MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

SLAM Using Both Points and Planes for Hand-Held 3D Sensors

Taguchi, Y.; Jian, Y-D; Ramalingam, S.; Feng, C.

TR2012-091 November 2012

Abstract

We present a simultaneous localization and mapping (SLAM) algorithm for a hand-held 3D sensor that uses both points and planes as primitives. Our algorithm uses any combination of three point/plane primitives (3 planes, 2 planes and 1 point, 1 plane and 2 points, and 3 points) in a RANSAC framework to efficiently compute the sensor pose. As the number of planes is significantly smaller than the number of points in typical 3D scenes, our RANSAC algorithm prefers primitive combinations involving more planes than points. In contrast to existing approaches that mainly use points for registration, our algorithm has the following advantages: (1) it enables faster correspondence search and registration due to the smaller number of plane primitives; (2) it produces plane based 3D models that are more compact than point-based ones; and (3) being a global registration algorithm, our approach does not suffer from local minima or any initialization problems. Our experiments demonstrate real-time, interactive 3D reconstruction of office spaces using a hand-held Kinect sensor.

International Symposium on Mixed and Augmented Reality (ISMAR)

SLAM Using Both Points and Planes for Hand-Held 3D Sensors

Yuichi Taguchi* Yong-Dian Jian† Srikumar Ramalingam* Chen Feng‡
*Mitsubishi Electric Research Labs (MERL) †Georgia Institute of Technology ‡University of Michigan

ABSTRACT

We present a simultaneous localization and mapping (SLAM) algorithm for a hand-held 3D sensor that uses both points and planes as primitives. Our algorithm uses any combination of three point/plane primitives (3 planes, 2 planes and 1 point, 1 plane and 2 points, and 3 points) in a RANSAC framework to efficiently compute the sensor pose. As the number of planes is significantly smaller than the number of points in typical 3D scenes, our RANSAC algorithm prefers primitive combinations involving more planes than points. In contrast to existing approaches that mainly use points for registration, our algorithm has the following advantages: (1) it enables faster correspondence search and registration due to the smaller number of plane primitives; (2) it produces plane-based 3D models that are more compact than point-based ones; and (3) being a global registration algorithm, our approach does not suffer from local minima or any initialization problems. Our experiments demonstrate real-time, interactive 3D reconstruction of office spaces using a hand-held Kinect sensor.

Index Terms: I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Range Data, Tracking

1 INTRODUCTION

Interactive 3D reconstruction has been useful for various applications in robotics, augmented reality, and computer vision. The emergence of inexpensive 3D sensors such as Kinect has made SLAM using 3D sensors more accessible [4, 6]. As newer and more exciting applications exploiting the potential of such 3D sensors are identified, the following two challenges become more important:

- **Fast and Accurate 3D Reconstruction:** The field of view and the resolution of 3D sensors usually result in a partial reconstruction of the entire scene. We need an accurate and fast registration algorithm that fuses the successive partial depth maps to model the entire scene.
- **Compact and Semantic Modeling:** The depth maps are usually noisy point clouds, which require a large memory and do not convey any semantic information.

In this paper, we address these challenges by proposing a registration algorithm that uses both points and planes as primitives. Using planes with points in the registration enables more efficient and accurate registration than using only points. The output of our system is registered point clouds as well as a plane-based representation of the scanned scene, as shown in Figure 1. The plane-based representation provides more compact and semantic information than point-based representations.

Local registration methods based on the iterative-closest point (ICP) algorithm [2, 6] are prone to local minima issues under fast sensor motion. Feature-based registration methods that solely depend on points [5, 1, 4] suffer from insufficient or incorrect correspondences in textureless regions or regions with repeated patterns. Plane-based methods [3, 7] suffer from degeneracy issues in scenes containing insufficient numbers of planes. For single-shot 3D sensors like Kinect, line correspondences are hard to obtain because they suffer from noisy or missing depth values especially around depth boundaries. Driven by these issues, our approach uses both points and planes to avoid the failure modes that are typical while using one of these primitives.

