Real-time 3D Perception System of Construction Equipment for Autonomous Mobile Robots

Duho Chung[a], Seunghun Im[a], Yohan Kim[a] and Hyoungkwan Kim[a]

[a]Department of Civil and Environmental Engineering, Yonsei University, Korea
E-mail jungduho1@yonsei.ac.kr, sunsetmorning1234@gmail.com, homez815@yonsei.ac.kr and hyoungkwan@yonsei.ac.kr

Abstract –

Automation using mobile robots at construction sites is gaining attention as a key technology to enhance productivity and safety. However, due to the unstructured nature of construction sites with dynamic elements like construction equipment, robots must perceive their positions and movements in real-time to adapt to workflow changes while performing tasks. To address this issue, this paper proposes a system that enables mobile robots to detect the spatial positions of construction equipment in real time. The system integrates a voxel-based deep learning detection model with a Simultaneous Localization and Mapping (SLAM) algorithm, allowing real-time detection of construction equipment within a consistent reference coordinate system using 3D point cloud data collected by the mobile LiDAR. Experimental results demonstrate that the system achieves an average detection accuracy of 74.51% (AP@R40, IoU 0.25) for excavators, dump trucks, graders, and rollers, demonstrating its capability to accurately recognize the spatial information of construction equipment. This approach highlights the potential for mobile robots to effectively perceive construction equipment in active construction sites, enabling safer and more efficient task execution.

Keywords –

Point Cloud Data; 3D Object Detection; Deep Learning; Vehicle Tracking; Autonomous Robot; Quadruped Robot.

1 Introduction

To address the persistent labor-intensive challenges in the construction industry, construction automation using mobile robots has emerged as a key technology to enhance productivity and safety on construction sites. Various robotic platforms, including drones and quadruped robots, are being explored for applications such as construction automation, 3D reconstruction, material assembly, and safety monitoring [1]. However, the low level of standardization and the dynamic nature of construction environments present significant challenges for robotic automation [2]. In active construction sites, mobile robots share space with dynamic obstacles like construction equipment. Each type of equipment follows distinct operational patterns, often working within a fixed radius or moving along predefined paths. Failure to recognize these machines can increase collision risks and lead to frequent, unnecessary path adjustments, disrupting workflow. To operate efficiently in dynamic construction environments, mobile robots must detect the location and type of construction equipment in real-time and track their movements over time.

Several studies have explored the automation of tasks using mobile robots on outdoor construction sites by employing sensors to identify and respond to obstacles. Kim et al. [3] automated laser scanning tasks with a mobile robot using initial map generated by a UAV (Unmanned Areial Vehicle) and addressed obstacles through a point cloud-based potential field. Asadi et al. [4] developed a mobile manipulator robot capable of grasping and transporting objects using vision-based environmental understanding, employing camera-generated point clouds for obstacle avoidance. Štibinger et al. [5] proposed a mobile manipulator robot for automated structural assembly, leveraging 3D point clouds to detect obstacles and perform navigation. Chung et al. [6] automated laser scanning of scaffold structures on outdoor construction sites without prior information, utilizing point clouds from a mobile LiDAR and IR camera to navigate around obstacles.

While these studies have demonstrated the automation of construction tasks using robots capable of avoiding dynamic obstacles in outdoor environments, there is a lack of research focused on accurately detecting construction entities and continuously recording their locations using point clouds to support robot operations. Studies such as [7] and [8] have explored situational awareness and interaction between mobile robots and construction workers, aiming to implement Human-

Robot Collaboration (HRC) systems. However, research on perception systems for mobile robots specifically designed to interact with construction equipment in 3D space remains significantly underexplored.

This study proposes a system that uses mobile LiDAR to detect the 3D positions of construction equipment in real-time and record them within a global coordinate frame. Figure 1 provides an overview of the proposed system, which integrates a voxel-based 3D object detection model with Simultaneous Localization and Mapping (SLAM). First, the system processes 3D point cloud data collected by the mobile LiDAR to detect various types of construction equipment, such as excavators, dump trucks, graders, and rollers, in real-time. The detected objects are then managed within a global coordinate frame through SLAM, enabling continuous tracking of construction equipment even in dynamic environments. The entire system is implemented on the Robot Operating System (ROS) framework, allowing seamless integration and real-time processing of sensor data acquired by the mobile robot. Using the proposed system, mobile robots can effectively perceive the locations of construction equipment on-site, providing a foundation for applications such as maintaining safe distances, identifying hazardous zones, and understanding work contexts.

