**ICCEPM 2024** https://dx.doi.org/10.6106/ICCEPM.2024.1073
The 10th International Conference on Construction Engineering and Project Management
*Jul. 29-Aug.1, 2024, Sapporo*

# Real-time Construction Progress Monitoring Framework leveraging Semantic SLAM

Wei Yi HSU [1], Aritra PAL [1]*, Jacob J. LIN [1], Shang-Hsien HSIEH [1]

[1] *Department of Civil Engineering, National Taiwan University, Taiwan,* E-mail address: apal@caece.net

**Abstract:** The imperative for real-time automatic construction progress monitoring (ACPM) to avert project delays is widely acknowledged in construction project management. Current ACPM methodologies, however, face a challenge as they rely on collecting data from construction sites and processing it offline for progress analysis. This delayed approach poses a risk of late identification of critical construction issues, potentially leading to rework and subsequent project delays. This research introduces a real-time construction progress monitoring framework that integrates cutting-edge semantic Simultaneous Localization and Mapping (SLAM) techniques. The innovation lies in the framework's ability to promptly identify structural components during site inspections conducted through a robotic system. Incorporating deep learning models, specifically those employing semantic segmentation, enables the system to swiftly acquire and process real-time data, identifying specific structural components and their respective locations. Furthermore, by seamlessly integrating with Building Information Modeling (BIM), the system can effectively evaluate and compare the progress status of each structural component. This holistic approach offers an efficient and practical real-time progress monitoring solution for construction projects, ensuring timely issue identification and mitigating the risk of project delays.

**Keywords:** construction progress monitoring, semantic SLAM, automation, robot

## 1. INTRODUCTION

In contemporary construction project management, Real-Time Automated Construction Progress Monitoring (ACPM) is pivotal for enhancing project efficiency and reducing the risk of delays [1][2]. Traditional progress monitoring methods rely on manual data collection and processing, which tends to delay issue identification and is labor-intensive. The advent of technologies like Semantic Simultaneous Localization and Mapping (Semantic SLAM) has made Real-Time ACPM feasible, addressing the critical limitations of manual monitoring methods, including the potential for human error and data processing inefficiencies [3]. The necessity for real-time data collection and processing in construction project management has become increasingly apparent. The delayed identification of construction issues inherent in traditional monitoring approaches leads to project delays and increases costs due to the need for rework.

This study introduces a real-time progress monitoring framework that leverages advancements in Semantic SLAM technology. This framework integrates cutting-edge semantic SLAM techniques with automated tools, such as robotic systems, for on-site inspections, providing a practical approach to monitoring construction progress. Furthermore, applying deep learning models enhances the system's ability to identify and classify different structural elements accurately [4]. This capability is crucial for maintaining the accuracy and reliability of progress monitoring [2]. Additionally, integrating with Building Information Models (BIM) offers more insights, such as comparing real-time identified tasks with 4D BIM scheduling to check for plan adherence.

The transition towards Real-Time ACPM, emphasized by integrating Semantic SLAM and automated robotic inspections, represents a pivotal advancement in construction project management. This approach can significantly enhance project efficiency, reduce costs, and improve the overall quality of construction projects. Building upon this conceptual foundation, the subsequent advancements and empirical validations further illustrate the evolution and practical application of Real-Time ACPM. The evolution of Real-Time ACPM represents a significant leap forward in construction project management, aiming to enhance efficiency and reduce project delays.

Integrating semantic Simultaneous Localization and Mapping (semantic SLAM) technologies and advanced deep learning models has revolutionized real-time monitoring practices. Studies by Yang et al. [2] and Pučko et al.[3] highlight the transformative impact of these technologies in addressing the limitations of manual monitoring methods, highlighting the crucial role of real-time data acquisition and processing in improving project management outcomes. Raut et al. [5] emphasized the importance of effective cost and time-monitoring techniques, advocating for an integrated approach that combines Average Index formulas and S-curves for more accurate project management. Similarly, Liu et al. [6] explored the benefits of near real-time 3D reconstruction for critical monitoring tasks, demonstrating the potential of LDSO-based methods in enhancing the rapidity and quality of construction monitoring. Halder et al. [7] presented an innovative approach to real-time progress monitoring through quadruped robots integrated with Augmented Reality (AR), showcasing a computational framework that leverages cloud-based solutions for remote navigation and progress monitoring. Jiang et al. [8] introduced a semi-automatic framework utilizing Scan-vs-BIM technology for real-time monitoring of bridge construction projects, highlighting the efficiency of integrating geometric information from as-built to virtual point clouds for more accurate monitoring.

