Detection and Recognition of Dinning Table Objects Using RGB-D Camera

Jakob Baumgartner
Faculty of Electrical Engineering,
University of Ljubljana
Tržaška 25, 1000 Ljubljana, Slovenia
jakob.baumgartner@student.uni-lj.si

Janez Perš
Faculty of Electrical Engineering,
University of Ljubljana
Tržaška 25, 1000 Ljubljana, Slovenia
janez.pers@fe.uni-lj.si

Abstract

In this paper, we present a real-time method for 3D bounding box object detection and recognition in dinning tabletop scenarios using RGBD cameras, with potential applications in robotics. Our approach combines color and depth information to accurately detect and classify a variety of dinning room objects. While the accuracy of our method is not the highest among all approaches, it is able to achieve good performance in a real-time setting, making it suitable for use in robotics applications. We evaluate the performance of our method on the SUN-RGBD dataset, a widely used benchmark for 3D object recognition, and demonstrate its real-time capabilities using Intel RealSense camera. The results of our experiments show that our method is able to achieve competitive accuracy while maintaining a high frame rate, making it a promising solution for object detection and recognition in robotics.

1. Introduction

RGBD (Red, Green, Blue, Depth) sensors[10] use a combination of traditional RGB (Red, Green, Blue) cameras, which capture colour information about an object, and depth sensors, which measure the distance from the sensor to the object. In the context of a dining table, RGBD sensors can be used to allow the robot to recognise and classify a variety of objects, including dishes, glasses, utensils and food items. By combining the colour information from the RGB cameras with the depth information from the depth sensors, the robot is able to create a detailed 3D model of the objects on the table, allowing it to understand their shape, size and orientation.

RGBD object detection and classification algorithms have a number of advantages over traditional object detection algorithms that rely solely on colour information[2]. One advantage of using RGBD information is that the algorithm can better handle occlusion and clutter in the scene. Occlusion occurs when an object blocks the view of another object, and clutter refers to the presence of multiple objects in the scene that may be difficult to distinguish. By using depth information, the algorithm can better understand the spatial relationships between objects and more accurately detect and classify objects even when they are partially occluded or surrounded by clutter. Example of occlusion on a dining table is when a plate is partially hidden behind a glass. Another benefit of using RGBD information is that it enables the algorithm to better distinguish between objects that are similar in colour but different in shape or size. This can be particularly useful when there are multiple objects of the same colour in the scene, as the algorithm can use the depth information to distinguish them based on their shape and size. In our case it can help us distinguish between colourful tablecloths as a background and objects sitting on them.

However, using RGBD object detection is still less common than using RGB object detection. One reason is that RGB sensors are more common and less expensive than RGBD sensors. As a result, there are more and better quality RGB than RGBD data sets. This has partially changed with the appearance of commercial sensors such as Microsoft Kinect and Intel Real-Sense. In addition, algorithms using RGBD data often require more computing power and resources to analyse the data, which can complicate implementation in real-time applications or on resource-constrained devices. In addition, RGBD data can be noisy and inaccurate, especially in low light or at long distances, which can affect the performance of the object detection algorithm. This can make it difficult to achieve the same level of accuracy and robustness as with the RGB object detection algorithms. This can be partially corrected with indoor artificial light, but is difficult under natural conditions.

We developed an algorithm that uses an RGBD sensor to detect bounding boxes around objects sitting on a dining table in real time. The algorithm also classifies the objects within these bounding boxes and accurately identifies the type of each object. This system is designed to work effectively in a dynamic environment, such as a dining room, where items may be frequently added to or removed from the table. The ultimate goal of this project is to develop a system that can assist in tasks such as identifying and organising objects on the dining table, or providing a user with information about the table's contents.

2. Related Work

There are a number of approaches to RGB object detection, including traditional computer vision techniques and machine learning-based approaches. CNNs are particularly effective in image recognition tasks and are widely used for object recognition in a variety of fields, including computer vision, image processing, and autonomous systems. Because of the smaller data dimension, algorithms using RGB data generally run faster than algorithms using RGBD data and are widely used in robotics. Some of the networks that currently get the best results in the Coco dataset are: EVA[4], DyHead[3] and Focal-L[14].

