Multiple Indoor Object Detection Using Depth Information

Jinchao Ye
Stanford University
Stanford CA 94305
jcye@stanford.edu

Second Author
Institution2
First line of institution2 address
http://www.author.org/~second

Abstract

RGB-D sensors, such as Microsoft Kinect sensor, can provide us depth information which might be useful to do multiple object detection. With this facility, we propose a new method to detect multiple objects more accurately in a single image with depth information. We first try to detect object on different depth layers. We also plan to estimate the conditional probability of whether an object exists given the existence of another object and their distance.

1. Introduction

1.1. Proposal

Multiple object detection is a fundamental as well as challenging topic in computer vision. Multiple objection detection can be used in image retrieval, automated driving, surveillance, etc. Compared with single objection detection, multiple object detection is more challenging due to occlusion, deformation and so on.

Meanwhile, recently available RGB-D sensors, such as Microsoft Kinect RGB-D sensor, are able to provide depth map of objects. Will it help multiple object detection? We think this makes it possible to do multiple object detection more accurately since depth information is useful for segmentation or handling occlusion.

We want to do multiple object detection using the depth information. Since the effective depth sensor range of Kinect is between 0.8m to 3.5m, we will narrow our problem on indoor objects. Firstly, we want to apply classic object detection algorithm [1] on the training dataset. In this way, we can get several object detectors. Secondly, using these object detectors and the depth map, we can detect objects on different depth layer and thus got more accurate results. Thirdly, we might also include depth information in training process. That means we can estimate the conditional probability of whether an object exist given the existence of another object and their distance.

We plan to use ImageNet as training database, and use Kinect to collect RGB-D images to test the performance of our algorithm. To evaluate our results, we will compare [1] with our results on the same color-only dataset. We might also compare our result with results from [2] or [3] on a small dataset collected by ourselves. They either did multiple object detection or dense object detection. More specifically, we will evaluate the precision and recall of different objects as well as the overall precision and recall.

References

[1] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 9, September 2010.

[2] Hema Swetha Koppula, Abhishek Anand, Thorsten Joachims, Ashutosh Saxena. Labeling 3D scenes for Personal Assistant Robots. In *RSS workshop on RGB-D cameras*, 2011

[3] Krystian Mikolajczyk, Bastian Leibe, Bernt Schiele. Multiple Object Class Detection with a Generative Model. In *CVPR*, 2006.

2. Appendix

My course project is part of a larger job in vision lab at Stanford.