These studies illustrate the ongoing advancements in ACPM, emphasizing the critical role of technology in overcoming traditional monitoring challenges and enhancing construction project management. The collective findings highlight a significant shift towards leveraging cutting-edge technologies for real-time data acquisition, processing, and monitoring, substantially improving construction project management practices.

## 2. REAL-TIME CONSTRUCTION MONITORING FRAMEWORK

For Real-Time ACPM, this study introduces a lightweight monitoring framework suitable for edge computing devices. The framework is divided into five main modules: data preprocessing, on-site data collection, semantic SLAM, progress status estimation, and output with visualization, as demonstrated in Figure 1.

### 2.1. Data Preprocessing

In the automation realm, data preparation and preprocessing are essential for system operation, particularly in applications involving robotics and edge computing devices where effective data processing significantly influences system efficiency. Our research begins by utilizing the Revit API to extract specific information from 4D BIM, concentrating on necessary details such as each element’s geometry and anticipated schedules. By selectively extracting relevant information, we avoid the inefficiency of importing entire models that are filled with abundant information. The BIM model used for input must possess a Level of Development (LOD) of 300 or higher to ensure the model contains sufficient information. Subsequently, the extracted data is inputted into the Robot Operating System (ROS), where the coordinate system of the BIM model automatically aligns with that of the robot, as illustrated in Figure 2. This targeted approach to data extraction and processing reduces the computational load and makes integrating BIM data with robotics and edge computing applications more straightforward, establishing a flexible and effective foundation for preparatory work.

### 2.2. Data Collection

Automating real-time data collection in construction site environments poses a significant challenge, leading to increasing research focusing on the application of robots in construction. The effectiveness of data collection depends on the SLAM algorithms. For instance, approaches based on RGB-D and LiDAR have distinct advantages. RGB-D sensors are suited for detail-rich indoor environments as they can capture surface details and color information for more precise object identification and spatial positioning [9]. On the other hand, LiDAR offers high-precision distance measurement, making it suitable for large-scale outdoor environments with stable light conditions [10]. Each approach provides different advantages based on the specific needs of the construction site. This study primarily employs an RGB-D-based approach because it can give real-time depth and color information, making segmentation more effective. Additionally, selecting the right robot is crucial. While there are studies on deploying robotic dogs or drones on construction sites, applying Unmanned Ground Vehicles (UGVs) is less common. The advantages of UGVs include their lower cost and smaller size, which allows them to avoid obstacles on construction sites. However, their limitation lies in less agility in environments with significant elevation differences. For UGVs, choosing the proper mode of mobility is essential; tracks are suitable for navigating rugged surfaces, while omnidirectional wheels support 360-degree movement. These features apply to many data collection scenarios on construction sites. Robots significantly enhance the automation in data collection [11][12]. Figure 3 shows the robot used in this study for collecting data. It is equipped with an RGB-D camera and LiDAR sensors. The robot's chassis utilizes omnidirectional wheels for movement within the construction site.

