International Journal of Advanced Smart Convergence Vol.9 No.3 192-198 (2020)
http://dx.doi.org/10.7236/IJASC.2020.9.3.192

IJASC 20-3-24

Object Detection with LiDAR Point Cloud and RGBD Synthesis Using GNN

Tae-Won Jung*, Chi-Seo Jeong**, Jong-Yong Lee*** and Kye-Dong Jung***

**Doctoral Student, Department of Realistic Convergence Contents KwangWoon University Graduate School, 20 Kwangwoon-ro, Nowon-gu, Seoul 01897, Korea*
***Master Student, Department of Information System KwangWoon University Graduate School, 20 Kwangwoon-ro, Nowon-gu, Seoul 01897, Korea*
****Professor, Ingenium College of liberal arts, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 01897, Korea*
E-mail : {onom ,xhichi, jyonglee, gdchung }@kw.ac.kr

Abstract

The 3D point cloud is a key technology of object detection for virtual reality and augmented reality. In order to apply various areas of object detection, it is necessary to obtain 3D information and even color information more easily. In general, to generate a 3D point cloud, it is acquired using an expensive scanner device. However, 3D and characteristic information such as RGB and depth can be easily obtained in a mobile device. GNN (Graph Neural Network) can be used for object detection based on these characteristics. In this paper, we have generated RGB and RGBD by detecting basic information and characteristic information from the KITTI dataset, which is often used in 3D point cloud object detection. We have generated RGB-GNN with i-GNN, which is the most widely used LiDAR characteristic information, and color information characteristics that can be obtained from mobile devices. We compared and analyzed object detection accuracy using RGBD-GNN, which characterizes color and depth information.

Keywords: *3D Point Cloud, Graph Neural Network, RGBD, LiDAR intensity, Depth camera*

1. Introduction

With the development of mobile devices, virtual reality and augmented reality can be experienced on smartphones and tablets [1]. The 3D point cloud is important information that recognizes objects in virtual reality and augmented reality using mobile devices. The camera and ToF (Time of Flight) of the mobile device allow you to acquire color images RGB data and can obtain depth data. It also recognizes the shape and color of objects similar to human vision, resulting in high object detection performance [2]. The conventional method of using sensors emits lasers to represent signals reflected from objects within the measurement range as point cloud data. LiDAR (Light Detection and Ranging) intensity and ToF are complementary to RGB cameras because they can accurately measure the distance from objects using the surface information of objects. The

Manuscript Received: July. 24, 2020 / Revised: July. 28, 2020 / Accepted: August. 5, 2020
Corresponding Author: gdchung@kw.ac.kr
Tel: +82-2-940-5288
Professor, Ingenium College of liberal arts, Kwangwoon University, Seoul, Korea

development of sensor convergence technology is actively taking place to enhance object detection performance by converging information from sensors [3]. The advantages and disadvantages of RGB and LiDAR are as follows. Because LiDAR uses lasers, it is possible to detect objects even in dark environments. However, in light-absorbing colors, such as black, the detection capability is low. Although RGB is heavily influenced by light, it clearly distinguishes objects from LiDAR.

In this paper, RGB and Depth map are used to apply to various fields of object detection as the only information obtained from mobile devices. Using GNN with information on LiDAR reflectivity, RGB, and RGBD characteristics in the 3D point cloud environment, the performance of i-GNN, RGB-GNN, and RGBD-GNN object detection performance is verified through comparison of performance after learning.

2. Related research

2.1 3D point cloud

The 3D point cloud is a collection of points belonging to a coordinate system. In a 3D coordinate system, the point is usually defined by the coordinates X, Y, and Z, and is often used to represent the surface of an object. The point cloud can be obtained with a 3D scanner. The 3D scanner automatically measures a number of points on the surface of an object and sometimes outputs the point cloud generated by it to a file. A set of points measured by the above machine is sometimes expressed as a 3D point cloud. These 3D point clouds are used in a variety of areas for automation, including CAD (Computer Aided Design) modeling, meteorology, quality testing, visualization, animation, rendering, and mass customization, and scanning sensors such as LiDAR (Light Detection and Ranging) are mainly used [4]. LiDAR is a technology for detecting objects and measuring distances using lasers, enabling accurate distance measurements from objects, including reflectivity information according to surface properties and distance information over time reflected. However, because it measures only reflected laser signals, it represents environmental information that is only included in the reflective area, which limits the actual environment to express all the information in the actual environment as the resolution of the data represented by the point cloud data is very small within 10% of the image data [5]. LiDAR can be divided into laser transmission and reception modules and signal processing modules and can be divided into ToF (Time of Flight) and PS (Phase shift) depending on the laser signal modulation method. The ToF method fires a laser to measure the distance by measuring the time the laser pulse signal is reflected back to the object ahead, and the PS method measures the distance by measuring the variation of the phase reflected on the object ahead by releasing continuously modulated lasers at a certain frequency [6]. The ToF-type RGBD camera can flood the infrared pattern to obtain the distance and color to the target pixel at the same time [7]. The use of ToF sensors for AR or VR environments in mobile devices is increasing.