Additionally, mobile LiDAR offers a significant advantage by covering a wide area while being less affected by occlusions, enabling the real-time acquisition of high-density 3D data [9]. Due to these characteristics, it is increasingly being integrated with SLAM for 3D reconstruction and safety management in construction sites [10]. The proposed system enables real-time perception of construction equipment movements using mobile LiDAR, which can be utilized to develop a comprehensive progress monitoring system for construction sites, including quality control of individual structures and process monitoring of construction equipment. Based on the potential of mobile LiDAR-based technology, this system supports mobile robot operations on construction sites and lays the foundation for establishing a comprehensive progress monitoring system for large-scale construction projects.

2 Methodology

The proposed system is designed to accurately identify the spatial positions of construction equipment in real time using 3D point cloud data collected by mobile LiDAR mounted on a robot. The mobile LiDAR continuously captures 3D point cloud data as the robot navigates the environment. This data is processed in real time by a voxel-based 3D object detection model, which detects construction equipment and generates bounding boxes. These detection results are then integrated with a SLAM algorithm, transforming them into a consistent global coordinate system, where they are recorded for further use.

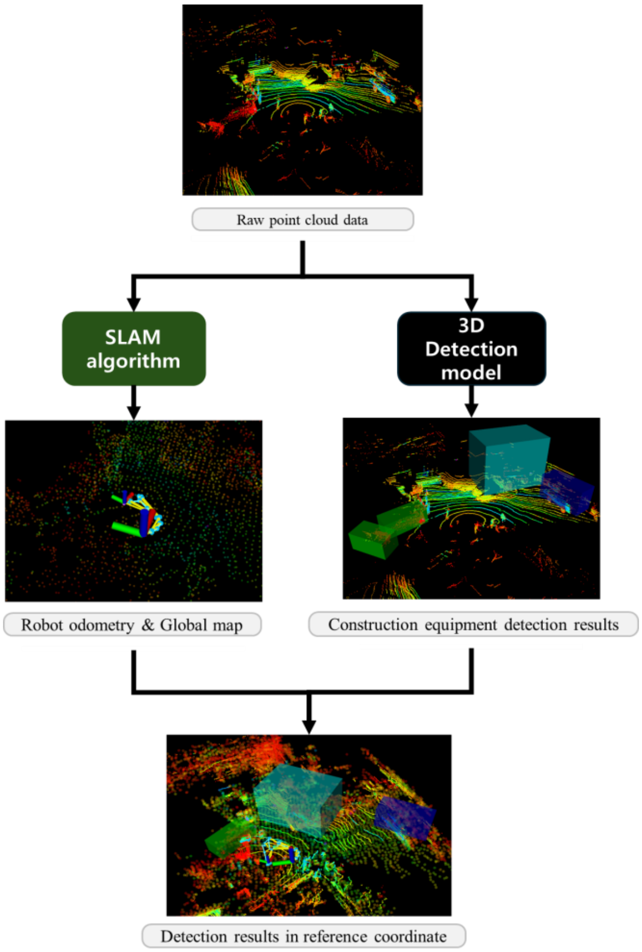

Figure 1. Overview of the proposed system

As shown in Figure 2, the mobile robot system comprises a quadruped robot (Unitree AlienGo), a mobile LiDAR (Velodyne VLP-32), an IMU sensor (Microstrain), and a mini-PC equipped with an NVIDIA GeForce RTX 2060 (Intel NUC 11 PHKi7C). This system was used to acquire the data necessary for model training and experimentation. The mini-PC was directly mounted on the quadruped robot along with a battery, enabling the entire proposed system to operate in a mobile environment. To construct a point cloud dataset for model training, the mobile robot was remotely operated to collect point cloud data at a construction site. During this process, the mobile LiDAR scanned excavators, dump trucks, graders, and rollers from various angles and distances, ensuring diversity in the training data.