While algorithms using RGBD data are not as well developed as algorithms using RGB data, there are a number of existing algorithms[12]. These can be roughly divided into outdoor and indoor algorithms. This is because indoors we can use sensors that give us dense depth images, while outdoors due to lighting conditions we have to use scanners that give us sparse depth images. Another way to split detection and classification algorithms is between traditional and deep learning methods. Algorithms in both categories mostly follow the same pipeline. First we create region suggestions and then classify them. Traditional methods rely on handcrafted functions, while deep learning methods replace one or more of the algorithm parts with deep neural networks. Despite extensive research, the performance of traditional methods is still far from that of the human visual system. Traditional methods often lose performance when dealing with crowded scenes and scenes with high occlusion. Deep learning methods have solved some or most of these problems. Objects that our algorithm recognises are similar to the objects contained in the data set SUN-RGBD[9] [1]. Deep learning models with the best performance in the data set based on mAP@0.25 are DeMF[13], CaGroup3D[11] and FCAF3D[6]. However, these algorithms are computationally intensive and cannot be executed in anywhere near real-time on normal computing devices. Since one of our goals is real-time execution, algorithms that are more comparable to ours are Frostum PointNet[5], Frostum VoxNet[7] and Frostum VoxNet v2[8]. These offer slightly worse but still impressive detection while operating at 5-10Hz.

References

[1] Sun-rgb dataset webpage. `https://rgbd.cs.princeton.edu/`. Accessed : 2022-1-1.

[2] *RGB-D Image Analysis and Processing*. Advances in Computer Vision and Pattern Recognition. Springer International Publishing, Cham, 2019.

[3] Xiyang Dai, Yinpeng Chen, Bin Xiao, Dongdong Chen, Mengchen Liu, Lu Yuan, and Lei Zhang. Dynamic head: Unifying object detection heads with attentions, 2021.

[4] Yuxin Fang, Wen Wang, Binhui Xie, Quan Sun, Ledell Wu, Xinggang Wang, Tiejun Huang, Xinlong Wang, and Yue Cao. Eva: Exploring the limits of masked visual representation learning at scale, 2022.

[5] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J Guibas. Frustum pointnets for 3d object detection from rgb-d data. 2018.

[6] Danila Rukhovich, Anna Vorontsova, and Anton Konushin. Fcaf3d: Fully convolutional anchor-free 3d object detection, 2021.

[7] Xiaoke Shen and Ioannis Stamos. Frustum voxnet for 3d object detection from rgb-d or depth images. (arXiv:1910.05483), Feb 2020. arXiv:1910.05483 [cs, eess].

[8] Xiaoke Shen and Ioannis Stamos. 3d object detection and instance segmentation from 3d range and 2d color images. *Sensors*, 21(44):1213, Jan 2021.

[9] Shuran Song, Samuel P. Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 567–576, 2015.

[10] Kyriaki A. Tychola, Ioannis Tsimperidis, and George A. Papakostas. On 3d reconstruction using rgb-d cameras. *Digital*, 2(33):401–421, Sep 2022.

[11] Haiyang Wang, Lihe Ding, Shaocong Dong, Shaoshuai Shi, Aoxue Li, Jianan Li, Zhenguo Li, and Liwei Wang. Cagroup3d: Class-aware grouping for 3d object detection on point clouds, 2022.

[12] Yangfan Wang, Chen Wang, Peng Long, Yuzong Gu, and Wenfa Li. Recent advances in 3d object detection based on rgb-d: A survey. *Displays*, 70:102077, Dec 2021.

[13] Hao Yang, Chen Shi, Yihong Chen, and Liwei Wang. Boosting 3d object detection via object-focused image fusion, 2022.

[14] Jianwei Yang, Chunyuan Li, Pengchuan Zhang, Xiyang Dai, Bin Xiao, Lu Yuan, and Jianfeng Gao. Focal self-attention for local-global interactions in vision transformers, 2021.