A Framework for Multi-scale Quantitative Defect Inspection of Concrete Bridge Surface Using Multi-view RGB-D Data

Sungwoo Cho[a], Seunghun Im[a], Taegyu Kim[a], Duho Chung[a], Ilhyoung Shin[b], Hyeongu Ji[b] and Hyoungkwan Kim[a*]

[a]Department of Civil and Environmental Engineering, Yonsei University, Korea
[b]Dangjin Branch, Korea Expressway Corporation, Korea
E-mail paul0524@yonsei.ac.kr, imseunghun4006@yonsei.ac.kr, tacoskywalker@yonsei.ac.kr, jungduho1@yonsei.ac.kr, niceguy@ex.co.kr, jeehw@ex.co.kr and hyoungkwan@yonsei.ac.kr

Abstract -

Robotic bridge inspection systems are emerging as crucial technologies to improve efficiency and safety in infrastructure maintenance, particularly in accessibility-constrained environments. However, the challenge of processing multi-view data and integrating detection, quantification, and report generation into a unified workflow remains unresolved. To address this challenge, this paper proposes an integrated inspection framework using multi-view RGB-D image sequences for a robotic bridge inspection. The framework supports multi-scale inspection by enabling both wide-area inspection and local-area inspection, while employing a two-stage voting mechanism to mitigate duplicate detections and false positives in 3D space. Experiments at a Precast Segmental Method (PSM) concrete box-girder bridge demonstrated that the framework successfully detected nine defect types, generated 2D inspection maps, and produced millimetre-scale quantification results. This work demonstrates the feasibility of an integrated robotic bridge inspection system that encompasses multi-view data analysis and quantification for practical workflows.

Keywords –

Bridge inspection; Multi-view integration; Defect detection; Quantification; Multi-scale inspection

1 Introduction

Surface defect inspection of concrete bridge structures is critical for ensuring structural safety and supporting maintenance decision-making. In South Korea, the proportion of bridges in service for over 30 years has been increasing [1], with surfaces such as box-girder interiors and exterior abutments representing typical hard-to-access regions due to their locations and structural characteristics [2]. Practical bridge inspection requires multi-scale approaches combining wide-area inspection of the overall structure and local-area inspection of defect-concentrated regions [3]. Beyond detection, defect measurement and quantification are essential components of concrete bridge inspection [4]. Obtaining quantitative measurements in physical units is essential for condition assessment and maintenance decision-making in practical bridge inspection. However, existing processes rely heavily on visual observation by inspectors, causing persistent issues including subjectivity, lack of consistency, and labour-intensive procedures [5]. Accordingly, there is growing need for automated inspection technologies providing objective and consistent quantitative results even in environments with limited accessibility.

In response to these demands, recent research has actively employed deep learning techniques for bridge surface defect detection and quantification [6, 7]. Torok et al. [8] investigated methods to identify global defect location and distribution through Structure-from-Motion (SfM), while Deng et al. [9] explored image-based Simultaneous Localization and Mapping (SLAM) approaches. Research on surface defect detection using Light Detection and Ranging (LiDAR) sensors has also been presented. Feng et al. [10] quantified crack geometry and location in 3D space through sensor fusion SLAM combining LiDAR and camera. Lin et al. [11] proposed surface defect detection methodologies utilizing 2D projection of LiDAR-acquired point clouds, demonstrating image and point cloud integration in defect inspection. However, these prior efforts have been primarily limited to specific defect types such as cracks and have not addressed comprehensive quantification across multiple defect types or inspection report generation. Furthermore, multi-scale inspection strategies encompassing both wide-area and local-area inspection have been largely overlooked.

Meanwhile, research on robot-based data acquisition systems for accessibility-constrained environments has also been conducted [12], and quadruped robots have demonstrated stable mobility and continuous image sequences acquisition across various environments [13]. However, these robotic data acquisition technologies and defect analysis algorithms have been developed independently, and methodologies to systematically address duplicate detections and false positives arising from multi-view image sequences remain insufficient.