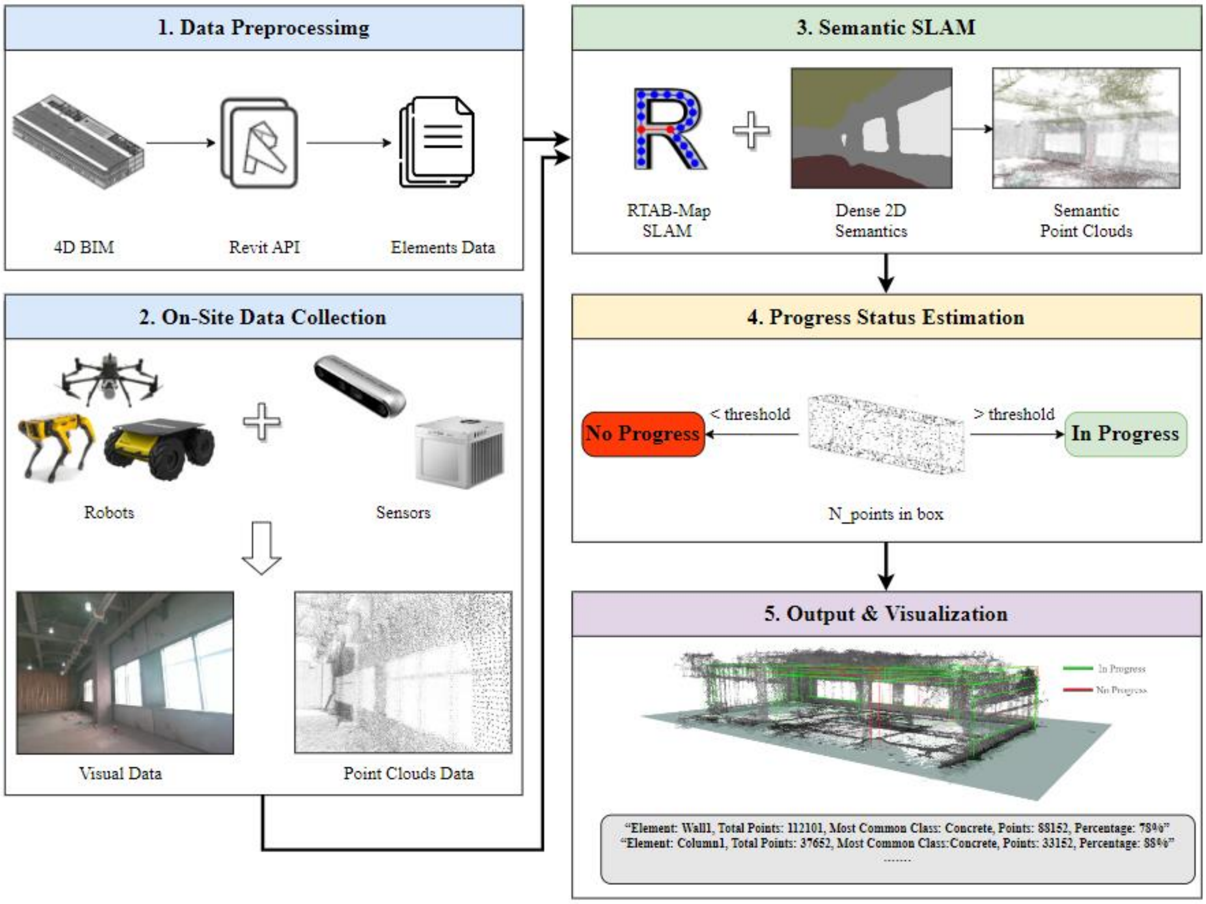


Figure 1. The system overview of the proposed framework.