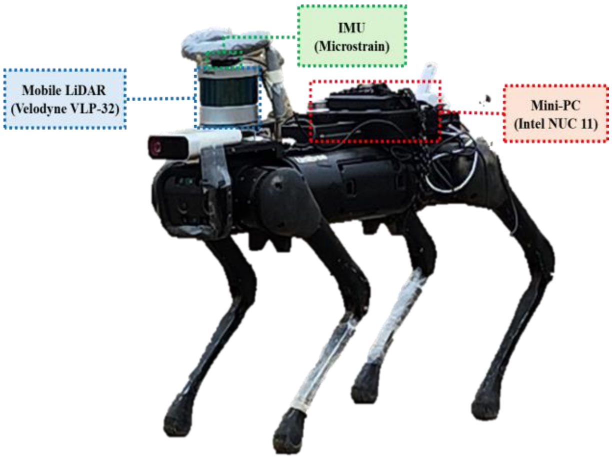

Figure 2. Hardware configuration.

For the mobile robot to understand and respond to the movements of construction equipment, it must rapidly and accurately identify their spatial positions despite limited computational resources. A voxel-based approach efficiently processes 3D point cloud data, significantly reducing computational complexity for large datasets. To achieve real-time detection of construction equipment in 3D point cloud data, we employed the voxel-based 3D object detection model VoxelNeXt [11]. VoxelNeXt utilizes a fully sparse convolutional network to directly predict 3D objects from voxel features, maximizing computational efficiency by eliminating the need for generating dense features after voxel transformation. This makes it particularly well-suited for mobile robot environments with limited computational resources. To train the model efficiently with limited data, VoxelNeXt was pre-trained on the KITTI dataset [12], and transfer learning was performed using this pre-trained model as a starting point. Figure 3 illustrates the architecture of VoxelNeXt. VoxelNeXt utilizes a six-stage Sparse CNN Backbone to process 3D point cloud data. Instead of generating a dense feature map, it applies Sparse Height Compression to create a BEV (Bird's Eye View) feature, preserving only the original data points in sparse format. This approach reduces unnecessary computations and ensures efficient object detection even with limited resources.

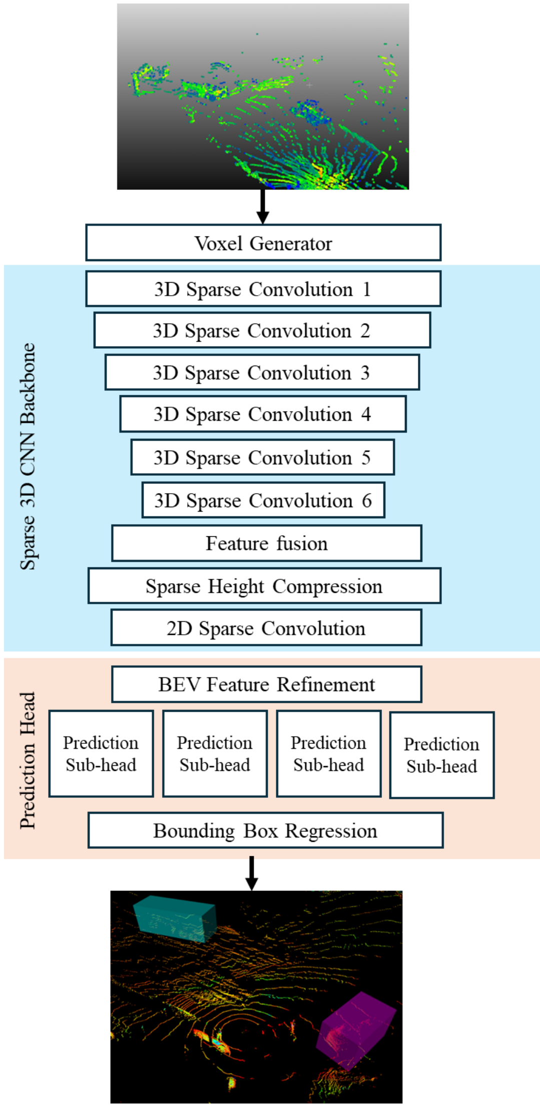

Figure 3. Structure of VoxelNeXt.