This paper proposes an integrated inspection framework for multi-scale defect detection, quantification, and report generation using bridge surface image sequences data. The proposed framework addresses the detection, classification, and quantification of nine defect types specified in Korean regulations [14], while mitigating duplicate detections and false positives. To validate the proposed framework, experiments were conducted using data acquired by a quadruped robot at an exterior abutment of a box-girder bridge. Additionally, a multi-scale inspection strategy was validated through experiments using interior diaphragm data captured by a handheld camera. These experiments demonstrated the practical applicability of the framework in real inspection environments.

2 Methodology

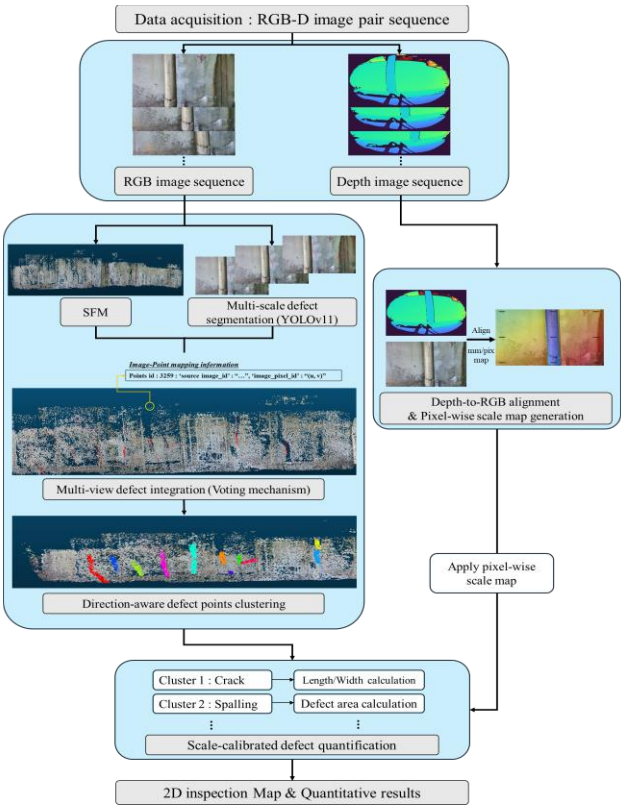


Figure 1. Overall Framework

Figure 1 shows the overall workflow of the proposed framework. The framework processes RGB-D image sequences, leveraging both 2D information from RGB images and 3D spatial information from depth data. Each processing module is described in detail in the following subsections.

2.1 SfM-based 3D reconstruction

SfM-based reconstruction is performed to integrate 2D RGB images into 3D space. The key advantage of SfM is that it establishes 2D-3D correspondence by identifying which pixel in which source image generated each point in the reconstructed point cloud. In overlapping regions between images, a single 3D point is mapped to multiple images, which forms the basis for multi-view integration. The proposed framework utilizes this 2D-3D correspondence to address duplicate detections and false positives arising from multi-view image sequences, as well as to support defect clustering and quantification. In addition, the reconstructed point cloud serves as the foundation for generating a 2D inspection map that visualizes the relative spatial distribution of defects within the inspected area.

2.2 Multi-scale defect detection

A YOLOv11x-seg model [15] is applied to each RGB image to detect and segment defects. The model was trained on nine types of surface defects in the dataset: crack, detachment, spalling, rebar exposure, material separation, damage, leak, lifting, and efflorescence. This paper used datasets from AI-hub [16]. All images are processed at the standard input resolution of 640 by 640 pixels. The model outputs a defect mask in polygon format for each detected defect, enabling the identification of defect type and boundary information within each RGB image.

To support both wide-area inspection and local-area inspection, the proposed framework provides two detection modes. Standard Mode uses the original image as input to the YOLOv11 model and detects defects directly. In contrast, Crop Mode divides the image into a grid of user-specified dimensions and performs YOLOv11 model for inference independently on each cropped image. To prevent defects from being split across crop boundaries, adjacent cropped images are overlapped by a specified number of pixels. For example, when using a 2 by 2 grid with 150-pixel overlap, defects detected in each cropped image can be geometrically merged. The continuous shape of defects that span multiple crops is then restored by computing the geometric union of overlapping defect polygons. While the grid size in Crop Mode can be selected flexibly without compromising detection performance, larger grids increase computational cost. Users can therefore choose the grid size based on site conditions and required detection precision.

