3D Environment Detection Using Multi-view Color Images and LiDAR Point Clouds

Bo-Tai Wu, Pei-Cian Li
Department of Electronic Engineering
National Taipei University of Technology
Taipei, Taiwan
E-mail: skystar@ntut.edu.tw

Jian-Hong Chen, Yen-Ju Li, Yu-Cheng Fan
Department of Electronic Engineering
National Taipei University of Technology
Taipei, Taiwan
E-mail: skystar@ntut.edu.tw

***Abstract*–We present a "3D environment detection using multi-view color images and LiDAR point clouds" in this paper. Combining "multi-view color images" and "LiDAR point clouds" to solve the problem of insufficient resolution of traditional LiDAR point clouds to further realize accurate three-dimensional environment detection and recognition. The system adopts eight cameras to capture multi-view panorama images and combines with the depth information of the LiDAR point clouds to establish high-resolution 3D color images and depth images for subsequent objects detection and segmentation. According to the 3D point clouds information and multi-view images, 3D dynamic environment preprocessing, data processing and data format conversion are performed, the neural network input signal is extracted. Then through the neural network training, to determine the three-dimensional environment, and cutting out obstacles such as cars and pedestrians.**

Keywords–3D Environment Detection, LiDAR System, Multi-view Images, Point Clouds

I. INTRODUCTION

In recent years, the techniques like internet of things, big data and artificial intelligence have developed rapidly. The automatic driving is the most attractive technique among them. Traffic is closely linked with our life. More convenient and secure transportation is the reason why we all work together, and automatic driving technique will change and improve existing traffic problems. At present, transportation has always been an important issue that every country attaches importance to the related research topics. Like the solution of the shortage of drivers, traffic jams, traffic safety and long driving are often a major safety problems. Long driving means fatigue driving and self-driving is not affected by the time. Even unmanned vehicles have changed everyone's work. No longer need taxi drivers, traffic police and traffic lights and it is more effective for traffic management. Its purpose is to make human life more convenient and safer.

Self-driving is an important trend. The most important thing to do by self-driving is to avoid collisions with obstacles, and LiDAR [1] can completely detect the surrounding environment. Using LiDAR to detect the road can be a very suitable tool for self-driving. However, we usually have various situations when driving on the road. It is hard to find an algorithm that will solve all the problems. Artificial intelligence has begun to flourish with the advent of Industry 4.0. Many products and applications are also related to artificial intelligence. Self-driving is a very typical example.

According to the current status of automatic driving, there are not many studies on using point cloud to classify objects and matching them to color images. At the same time, the research of matching correction between point cloud map and color image has not been developed yet. Therefore, we propose a matching algorithms and artificial intelligence algorithms to improve the accuracy and immediacy of object classification.

In this paper, we present 3D environment detection using multi-view color images and LiDAR point clouds. We will discuss proposed method and system in Section II. Then, the experimental results are discussed in Section III. Finally, the conclusion is addressed in Section IV.

II. PROPOSED METHOD

Generally, the object classification algorithm is used for the color images. However, when the self-driving system detects surrounding objects, LiDAR provides related information as the first-hand data. Therefore, when the object classification algorithm is applied on the point cloud information for analysis, it will achieve faster judgment. In recent years, the rise of deep learning research has also led to the rapid evolution of object classification algorithms, then we proposed a convolutional neural networks to train the marked point cloud data for object segmentation and classification.

A. Datasets

In order to train the point cloud data, we have to prepare a well-marked ground truth dataset. The KITTI database [2] provides a wealth of point cloud data and ground truth of road marking information which is shown in figure 2. In addition, KITTI provides a well-calibration formula for matching the point cloud data and color image. The parameter x, y, z is the LiDAR three-dimensional coordinates and u, v is color image two-dimensional coordinates. Then we match the LiDAR point cloud data and well-marked color image through formula (1) and get a marking point cloud data. Then we use the convolutional neural network to train and classify the data.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

B. Network Architecture

The neural network is usually begun with the convolution layer and followed a pooling layer for the purpose of increasing the receptive fields and reducing the time of