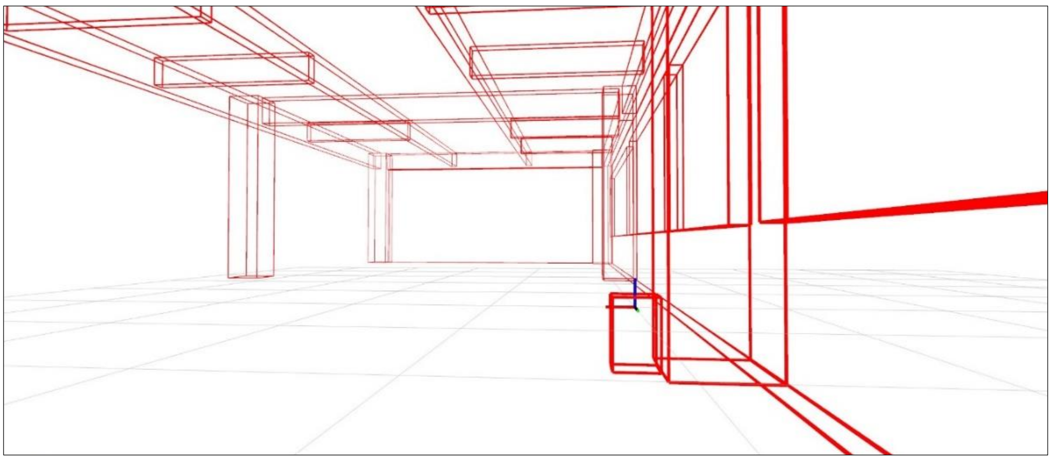

Figure 2. The alignment of the BIM model with the robot's coordinates.


Figure 3. The robot used in this study for on-site data collection.

### 2.3. Semantic SLAM

Semantic SLAM is a technology that combines SLAM with visual image recognition, primarily tackling three conceptual questions: “Where am I?” “Where are the objects around me?” and “What are the objects around me?”. The solutions to these questions correspond to three core technologies: Localization, Mapping, and Semantics. This technology effectively addresses the technical challenges encountered in this study, namely robot navigation and the generation of 3D semantic point clouds. SLAM works by iteratively estimating the position of a robot and updating the map of the environment simultaneously. It utilizes sensor data to identify distinctive features in the surroundings, using these as reference points to build and refine a map while tracking the robot’s location within it. This paper introduces RTAB-Map, a Graph-Based SLAM approach using an incremental appearance-based loop closure detector [13], which assesses whether a new image matches a previous location. It incorporates a memory management strategy for efficient real-time mapping and localization in extensive environments. Localization and mapping assist the robot in autonomous navigation and in defining the spatial relationships with surrounding components, while Semantics help us understand what these components are. Generating 3D semantic point clouds involves fusing 2D dense semantics from deep learning models with 3D point clouds at each frame. At the same time, the SLAM system provides the robot’s position at each frame, thereby creating a 3D semantic map. Figure 4 presents the 3D points cloud results generated by applying semantic SLAM. It processes image input from RGB-D cameras and outputs 2D semantic segmentation of images. It then fuses semantic segmentation results with depth point clouds. Finally, the pose output by the SLAM system is utilized to determine the position of each semantic point cloud. Here, different structural elements can be easily identified by their semantic labels.

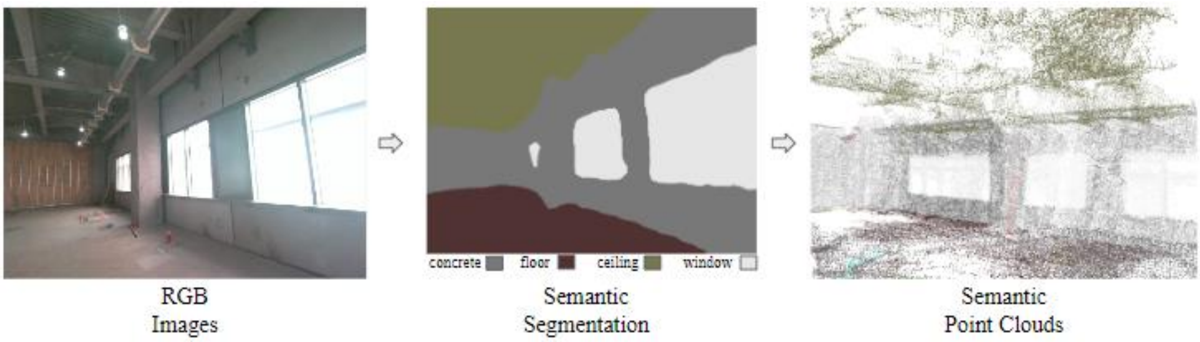


Figure 4. The process of generating semantic point clouds.