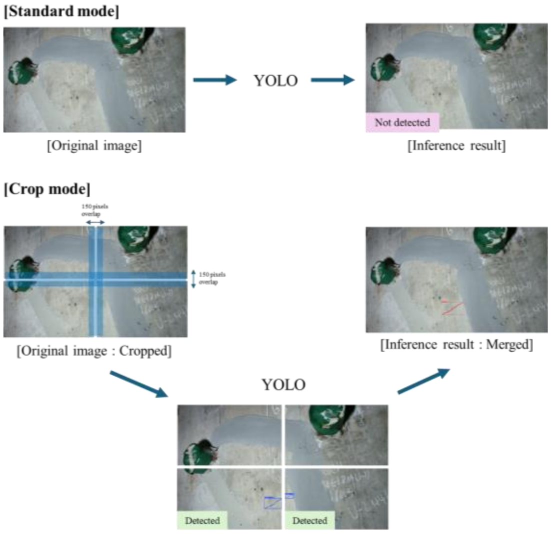


Figure 2. Standard Mode and Crop Mode for multi-scale defect detection (2×2 grid example)

2.3 Pixel-wise scale map generation

Since SfM-based reconstruction is inherently scale-free, depth information is required to convert pixel-based measurements into real-world physical units. To address the perspective-induced scale variation where the same physical dimension is represented by different numbers of pixels depending on the camera-to-surface distance, a pixel-wise scale map is generated for each image using depth information obtained from the RGB-D camera. This scale map computes the physical dimension per pixel (in millimetres per pixel) at each location in the image based on the depth value and camera focal length.

First, the depth image is aligned to the coordinate frame of the corresponding RGB image. This process is performed through coordinate-transformation and resolution matching using the extrinsic parameters between the RGB and depth cameras and the intrinsic parameters of each camera.

Each pixel (u_d, v_d) in the depth image and its corresponding depth value Z_d are back-projected to a 3D point $\mathbf{P}_d = (X_d, Y_d, Z_d)^\top$ in the depth camera coordinate frame using the depth camera intrinsic parameters. The depth camera intrinsic parameter matrix $\mathbf{K}_d$ is defined as:

$$\mathbf{K}_d = \begin{bmatrix} f_{x,d} & 0 & c_{x,d} \\ 0 & f_{y,d} & c_{y,d} \\ 0 & 0 & 1 \end{bmatrix}$$

where $f_{x,d}$, $f_{y,d}$ are the focal lengths of the depth camera and $c_{x,d}$, $c_{y,d}$ represent the principal point. The back-projection process is expressed as:

$$\begin{bmatrix} X_d \\ Y_d \\ Z_d \end{bmatrix} = Z_d \cdot \mathbf{K}_d^{-1} \begin{bmatrix} u_d \\ v_d \\ 1 \end{bmatrix}$$

The 3D point $\mathbf{P}_d$ expressed in the depth camera coordinate frame is then transformed to the RGB camera coordinate frame. This transformation is performed using the extrinsic parameters between the depth and RGB cameras, consisting of a rotation matrix $\mathbf{R}_{d\to r}$ and a translation vector $\mathbf{t}_{d\to r}$:

$$\mathbf{P}_r = \mathbf{R}_{d\to r}\, \mathbf{P}_d + \mathbf{t}_{d\to r}$$

where $\mathbf{R}_{d\to r} \in \mathbb{R}^{3\times 3}$ is the rotation matrix that rotates the depth camera coordinate frame to the RGB camera coordinate frame and $\mathbf{t}_{d\to r} \in \mathbb{R}^3$ is the translation vector representing the relative position difference between the two cameras.

