IJIBC 21-1-7

AR Anchor System Using Mobile Based 3D GNN Detection

Chi-Seo Jeong[1], Jun-Sik Kim[2], Dong-Kyun Kim[2], Soon-Chul Kwon[3] and Kye-Dong Jung[4*]

[1]*Master Course, Graduate School of Smart Convergence, Kwangwoon University, Korea*
[2]*Bachelor Course, Department of Electronic Engineering, Kwangwoon University, Korea*
[3]*Professor, Graduate School of Smart Convergence, Kwangwoon University, Korea*
[4*]*Professor, Ingenium College of liberal arts, Kwangwoon University, Korea*

E-mail : { xhichi, wnstlr5602, notou, ksc0226, gdchung }@kw.ac.kr

Abstract

AR (Augmented Reality) is a technology that provides virtual content to the real world and provides additional information to objects in real-time through 3D content. In the past, a high-performance device was required to experience AR, but it was possible to implement AR more easily by improving mobile performance and mounting various sensors such as ToF (Time-of-Flight). Also, the importance of mobile augmented reality is growing with the commercialization of high-speed wireless Internet such as 5G. Thus, this paper proposes a system that can provide AR services via GNN (Graph Neural Network) using cameras and sensors on mobile devices. ToF of mobile devices is used to capture depth maps. A 3D point cloud was created using RGB images to distinguish specific colors of objects. Point clouds created with RGB images and Depth Map perform downsampling for smooth communication between mobile and server. Point clouds sent to the server are used for 3D object detection. The detection process determines the class of objects and uses one point in the 3D bounding box as an anchor point. AR contents are provided through app and web through class and anchor of the detected object.

Keywords: *Augmented Reality, Smart Phone, Point Cloud, Deep Learning, GNN, 3D Object Detection*

1. Introduction

In the past, AR service required high-performance equipment, but it has become more accessible through the improved performance of mobile devices [1]. With faster and more stable access to wireless Internet than in the past, AR contents can be provided in real-time from a server without the need to pre-install [2]. To provide an AR service, accurately detect the object is important for the service. There are methods for recognizing objects in 2D and methods for recognizing using 3d, such as point clouds. A point cloud constructs a 3d space by representing the depth of an object through points and is obtained using images and depth maps to create a point cloud [3]. RGB information can be obtained through camera images and depth information of space can be obtained through a depth map. It has spatial information compared to 2D, making it easy to

Manuscript Received: November. 15, 2020 / Revised: November. 18, 2020 / Accepted: November. 23, 2020
Corresponding Author: gdchung@kw.ac.kr
Tel: +82-2-940-5288
Professor, Ingenium College of liberal arts, Kwangwoon University, Seoul, Korea

distinguish objects attached in 2D. Besides, when using ToF, it is possible to create a point cloud because depth information can be found even in a dark environment [4]. In this paper, we generate point clouds using information from mobile devices. The generated point cloud is downsampled and sent to the server and converted to graphs for efficient recognition. The converted graph is used as input to the GNN detection system. GNN generates a type of object and a three-dimensional bounding box and specifies one vertex of the bounding box as an anchor point. The specified Anchor point is converted to the camera coordinate system currently being filmed and provides AR content in real-time for the recognized object [5].

2. Related work

2.1 AR Anchor

Holding an accurate anchor on space or plane is important to provide AR to the desired location. An anchor is a reference point indicating a position to float a virtual object in space and can be simply expressed in a spatial coordinate system [6]. In AR implementation, when an object needs to have directionality, it is a concept that includes not only the x, y, and z coordinates in space but also the rotation value of each axis [7]. The method of holding the anchor is largely divided into 4 types. The first is Image Anchor, which recognizes the image as a marker, and the second is Ground Anchor, which recognizes the floor surface and implements it. The third is Point Cloud Anchor, which is generated based on the characteristics of the point cloud, and the last is Object Anchor, which is generated by object recognition. It is appropriate to use an object anchor to provide AR services for a specific object. In the past, the method of estimating the anchor point by extracting the feature point has been used a lot, but recently, the method of using the anchor point by combining it with deep learning is also widely used.

