2D Object Detection Method based on LiDAR Point Cloud

Jung-Eun Lee*, Byeong-Woo Kim**
*Dept. of Electrical, Electronic and Computer Engineering, University of Ulsan
**Dept. of Electrical Engineering, University of Ulsan
e-mail:rotnn429@ulsan.ac.kr

LiDAR 데이터를 이용한 2D 객체 인식 방법 연구

이정은*, 김병우**
*울산대학교 전기전자컴퓨터공학과
**울산대학교 전기공학부

Abstract

In recent years, the need for autonomous driving is emerging. Consequently, researchers are focusing on developing technologies for autonomous driving. Among various these technologies, object detection is a core technology. Furthermore, the research on object detection using artificial intelligence and machine learning based on camera sensor data is being actively conducted. However, the camera sensor has various limitations, which seriously affect autonomous driving. To overcome the limitations of the existing camera-based object detection method, this paper proposes an object detection method using a LiDAR sensor. 3D data can be obtained from the lidar sensor. Using this kind of data for object detection requires a lot of calculations. Therefore, data preprocessing is required for converting 3D data into 2D data. In this paper, we propose a CNN-based object detection method from the converted LiDAR data.

1. Introduction

In recent years, the need for autonomous driving is emerging. Consequently, researchers are focusing on developing technologies for autonomous driving. Among various these technologies, object detection is a core technology. Furthermore, the research on object detection using artificial intelligence (AI) and machine learning based on camera sensor data is being actively conducted. In particular, many studies on object detection using open computer vision (OpenCV) or deep learning based on camera images are underway such as SSD, Yolo [1-3]

Tesla, a representative autonomous vehicle maker, showed high technology using only camera sensors and deep learning. However, camera sensors have limitations, such as their effect on light, which sometimes reduces autonomous driving performance. An example is the Tesla accident. It was an accident where the white truck in front was mistakenly detected as the sky and collided.

To overcome the limitations of the existing camera-based object detection methods, this paper proposes an object detection method using a LiDAR sensor. The data in the form of a point cloud comes out of the LiDAR sensor, which is 3D data and requires a lot of computation to detect objects through deep learning. Therefore, in this paper, 3D data is converted into 2D data, and objects are detected using Convolution Neural Network (CNN), an algorithm that detects objects in representative 2D data. In this paper, the CNN algorithm is the Yolo algorithm, which is known to be the fastest and most accurate recently.[2]

2. Related Works

There are many ways to detect objects based on camera images. Object detection was classically based on OpenCV. This method has many limitations in that it is difficult to find the feature points as well as to derive all the feature points of the geometric object mathematically. To supplement this, machine learning-based object detection algorithms are representatively such as RCNN, SSD, Yolo, etc. This is a method to find features that are difficult to find with the naked eye through learning. For object detection, a camera video image is used, which uses RGB pixel values. However, the camera sensor has the following limitations. 1. Weak to the influence of weather such as day and night, rain, and fog. 2. There is a blur light when entering or exiting the tunnel. 3. There is distortion caused by the camera lens. 4. Blind spots due to the limitation of the measuring range. 5. No distance information. In particular, there are many problems due to the first and fifth limitations. Therefore, research is being conducted on a method of fusion of other sensors such as LiDAR and Radar to deal with those.

In this paper, object detection was performed based on the LiDAR sensor to solve the above limitations. However, the data in LiDAR is a matrix summarizing the depth of each point. In addition, for convenience of operation, point cloud data was converted into image data, and this was used for CNN to detect objects required for autonomous driving. (Point Cloud (Range or XYZ -> Image (RGB))

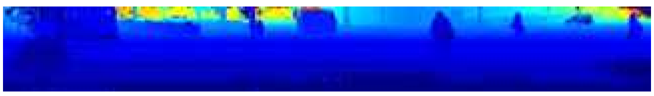
3. Data Preprocessing

In this case, batch learning is used, so the dataset is an important factor. Several organizations such as Kaggle, Dacon, KITTI, and Deep Drive provide data sets. We use the KITTI data set, which is widely used in mobile robotics and autonomous driving among several data sets. KITTI provides the following data: Calibration, Velodyne (lidar), label, image, planes, Row data.

We checked the distribution of the dataset. The majority of data are images of cars. Thus, it shows data imbalance. We can guess that the car can detect well, but not sure about other things. Data preprocessing is required for LiDAR-based 2D object detection. We have three kinds of methods for making the image from LiDAR data. We will explain in more detail below.

2.1 Panoramic image

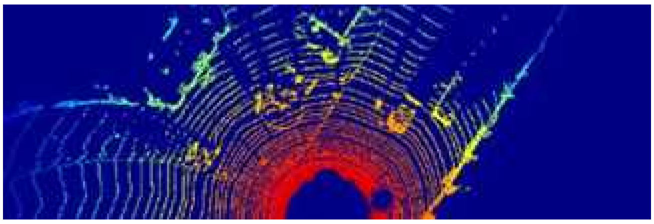
A panoramic image is an image of the entire point cloud. By creating an image for the entire 360 degrees, it is possible to detect objects in all directions around the vehicle. However, geometrical distortion occurs while creating an image, which may cause a slight difference from the shape of the actual object.