In environments where both mobile robots and construction equipment are in motion, integrating detection results into a consistent reference coordinate system is essential for long-term perception of construction equipment positions. To achieve this, the proposed method integrates the SLAM algorithm with VoxelNeXt to process detection results. SLAM is an algorithm that simultaneously maps the environment around a sensor and estimates the sensor's position using the sensor data. The proposed method utilizes the LIO-SAM [13] algorithm, which combines mobile LiDAR and IMU data to generate precise maps. By utilizing odometry information calculated by LIO-SAM, the detected bounding boxes are transformed into the reference coordinate system, enabling accurate localization of construction equipment on the global map of the construction site. Figure 4 illustrates the process of registering detected equipment in the global coordinate frame of LIO-SAM using VoxelNeXt.

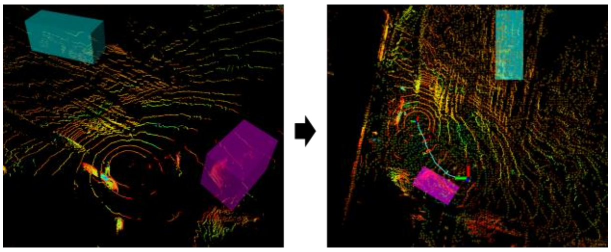

Figure 4. Transformation of detection results: from local coordinates (left) to global map using LIO-SAM (right)

3 Experiments and Results

In this study, a quadruped robot was remotely operated at four large-scale construction sites to collect mobile LiDAR data of construction equipment for VoxelNeXt model training. While navigating active construction sites with moving equipment, the quadruped robot recorded continuous point cloud data. The collected data was captured at one frame per three seconds and converted into PCD file format to construct the dataset.

During the data labelling process, only construction equipment within 40 meters of the robot was selected for annotation. Objects that were visually unidentifiable due to occlusion or motion blur were excluded from labelling. Figure 5 presents examples of the collected data from each site. A total of 1,246 frames of point cloud data were used for training and validation, while 261 frames of point cloud data were reserved for testing. This study focused on training the model to recognize four classes of construction equipment: excavators, dump trucks, graders, and rollers. Detailed class-wise data distribution is provided in Table 1.

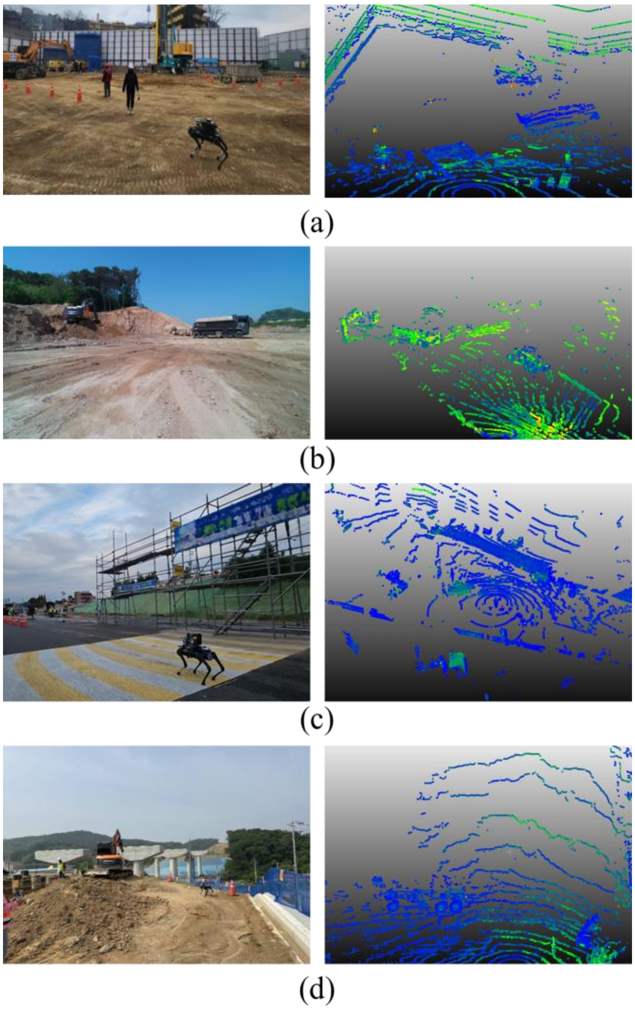

Figure 5. Picture of the site (left) and examples of point cloud data collected at each site(right); Site 1, 2 and 3 were used for training and validation, and Site 4 was used for testing; (a) Site 1, (b) Site 2, (c) Site 3, (d) Site 4.