2.2 WebXR

WebXR is a framework for providing AR services through a web browser [8]. Generally, it is necessary to download applications suitable for the operating system in advance to receive AR or VR services on mobile devices. In this case, the disadvantage is that it is less accessible because it uses the mobile device storage and needs to store AR contents necessary for the service in advance on the device. WebXR provides the service through a web page, so there is no need to receive additional applications or content, and it provides it in real-time. Previously, it was difficult to provide AR content through web pages because of its large capacity, but real-time service became possible due to the development of communication technology. In the case of AR, camera permissions are obtained and used through a web browser, and the service is provided through a coordinate system created by WebGL.

2.3 3D Graph Neural Network

The use of CNN (Convolution Neural Networks) is typical for 3D object detection using deep learning. For 3D CNN, we construct a rectangular grid space based on voxel and extract feature maps via the convolution process. Although there are advantages of identifying regional and overall characteristics depending on the kernel size, there is a disadvantage of having to perform operations to space based on grid and convolution for each kernel size to extract features [9]. 3D GNN generates an edge list by selecting points to use as a vertex in a graph by converting the point cloud into a graph [10]. In learning, there is no need for computation on empty spaces because feature maps are aggregated around the vertex and feature operations are performed with the vertex, and point operations are performed. It also has the advantage of being computationally wasteful because it recycles existing constructed structures rather than creating new graph structures to extract features.

Figure 1 illustrates the mobile AR service process [11]. It scans objects through mobile devices and generates 3D point clouds via RGB images and depth maps obtained through scans. We transform point clouds into graphs, obtain anchor points through the GNN detection process, and provide AR content tailored to objects.

