On-site Semantic Mapping and Waypoint Planning for Autonomous Aerial Bridge Monitoring

Yohan Kim[1], Sunwoong Paik[1], and Hyoungkwan Kim[1]

[1]School of Civil and Environmental Engineering, Yonsei University, Republic of Korea.
E-mail: homez815@yonsei.ac.kr, tjsdnd4066@yonsei.ac.kr, and hyoungkwan@yonsei.ac.kr

Abstract –

Effective monitoring of aging bridges is critical for ensuring their safety and maintenance. This study introduces a framework for on-site autonomous aerial bridge monitoring using sensor fusion and SLAM (Simultaneous Localization and Mapping). The proposed method utilizes a lightweight LiDAR sensor and a mini PC onboard a drone to generate real-time 3D semantic maps and flight waypoints. YOLOv8-based image segmentation is employed to identify bridge components, achieving a mean Average Precision (mAP50-95) of 86.6% across test data. Segmentation requires less than 10 milliseconds per frame, while processing LiDAR point clouds takes less than 1 second per frame. Waypoint generation based on the semantic map is completed in under 3 seconds. These results demonstrate the framework's capability to deliver precise and reliable on-site monitoring. This system provides a significant advancement in autonomous aerial bridge inspection by enabling efficient and real-time operation.

Keywords –

Autonomous Drone; Bridge Monitoring; Semantic SLAM; Sensor Fusion

1 Introduction

The need for rapid and precise monitoring of aging bridges has become increasingly critical, and drones have been demonstrated as effective tools for this purpose in various studies [1]. Particularly, autonomous drone navigation minimizes errors caused by subjective human judgment and ensures consistent data quality, making it a promising technology. However, traditional drone-based inspections often rely on predefined flight paths and manually generated waypoints, which can lead to incomplete coverage and inefficiencies in dynamic environments. Additionally, many existing approaches depend on pre-built 3D models, which are often unavailable for aging infrastructure or may become outdated due to structural modifications.

Simultaneous Localization and Mapping (SLAM) is a core technology in autonomous navigation, enabling robots or drones to perceive their surroundings and estimate their position while simultaneously creating a map. SLAM can be categorized into Visual SLAM and LiDAR SLAM. Visual SLAM uses cameras to extract features and reconstruct 3D environments, but its performance degrades in environments with poor lighting or limited texture. On the other hand, LiDAR SLAM generates dense 3D point clouds, offering precise distance measurements and real-time mapping capabilities [2], making it particularly advantageous for large-scale infrastructure inspection.

For autonomous bridge monitoring, maps should include semantic information in addition to geometric data. To achieve this, 3D semantic segmentation techniques can be applied, using either traditional or learning-based approaches. Traditional methods are computationally efficient but lack generalization and accuracy, while learning-based methods provide high accuracy but require significant computational resources.

In this study, a drone equipped with a lightweight LiDAR sensor and a mini PC is utilized to generate high-precision maps in real time, eliminating the need for pre-existing models and enabling adaptive waypoint generation based on continuously updated spatial information. By integrating semantic information with LiDAR SLAM, the proposed approach enhances scene understanding, allowing drones to differentiate bridge components and optimize flight paths accordingly. Camera-LiDAR sensor fusion is employed to achieve fast and accurate 3D bridge segmentation, enabling autonomous navigation with improved structural awareness.

As shown in Figure 1, the proposed method comprises three stages. First, bridges are identified using image segmentation based on deep learning. Second, semantic SLAM is performed by assigning classes to bridge point clouds through camera-LiDAR sensor fusion. Finally, waypoints are generated to provide autonomous flight paths. This approach enables effective aerial monitoring even for aging bridges without pre-existing 3D models.

