Fast Indoor Structure Analysis of Single RGBD Images

Junho Jeon
POSTECH
zwitterion27@postech.ac.kr

Daehoon Yoo
Korea Military Academy
bbabbol@postech.ac.kr

Seungyong Lee
POSTECH
leesy@postech.ac.kr

Abstract

This paper presents a novel method for estimating the spatial layout and objects of an indoor scene simultaneously from a Kinect RGBD image. By exploiting the axis-aligned nature of man-made indoor structures, we first extract rectilinear polygons to construct a candidate set for room and object cuboids. We then jointly estimate the spatial room layout and objects by considering their mutual arrangements. Using a mixed integer program to select optimal cuboids that properly represent the room configuration, our method can estimate the indoor structure within a half second. Experimental results demonstrate that our method can estimate axis-aligned indoor scene structures from RGBD images quickly and reliably.

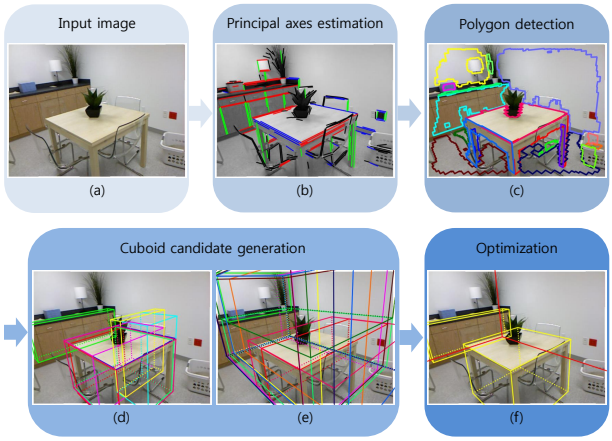


Figure 1. Overview of proposed method

1. Introduction

Estimating indoor scene structures from single images is an important task in scene understanding, and has been extensively studied in computer vision community [2, 6]. Recently, with wide spread of low-cost depth cameras such as Microsoft Kinect, it has become easy to obtain rough geometry of the scene as a depth image. Several approaches [3, 4] have been proposed to utilize depth images, where they represent objects using cuboids to approximate convex shapes in given RGBD images.

In this paper, we analyze the indoor structure as an arrangement of room and objects using the boxes in the box model [2]. To represent the indoor structure, we utilize the well known Manhattan world assumption that the indoor scene has the axis-aligned property. We assume that structural objects, such as walls and ceilings, are basically orthogonal and parallel to the ground, respectively, and most objects, such as desks and beds, are aligned to these structural objects. By exploiting this axis-aligned nature of indoor scenes, we propose a fast structure analysis method that estimates the indoor scene structure represented as a combination of room and object cuboids from an input single RGBD image. We focus on the mutual arrangement of the room and object cuboids to choose the optimal configuration of the scene. With our real-time rectilinear polygon detection and polygon-based cuboid generation algorithms, we achieve significant reduction of the computation time, reaching less than 0.5 seconds in total. Our polygon-based approach also provides robustness against depth noise.

2. Approach

To exploit the axis-aligned structure of an indoor scene, we first estimate the principal axes of the scene from the input RGB-D image (Fig. 1b). We used the work of [5] to estimate principal axes in the scene using line segments. We slightly modified the energy function of the previous method [5] to utilize the depth information. The modified energy function encourages the estimated principal axes to reflect the surface normal distribution of the scene.

2.1. Rectilinear polygon detection

We detect rectilinear polygons that are aligned to the principal axes. Our target polygon consists of coplanar 3D points. We construct coplanar point sets using point normals and point distributions along the principal axes. We then perform a regular grid analysis on each set of coplanar points and apply a sweep line algorithm [1] to extract the connected boundaries of polygons (Fig. 1c).

2.2. Cuboid candidate generation

Each pair of detected polygons is used to generate a cuboid candidate. There are two types of candidates to represent the indoor structure; object and room. We generate disjoint cuboid candidates by examining the potential cuboids with simple reasoning. First, as the camera is located outside of the objects and inside of the room, faces of the object and room cuboids should form convex and concave edges from the camera view, respectively. Second, the distances between faces of object cuboids should be small, but the distances between faces of the room cuboid can be large because of occlusions. These rules are also used to filter out false candidates, e.g., cuboids floating in the air.

