications. This is evidenced by the recent effort of Xiao et al. to reconstruct a large number of indoor scenes. Due to the unreliability of automatic scene reconstruction pipelines, the authors resorted to manual labeling to establish correspondences between different views. ("existing automatic reconstruction methods are not reliable enough for our purposes." [4])

In this work, we present a fully automatic scene reconstruction pipeline that matches the reconstruction quality obtained with manual assistance by Xiao et al. and significantly exceeds the accuracy of prior automatic approaches to indoor reconstruction. An example reconstruction produced by our approach is shown in Figure 1. Our pipeline is geometric: pairs of local scene fragments are registered and a global model is constructed based on these pairwise alignments. A critical weakness of such pipelines that we address is the low precision of geometric registration. Geometric registration algorithms are error-prone due to sensor noise, which leads to aliasing of fine geometric details and inability to disambiguate different locations based on local geometry. The difficulty is compounded by the necessity to register loop closure fragments that have low overlap. In practice, false pairwise alignments can outnumber correctly aligned pairs.

Our approach resolves inconsistencies and identifies correct alignments using global optimization based on line processes [1]. The advantage of the line process formulation is that the optimization objective retains a least-squares form and can be optimized by a standard high-performance least-squares solver. We show that this framework is extremely effective in dealing with pairwise registration errors. Our implementation automatically prunes false pairwise alignments even when they significantly outnumber correct ones. Extensive experiments demonstrate that our approach substantially increases reconstruction accuracy.

Our work contains a number of supporting contributions of independent interest. First, we provide infrastructure for rigorous evaluation of scene reconstruction accuracy, augmenting the ICL-NUIM dataset [2] with challenging camera trajectories and a realistic noise model. Second, we perform a thorough quantitative evaluation of surface registration algorithms in the context of scene reconstruction; our results indicate that well-known algorithms perform surprisingly poorly and that algorithms introduced in the last few years are outperformed by older approaches. Third, in addition to accuracy measurements on synthetic scenes we describe an experimental procedure for quantitative evaluation of reconstruction quality on real-world scenes in the absence of ground-truth data.

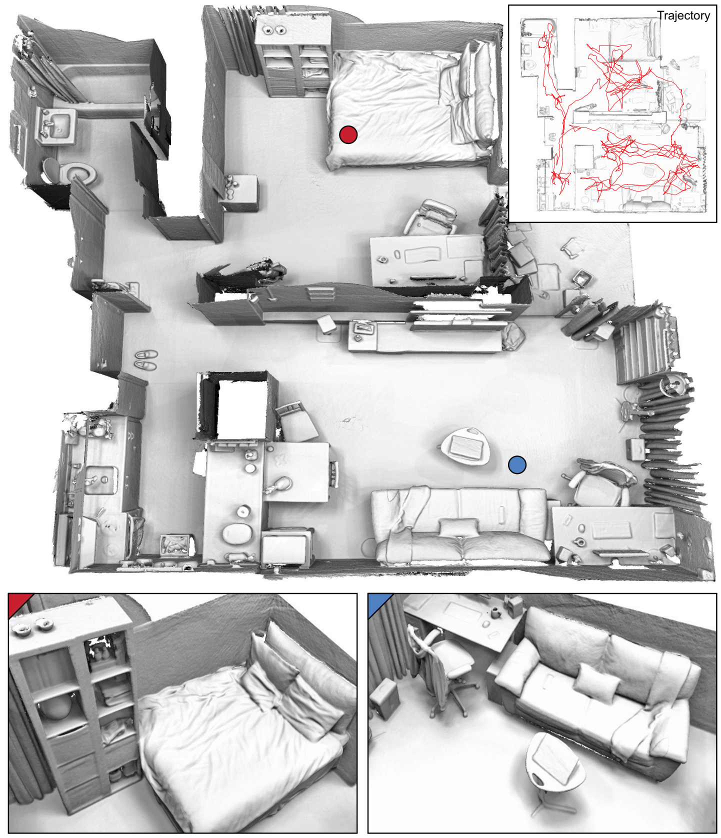

Figure 1: A complete apartment reconstructed by the presented approach.

[1] Michael J. Black and Anand Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *IJCV*, 19(1), 1996.

[2] Ankur Handa, Thomas Whelan, John McDonald, and Andrew J. Davison. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In *ICRA*, 2014.

[3] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *ISMAR*, 2011.

[4] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. SUN3D: A database of big spaces reconstructed using SfM and object labels. In *ICCV*, 2013.

[5] Qian-Yi Zhou and Vladlen Koltun. Color map optimization for 3D reconstruction with consumer depth cameras. *ACM Transactions on Graphics*, 33(4), 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.