Depth Camera Tracking with Contour Cues

Qian-Yi Zhou, Vladlen Koltun
Intel Labs

Abstract. We present an approach for tracking camera pose in real time given a stream of depth images. Existing algorithms are prone to drift in the presence of smooth surfaces that destabilize geometric alignment. We show that useful contour cues can be extracted from noisy and incomplete depth input. These cues are used to establish correspondence constraints that carry information about scene geometry and constrain pose estimation. Despite ambiguities in the input, the presented contour constraints reliably improve tracking accuracy. Results on benchmark sequences and on additional challenging examples demonstrate the utility of contour cues for real-time camera pose estimation.

Introduction. Tracking self-motion is a primary function of visual perception in animals. In computer vision, the corresponding problem of visual odometry underlies a host of applications and has been extensively studied. Our work concerns depth cameras, which are increasingly utilized in computer vision systems. Our goal is to improve the accuracy of depth camera tracking, particularly in challenging scenarios that currently lead to odometry drift.

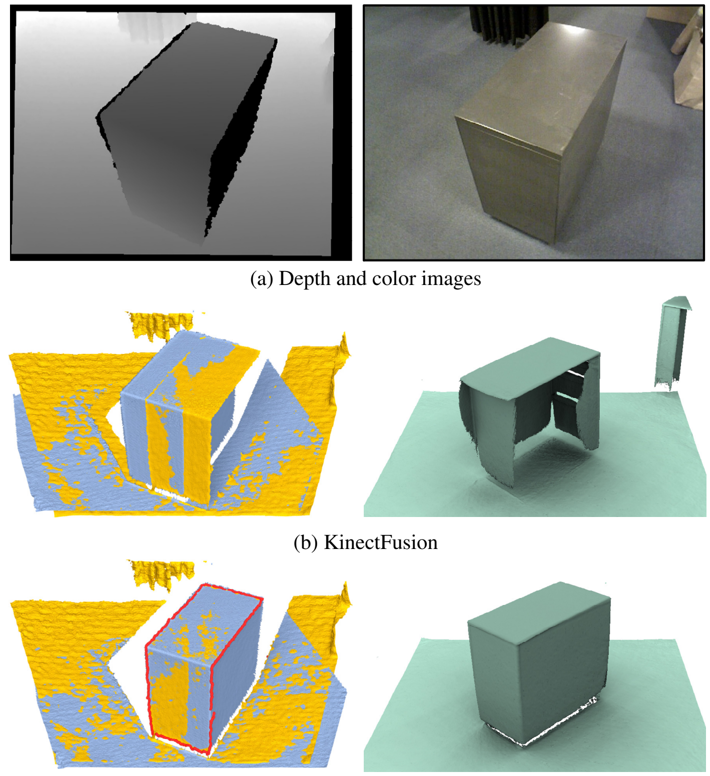
The influential KinectFusion system [3] demonstrated real-time depth camera tracking and dense scene reconstruction by registering incoming depth images to a volumetric representation of the scene. Our work extends these ideas by integrating occluding contours into the optimization objective and showing that explicit handling of contours can lead to significant gains in tracking accuracy. A different extension of the KinectFusion approach was developed by Bylow et al. [1], who derived a principled optimization algorithm but did not track occluding contours. Our experiments demonstrate that contour tracking has significant benefits.

A number of odometry systems use both depth and color images [2, 5]. Our work aims to maximize tracking accuracy without relying on a color image stream. One reason is that some depth cameras are not accompanied by color cameras. Another is that even if a color camera is present, its viewpoint is different and its shutter may not be perfectly synchronized with the depth camera. Finally, we aim for systems that function even in minimal lighting.

Our approach is based on tracking occluding contours and using contour cues to constrain registration. This addresses a common failure mode of geometric registration approaches based on the iterative closest point (ICP) algorithm and its variants, namely instability in the presence of smooth surfaces. This is illustrated in Figure 1, which shows a cabinet being imaged by a depth camera. In some aspects, the cabinet is seen as a collection of large planar surfaces that cause geometric alignment to slip and camera tracking to drift. This behavior can be easily observed in practice and is also apparent in benchmark odometry sequences [4]. Our solution integrates contour constraints into the registration objective, stabilizing camera tracking in challenging scenarios.

Occluding contours were considered a primary source of information in the early days of computer vision. They are now used in state-of-the-art multi-view stereo systems to inform shape reconstruction given calibrated camera parameters. Our work leverages contour cues in a real-time tracking system that operates on high frame-rate depth image streams.

Experimental results on challenging input sequences demonstrate that our formulation significantly improves depth camera tracking accuracy.



(a) Depth and color images

(b) KinectFusion

(c) Our approach

Figure 1: (a) Depth and color images from an input sequence. (b) Surface registration slips on planar surfaces, leading to tracking drift and reconstruction failure. (c) Our approach establishes contour constraints that stabilize real-time camera tracking (red). The color image (a, right) is shown for clarity and is not used by either approach.

[1] Erik Bylow, Jürgen Sturm, Christian Kerl, Fredrik Kahl, and Daniel Cremers. Real-time camera tracking and 3D reconstruction using signed distance functions. In *RSS*, 2013.

[2] Christian Kerl, Jürgen Sturm, and Daniel Cremers. Robust odometry estimation for RGB-D cameras. In *ICRA*, 2013.

[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] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In *IROS*, 2012.

[5] T. Whelan, H. Johannsson, M. Kaess, J.J. Leonard, and J.B. McDonald. Robust real-time visual odometry for dense RGB-D mapping. In *ICRA*, 2013.

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