A cuboid detection algorithm with computer vision methods

Kun Huang
huangkun@shanghaitech.edu.cn

January 20, 2017

Abstract

In robotics and Virtual Reality, there's a need to detect and understand the structure of the scene. And a cuboid is a regular shape containing geometry information about the scene. With the cuboid detection results, the robot can better interact with the scene environment, even calibrate the camera. Recently, computer vision methods make it possible to have an accurate detection via end to end pipeline. In this project, we take a two-stages process: first localizing the object and detecting vertex within the localization area.

Introduction

To detect vertex of a cuboid, recent computer vision methods first detect a bounding box of the region of the cuboid and then find the vertex within the bounding box in a finer scale. Of couse, in real senario, there could be more than one cuboid shape. For example, in a common living room, we have cabinet, bed, etc. In such complicated cases, the pipeline of Faster RCNN [5] is able to handle it. Faster RCNN takes a grid search strategy, which search every grid of the M by N feature map, and predict a cuboidness score along with a possible bounding box coordinates. And a second stage detection is applied for detecting the vertex within a bounding box with high cuboid score in a finer scale. A second part of this project is to apply the cuboid detection algorithm for a robotics senario, where we use cuboid vertex as a reference frame and estimate the environment parameter relative the reference frame.

State-of-the-art

Recent cuboid detection methods include [2, 6]. Though the pipeline of [2] is basically Faster RCNN [5], but it's the first to apply deep learning techniques to cuboid detection. This paper largely follows the localization plus finer-scale vertex alignment pipeline. Aside

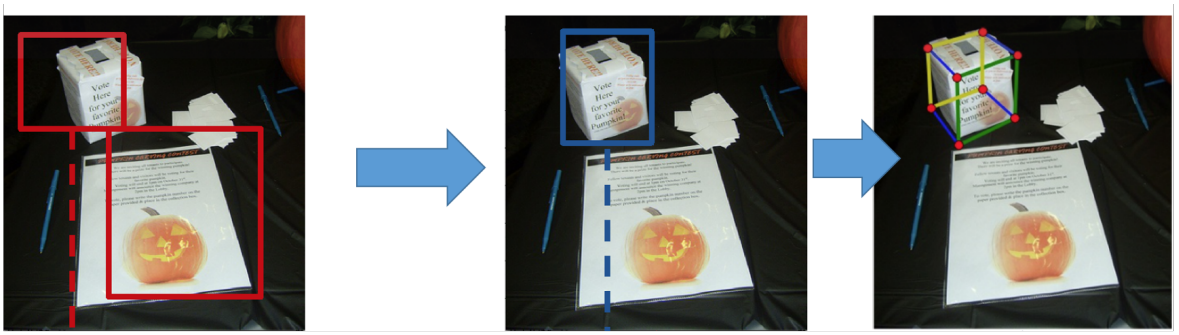

Figure 1: Pipeline of recent vision method for detection cuboid vertex.

from the cuboid bounding box regression, the target also includes vertex location estimation. So the final error function is.

$$L_{\text{all}} = L_{\text{bbox}} + L_{\text{vertex}} \tag{1}$$

, in which bbox means bouding box as shown in 1, which surrounds the object with a rectangle, represented by (x, y, h, w) in math. Beside Faster RCNN [5], we could also consider other object detection methods, such as more recent work [3, 1], and real-time object detection [4] for robotic applications.

Experiment

To measure how well our algorithm does in this task, there're three aspects: a) Whether the detected bounding box truely contains a cuboid. b) The offset of the detected bounding box relative to the groundtruth. c) The offset of the cuboid vertex (8 vertex) relative to the groundtruth location. . Currently, our algorithm can detect the bounding box location quite well. See Fig 2 for some examples. Cases include single cuboid and multiple cuboids. And the vertex location detection part is still being trained, and will be finished in a short time (a little buggy).

Conclusion

The algorithm is effective though, but far from perfect for application in real scene, we can still improve the algorithm to be more accurate. This will be our future work.

Figure 2: Examples of cuboid localization, including both single cuboid case and multiple cuboid case.

References

[1] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. *CoRR*, abs/1605.06409, 2016.

[2] Debidatta Dwibedi, Tomasz Malisiewicz, Vijay Badrinarayanan, and Andrew Rabinovich. Deep cuboid detection: Beyond 2d bounding boxes. *CoRR*, abs/1611.10010, 2016.

[3] Tsung-Yi Lin, Piotr Dollr, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. *CoRR*, abs/1612.03144, 2016.

[4] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector.

[5] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. 2015.

[6] Jianxiong Xiao, Bryan Russell, and Antonio Torralba. Localizing 3d cuboids in single-view images. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 746–754. Curran Associates, Inc., 2012.