The 3D point $\mathbf{P}_r = (X_r, Y_r, Z_r)^\top$ which has been transformed into the RGB camera coordinate frame, is projected onto the RGB image plane using the RGB camera intrinsic parameter matrix $\mathbf{K}_r$:

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \frac{1}{Z_r} \mathbf{K}_r \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix}$$

Through this process, each pixel in the depth image is mapped to a corresponding pixel (u_r, v_r) in the RGB image coordinate frame, establishing pixel-level correspondence between the RGB and depth images.

From the aligned depth map, the physical scale $s_x(u_s, v_s)$, $s_y(u_s, v_s)$ at each pixel is computed using the depth value $d(u_d, v_d)$ (in millimetres) and the RGB camera focal lengths f_x , f_y :

$$s_{x\ or\ y}(u_s, v_s) = \frac{d(u_d, v_d)}{f_{x\ or\ y}}$$

The pixel-wise scale map generated in this manner is created for each RGB image and is subsequently used in the defect quantification stage to convert pixel-based measurements into physical units.

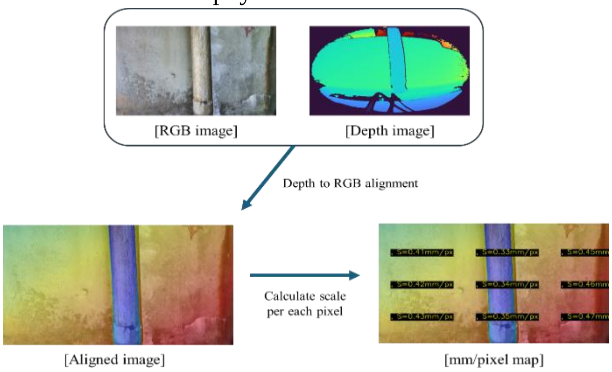


Figure 3. Scale map generation process

2.4 Multi-view defect integration in 3D space via two-stage voting mechanism

Defect polygons detected by the YOLOv11 model in each RGB image are integrated into 3D space through projection onto the point cloud generated by SfM. Using the 2D-3D correspondence established in the SfM database, each 3D point is back-projected to all RGB images in which it is visible. If the projected pixel coordinate falls inside a detected defect polygon, the point is designated as a defect point candidate. Through this process, each 3D point receives semantic information regarding defect presence and defect class from multiple RGB images. Since each 3D point uniquely represents a specific spatial location, the same defect can be represented as a single entity in 3D space even when observed in multiple images or captured in fragmented views.

During the projection process, false detections from the defect segmentation model must be considered. When acquiring numerous image sequences for inspection, false detections can occur frequently due to variations in lighting conditions and viewing angles. False detections can manifest in two forms: false positives where non-defect regions are incorrectly identified as defects, and misclassifications where defect classes are incorrectly predicted. To mitigate these issues, a two-stage voting mechanism is designed to leverage the multi-view nature of robotic inspection.

In Stage 1, the Detection Voting stage, the presence or absence of a defect at each point is determined. Each 3D point P_j has an associated set of images $\mathcal{I}_j$ in which it is observed. For each image $I_i \in \mathcal{I}_j$, the 3D point P_j is projected onto the image, and it is verified whether the projected pixel location falls inside any defect mask (including all defect classes). The defect presence voting ratio $v_d(P_j)$ for point P_j is defined as:

$$v_d(P_j) = \frac{1}{|\mathcal{I}_j|}\sum_{I_i \in \mathcal{I}_j} \mathbb{1}_{mask}(I_i, P_j)\ ,\ \tau_1 = 0.5$$

where $|\ \mathcal{I}_j\ |$ is the number of images in which point P_j is observed, and $\mathbb{1}_{mask}(I_i, P_j)$ is an indicator function that returns 1 if the projected location of point P_j in image I_i falls inside any defect mask, and 0 otherwise. In the proposed framework, a threshold τ_1 is applied, and only points satisfying $v_d(P_j) \geq \tau_1$ are selected as defect-present 3D points.