2.3. Formulation and optimization

The indoor structure is represented as an arrangement of object and room cuboid candidates. Based on 3D volumetric and camera geometric reasoning, we impose the following assumptions on the optimal arrangement:

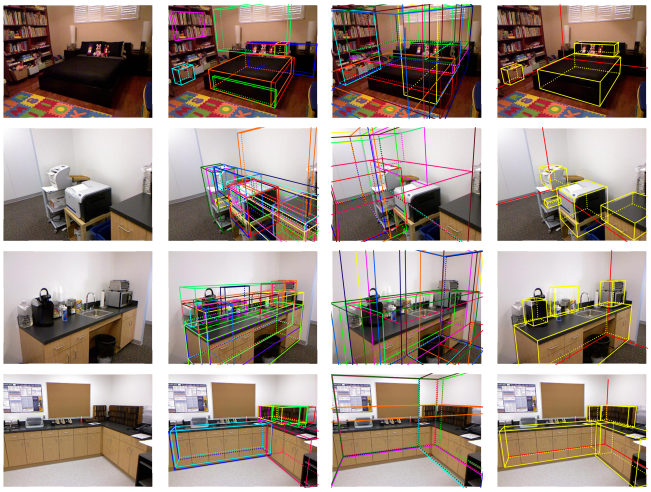
1. 3D points inside the projected silhouette of a cuboid should be contained in or on the cuboid.
2. Edges of the projected silhouette of a cuboid should match with vanishing line segments in the input image.
3. There is no intersection between object cuboids.
4. At least one face of an object cuboid should touch a face of the room cuboid.
5. There exists only one room cuboid.

Inspired by [4], we formulate our optimal candidate set selection problem as a 0-1 integer programming. Considering not only objects but also arrangements between objects and the room, our energy function consists of object, room, object-to-object, and object-to-room constraints. We minimize the energy function by selecting the optimal set of object and room cuboids using the branch-and-bound solver provided by IBM CPLEX optimizer.

3. Experiments and Conclusion

We implemented the proposed method mainly using C++, and tested our implementation on Intel Core i7-4790K CPU. For a 640×480 image, the entire procedure took less than 0.5 seconds. Principal axes estimation usually took about 200 ms, which may be easily accelerated using additional sensor data. Rectilinear polygon detection took less than 20 ms and the optimization took 20$\sim$100 ms, depending on the size of the candidate set.

We evaluated the proposed method with a set of RGBD images selected from the NYU dataset [7]. We selected target images from the dataset based on our setting where the scene satisfies Manhattan world assumption and consists of cuboid-like objects and room. Fig. 2 shows some of our experimental results. The proposed method could detect spatial layout of the room and cuboid-like objects quickly and reliably. Figs. 2b and 2c show that our cuboid candidate generation step constructs small sets of room and object candidates.



(a) Input (b) Object candidates (c) Room candidates (d) Result

Figure 2. Experimental results of the proposed method

The fast speed would allow our method to be useful for various scene understanding applications, such as interactive indoor scene reconstruction and large-scale indoor scene structure analysis.

Acknowledgements This work was supported in part by ICT/SW Creative Research Program (NIPA-2014-H0510-14-1007) and National Research Foundation of Korea grant (NRF-2014R1A2A1A11052779, NRF-2010-0019523).

References

[1] M. V. Anglada, N. P. Garcia, D. Ayala, and J. Martinez. Efficient algorithms for boundary extraction of 2d and 3d orthogonal pseudomanifolds. *Graphical Models*, 74(3):61–74, 2012.

[2] V. Hedau, D. Hoiem, and D. Forsyth. Thinking Inside the Box : Using Appearance Models and Context Based on Room Geometry. *In Proc. ECCV*, pages 224–237, 2010.

[3] H. Jiang. Finding approximate convex shapes in rgbd images. *In Proc. ECCV*, pages 582–596, 2014.

[4] H. Jiang and J. Xiao. A Linear Approach to Matching Cuboids in RGBD Images. *In Proc. CVPR*, pages 2171–2178, 2013.

[5] H. Lee, E. Shechtman, J. Wang, and S. Lee. Automatic upright adjustment of photographs with robust camera calibration. *IEEE TPAMI*, 36(5):833–844, 2014.

[6] A. Schwing and R. Urtasun. Efficient exact inference for 3D indoor scene understanding. *In Proc. ECCV*, 1, 2012.

[7] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. *In Proc. ECCV*, pages 746–760, 2012.