A Linear Approach to Matching Cuboids in RGBD Images

Hao Jiang
Boston College
hjiang@cs.bc.edu

Jianxiong Xiao
Massachusetts Institute of Technology
jxiao@csail.mit.edu

Abstract

We propose a novel linear method to match cuboids in indoor scenes using RGBD images from Kinect. Beyond depth maps, these cuboids reveal important structures of a scene. Instead of directly fitting cuboids to 3D data, we first construct cuboid candidates using superpixel pairs on a RGBD image, and then we optimize the configuration of the cuboids to satisfy the global structure constraints. The optimal configuration has low local matching costs, small object intersection and occlusion, and the cuboids tend to project to a large region in the image; the number of cuboids is optimized simultaneously. We formulate the multiple cuboid matching problem as a mixed integer linear program and solve the optimization efficiently with a branch and bound method. The optimization guarantees the global optimal solution. Our experiments on the Kinect RGBD images of a variety of indoor scenes show that our proposed method is efficient, accurate and robust against object appearance variations, occlusions and strong clutter.

1. Introduction

Finding three-dimensional structures and shapes from images is a key task in computer vision. Nowadays, we can obtain reliable depth map using low cost RGBD cameras from the digital consumer market, e.g. Microsoft Kinect, Asus Xtion and Primesense. Just like digital cameras that capture raw RGB data, these devices capture raw depth maps along with RGB color images. RGBD images provide a pointwise representation of a 3D space. We would like to extract structures from such data.

Recently, there are a few heroic efforts in extracting structures in RGBD images, e.g. [13, 12]. However, most of these approaches group pixels into surface segments, i.e. the counterpart of image segmentation for RGB images. Although some noteworthy studies [8] infer support relations in scenes, there is still very little volumetric reasoning used in the 3D space, which should be even more important as the depth is available, than pure image information with which volumetric reasoning is well studied [3, 9]. In this paper, we design an efficient algorithm to match cuboid structures in an indoor scene using the RGBD images, as illustrated in Fig. 1.

A cuboid detector has many important applications. It is a key technique to enable a robot to manipulate box objects [1]. Cuboids also often appear in man-made structures [3, 6, 9]. A cuboid detector thus facilitates finding these structures. Detecting cuboids from RGBD images is challenging due to heavy object occlusion, missing data and strong clutter. We propose an efficient and reliable linear method to make a first step towards solving this problem. Our contribution is a novel linear approach that efficiently optimizes multiple cuboid matching in RGBD images. The proposed method works on cluttered scenes in unconstrained settings. Our experiments on thousands of images in the NYU Kinect dataset [8] and other images show that the proposed method is efficient and reliable.

2. Method

We optimize the matching of multiple cuboids in a RGBD image from Kinect. Our goal is to find a set of cuboids that match the RGBD image and at the same time satisfy the spatial interaction constraint. We construct a set of cuboid candidates and select the optimal subset. The cuboid configuration is denoted by $x$. We formulate cuboid matching into the following optimization problem,

$$\min_{x} \{U(x) + \lambda P(x) + \mu N(x) - \gamma A(x) - \xi O(x)\}$$  \hspace{1cm} (1)

s.t. Cuboid configuration $x$ satisfies global constraints.
Here $U(x)$ is the unary term that quantifies the local matching costs of the cuboids, $P(x)$ is a pairwise term that quantifies the intersection between pairs of cuboids, $N(x)$ is the number of matched cuboids in the scene, $A(x)$ quantifies the covered area of the projected cuboids on the image plane, and $O$ penalizes the occlusions among the cuboids. $\lambda$, $\mu$, $\gamma$ and $\xi$ control the weight among different terms. By minimizing the objective function, we prefer to find the multiple cuboid matching that has low local matching cost, small object intersection and occlusion, and covers a large area in the image with a small number of cuboids. Besides the soft constraints specified by the objective function, we further enforce that the optimal cuboid configuration $x$ satisfies hard constraints on cuboid intersection and occlusion. This optimization is a combinatorial search problem, and we propose an efficient linear solution.

**References**