### 2.4. Progress Status Estimation

As the 3D semantic map automatically gets aligned with the BIM, we use the method of occupancy check to estimate the progress status of the structural elements of the construction project. First, the bounding boxes of individual elements are extracted from BIM and expanded by a threshold value to accommodate minor errors. Subsequently, the total number of points within each bounding box is calculated. If the number of points exceeds a predetermined threshold, it indicates that the construction of the components has been completed; otherwise, it signifies no progress [14]. Accordingly, elements are color-coded for progress visualization. Furthermore, we also analyze the number of point clouds in the most common segmentation class to gain advanced information on the progress status.

## 3. EXPERIMENTS AND RESULTS

The framework proposed in this study was implemented and deployed on robots and edge computing devices for real-world testing at construction sites.

### 3.1. Case Study

We selected a hospital building's construction site for testing, as illustrated in Figure 5. As shown in the figure, we applied our method to a specific site portion. The floor area of this section was approximately 385 square meters, and the structural components included columns, beams, walls, and slabs, totaling 30 in number.

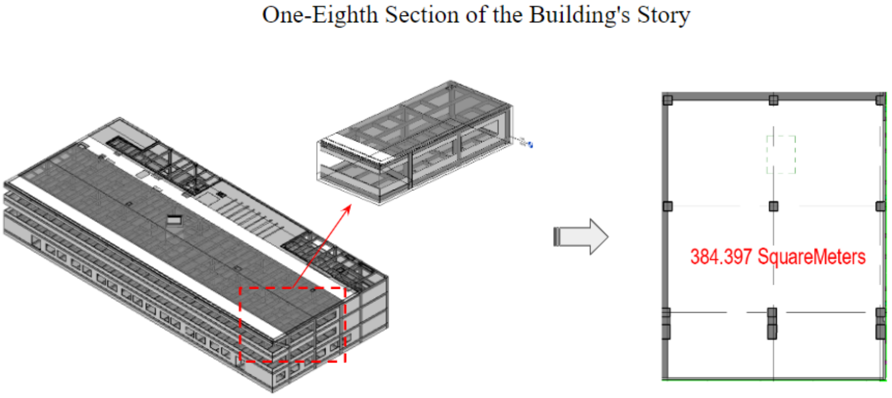


Figure 5. The information from the case study.

### 3.2. Semantic SLAM

In our SLAM system, RTAB-Map was employed, utilizing loop closure detection to enable real-time and accurate large-scale mapping. Figure 6 displays the output of our SLAM system using RTAB-Map, illustrating a detailed 2D map and the path trajectory, along with a 3D point clouds representation. It showcases the system's ability to dynamically generate a comprehensive 2D map and path in real time. The figure highlights the map's completeness and the path's clarity, demonstrating the system's effectiveness in mapping environments dynamically. It visually demonstrates the system's capacity to produce detailed spatial representations, crucial for understanding and navigating the mapped environment.