2.2 Graph Neural Network

Graphs are data structures consisting of dots and edges and are mainly used to analyze data that represent connections or interactions. Complex problems can also be simplified into simpler expressions or addressed from a different perspective [8]. GNN (Graph Neural Network) refers to the artificial neural network used in graph structures [9]. Artificial neural networks are usually given inputs in vector or matrix form, whereas in the case of GNN, the input is characterized by a graph structure. GNN receives graph structure and feature information for each node by input. Based on the feature information received as input and the neighbor information that appears within the graph, the vector embedding for each node is obtained as output results. On one layer of GNN, each node uses information from its neighbors on the graph and its own information to create an embedding.

$$E = \left\{ (p_i, p_j) \;\middle|\; \left\| x_i - x_j \right\|_2 < r \right\} \tag{1}$$

When defining a point cloud with N points as $P = \{p_1, ..., p_n\}$, the i point p_i consists of x_i with three-dimensional coordinates. Expression (1) creates edge E by connecting a neighbor point whose distance between $x_i$and x_j, which are the three-dimensional coordinates of any point p_i, p_j, and x_j, is less than the fixed radius r. Use this process to generate a graph $G = (P, E)$ consisting of P and E.

$$\begin{aligned} v_i^{t+1} &= g^t\left(\rho\left(\left\{e_{ij}^t \mid (i,j) \in E\right\}\right), v_i^t\right) \\ e_{ij}^t &= f^t(v_i^t, v_j^t) \end{aligned} \tag{2}$$

Expression (2) is a formula that shows the vertex update process of GNN. The feature of the edge and the feature of the vertex are aggregated to update the feature of the vertex. where $e^t$and v^tmean the edge and vertex characteristics in the t-t iteration. The function $f^t(.)$ calculates the edge characteristic between two nodes. The function $\rho(.)$ indicates the process of aggregating the edge characteristics for each vertex. The function $g^t(.)$ updates the aggregate properties to generate v^{t+1}. This process is carried out as many times as the vertex updates.

In the 3d object detection field, many methods are used to perform detection using 3D Point Cloud. The CNN (Convergence Neural Network) is an effective method in 2D. However, unlike images, point clouds have uneven spacing because there is an empty space between points. Therefore, there is an empty space in the grid space for CNN processing, which causes wasted computations in the context of the convolution operation. PointNet techniques are also effective methods for object detection, but they should be sampled and grouped for each calculation process for detection. Grouping and sampling large point clouds every time is a waste of resources. By comparison, GNN uses only the characteristic information that each point has using graph structure, thus reducing waste in computation. It can also prevent waste of resources by continuously updating the characteristics using the first generated graph rather than a grouping or sampling them every time to extract them [10].

3. Data Preprocessing for Supervised Learning

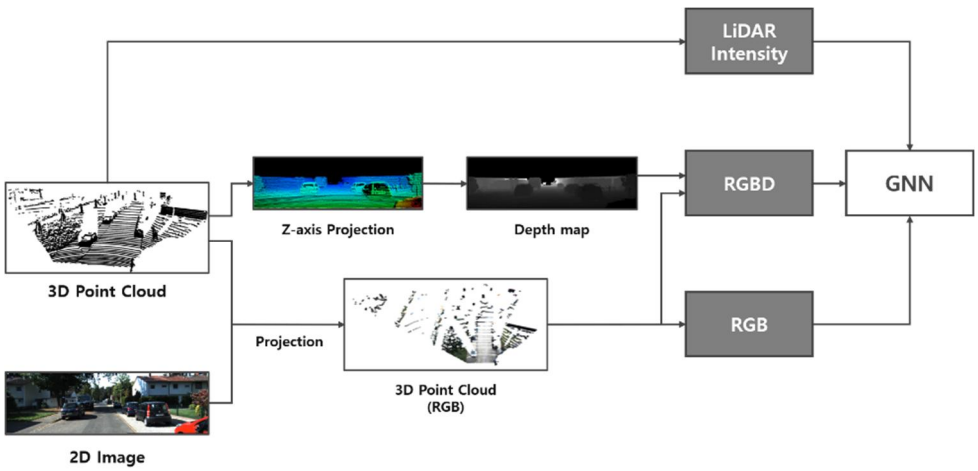

Figure 1. Point Feature for Object Detection System (GNN)