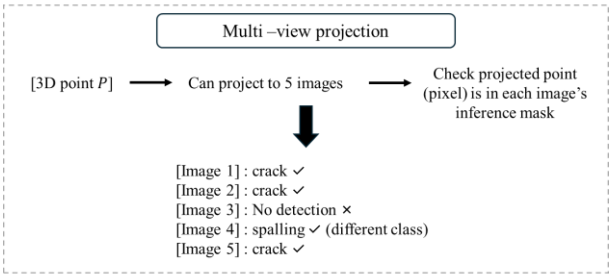
In Stage 2, the Class Voting stage, the final defect class is determined for each defect-present 3D point selected from Stage 1 by aggregating the observation frequency for each defect class. The voting ratio $v_c(P_j, c)$ for point P_j to be classified as class c is defined as:

$$v_c(P_j, c) = \frac{1}{|\mathcal{I}_j|}\sum_{I_i \in \mathcal{I}_j} \mathbb{1}_{mask}(I_i, P_j, c)$$

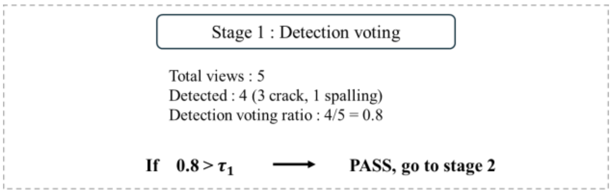
where $\mathbb{1}_{mask}(I_i, P_j, c)$ returns 1 if the projected location of point P_j in image I_i, falls inside a defect mask of class c, and 0 otherwise. The final defect class is determined through the following condition, which eliminates cases where defect classes are inconsistently predicted:

$$v_c(P_j, c)\ \geq\ \tau_2\ ,\ \ \tau_2\ =\ 0.5$$

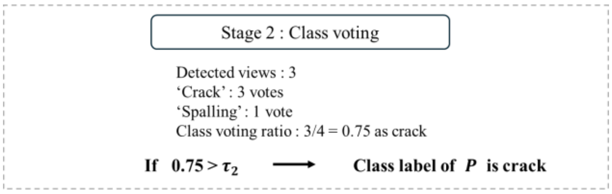
This two-stage voting mechanism mitigates false detections and misclassifications by labelling each point in the point cloud with a defect class based on the projection results. Through this process, multi-view defects are integrated into a unified 3D representation. Figure 4 illustrates this process.



(a)



(b)



(c)

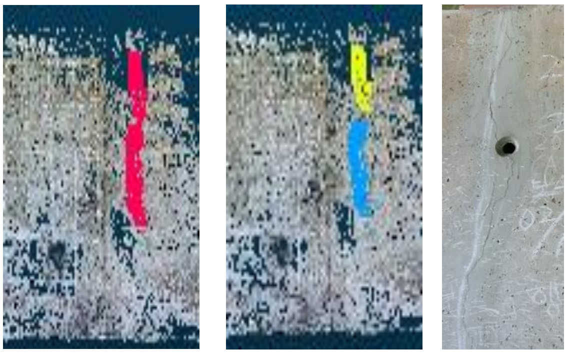
Figure 4. Two-stage voting mechanism example; (a) multi-view projection, (b) detection voting, (c) class voting

2.5 Direction-aware defect points clustering

The 3D defect points selected through voting stages must be clustered into individual defect instances. For defects such as cracks, conventional distance-based clustering using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) can erroneously merge intersecting cracks into a single cluster. To address this issue, a direction-aware clustering strategy that leverages local orientation information is applied.

First, density-based primary clustering is performed using DBSCAN to remove noise and extract candidate defect regions. Second, the local orientation of points within each cluster is estimated using Principal Component Analysis (PCA) and point groups exhibiting large orientation differences are separated to split intersecting defects into independent instances. Finally, remerging is performed based on centroid distance and principal component orientation similarity to restore contiguous defects that may have been inadvertently separated. This direction-aware clustering approach overcomes the limitations of simple distance-based clustering, enabling more accurate clustering of complex defect geometries in multi-view environments. Figure 5 illustrates the differences between clustering methods.

Through this process, the final defect clusters can be visualized in 3D space, and this information is subsequently utilized to generate a 2D inspection map. This 2D inspection map provides the spatial locations of defects within the structure and can be used in practical inspection workflows.