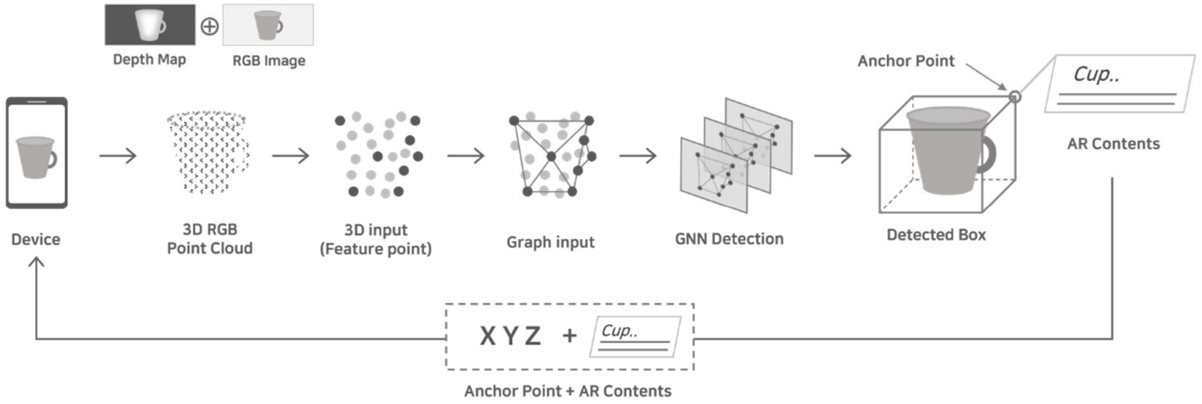


Figure 1. Process of mobile AR

3. Design of Proposal System

Figure 2 shows the learning process for providing AR services. Point Cloud Generator creates a point cloud with a Depth map created on the mobile. The generated point cloud is moved to the Data Labeling Processor through process ① to proceed with labeling. The position and size of the box that contains the object are manually set and converted into a label file for learning. The generated label and point cloud are passed through process ② to the Graph Generator of the GNN Model Generation Layer. The GNN Detection Processor consists of a Graph Generator that converts point clouds into graphs, a GNN Learning Processor with a learning layer, and a Model Saver that stores the final model from learning. Graph Translator splits the entered point cloud into spaces by voxel units and extracts key points to be used as nodes in the graph in each voxel. Key points are carried out by randomly picking one point from the point clouds included in the voxel to prevent overfitting during learning. It generates an edge list that represents the connection relationship of the point cloud around the generated key point and is passed to the GNN Learning Processor through process ③. The GNN Learning Processor module aggregates the unique RGB characteristics of key points into vertices which are the center of each graph through graph operation. Perform classification and box detection through the feature map of the implicated key point. As a certain level of learning progresses, the Model Saver is called through process ④ to save the corresponding learning model. The stored model is passed to the Box Recombination module via process ⑤, where the predicted box for each key point is confirmed as one box per object. The point in the upper right corner of the fixed object is transferred to the Anchor Translator through process ⑥ and converted to the anchor point for AR service.

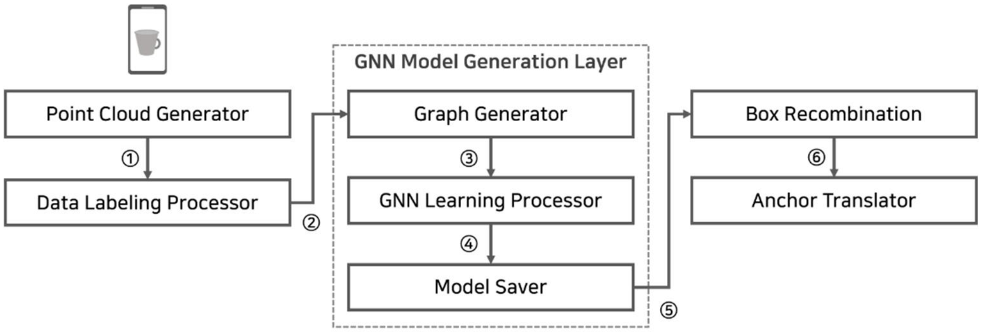


Figure 2. GNN learning flow

Figure 3 shows the overall framework with modules that compose the AR service. The information obtained through the mobile device is converted into a point cloud by entering the Point Cloud Generator in the hybrid app through the ⓐ process. The downsampled point cloud for mobile devices is provided to the GNN Processor on the server through the ⓑ process. The GNN Processor converts the point cloud into a graph and performs object recognition. Object class and box are derived through object detection. It is transferred to AR Content Mapper through the ⓒ process. AR Contents Mapper has AR contents, maps AR contents suitable for the recognized object, and designates a point at the top right of the detection box as an anchor. The designated anchor and AR contents are delivered to the hybrid app through the ⓓ process and are converted into points in the camera coordinate system for AR service by the anchor translator. Finally, the service is provided to users through an AR framework, such as WebXR, through the ⓔ process.