In this paper, the whole processing process was constructed as shown in Figure 1 to compare object detection performance for each characteristic [10]. We used LiDAR reflective strength, RGB and RGBD as

characteristic information. LiDAR reflectance strength used point cloud. The RGB information was extracted from the 2D image and the Depth information was generated using the Z-axis value of the point cloud.

- RGB Preprocessing

The LiDAR point cloud in the KITTI dataset generally does not have color information. Therefore, a 3D RGB point cloud projected to the LiDAR point cloud was created based on the 2D RGB image as shown in Figure 2. It also matched the position and angle of the 2D image camera.

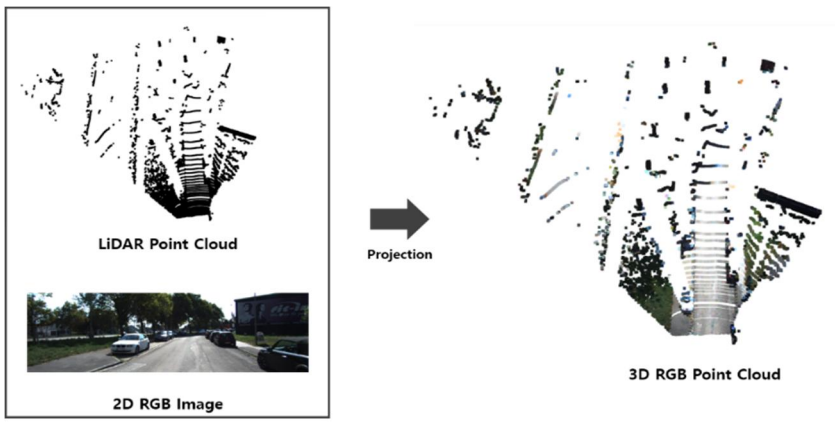

Figure 2. Generate RGB Point Cloud

- RGBD Preprocessing

The LiDAR Point Cloud of the KITTI dataset had no depth map information, so it projected based on camera points and 2D images to create an RGB point cloud depth map. The KITTI data set provides reflected laser signals with three-dimensional coordinate values (x, y, z) and reflectivity information (r), which means the strength of the reflected signals according to the roughness, color, and material of the reflecting surface of the ground and objects [11]. Object detection using this is divided into the detection of objects by using three-dimensional coordinate values or projecting them into two-dimensional spaces in the top view or the front view. Object detection using top views is easy to extract the direction of progress and the speed of motion of the vehicle, but the computation process of object detection is complex, where object detection using RGB cameras and the same front view as the driver's view is simple compared to object detection using top views [12]. In this paper, depth information was generated in a 2D RGB image to compare the performance of RGBD-GNN.

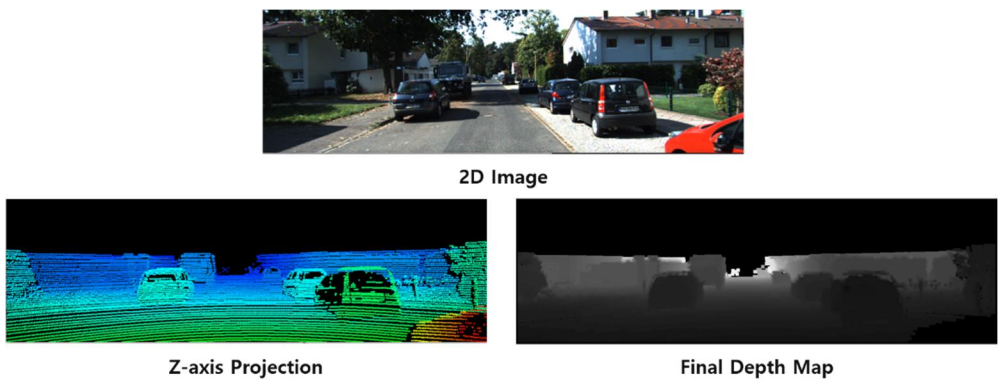

Figure 3. Generate Depth map

For the Depth map, as shown in Figure 3, the Z-value of the point cloud was projected relative to the camera point, then the grid size was adjusted to generate the average value, and the Depth Map was also created to match the resolution of the 2D image.