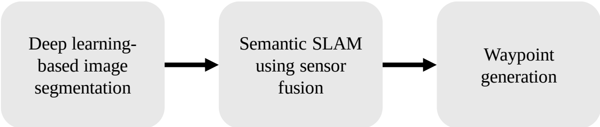

Figure 1. Overview of proposed method

2 Related Works

Most existing studies on waypoint generation and path planning for autonomous drone navigation rely on pre-existing 3D models, such as point cloud, CAD, or BrIM/BIM models. Bolourian and Hammad [3] utilize BrIM/BIM models to define key observation points and apply ray-tracing-based visibility analysis to optimize drone flight paths. Zhao et al. [4] employ a two-stage optimization approach, first determining waypoint locations using genetic algorithms and then refining camera poses through a greedy algorithm to maximize coverage efficiency. Phung et al. [5] model drone path planning as a traveling salesman problem (TSP) and solve it using discrete particle swarm optimization (DPSO), optimizing waypoint sequences to minimize flight distance while ensuring complete coverage. Lin et al. [6] integrate 2D satellite maps for flight planning, enabling flexible data collection missions.

While these methods effectively plan drone flight paths, many aging bridges lack accurate 3D models, and structural changes due to maintenance or disasters may render existing models outdated. To overcome these limitations, some studies have explored on-site mapping approaches that do not rely on pre-existing models. For instance, Narazaki et al. [7] focuses on the rapid post-earthquake inspection of reinforced concrete railway viaducts, specifically targeting damage assessment of columns. To achieve this, their approach generates waypoints based on a progressively built sparse point cloud, enabling drones to autonomously inspect structural elements. While various mapping methods, including SLAM, were considered, their experiments primarily relied on Structure from Motion (SfM) for 3D reconstruction. SfM requires acquiring a sufficient number of images before generating a usable model, which introduces a delay and limits real-time waypoint generation.

This study introduces a real-time LiDAR SLAM-based approach that generates high-precision 3D maps and enables drone-based waypoint generation without relying on pre-existing models. By leveraging LiDAR-generated point clouds with semantic information, our method ensures accurate localization and detailed structural representation, enabling drones to operate based on continuously updated spatial information. This approach facilitates efficient aerial monitoring even for aging bridges without pre-existing 3D models, providing a reliable framework for autonomous navigation in environments where structural changes from maintenance or disasters may render existing models inaccurate or obsolete.

3 Methodology

3.1 Deep Learning-based Image Segmentation

Accurate identification of bridges is a prerequisite for monitoring using autonomous drone navigation. However, processing 3D point clouds in real-time during drone flight is challenging due to computational resource limitations. To address this, the proposed method utilizes deep learning-based image segmentation and sensor fusion to recognize bridges. A real-time deep learning model segments image data captured by the drone's camera into superstructure and substructure. The