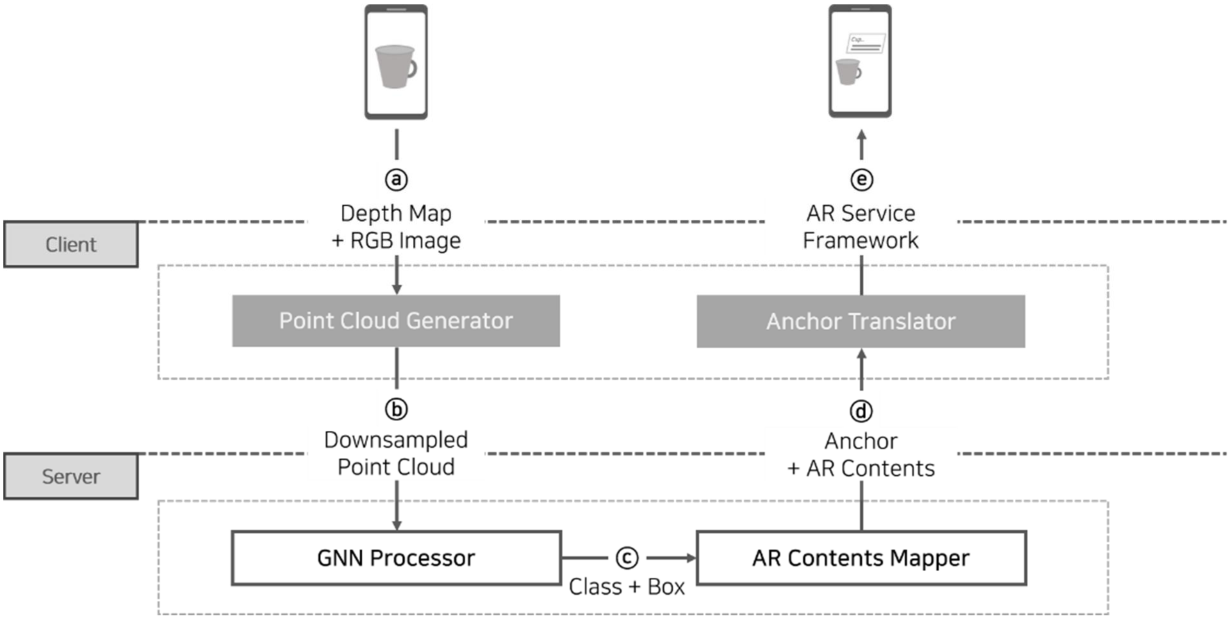


Figure 3. AR service framework

4. Example of Applying

4.1 AR Service based on GNN Detection

For the test, we used a Sony Xperia 1i product, a mobile device equipped with ToF. Depth information from the ToF sensor was used through Android's Camera 2 API. To use color information, we created an RGB point cloud using an RGB image from a camera. Figure 4 shows the depth map and RGB image extracted through the device. After creating a point cloud in a mobile device using the two data, downsampling was performed. A total of 100 datasets were created from the data, and 80 for training and 20 for testing.

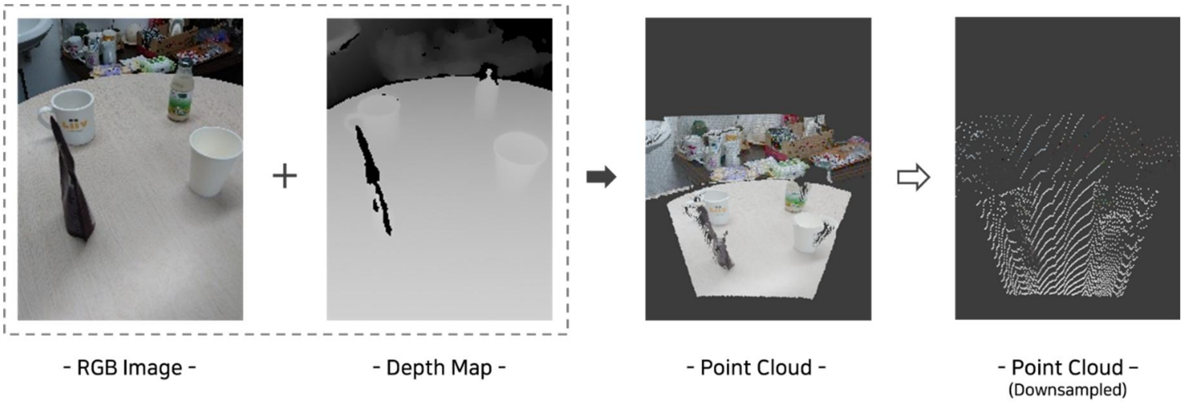


Figure 4. RGB point cloud (using mobile)

Figure 5 is a framework configured for actual implementation. The point cloud created through the depth map and RGB image was downsampled and transmitted to the server for smooth communication with the server. The point cloud was transmitted to the server in binary form using socket communication and saved as a ply file. File Checker determines whether or not a file is saved and proceeds with GNN Detection. For prediction through GNN, a pre-trained model was used. If the learned object is detected, the anchor point can be saved as a file. We checked whether the anchor point file saved in the web server was saved or not, and implemented AR using AR contents that fit the anchor point and object.