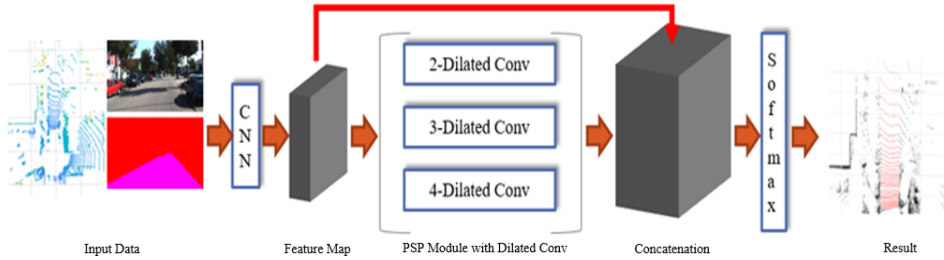


Figure 1 Proposed neural network architecture

calculation. However, the pooling layer will cause excessive message loss, resulting in the problem of distortion when restored. Dilated convolution layer [3] is a transformation of convolution layer (Formula (2)). *F* is the image input, *k* is the kernel function, *l* is the expansion rate.

$$(F *_l k)(\mathrm{p}) = \sum_{s+lt=p} F(s) \times k(t) \tag{2}$$

The advantage of dilated convolution is that it can increase the range of perception without reducing the picture size and using fewer parameters.

Our architecture (figure 1) is refer to the pyramid scene parsing network (PSP) architecture which is also known as PSP net [4]. PSP net improves the structure of the previous FCN-8s structure. After the convolution layer, the pooling layer is not by the way of addition, but through the way of concatenation. The experiment results of the combination of dilated convolution and PSP net are confirmed that perform better than the previous FCN structure.

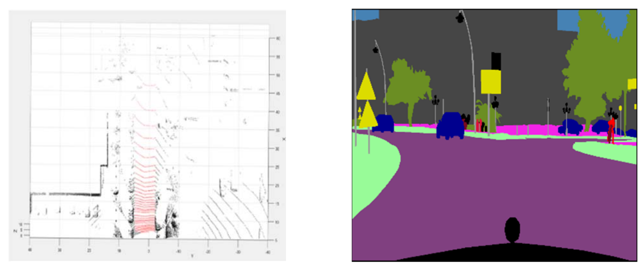
III. Experimental Results

We use the point cloud data from the KITTI database that adopts 64 channel LiDAR. Its scanning range is up to 120 meters. Vertical scanning angle is from +2.0 to -24.9 degrees and horizontal scanning angle is from 0 to 360 degrees. Its rotational speed is 5-20 rotations per second. The error range is about 2 centimeters.

Our convolutional neural network (CNN) is trained efficiently during 20K iterations and batch size is 5. The learning rate is set to 1×10^{-3} and the convolution filter size is set to 3 by 3. After well-trained, we can obtain a point cloud segmentation network which takes about 1 seconds to process 1 million points. Experimental results show this network architecture can segment objects in real-time. The segmentation result is shown in figure 3. We can see that the road is marked as red color, and it is accurately segmented.

IV. Conclusion

We present a method that combines LiDAR point cloud map and color image. We use the well-marked color images to train the driving road recognition and use the recognition results to mark object on point cloud map. Then we through the neural network training to determine the three-dimensional dynamic environment and cut out obstacles such as cars and roads. Afterward, we perform "3D environment detection system using multi-view color images and LiDAR point clouds" and further realize the goal of autonomous vehicles.



(a) (b)

Figure 2 (a) Road marking map; (b) color marker image.

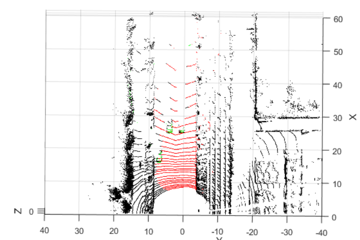


Figure 3 Result with marked road and car

Reference

[1] Y. Zhang, J. Wang, X. Wang, C. Li, and L. Wang, "A Real-Time Curb Detection and Tracking Method for UGVs by Using a 3D-LiDAR Sensor," in *IEEE Conference on Control Applications (CCA)*, Sydney, Australia, Sept. 21-23, 2015, pp. 1020-1025.

[2] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision Meets Robotics: The KITTI Dataset," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231-1237, 2013.

[3] J. Li, Y. Wu, J. Zhao, L. Guan, C. Ye, and T. Yang, "Pedestrian Detection with Dilated Convolution, Region Proposal Network and Boosted Decision Trees," in *International Joint Conference on Neural Networks (IJCNN)*, May 14-19, 2017.

[4] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network, " in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.