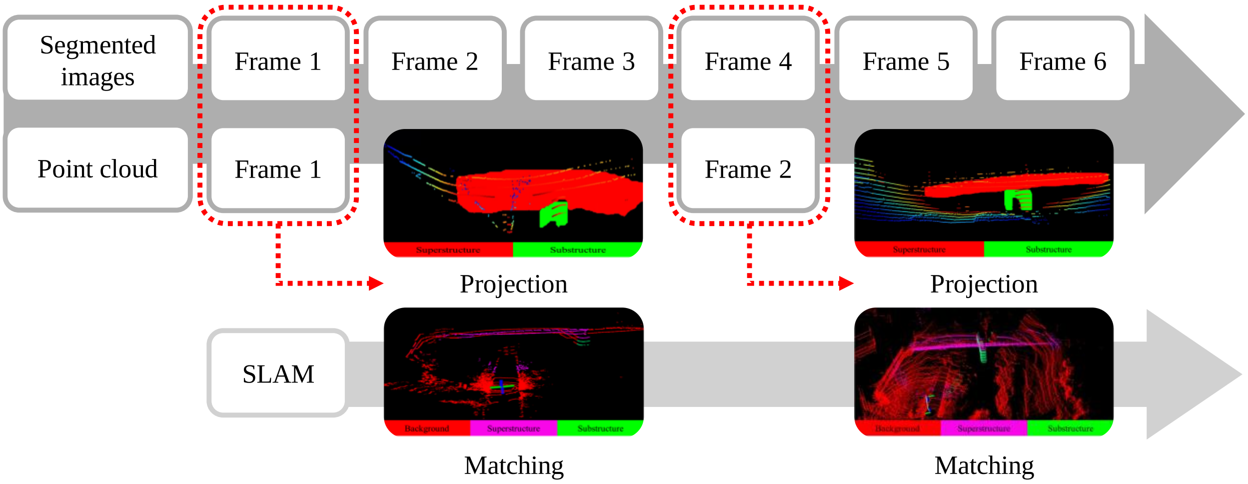

Figure 2. Process of semantic mapping using sensor fusion and SLAM

segmentation results are then used to assign classes to point clouds within the camera view, enabling the system to differentiate between bridge components effectively. This classification is crucial for optimizing waypoint generation and ensuring accurate inspection of critical structural elements.

3.2 Semantic Mapping Using Sensor Fusion and SLAM

To project 3D point clouds acquired by LiDAR onto images, precise calibration between the LiDAR and the camera is required. Calibration aligns the camera's 2D images and the LiDAR's 3D data within the same coordinate system. This involves estimating the camera's intrinsic parameters to correct distortions and matching features between LiDAR point clouds and camera images to calculate the transformation matrix (extrinsic parameters). In this study, a target-based approach using a checkerboard was employed to perform feature matching between the point clouds and images. The checkerboard provides reliable and consistent feature points, ensuring high calibration accuracy.

Once alignment is completed, the segmented images are projected onto the point clouds. The timestamps of the images and point clouds are compared to find the closest frames for matching. Since the number of images is typically greater than that of point clouds within the same duration, the nearest image to each point cloud is selected. Points within the camera's field of view are projected onto the image, and the class of the corresponding pixel is assigned to the points. This ensures that each point in the cloud is semantically labeled with a category, such as superstructure or substructure, and is subsequently utilized in the SLAM and waypoint generation processes. Figure 2 shows the process of semantic mapping in this study.

The LiDAR-Inertial SLAM algorithm integrates data from LiDAR and the Inertial Measurement Unit (IMU) to achieve high-precision real-time localization and map generation. The IMU compensates for motion between LiDAR frames, while LiDAR matching corrects the IMU's drift. This fusion ensures stable performance even during fast movements, rotations, or vibrations of the drone. In this study, calibration between the LiDAR and the drone's IMU was conducted to obtain the transformation matrix, aligning the classified point clouds with the IMU data to generate a real-time semantic SLAM map.

3.3 Waypoint Generation

The generated semantic SLAM map is processed to extract bridge point clouds, which are used as the basis for waypoint generation. First, the point cloud is downsampled to reduce computational load, and noise is removed to enhance the quality of the extracted data. Then, the RANSAC algorithm is applied to extract vertical planes of the bridge, which represent key structural elements such as piers or deck. These vertical planes are critical for determining the spatial boundaries of the bridge and for planning inspection routes.

Second, XY histograms are generated for both superstructure and substructure points, and empty bins around the structure are identified. The center of these bins is used to define the X and Y coordinates, while the Z coordinate is added considering the structure's height to define 3D waypoints. The use of histograms ensures that the generated waypoints are distributed evenly around the structure, optimizing the drone's flight path. This process enables the drone to plan flight paths for efficiently inspecting the superstructure and substructure while maintaining consistent coverage of critical areas.

4 Experiment

4.1 Implementation Details

As shown in Figure 3, the data acquisition platform was built by equipping a drone (DJI Matrice 300 RTK) with a mini PC (Intel NUC) and a LiDAR sensor (Velodyne VLP-16). The camera, LiDAR, and the drone's built-in IMU data were transmitted in ROS (Robot Operating System) message format and stored as bag files. Each message included a timestamp, which was utilized for matching point clouds with images. Data collection was conducted at two bridges in South Korea. Image segmentation was performed in real-time using YOLOv8 [8].