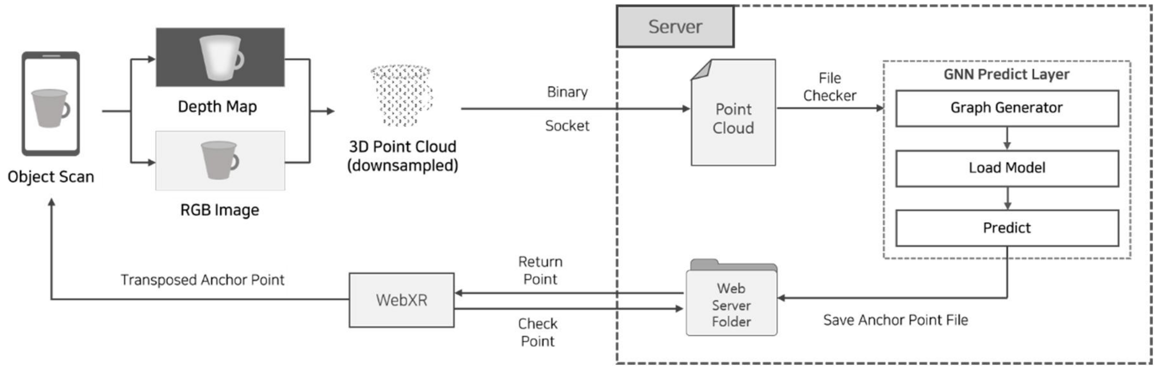
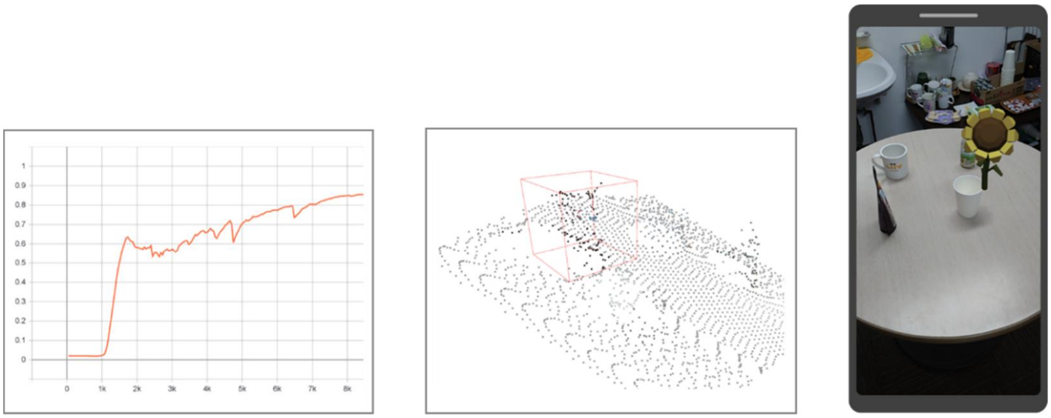


Figure 5. Implemented framework

Figure 6 is the result of detecting the cup to provide service from the input point cloud. Based on the result, the sunflower model, which is WebXR's basic AR content, is displayed on the cup by using Hit Test function.



- Learning Result (mAP) -

- Detection Result -

- WebXR -
(Hit Test)

Figure 6. Detection result and WebXR implementation

4.2 AR System Comparison

Table 1. AR system comparison of implementation

	Proposed GNN-based AR System	AR Framework (WebXR) based AR System
Input Data Generation	Mobile device (RGB + ToF)	Mobile device (RGB)
AR Service Providing Method	GNN object detection	User’s hand touched event
Key Point Extraction	Point cloud downsampling	Feature point extraction
Anchor Method	Object anchor using object detection	Ground anchor using plane detection

Table 1 compares the general AR framework and the proposed system. In the case of input data, both methods use data generated by mobile devices. However, the GNN-based AR system uses a ToF sensor because it is based on a point cloud. The proposed system provides AR services through detection, and the AR framework provides users' touch events. In the case of a GNN-based AR system, key point extraction is used by downsampling the point cloud, and the AR framework uses feature point extraction. Lastly, the anchor point uses object detection-based object anchors for the proposed system, and the ar framework uses plane detection-based ground anchors.

