CubeSLAM: Monocular 3-D Object SLAM

Shichao Yang, Sebastian Scherer

Abstract— In this paper, we present a method for single image three-dimensional (3-D) cuboid object detection and multiview object simultaneous localization and mapping in both static and dynamic environments, and demonstrate that the two parts can improve each other. First, for single image object detection, we generate high-quality cuboid proposals from two-dimensional (2-D) bounding boxes and vanishing points sampling. The proposals are further scored and selected based on the alignment with image edges. Second, multiview bundle adjustment with new object measurements is proposed to jointly optimize poses of cameras, objects, and points. Objects can provide long-range geometric and scale constraints to improve camera pose estimation and reduce monocular drift. Instead of treating dynamic regions as outliers, we utilize object representation and motion model constraints to improve the camera pose estimation. The 3-D detection experiments on SUN RGBD and KITTI show better accuracy and robustness over existing approaches. On the public TUM, KITTI odometry and our own collected datasets, our SLAM method achieves the state-of-the-art monocular camera pose estimation and at the same time, improves the 3-D object detection accuracy.