YOLOv8 offers models of various sizes: nano, small, medium, large, and x-large. In this study, YOLOv8n (nano model), which is suitable for real-time processing on CPUs, was selected. The model was pretrained on the COCO dataset and fine-tuned using images captured by the drone camera. Of the total 1,032 images, 722 were used as training data, and 310 as testing data. Figure 4 illustrates an example of annotations.

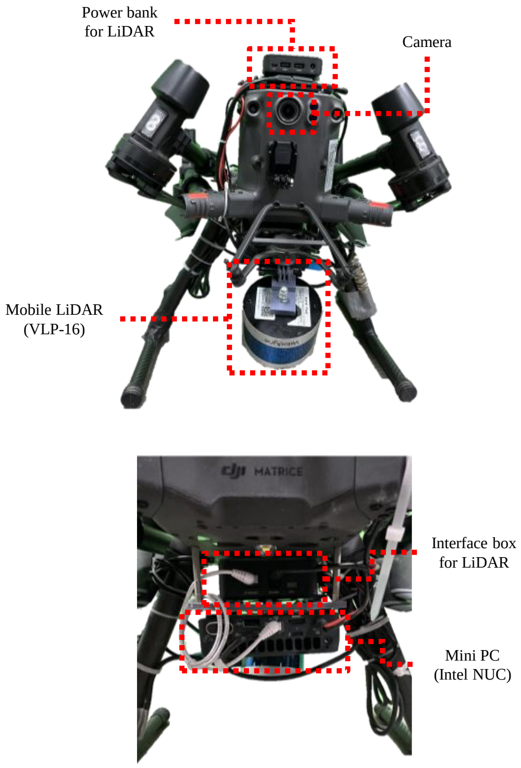

Figure 3. Hardware configuration

The LIO-SAM [9] algorithm was employed for SLAM, organizing sensor data into graph structures and optimizing them to achieve precise localization and mapping. The IMU data was utilized to compensate for the drone's motion, ensuring accurate alignment and consistency within the generated point cloud map. For waypoint generation, it was assumed that the bridge points were sufficiently accumulated on the map, with the initial flight performed using GPS-based or manual control. The map was used for waypoint generation once the bridge points exceeded 100,000 in number.

Semantic mapping, waypoint generation, and the training and testing of the YOLOv8 model were performed on a computer running Ubuntu 18.04 with an Intel Core i7-10700KF CPU and an Nvidia GeForce RTX 3090 GPU.

4.2 Experimental Results

Table 1 and Figure 5 present the quantitative results of image segmentation and examples of segmented images, respectively.

Table 1. Quantitative result of image segmentation

Class	mAP(50-95)
Superstructure	88.1%
Substructure	85.1%
All	86.6%

Figure 5. Visualization of segmented image

The segmentation results achieved an overall mAP(50-95) of 86.6%, where mAP(50-95) refers to the mean Average Precision calculated over IoU (Intersection over Union) thresholds ranging from 0.50 to 0.95 at 0.05 intervals. This metric provides a comprehensive evaluation of the model's detection performance, considering both loose and strict matching criteria. The slightly higher accuracy for superstructure (88.1%) compared to substructure (85.1%) suggests that superstructure features may be more distinctive or easier to identify in the captured images. This highlights the robustness of the YOLOv8 model in recognizing bridge components under varying conditions, though substructure classification still has room for improvement.

Figure 6 shows the projection of point clouds onto images using the transformation matrix obtained through camera-LiDAR calibration, as well as the generated semantic SLAM map. This figure confirms that the LiDAR data has been accurately projected onto the image and that class information is being appropriately incorporated during the mapping process. The process of waypoint generation are visualized in Figure 7.