(a) (b) (c)

Figure 5. Differences by clustering method; (a) DBSCAN clustering, (b) proposed clustering, (c) original crack RGB image.

2.6 Scale-calibrated defect quantification

Using the pixel-wise scale map established in Section 2.3, each clustered defect instance is quantified in millimetres. Since point cloud cluster data lacks clear boundary definition and exhibits sparsity, leading to potential measurement errors, defects are quantified using the RGB images associated with points in each cluster through the 2D-3D correspondence. As illustrated in Figure 6, each 3D point within a defect cluster can be projected to multiple RGB images through the 2D-3D correspondence. By aggregating these projection relationships, the image most frequently associated with the cluster points is selected as the measurement source. This image selection process is independently performed for each defect cluster, ensuring that every cluster is quantified from its own most representative image. The selected image's pixel-wise scale map and defect mask are then loaded, and different measurement strategies are applied depending on the defect type.

2.6.1 Crack measurement: length and width

For cracks, the segmentation mask is narrow, requiring measurement of length and width, but direct calculation from the mask is impractical. Therefore, the single-pixel-width crack centreline skeleton extraction and measurement algorithm proposed by [6] is adopted to measure length and width. However, [6] presented a measurement approach for single images. In the proposed framework, the image with the highest number of point observations from the 3D crack cluster is selected, and measurement is conducted on that image. Furthermore, while [6] used a fixed scale factor, the proposed approach applies the spatially varying scale map to convert pixel-based measurements into real-world units, thereby reducing measurement errors arising from depth changes and perspective effects.

2.6.2 Non-crack defect measurement: Area

For non-crack defects such as detachment, spalling, and rebar exposure, area is used as the primary quantification metric. In the selected image for the defect instance, the physical area of each pixel within the defect mask is computed by multiplying the scale values in the u and v directions from the pixel-wise scale map, and the total defect area is obtained by summing these individual pixel areas in square millimeters. Unlike cracks, non-crack defects do not require single-pixel-width precision, allowing the segmentation mask to be used directly.

The quantification results for each defect are stored in JSON format. Using the 2D inspection map and defect quantification results, the final inspection report is generated.

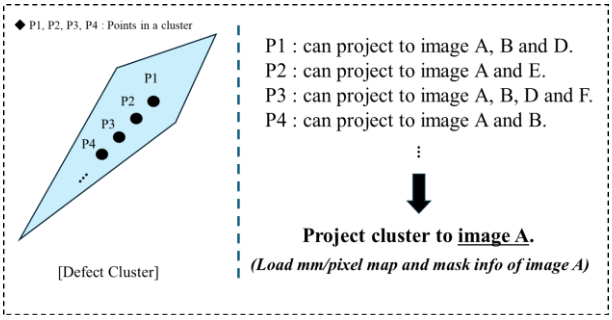


Figure 6. Scenario of image selection for defect cluster quantification

3 Experiments and Results

3.1 Data acquisition

For the experiments, data were collected using an RGB-D camera (Femto Bolt by Orbbec). In robot-based acquisition, the RGB-D camera was mounted on a quadruped robot (Go2 Edu by Unitree Robotics), with a mini-PC (Intel NUC 11) also equipped to store and manage the captured RGB-D image sequences (Figure 7(a)). In handheld acquisition, image sequences were acquired using the same camera model.

RGB and depth image sequences were captured as synchronized pairs. To ensure overlap for SfM-based 3D reconstruction, pairs were acquired at 0.5-second intervals. RGB-D images were captured at different resolutions per camera specifications (RGB: 3840 by 2160 pixels, depth: 512 by 512 pixels). Additionally, the RGB and depth lenses had different Fields of View (FOV). These characteristics and the extrinsic and intrinsic parameters were incorporated into the scale map generation described in Section 2.3.