Table 1. Class-wise dataset distribution

Class	Train	Validation	Test
Excavator	877	175	95
Dump Truck	564	121	71
Grader	199	72	117
Roller	125	32	118
PCD files	1000	246	261

The training of the construction equipment detection model was conducted on an Ubuntu 20.04 environment. Transfer learning was performed using pre-trained weights from the KITTI dataset, with the model trained for 50 epochs using a batch size of 4. The training process employed an AdamW-based one-cycle learning schedule, where the learning rate increased to a maximum of 0.01 during the first 40% of training and decreased over the remaining 60%. These settings were based on the default VoxelNeXt configuration for training on the KITTI dataset.

During the transfer learning process, fine-tuning was conducted under four different configurations, based on the extent to which the 3D sparse convolution layers in the backbone were frozen. These layers correspond to "3D Sparse Convolution 1–6" in Figure 3. 1) Full fine-tuning was performed, with no layers frozen. 2) Only the first two 3D sparse convolution layers were frozen; the rest of the backbone and the detection head were trained. 3) Only the first four 3D sparse convolution layers were frozen; the remaining layers and the detection head were trained. 4) All 3D sparse convolution layers were frozen, while the detection head and subsequent 2D layers remained trainable. The results showed that full fine-tuning without freezing achieved the highest average class performance, which was used as the basis for selecting the final model configuration.

On the test set, the 3D bounding box accuracy based on AP@R40 IoU 0.25 was recorded as follows: dump truck (79.13%), excavator (70.81%), grader (68.74%), and roller (79.36%). Here, R40 (Recall 40) is an evaluation metric used in the KITTI dataset, where Average Precision (AP) is computed based on 40 recall points. The detailed performance evaluation results are presented in Table 2, while Figure 6 visually illustrates the detection results for each class.

To verify whether the proposed system can reliably perceive construction equipment in real-time environments using data acquired by an actual mobile robot, a live experiment was conducted using sequential point cloud data collected from Site 4. The results demonstrated that VoxelNeXt, implemented within the proposed ROS-based system, detected construction equipment at an average speed of 2.8 Hz on the mini-PC. The detection results were continuously integrated with LIO-SAM in real time, allowing the positions of construction equipment to be sequentially localized on the global map. Figure 7 illustrates the detection results of a continuously moving grader and roller over a 60-second period, recorded on the global map in the test environment. These results demonstrate that the proposed system effectively detects and records the spatial information of construction equipment even in real-time operational conditions.

Table 2. The performance of detection model

Class	AP (IoU > 0.5)		AP (IoU > 0.25)	
	BEV	3D	BEV	3D
All Layers Trainable (No Freezing)				
Excavator	66.59	41.97	70.81	70.81
Dump Truck	78.97	47.90	79.31	79.13
Grader	68.45	50.02	70.14	68.74
Roller	75.62	41.37	80.13	79.36
Freeze 3D Sparse Convolution 1-2				
Excavator	54.05	39.60	55.48	55.48
Dump Truck	82.60	45.21	83.59	81.08
Grader	69.85	53.66	75.99	73.92
Roller	73.14	39.01	77.76	76.41
Freeze 3D Sparse Convolution 1-4				
Excavator	42.38	38.32	54.82	53.12
Dump Truck	83.73	43.20	84.08	82.30
Grader	70.22	51.40	75.45	72.77
Roller	65.67	29.02	74.41	74.41
Freeze All 3D Sparse Convolution				
Excavator	20.63	12.45	28.94	28.27
Dump Truck	85.37	36.93	86.38	81.98
Grader	43.80	33.42	53.65	47.59
Roller	70.97	14.55	82.12	61.19

