Literature and Product Review on Detecting 3D objects in point clouds

Bingquan Cai

1. Introduction

A point cloud is a set of data points in space. The points may represent a 3D shape or object. Each point position has its set of Cartesian coordinates (X, Y, Z).[1] The point cloud contains a wealth of information, including 3D coordinates X, Y, Z, color, classification values, intensity values, time, and so on. In my opinion, the point cloud can atomize the real world, and the real world can be restored through high-precision point cloud data.

Point clouds are not obtained by ordinary cameras, but by 3D imaging sensors, such as binocular cameras, 3D scanners, RGB-D cameras, and so on. At present, the mainstream RGB-D cameras include Microsoft's Kinect series, Intel's Realsense series, structure sensor, etc. Point clouds can be created from scanned RGB-D images and the intrinsic parameters of the scanned camera by calibrating the camera and calculating real-world points (x, y) using the camera's intrinsic parameters. Thus, RGB-D images are grid-aligned images, while point clouds are more sparse structures. In addition, better methods to obtain point clouds include LiDAR laser detection and measurement, which are mainly acquired via satellite-based, airborne, and ground-based methods.

Point clouds can express the spatial outline and specific location of objects, we can see the shape of streets, houses, and the distance of objects from the camera is also known. Secondly, the point cloud itself is independent of the viewpoint, and can be arbitrarily rotated to observe a point cloud from different angles and directions, and different point clouds can be directly fused if they are under the same coordinate system. Therefore, point cloud data can often be used for detecting 3d objects.

2. Application

Nowadays, there are many applications that use the point cloud data for detecting 3d objects.

a. Autonomous Driving

Autonomous driving is a highly discussed topic nowadays, and point clouds data can be applied to it for the detection part. Especially when driving at night, the camera does not present the image very well, but if we use LiDAR laser to obtain point clouds data, we can identify the object better.

b. Urban Planning

The data detected by point clouds can help us to carry out urban planning. For instance, we can detect objects on the side of road to determine whether streetlights or trees block the signs of the road; we can use the data to determine if the distance between the house and the road is too close and requires additional acoustic glass.

c. Archaeological Artifact Detection

In archaeological research, we may dig up many archaeological artifacts that are not convenient for actual detection due to their material and characteristics. Therefore, we can use cameras to get the point cloud data of them and analyze it in the computer.

3. Literature Review

I am interested in using point clouds for 3d objects detection and applying it to autonomous driving,

so I do research on approaches relative to the detection during autonomous driving.

a. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud

This article proposes PointRCNN for 3D object detection from raw point cloud. The whole framework is composed of two stages: stage-1 for the bottom-up 3D proposal generation and stage-2 for refining proposals in the canonical coordinates to obtain the final detection results.[2]

Instead of generating proposals from RGB image or projecting point cloud to bird's view or voxels as previous methods do, the stage-1 sub-network directly generates a small number of high-quality 3D proposals from point cloud in a bottom-up manner via segmenting the point cloud of the whole scene into foreground points and background.

The stage-2 sub-network transforms the pooled points of each proposal to canonical coordinates to learn better local spatial features, which is combined with global semantic features of each point learned in stage-1 for accurate box refinement and confidence prediction.

The PointRCNN method used in this article has some highlights. First, we will generally do the data processing on the point cloud before it could be used to the further steps. However, the PointRCNN method can use raw point cloud for object detection and thus reduce many steps of data processing. Second, it reduces the search range of the detection box by using the foreground point segmentation. Third, the stage-2 network refines the proposals in the canonical coordinate by combining semantic features and local spatial features.

b. Multi-View 3D Object Detection Network for Autonomous Driving

This paper proposes Multi-View 3D networks (MV3D), a sensory-fusion framework that takes both LIDAR point cloud and RGB images as input and predicts oriented 3D bounding boxes. It encodes the sparse 3D point cloud with a compact multi-view representation. The network is composed of two subnetworks: one for 3D object proposal generation and another for multi-view feature fusion. The proposal network generates 3D candidate boxes efficiently from the bird's eye view representation of 3D point cloud. It designs a deep fusion scheme to combine region-wise features from multiple views and enable interactions between intermediate layers of different paths.[3]

Most of the current methods voxelize the 3D point cloud, extract the structural features and send them to the SVM or neural network for classification, which is computationally expensive. In this paper, the author encodes the 3D point cloud into a multi-view feature map and applies it to a region-based multi-modal representation. Also, this paper designs a region-based fusion network that effectively combines features from multiple views to jointly classify candidate regions and do directional 3D box regression.

4. Open-Source Research

a. KITTI

The KITTI dataset, founded by the Karlsruhe Institute of Technology and the Toyota Technology Institute at Chicago, is the largest international dataset for the evaluation of computer vision algorithms in autonomous driving scenarios. The dataset is used to evaluate the performance of computer vision technologies such as stereo, optical flow, visual odometry, 3D object detection and 3D tracking in an in-vehicle environment.

The entire dataset consists of 389 pairs of stereo images and optical flow maps, 39.2 km visual

ranging sequences, and over 200k images of 3D labeled objects sampled and synchronized at 10 Hz. Overall, the original dataset was classified into 'Road', 'City', 'Residential', ' Campus' and 'Person'. For 3D object detection, the labels are subdivided into car, van, truck, pedestrian, pedestrian(sitting), cyclist, tram, and misc components.

KITTI: http://www.cvlibs.net/datasets/kitti/

Copyright: 'All datasets and benchmarks on this page are copyright by us and published under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License. This means that you must attribute the work in the manner specified by the authors, you may not use this work for commercial purposes and if you alter, transform, or build upon this work, you may distribute the resulting work only under the same license.'

b. The Waymo Open Dataset

The Waymo Open Dataset is provided by Waymo, the self-driving company owned by Google parent company Alphabet. The Waymo Open Dataset is comprised of high-resolution sensor data collected by Waymo self-driving cars in a wide variety of conditions.

The sensors in the Waymo dataset include five lidars, five cameras, and better synchronization between lidars and cameras. More importantly, the Waymo dataset contains 3,000 driving segments of 16.7 hours in total, with an average length of about 20 seconds per segment. The entire dataset contains 600,000 frames, with approximately 25 million 3D bounding boxes and 22 million 2D bounding boxes. In addition, the Waymo Open Dataset has been greatly enhanced in terms of dataset diversity, covering different weather conditions, different times of day and night, different locations in downtown and suburban areas, different road objects such as pedestrians and bicycles, and so on.

The Waymo Open Dataset: https://waymo.com/open/

License: https://waymo.com/open/terms

5. References

[1] https://en.wikipedia.org/wiki/Point_cloud
[2] https://arxiv.org/abs/1812.04244v2
[3] https://arxiv.org/abs/1611.07759