In waypoint generation, the bin size of the XY histogram was set to 2 meters, considering the size of the drone and its distance from the bridge. Similarly, the Z-axis interval was also set to 2 meters. Waypoints were initially generated at the midpoint height of the structure, and additional waypoints were generated above and below, considering the structure's height. The generated waypoints were verified against the structural geometry to ensure complete coverage of critical areas. Figure 8 shows the generated waypoints on the SLAM map.

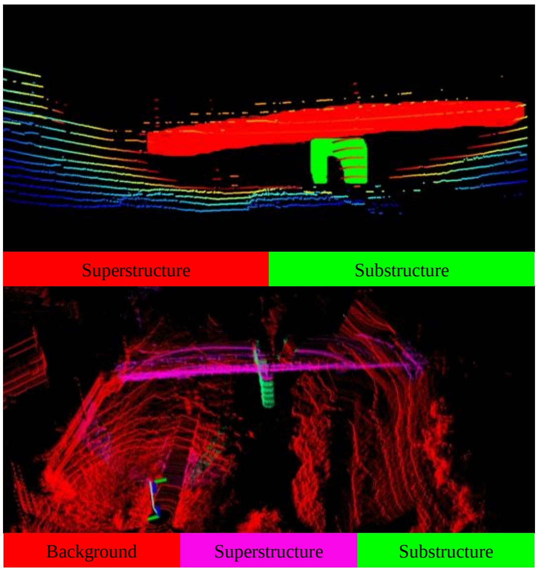

Figure 6. Visualization of projection and mapping

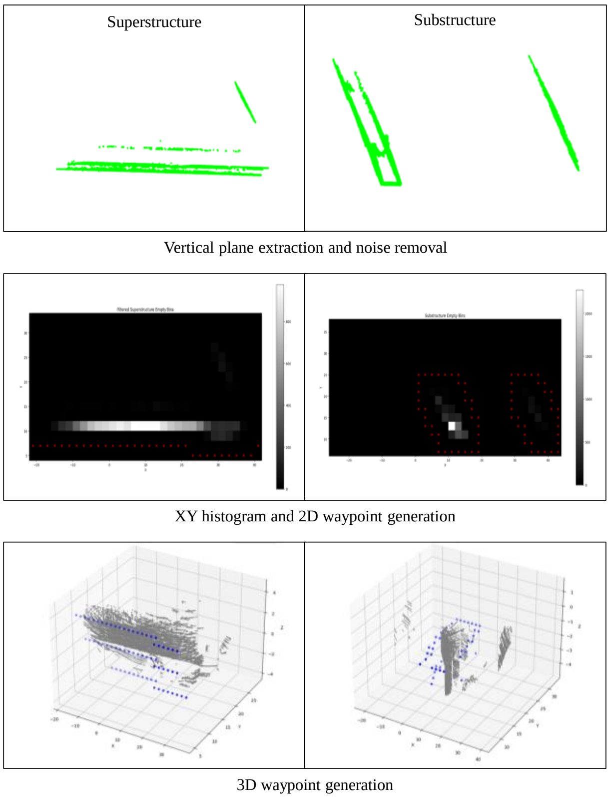

Figure 7. Visualization of waypoint generation result

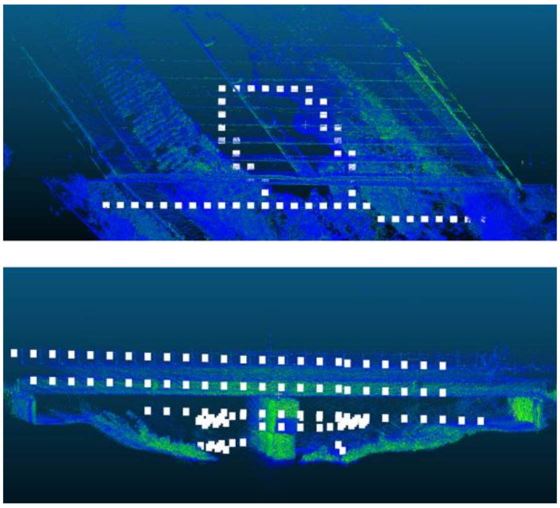
Figure 8. Visualization of generated waypoints on the SLAM map