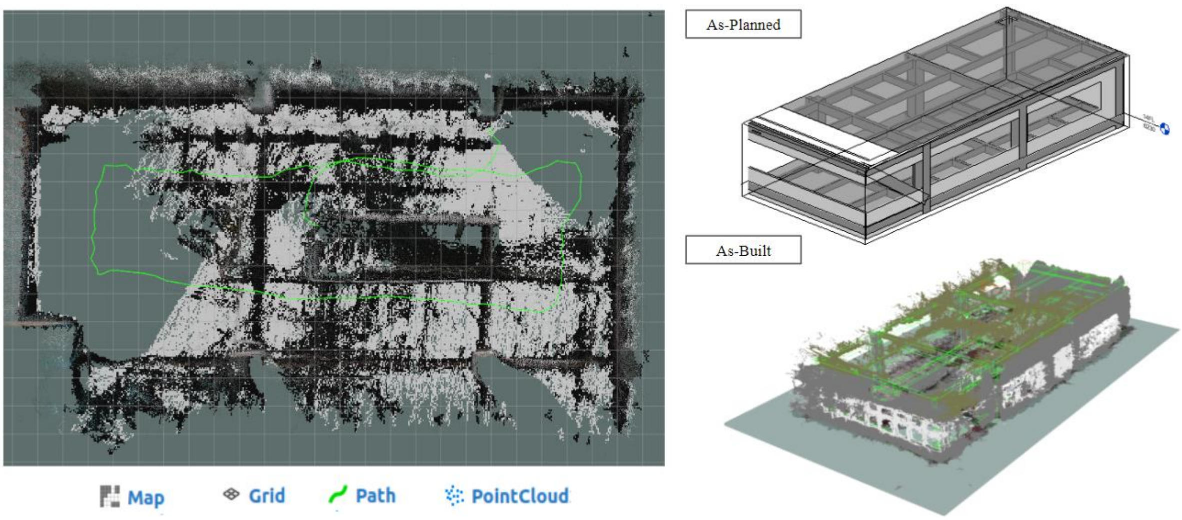


Figure 6. The result of semantic SLAM.

For semantic segmentation, we used MobileNetV2Dilated, a model distinguished by its efficiency and lightweight architecture, which aligns with our requirement for real-time processing capabilities. Figure 7 below depicts the MobileNetV2Dilated model architecture. The model architecture builds upon the efficient inverted residual structure introduced in MobileNetV2. It commences with a standard convolutional layer with a 3x3 kernel and a stride of 2 for initial spatial reduction and feature extraction. This is succeeded by a sequence of inverted residual blocks, characterized by their use of depthwise separable convolutions that offer a balance between lightweight models and representational power. Each block is described by an expansion factor (t), the number of output channels (c), the number of times the block is repeated (n), and the stride (s) for downsampling. From the third block onwards, the model employs dilated convolutions with a rate of 2 to enlarge the receptive field without increasing the computational burden excessively. As the layers deepen from the fifth block, the dilation rate is increased to 4, facilitating a broader context aggregation without resolution loss. The architecture culminates in a classifier, which integrates a dropout layer for regularization followed by a linear layer responsible for the final prediction output. This model demonstrated its real-time operational capabilities by running at 6.7 FPS, showcasing high accuracy in segmenting various structural elements such as concrete, floor, ceiling, and windows, with a mean pixel accuracy of 0.80 and a Mean IoU of 0.71. The detailed results, including IoU, pixel accuracy, precision, and recall, are shown in Table 1.

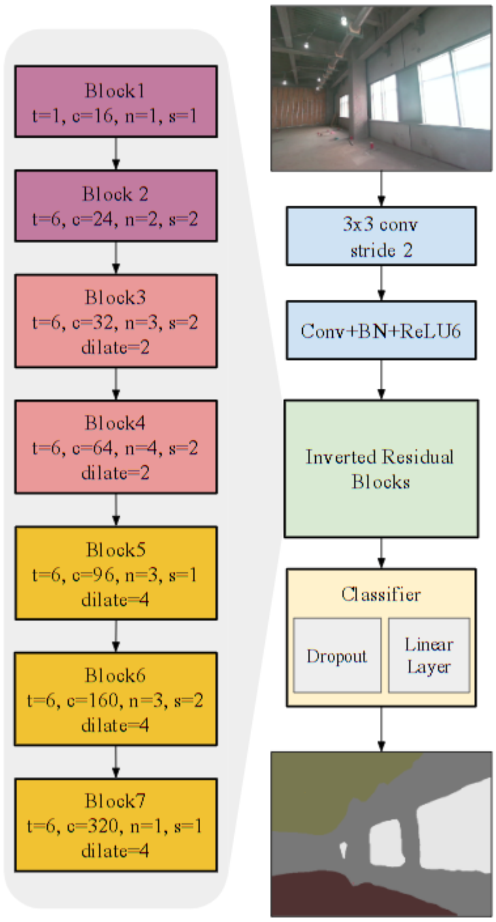


Figure 7. Network architecture of MobileNetV2Dilated.