[Fig. 1] Panoramic image from LiDAR

2.2 Top-view image

The top-view image is an image from the bird-eye view. It has the advantage that it is easy to detect the presence or absence of an object. However, since the z-axis information of the object is lost while converting to 2D, it is difficult to identify the type of object.



[Fig. 2] Top-view image from LiDAR

2.3 Projection image

Projection image is the image overlaid point cloud on-camera image. By using the projection image, the detection range of the object will be reduced. However, this kind of image doesn't need to make labels due to the available labels in the camera image. Therefore, it saves time for constructing the training dataset. In addition, since the object looks similar to that seen by the human eye, there is an advantage in that it is easy to check later.

[Fig. 3] Projection image from LiDAR

This paper built a dataset by selecting projection images to save time. In addition, the projection images were created using the calibration information and the coordinate transformation information of the sensor. A total of 7,481 images were generated through transformation, and the ratio was divided into 8:2 for training and validation.

4. Object Detect

In this paper, We choose the Yolo algorithm to detect objects in LiDAR images. As an abbreviation of You Only Look Once, It is a Convolutional Neural Network(CNN) designed to perform object detection. The biggest feature of this is to implement Unified Detection, which simultaneously executes Bounding Box Coordinate and Classification through the same neural network structure.[2] When the Lidar image preprocessed in Chapter 3 is input to CNN, We can get the position and type of the object.

To evaluate the performance of object detection, mAP and F1 score were used as indicators. The mAP and F1 scores evaluate the accuracy in the relationship between Recall and Precision. The mAP indicator represents the average of Precision according to the Recall value, and the F1 score represents the harmonic average of Recall and Precision.

[Table 1] Performance evaluation for object detection

mAP(@0.50)	mAP(@0.75)	F1-score(@0.50)	F1-score(@0.75)
0.905	0.586	0.94	0.79

[Table 2] Object detection accuracy of classes(@0.50)

Class	Car	Van	Truck	Pedestrian	Person sitting	Cyclist	Tram
ap[%]	95.81	98.49	99.39	78.71	63.47	92.54	98.69

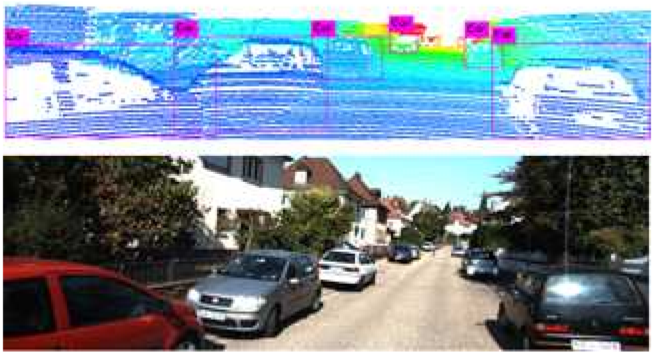
Table1 illustrates the result of object detection. Detection accuracy was 90.5% and 58.6%, respectively. From Table 2, We can realize that high accuracy was confirmed for most objects such as cars. Cognitive performance was relatively low for certain objects such as sitting people and pedestrians. Data imbalance was confirmed by comparing the proportions of the data before training. Also, since the size of the corresponding object is smaller than that of other objects, it does not have sufficient information for training.

5. Conclusion

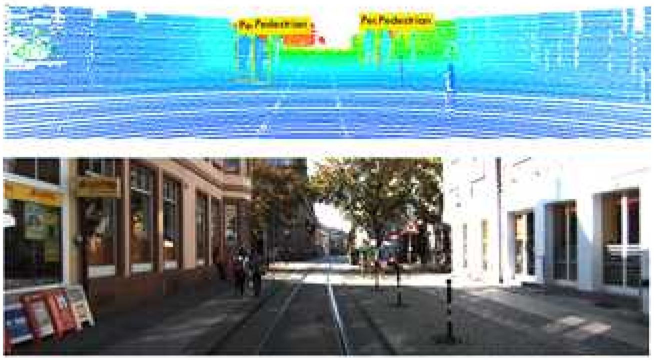
In this paper, LiDAR-based object detection was implemented to supplement the limitations of the camera-based object detection method. Since LiDAR has 3D information, the two-dimensional transformation was performed by preprocessing. Firstly, the projection images were generated through preprocessing. Finally, the training and the testing were completed by the yolov3 algorithm. As we can clearly see in the results that the proposed method can replace the object detection method based on the camera.

Acknowledgement

This research was financially supported by the Ministry of Trade, Industry and Energy(MOTIE) and Korea Institute for Advancement of Technology(KIAT) through the National Innovation convergence complex R&D program and the OEM demonstration cluster construction project to support autonomous vehicle parts suppliers. [Project Number: 1415173808, P0014572]



(a) Image 000008(Car)



(b)Image 000043(Pedestrain)

(c) Image 000142(Complex)

[Fig. 4] Projection image from LiDAR

References

[1] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[2] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[3] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.