To validate the effectiveness of the generated waypoints, a coverage evaluation was conducted using the whole bridge point cloud acquired through manual flight. The number of bridge points covered from each waypoint was measured, considering the LiDAR sensor's field of view (FOV). The visibility assessment incorporated a vertical FOV of 30° (-15° to 15°) and a horizontal FOV of 360°. Additionally, the detection range was limited to a maximum of 10 meters to exclude excessively distant points.

In this experiment, coverage was calculated based on a voxel-based approach. The space was divided into voxels of 0.5m size, and only those containing a certain minimum number of points were considered occupied. Using this information, a ray-casting technique was applied at each waypoint to evaluate visibility. During the ray-casting process, if the line of sight from a waypoint was blocked by bridge components (e.g., girders, decks, or columns) or other obstacles acting as occluding elements, the obstructed area was considered not visible and thus not covered.

As summarized in Table 2, 144 waypoints were generated for Bridge 1, covering 119,956 bridge points, which accounts for approximately 41.21% of the total 291,064 points. Similarly, 113 waypoints were generated for Bridge 2, covering 133,837 bridge points, corresponding to 30.19% of the total 443,280 points.

Table 2. Coverage of bridge point cloud from waypoints

Bridge	Way points	Covered points	Total points	Coverage ratio (%)
1	144	119,956	291,064	41.21
2	113	133,837	443,280	30.19

These results demonstrate that the generated waypoints can effectively serve as an initial exploration path, even with data acquired only from one side of the bridge. The coverage ratios of Bridges 1 and 2 indicate that despite being derived from a limited initial map, the waypoints provided visibility to a significant portion of the bridge structure. As the drone follows these waypoints, it progressively captures more of the bridge, accumulating additional information about the structure. This allows for the continuous refinement of waypoints, enabling a gradual expansion of coverage until the entire bridge is mapped.

The execution time analysis in Table 3 demonstrates the feasibility of on-site processing. Image segmentation requires only ~10 milliseconds per frame, making it highly suitable for on-the-fly data processing during drone flights. Point cloud processing, which includes projection and semantic labeling, takes ~1 second per frame, reflecting efficient handling of LiDAR data. Waypoint generation is completed in ~3 seconds, ensuring that the overall process from image capture to waypoint planning is achieved within ~4 seconds. This performance supports the system's applicability for on-site operations, allowing for timely and autonomous bridge inspection.

Table 3. Execution time for each process

Process	Time
Image segmentation	~10 milliseconds
Point cloud processing	~1 second
Waypoint generation	~3 seconds
Total	~4 seconds

5 Discussion

The experimental results demonstrate that unmanned and autonomous monitoring of bridges is feasible even in the absence of prior information. The proposed method achieved real-time and highly accurate bridge detection through deep learning-based image segmentation, while sensor fusion and SLAM facilitated the rapid creation of a 3D map containing the necessary bridge information for waypoint generation. The overall processing time was sufficiently short, confirming the method's applicability for on-site operations. However, the proposed method has certain limitations that require further improvement in future research.

First, the process of incorporating image segmentation results into the point cloud needs refinement. Currently, the segmentation result of the single image closest to the point cloud's timestamp is used. If the selected image's segmentation is inaccurate, the error could propagate to waypoint generation. To address this, multiple images could be utilized simultaneously, leveraging only regions identified consistently over a certain number of images, or by evaluating the reliability of segmentation results and selectively incorporating them.