**Table 1.** Class-wise performance of the MobileNetV2Dilated model

| **Class** | **IoU** | **Pixel Accuracy** | **Recall** | **Precision** |
|---|---|---|---|---|
| Wall & Column | 0.7271 | 0.7604 | 0.8786 | 0.7381 |
| Floor | 0.7552 | 0.8452 | 0.9229 | 0.7293 |
| Ceiling | 0.7949 | 0.9154 | 0.9116 | 0.8169 |
| Window | 0.5755 | 0.6874 | 0.7339 | 0.6950 |
| Overall | **0.7132** | **0.8021** | **0.8618** | **0.7448** |

### 3.3. Progress Status Estimation

For occupancy checks [14], this study calculated the points count within the bounding box of each element. For any element, if the count exceeded 1000 pts/m$^2$, it was indicated as “In progress”; otherwise, it was labeled as “No progress.” Semantic point clouds assist in parsing more detailed progress information. Specifically, when multiple categories of semantic point clouds are present within each bounding box, the category that constitutes the majority is used as a representative of the progress status for that element. Consequently, when walls and columns are predominantly identified as 'Wall & Column' and beams as 'Ceilings,' the structural work is deemed completed. Figure 8 illustrates the visualization of the Progress Status Estimation, where green represents “In progress” and red indicates “No progress.” Most components were correctly identified with their accurate progress status. However, there were misidentifications for a column and a beam. These errors are primarily attributed to inherent errors in SLAM and constraints associated with the camera's elevation angle and field of view.


| Component | Progress Status: In progress | Progress Status: No progress | Class |
|---|---|---|---|
| Wall | 3 | 0 | Wall & Column |
| Column | 5 | 1 | Wall & Column |
| Beam | 17 | 1 | Ceilings |

Figure 8. The visualization of progress status.

## 4. CONCLUSION

This paper proposes a real-time automatic construction progress monitoring framework. The framework is deployed on a robotic system with edge computing devices and tested in a real-world scenario. The framework uses 4D BIM and visual data collected from the site as input, processes these data through semantic SLAM to generate 3D semantic point clouds, and compares the spatial relationship between the point clouds and BIM components to ascertain progress status. Semantic SLAM enables robots to autonomously navigate and recognize construction site conditions in real-time, addressing the delay in progress analysis due to the limitations of existing ACPM methods. This real-time ACPM approach significantly reduces the risk of delayed identification of critical construction issues while enhancing the data collection automation level.

In its present form, the proposed approach is suitable for element-level progress monitoring. However, future research will investigate the applicability of real-time ACPM for activity-level progress monitoring. The framework faces limitations such as the robot’s navigation challenges in cluttered or uneven terrains, affecting data collection efficiency. Large-scale site monitoring may also surpass the memory capabilities of a single robot, posing constraints on operational scope. The framework’s current dependency on the inherent accuracy of SLAM poses challenges, as the inherent errors in SLAM are likely to be magnified with the expansion of the construction site area. This magnification of errors underscores the need for advanced calibration or integration strategies, especially in larger operational environments. Future efforts will optimize the proposed framework and evaluate its feasibility and performance through additional case studies. In terms of optimization, attempts will be made to extract more information from BIM models and integrate it with the SLAM system to improve SLAM performance. Regarding semantics, different deep learning models will be explored to assess their balance in speed and accuracy. Geometric statistical methods will also be employed to filter out noise from the semantic point clouds.

This study advances construction progress monitoring by integrating cutting-edge technologies that improve real-time data processing and analysis. The proposed framework addresses the current challenges in the field and opens new avenues for research and development in construction management, aiming for more efficient, accurate, and automated construction processes.

## ACKNOWLEDGMENT

The support of the National Science and Technology Council (NSTC), Taiwan, is gratefully acknowledged. This work was supported by the NSTC under Grant No. 112-2221-E-002 -127 -MY3.

## REFERENCES

[1] A. Pal, J. J. Lin, S.-H. Hsieh, and M. Golparvar-Fard, "Automated vision-based construction progress monitoring in built environment through digital twin," *Dev. Built Environ.*, vol. 16, p. 100247, 2023, doi: 10.1016/j.dibe.2023.100247.