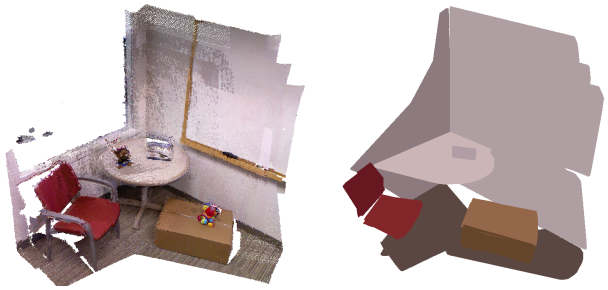

Figure 1: A 3D model reconstructed from a sequence in Figure 3. Our system not only generates registered 3D point clouds (left), but also reconstructs a scene as a set of planes (right). Note that the plane-based representation is obtained from plane landmarks, generated in our real-time SLAM algorithm (not in post-processing). In this model, the number of keyframes registered is 39, and the numbers of point and plane landmarks are 6912 and 9, respectively.

Our main contributions include: (1) an efficient RANSAC-based registration algorithm using any combination of three 3D point/plane primitives; (2) a bundle adjustment framework using both points and planes; and (3) a real-time SLAM system using the proposed techniques with a hand-held Kinect sensor.

2 SYSTEM OVERVIEW

Figure 2 shows the flow chart of our SLAM system. The input to the system is a pair of a color image and a depth map (RGB-D data). Our system extracts *measurements* from the input data and generates/updates *landmarks* in a global map. This is done by extracting points and planes from the incoming data and registering them with point/plane landmarks in the map, generated using the previous measurements. Here we briefly describe each component of our system.

Measurement Extraction: Our system extracts 2D keypoints (SURF) from each color image and back-projects them using the corresponding depth map to obtain 3D point measurements. We use a RANSAC-based plane fitting algorithm to extract plane measurements from the 3D point cloud generated from the depth map.

Registration: The pose of the current frame is computed by registering the measurements with respect to the landmarks in the map. We developed a closed-form registration algorithm using both point-to-point and plane-to-plane correspondences by combining solutions developed for individual cases using either point-to-point [5, 1] or plane-to-plane correspondences [3]. The algorithm is applicable to 3 or more correspondences; thus it can be used to generate hypotheses using the minimum number of 3 point/plane primitives in our RANSAC procedure, as well as to refine the camera pose using all inliers. For efficient and accurate registration, our RANSAC procedure prioritizes plane primitives over point prim-

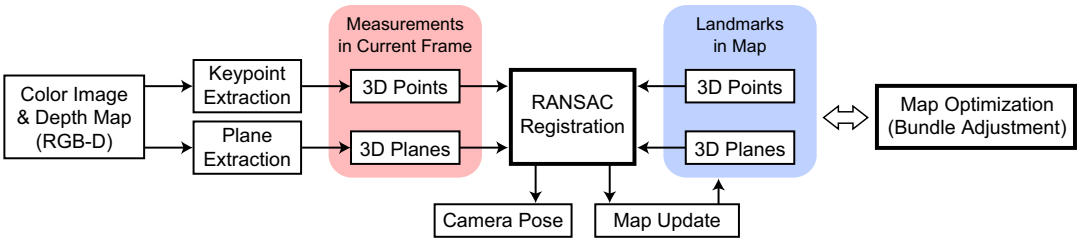

Figure 2: Overview of our SLAM system. The system extracts 3D points and planes as measurements from each frame, and registers them with the 3D point/plane landmarks in the map to compute the current camera pose. The landmarks in the map are updated using the registered measurements. The landmarks are periodically optimized with our bundle adjustment algorithm using both points and planes.

itives; i.e., it prefers the primitive combination in the order of 3 planes, 2 planes and 1 point, 1 plane and 2 points, and 3 points. This is because (1) the number of planes in the data is typically smaller than the number of points and (2) planes generated by many supporting points are less noisy than individual points, leading to more accurate registration. We also use several geometric invariants, such as the distance between points and the angle between planes, to prune false correspondences using an interpretation tree [3].

Map Update: In the global map, our system maintains point and plane landmarks generated from several *keyframes*. Our system adds the current frame as a keyframe to the map only if its pose is sufficiently different from previous keyframes.