Second, the waypoint generation process requires additional research. This study generated waypoints based on an initial map containing over 100,000 accumulated bridge points, but this map can be continuously updated during the drone's flight. Since the initial map was created using data collected from only one side of the bridge, the opposite side and substructure may not have been sufficiently represented. As the drone collects new data while following the initial waypoints, a dynamic approach could be employed to update the map and generate additional waypoints. For instance, after the drone passes a certain number of waypoints, the map could be updated to reflect new areas, and waypoints could be dynamically generated for further exploration.

Finally, it is essential to ensure the generalizability of the proposed method. While this study demonstrated meaningful results by applying the approach to two bridges, further validation is needed across a wider range of bridge types and environmental conditions.

Addressing these improvements would enable the proposed method to evolve into a more robust and accurate drone-based bridge monitoring system, adaptable to various environments.

6 Conclusion

This study proposed a drone-based monitoring system to address the absence of 3D models and structural changes in aging bridges. By integrating YOLOv8-based image segmentation and camera-LiDAR sensor fusion, bridge point clouds were segmented in real time, while LiDAR-Inertial SLAM enabled stable localization and continuous 3D map generation. The generated semantic SLAM map facilitated waypoint planning, allowing the drone to autonomously inspect bridges. The experimental results demonstrated that the proposed method enables adaptive bridge monitoring and autonomous flight even with limited prior information.

This study focused on adapting and integrating existing techniques into a unified system specifically designed for drone-based bridge inspection. By combining LiDAR SLAM, semantic segmentation, and waypoint optimization, the proposed framework overcomes challenges such as occlusions, real-time adaptability, and the lack of pre-existing 3D models. This integration ensures flexible and efficient coverage of bridge structures, highlighting the practical applicability of this approach in real-world scenarios.

Future research will focus on enhancing waypoint generation by incorporating multi-frame segmentation results and refining real-time map creation and path planning. Additionally, validating the system across a diverse range of bridge types will help assess its generalizability. Further optimization efforts will aim to ensure stable operation in resource-constrained environments, enabling more efficient deployment in large-scale infrastructure monitoring. This study contributes to the advancement of autonomous drone technologies for ensuring the safety and maintenance of aging bridges, demonstrating that existing algorithms can be effectively tailored to meet the specific demands of structural inspection.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Ministry of Education (RS-2018-NR031077) and the Ministry of Science and ICT (RS-2024-00457308).

References

[1] P. Huthwohl, R. Lu, and I. Brilakis. “Challenges of bridge maintenance inspection.” In Proceedings of the International Conference on Computing in Civil and Building Engineering, pages 51–58, 2016.

[2] A. Yarovoi and Y. K. Cho. "Review of simultaneous localization and mapping (SLAM) for construction robotics applications." Automation in Construction, 162:105344, 2024.

[3] N. Bolourian and A. Hammad. "LiDAR-equipped UAV path planning considering potential locations of defects for bridge inspection." Automation in Construction, 117:103250, 2020.

[4] Y. Zhao, B. Lu, and M. Alipour. "Optimized structural inspection path planning for automated unmanned aerial systems." Automation in Construction, 168:105764, 2024.

[5] M. D. Phung, C. H. Quach, T. H. Dinh, and Q. Ha. "Enhanced discrete particle swarm optimization path planning for UAV vision-based surface inspection." Automation in Construction, 81:25–33, 2017.

[6] J. J. Lin, A. Ibrahim, S. Sarwade, and M. Golparvar-Fard. “Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3D mapping, defect detection, analysis, and reporting.” Journal of Computing in Civil Engineering, 35(2):04020064, 2021.

[7] Y. Narazaki, V. Hoskere, G. Chowdhary, and B. F. Spencer Jr. “Vision-based navigation planning for autonomous post-earthquake inspection of reinforced concrete railway viaducts using unmanned aerial vehicles.” Automation in Construction, 137:104214, 2022.

[8] YOLOv8. Online: https://yolov8.org/, Accessed: 24/12/2024.

[9] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus. "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping." In Proc., 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5135–5142, 2020.