5. Conclusion

The development of mobile devices and wireless internet has become a platform suitable for providing AR content to users. The improved performance of mobile devices made it possible to provide augmented reality, and it was possible to provide real-time, rather than pre-storing, AR content through high-speed wireless internet. Besides, by mounting sensors such as ToF, more accurate spatial information can be used, thereby providing high-quality AR services. Therefore, in this paper, we designed a system for mobile AR service using GNN. The proposed system provides AR through Object detection. A point cloud was created using a sensor of a mobile device, downsampling was performed and then converted into a graph. At this time, the

RGB image from the camera was also used to detect objects of specific colors. The point cloud converted into a graph was object detection using GNN, class identification, and bounding box generated. The Anchor point is created using the information and AR service suitable for the detected object is provided. The features of the designed system are as follows.

First, as mobile devices have recently been equipped with sensors such as ToF, we designed the system to use those sensors. Second, point clouds generated by mobile devices were downsampled and transmitted for stable and smooth communication with the server. Third, we graph the point cloud through a detection system utilizing GNN, enabling efficient aggregation of information in the point cloud.

In the future, we plan to increase accuracy by reducing the amount of computation and improving the GNN system by using feature points rather than simple point clouds. Also, the database should be configured to support multiple clients and objects to provide services.

Acknowledgement

This research was supported in 2020 by The Ministry of Education, Republic of Korea, under the Program of university for innovation and Transformation.

References

[1] HONG, Dong-Pyo, et al. A Study on the Mobile Augmented Reality system trends. Communications of the Korean Institute of Information Scientists and Engineers, 2008, 26.1: 88-97.

[2] JEON, Jong-hong; LEE, Seung-yoon. The trend of standardization of mobile augmented reality technology. Electronic Communication Trend Analysis, 2011, 26.2: 61-74.

[3] LI, Xingdong, et al. Generating colored point cloud under the calibration between TOF and RGB cameras. In: 2013 IEEE International Conference on Information and Automation (ICIA). IEEE, 2013. p. 483-488. DOI: doi.org/10.1109/ICInfA.2013.6720347

[4] Jung, Tae-Won; Jeong, Chi-Seo, et al. Object Detection with LiDAR Point Cloud and RGBD Synthesis Using GNN. International Journal of Advanced Smart Convergence, 2020, 9.3 p. 192–98. DOI: doi.org/10.7236/IJASC.2020.9.3.192

[5] COUCLELIS, Helen, et al. Exploring the anchor-point hypothesis of spatial cognition. Journal of environmental psychology, 1987, 7.2: 99-122.

[6] LEE, Yongjae, et al. Unified Representation for XR Content and its Rendering Method. In: The 25th International Conference on 3D Web Technology. 2020. p. 1-10. DOI: doi.org/10.1145/3424616.3424695

[7] LEE, Jaehyun, et al. A Study on Virtual Studio Application using Microsoft HoloLens. International journal of advanced smart convergence, 2017, 6.4 p. 80-87. DOI: doi.org/10.7236/IJASC.2017.6.4.12

[8] MACLNTYRE, Blair; SMITH, Trevor F. Thoughts on the Future of WebXR and the Immersive Web. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE, 2018. p. 338-342. DOI: doi.org/10.1109/ISMAR-Adjunct.2018.00099

[9] SHI, Shaoshuai, et al. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. p. 10529-10538. DOI: doi.org/10.1109/CVPR42600.2020.01054

[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. DOI: doi.org/10.1109/CVPR42600.2020.00178

[11] Jeong, Chi-Seo, et al. Deep Learning-Based 3D Object Recognition System for Mobile AR. In: 8th International Symposium on Advanced & Applied Convergence, 2020. p. 40-42.