4. Experimental Results

The KITTI dataset used in the evaluation in this paper is a vehicle equipped with sensors such as RGB camera and LiDAR, which is extracted from urban areas and consists of 7481 sequence learning data. Out of 200 car datasets, 100 were used for learning and 100 were used for testing. The algorithms of the neural network for learning are Ubuntu 16.04, GTX 1080 Ti (11GB) for workstations that use GNN, and Cuda V10.0, Cudn 4.5.3, and Opencv 4.2.0 for GPUs. The number of epochs was set at 1000 and each GNN model was learned. Object detection experiments using LiDAR reflective information (i-GNN), RGB characteristic information (RGB-GNN), and RGB depth characteristic information (RGBD-GNN) were conducted. In Object detection, the accuracy of the model can be assessed through an mAP (mean Average Precision), the higher the accuracy, and the lower the accuracy. As a result of the comparative evaluation experiment, LiDAR is a unique characteristic of learning. Similar results were found when comparing reflective information with the mAP of RGB and RGBD. Comparing RGB and RGBD mAP, we found that each mAP was similar but slightly increased within the margin of error.

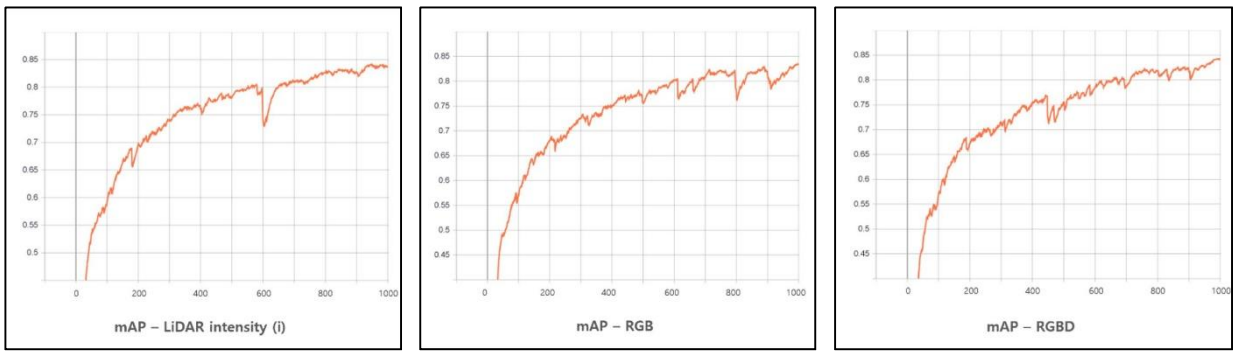

Figure 3. mAP (LiDAR Intensity, RGB, RGBD)

Figure 3 shows the mAP results of object detection with LiDAR reflective strength information and RGB characteristic information. Learning LiDAR reflection strength information and RGB characteristic information, comparing object-aware mAP, formed a similar type of graph within the margin of error. It has been confirmed that object detection can be done with RGB and RGBD characteristic information without using LiDAR reflection strength.

Table 1. Total Loss comparison of 3D object detection

Feature	Epoch 250	Epoch 500	Epoch 750	Epoch 1000
I (LiDAR)	0.1378	**0.1006**	**0.0850**	0.0760
RGB	0.1416	0.1146	0.0885	0.0771
RGBD	**0.1356**	0.1183	0.0873	**0.0745**

Table 1 shows the total loss results used as characteristic information for LiDAR strength i-RGB and RGBD. Total loss is a measure of both cls loss, loc loss, and reg loss, and the result is a comparison of total loss learned

using the entire data Epoch 250, Epoch 500, Epoch 750, and Epoch 1000 times. In Epoch 250 the total loss of RGBD was the lowest, and in Epoch 500 and 750, LiDAR characteristic information i was also similar in learning of RGBD characteristic information in Epoch 1000.

Table 2. mAP comparison of 3D object detection by class

Feature	Background	Car(vertical)	Car(horizontal)	DontCare
I (LiDAR)	**0.9986**	**0.7676**	**0.8089**	0.5472
RGB	0.9985	0.6885	0.7818	0.5509
RGBD	0.9919	0.7371	0.7843	**0.5528**

Table 2 shows the mAP results by class as a dataset for testing. Classes were divided into four categories. The background is the part excluding Object and Car is composed of vertical and horizontal according to the front and rear sides. DontCare means that the object is too small to judge or has truncation. The test results showed that for the Car class corresponding to the Object, all classes except DontCare had the highest accuracy when using the i characteristic information. For the RGB and RGBD characteristics, RGBD had similar accuracy on the side of the car, but for the front and rear sides of the car, RGBD had slightly higher accuracy.