Map Optimization: To refine the point and plane landmarks as well as the poses of keyframes, our system runs bundle adjustment by computing the Jacobian using both point-to-point and plane-to-plane correspondences. The bundle adjustment is done in a separate thread asynchronously from the main thread for camera tracking.

3 Experimental Results

We use a Kinect sensor that provides color images and depth maps at a resolution of 640×480 pixels for our real-time SLAM system. Figure 3 shows results for an indoor office room sequence. Figures 3(a) and (b) show examples of the color images and depth maps from the sequence. Figure 3(c) depicts plane extraction results from the corresponding depth maps. Figure 3(d) shows snapshots of our interactive visualization of the SLAM result. Results for the entire sequence are available in the supplementary video.

A 3D model reconstructed from the sequence is shown in Figure 1. In addition to the registered point clouds, our system provides reconstructed plane landmarks as a plane-based representation of the scene. Currently our system runs at 3 frames per second on a standard PC with Intel Core i7-950 processor.

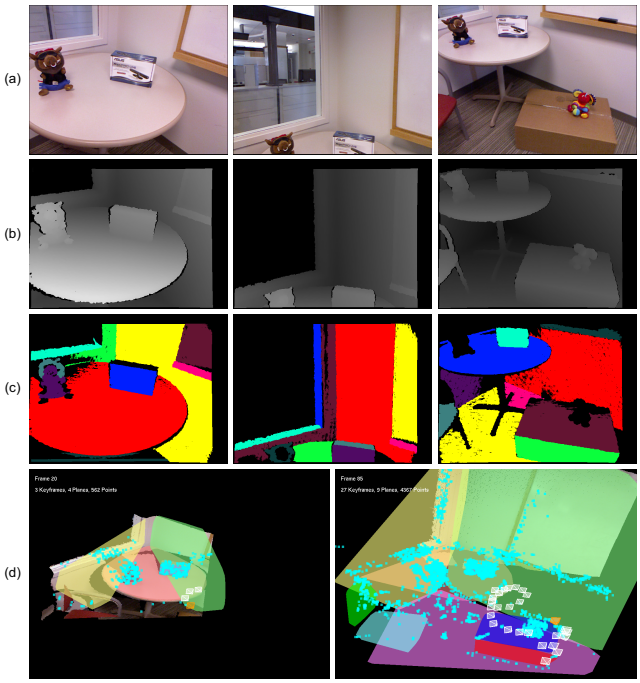

Figure 3: An example of real-time 3D reconstruction using a hand-held Kinect. (a) Color images and (b) depth maps from the captured sequence. (c) Our system performs plane fitting and segmentation to extract plane measurements from the depth maps. Each plane measurement is described with different colors. (d) Snapshots of our interactive visualization system. Plane landmarks (transparent polygons with different colors) and point landmarks (cyan points) are superimposed on the current frame (colored point cloud), demonstrating the correct registration. White camera icons represent the poses of keyframes, while the orange one shows the current pose.

4 Conclusions

We have shown a real-time SLAM system for hand-held 3D sensors that uses both point and plane primitives for registration. This hybrid approach enables faster and more accurate registration than using only points. Our system generates a 3D model as a set of planes, which provides more compact and semantic information of the scene than point-based representations.

Acknowledgments: We thank Jay Thornton, Amit Agrawal, and Tim K. Marks for their helpful comments and feedback. This work was supported by and done at MERL. Yong-Dian Jian and Chen Feng contributed to the work while they were interns at MERL.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5):698–700, May 1987.

[2] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, Feb. 1992.

[3] W. E. L. Grimson and T. Lozano-Pérez. Model-based recognition and localization from sparse range or tactile data. MIT AI Lab, A. I. Memo 738, Aug. 1983.

[4] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. In *Proc. Int'l Symp. Experimental Robotics (ISER)*, Dec. 2010.

[5] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4(4):629–642, Apr. 1987.

[6] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *Proc. IEEE Int'l Symp. Mixed and Augmented Reality (ISMAR)*, Oct. 2011.

[7] K. Pathak, A. Birk, N. Vaškevičius, and J. Poppinga. Fast registration based on noisy planes with unknown correspondences for 3-D mapping. *IEEE Trans. Robotics*, 26(3):424–441, June 2010.