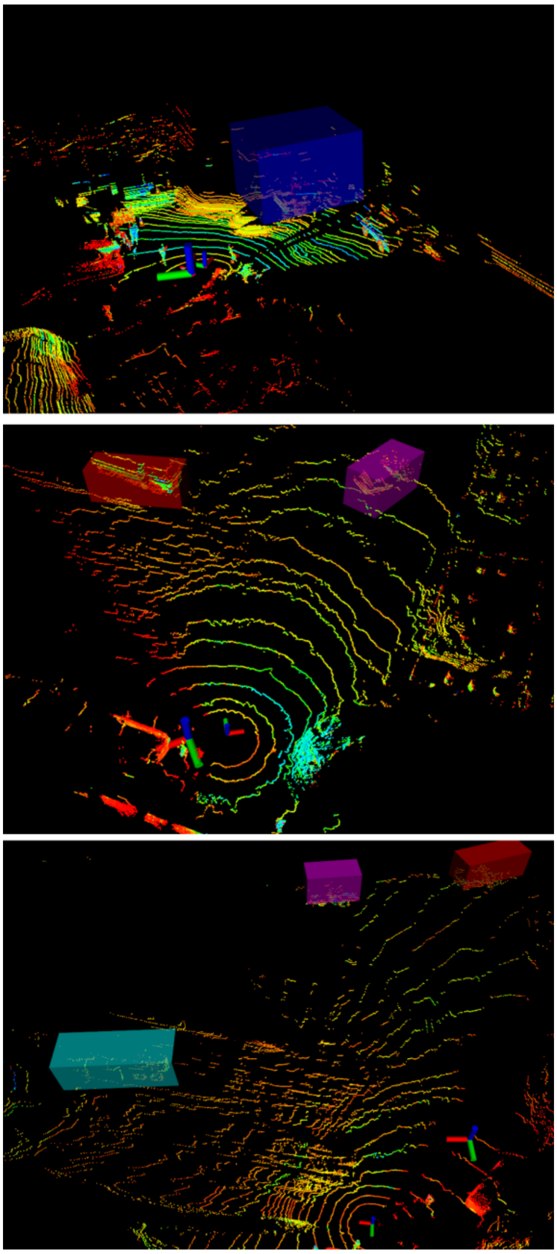

Figure 6. Visual examples of the detection outputs at test dataset: blue boxes represent excavator, red boxes represent truck; cyan boxes represent grader, and pink boxes represent roller.

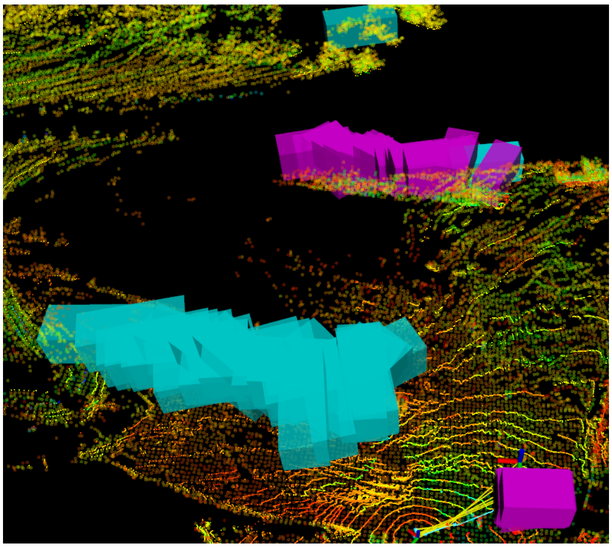

Figure 7. Visualization of grader(cyan) and roller(pink) bounding boxes recorded on the global map at Site 4.

4 Discussion

The experimental results confirmed that the proposed methodology can autonomously perceive construction equipment in real-world construction site data. By leveraging VoxelNeXt, the system demonstrated the ability to detect construction equipment in real time across a wide area and integrate the detection results with LIO-SAM, effectively mapping their positions within a consistent coordinate system. Furthermore, the system successfully recorded equipment positions continuously, enabling long-term perception of their movements and operational ranges. This capability supports mobile robot operations on construction sites and highlights the system's potential as foundational technology for comprehensive progress monitoring in large-scale construction projects.

However, the visually sparse nature of mobile LiDAR point clouds presents challenges in maintaining consistent detection performance, particularly in complex environments. In construction sites where structures and equipment are densely intermingled, achieving robust detection necessitates further research. These sparsity issues become more pronounced when applying technologies originally developed for autonomous driving to complex off-road environments, such as construction sites.

As shown in Table 2, when fine-tuning with more backbone layers frozen from the pretrained model (trained on the KITTI dataset), detection performance remained high for classes similar to those in KITTI, such as dump trucks. In contrast, performance significantly declined for equipment structurally different from KITTI objects, such as excavators. This demonstrates a clear domain gap between autonomous driving datasets and construction site data, indicating that structural diversity of construction equipment and environmental complexity directly influence detection performance.