5. Conclusion

ToF sensors are increasingly being installed for AR and VR technologies in mobile devices. Therefore, in a mobile environment, it is appropriate to use RGB characteristic information and depth map instead of characteristic information obtained from a LiDAR sensor. In this paper, we compare object detection accuracy in the mobile device environment using LiDAR, RGB, and RGBD information. These characteristics were learned using GNN on a 3D point cloud base. The RGB characteristics for learning GNN created a 3D RGB point cloud by projecting the 2D image of the KITTI dataset into the 3D point cloud, and the RGBD characteristics also extracted depth information from the LiDAR point cloud.

As a result of the experiments learned with GNN, similar object detection accuracy was confirmed within the margin of error when the RGB and RGBD characteristics were compared with object detection using LiDAR characteristic information.

In the future, we plan to check the accuracy of object detection by fusion of RGB information, which has high resolution but is sensitive to external environmental factors, and ToF information, which has little effect on external environmental factors, but low resolution. Also, we continue to improve the object detection performance by improving the algorithm and RGBD GNN model to extract depth information without using the ToF sensor in mobile devices.

Acknowledgement

The work reported in this paper was conducted during the sabbatical year of Kwangwoon University in 2020.

References

[1] Budiman, Sutanto Edward; Lee, Suk-Ho. Virtual Reality Image Shooting for Single Person Broadcasting with Multiple Smartphones. International Journal of Internet, Broadcasting and Communication, 2019, 11.2: 43-49. DOI: https://doi.org/10.7236/IJIBC.2019.11.2.43

[2] Yi, Chuho; Cho, Jungwon. A Real-time Plane Estimation in Virtual Reality Using a RGB-D Camera in Indoors. Journal of Digital Convergence, 2016, 14.11: 319-324.
DOI: https://doi.org/10.14400/JDC.2016.14.11.319

[3] Kim, Jinsoo; Cho, Jungho. YOLO-based real-time object detection through RGB image and LiDAR point cloud synthesis. Journal of the Korean Society of Information Technology, 2019, 17.8: 93-105.
DOI: https://doi.org/10.14801/jkiit.2019.17.8.93

[4] Kim, J.; Kwon, K. K.; Lee, Su In. Trends and applications on LiDAR sensor technology. Electronics and Telecommunications Trends, 2012, 27.6: 134-143.

[5] Premebida, Cristiano, et al. High-resolution lidar-based depth mapping using bilateral filter. In: 2016 IEEE 19th international conference on intelligent transportation systems (ITSC). IEEE, 2016. p. 2469-2474.
DOI: https://doi.org/10.1109/ITSC.2016.7795953

[6] Robinson, Rod, et al. Infrared differential absorption Lidar (DIAL) measurements of hydrocarbon emissions. Journal of environmental monitoring, 2011, 13.8: 2213-2220.
DOI: https://doi.org/10.1039/c0em00312c

[7] Scarselli, Franco, et al. The graph neural network model. IEEE Transactions on Neural Networks, 2008, 20.1: 61-80. DOI: https://doi.org/10.1109/TNN.2008.2005605

[8] Teney, Damien; Liu, Lingqiao; Van Den Hengel, Anton. Graph-structured representations for visual question answering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1-9. DOI: https://doi.org/10.1109/CVPR.2017.344

[9] Kwon, Soon Chul, et al. A Study on Depth Information Acquisition Improved by Gradual Pixel Bundling Method at TOF Image Sensor. International Journal of Internet, Broadcasting and Communication, 2015, 7.1: 15-19. DOI: https://doi.org/10.7236/IJIBC.2015.7.1.15

[10] Shi, Weijing; Rajkumar, Raj. Point-gnn: Graph neural network for 3d object detection in a point cloud. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. p. 1711-1719.

[11] A project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago, "The KITTI Vision Benchmark Suite", http://www.cvlibs.net/datasets/kitti/

[12] Kim, Jeong-Hwan; Shin, Yong-Hyeon. A Study on Deep Learning-based Pedestrian Detection and Alarm System. The Journal of The Korea Institute of Intelligent Transport Systems, 2019, 18.4: 58-70.
DOI: https://doi.org/10.12815/kits.2019.18.4.58