[2] J. Yang, A. Wilde, K. Menzel, M. Z. Sheikh, and B. Kuznetsov, "Computer Vision for Construction Progress Monitoring: A Real-Time Object Detection Approach," *IFIP Adv. Inf. Commun. Technol.*, vol. 688 AICT, pp. 660–672, 2023, doi: 10.1007/978-3-031-42622-3_47.

[3] Z. Pučko, N. Šuman, and D. Rebolj, "Automated continuous construction progress monitoring using multiple workplace real time 3D scans," *Adv. Eng. Informatics*, vol. 38, no. October 2017, pp. 27–40, 2018, doi: 10.1016/j.aei.2018.06.001.

[4] A. Pal and S. H. Hsieh, "Deep-learning-based visual data analytics for smart construction management," *Autom. Constr.*, vol. 131, no. August, p. 103892, 2021, doi: 10.1016/j.autcon.2021.103892.

[5] S. Raut, S. S. Pimplikar, and K. Sawant, "Effect of Project Cost and Time Monitoring on Progress of Construction Projct," *Int. J. Res. Eng. Technol.*, vol. 02, no. 12, pp. 796–800, 2013, doi: 10.15623/ijret.2013.0212132.

[6] Z. Liu, D. Kim, S. Lee, L. Zhou, X. An, and M. Liu, "Near Real-Time 3D Reconstruction and Quality 3D Point Cloud for Time-Critical Construction Monitoring," *Buildings*, vol. 13, no. 2, pp. 1–19, 2023, doi: 10.3390/buildings13020464.

[7] S. Halder, K. Afsari, J. Serdakowski, S. Devito, M. Ensafi, and W. Thabet, "Real-Time and Remote Construction Progress Monitoring with a Quadruped Robot Using Augmented Reality," *Buildings*, vol. 12, no. 11, p. 2027, 2022, doi: 10.3390/buildings12112027.

[8] Z. Jiang, X. Shen, M. H. Ibrahimkhil, K. Barati, and J. Linke, "Scan-Vs-Bim for Real-Time Progress Monitoring of Bridge Construction Project," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 10, no. 4/W3-2022, pp. 97–104, 2022, doi: 10.5194/isprs-annals-X-4-W3-2022-97-2022.

[9] P. Kim, J. Chen, J. Kim, and Y. K. Cho, "SLAM-driven intelligent autonomous mobile robot navigation for construction applications," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10863 LNCS, no. January, pp. 254–269, 2018, doi: 10.1007/978-3-319-91635-4_14.

[10] L. Marcy and I. Iordanova, "Slam and Beacon Data for Automation of Indoor Construction Progress Tracking," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1218, no. 1, p. 012009, 2022, doi: 10.1088/1757-899x/1218/1/012009.

[11] K. Asadi *et al.*, "Building an integrated mobile robotic system for real-time applications in construction," *ISARC 2018 - 35th Int. Symp. Autom. Robot. Constr. Int. AEC/FM Hackathon Futur. Build. Things*, no. Isarc, 2018, doi: 10.22260/isarc2018/0063.

[12] C. Feng, N. Fredricks, and V. R. Kamat, "Human-robot integration for pose estimation and semi-autonomous navigation on unstructured construction sites," *ISARC 2013 - 30th Int. Symp. Autom. Robot. Constr. Mining, Held Conjunction with 23rd World Min. Congr.*, vol. 48109, pp. 1317–1325, 2013, doi: 10.22260/isarc2013/0148.

[13] M. Labbé and F. Michaud, "RTAB-Map as an Open-Source Lidar and Visual SLAM Library for Large-Scale and Long-Term Online Operation," *Journal of Field Robotics*, vol. 36, p. 416446, 2019, doi: 10.1002/rob.21831.

[14] A. Pal, J. J. Lin, S.-H. Hsieh, and M. Golparvar-Fard, "Activity-level construction progress monitoring through semantic segmentation of 3D-informed orthographic images," *Autom. Constr.*, vol. 157, p. 105157, 2024, doi: 10.1016/j.autcon.2023.105157.