Therefore, acquiring construction-specific datasets is essential. Given the current lack of standardized 3D point cloud datasets for construction equipment, generating realistic, variability-inclusive virtual data is an effective alternative for dataset augmentation. This approach can significantly enhance data diversity, improve the generalization capability and robustness of detection models, and ultimately strengthen detection reliability in complex construction environments.

5 Conclusion

This paper proposed a system capable of perceiving the 3D positions of construction equipment in real time using a mobile LiDAR. The system integrates a voxel-based real-time object detection model with a SLAM algorithm to efficiently process 3D point cloud data collected by the mobile LiDAR and consistently map detected results in a global coordinate system. Experimental results demonstrated that the system achieves detection accuracies of 79.13% (dump truck), 70.81% (excavator), 68.74% (grader), and 79.36% (roller) based on AP@R40 with 3D IoU 0.25, proving its capability to reliably recognize the spatial locations of construction equipment. Additionally, experiments confirmed that mobile LiDAR-based detection can effectively identify and track a wide range of construction equipment in dynamic construction environments in real time. Future work will focus on enhancing detection performance and reliability by augmenting the dataset with realistic, simulation-based virtual data, thereby improving the generalization capability and robustness of the detection model. Through these advancements, the proposed system has the potential to support efficient mobile robot operations on construction sites and serve as a foundational technology for automated progress monitoring of large-scale construction projects using mobile LiDAR.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (No.RS-2024-00457308) and the "National R&D Project for Smart Construction Technology (RS-2020-KA156488)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

References

[1] L. Zeng, S. Guo, J. Wu, & B. Markert. Autonomous mobile construction robots in built environment: A comprehensive review. Developments in the Built Environment, 19: 100484, 2024.

[2] B. Xiao, C. Chen, & X. Yin. Recent advancements of robotics in construction. Automation in Construction, 144: 104591, 2022.

[3] P. Kim, J. Park, Y. K. Cho, & J. Kang. UAV-assisted autonomous mobile robot navigation for as-is 3D data collection and registration in cluttered environments. Automation in Construction, 106: 102918, 2019

[4] K. Asadi., V. R. Haritsa, K. Han, & J. P. Ore. Automated object manipulation using vision-based mobile robotic system for construction applications. *Journal of Computing in Civil Engineering*, *35*(1): 04020058, 2021.

[5] P. Štibinger, G. Broughton, F. Majer, Z. Rozsypálek, A. Wang, K. Jindal, A. Zhou, & M. Saska, Mobile manipulator for autonomous localization, grasping and precise placement of construction material in a semi-structured environment. *IEEE Robotics and Automation Letters*, *6*(2): 2595-2602, 2021.

[6] D. Chung, J. Kim, S. Paik, S. Im, & H. Kim. Automated system of scaffold point cloud data acquisition using a robot dog. *Automation in Construction*, *170*: 105944, 2025.

[7] Y. Liu, M. Habibnezhad, & H. Jebelli. Brainwave-driven human-robot collaboration in construction. Automation in Construction, 124: 103556, 2021.

[8] Y. Kim, S. Kim, Y. Chen, H. Yang, S. Kim, S. Ha, M. Gombolay, Y. Ahn, & Y. K. Cho. Understanding human-robot proxemic norms in construction: How do humans navigate around robots? *Automation in Construction*, *164*: 105455, 2024.

[9] D. Hauser, C. Glennie, & B. Brooks. Calibration and accuracy analysis of a low-cost mapping-grade mobile laser scanning system. Journal of Surveying Engineering, 142(4), 04016011, 2016

[10] J. Kim, J. Kim, S. Paik, & H. Kim. Point cloud registration considering safety nets during scaffold installation using sensor fusion and deep learning. *Automation in Construction*, *159*: 105277, 2024.

[11] Y. Chen, J. Liu, X. Zhang, X. Qi, & J. Jia. (2023). Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 21674-21683).

[12] A. Geiger, P. Lenz, C. Stiller, & R. Urtasun. Vision meets robotics: The kitti dataset. *The*

international journal of robotics research, *32*(11), 1231-1237, 2013.

[13] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, & D. Rus. “Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping,” in *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE: pp. 5135–5142, 2020.