3.2 Experiment Site

Images were acquired at an exterior abutment and an interior diaphragm of a Precast Segmental Method (PSM) concrete box-girder bridge (Figure 7 (b), (c)). Both sites represent different inspection scales within the multi-scale framework. At the abutment site, requiring wide-area inspection, the entire wall surface was captured using a quadruped robot, acquiring 698 RGB-D image pairs. At each position, images were taken at three camera orientations (upward, horizontal, downward tilt) from 50 to 70 centimetres standoff distance, advancing less than one metre between positions with five-second stationary periods to prevent motion blur. The diaphragm site represents a critical area where load-bearing prestressing tendons run through, requiring local-area inspection. A total of 212 RGB-D image pairs were acquired using a handheld camera. Since the diaphragm is inside the bridge structure, images were acquired with supplementary lighting due to limited natural lighting.

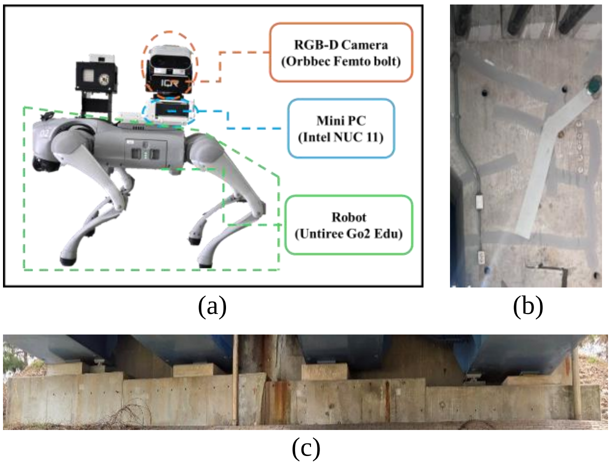


Figure 7. (a) Hardware configuration on quadruped robot, (b) diaphragm site, (c) abutment site

3.3 Results

3.3.1 Abutment site

Figure 8 presents the wide-area inspection results obtained by applying the proposed framework to the exterior abutment. Figure 8(a) shows the 3D point cloud of the abutment reconstructed through SfM using the RGB-D image sequences acquired by the quadruped robot. This reconstruction enables verification of the overall geometry and inspection coverage of the target structure at a structure-wide scale. Figure 8(b) presents the integrated defect information obtained through multi-view integration and clustering of detections from multi-view image sequences. While defects may be observed repeatedly or captured in fragmented views across different viewpoints, the proposed framework consolidates them into single defect instances, mitigating duplicate detections. Different colours are assigned to each cluster for visualization purposes.

Figure 8(c) shows the 2D inspection map generated based on the identified 3D defect instances. The map visualizes defect locations and types in a structured format, enabling clear visualization of wide-area inspection results. Figure 8(d) presents quantification results for each defect type in tabular form.

To quantitatively evaluate the detection performance, 33 unique defects were manually identified across the abutment surface and used as Ground Truth (GT). YOLO alone produced 1,762 detection masks across 689 images, where duplicate detections from multi-view image sequences make it impossible to determine the actual number of unique defects without 3D projection. When detections were projected into 3D space and clustered without the voting mechanism, the framework achieved a Precision of 0.429, Recall of 0.758, and F1 Score of 0.556. With the multi-view voting mechanism applied, Precision improved to 0.889 with an

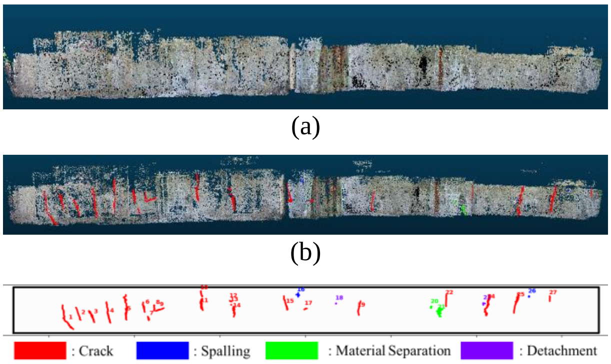
F1 Score of 0.781, demonstrating that the voting mechanism substantially reduces false positives while maintaining comparable Recall (0.697).

3.3.2 Diaphragm site

The diaphragm site represents a localized surface and requires finer inspection detail. Figure 9 demonstrates the effectiveness of the multi-scale inspection strategy at the interior diaphragm of the PSM box-girder bridge.

Figure 9 (a) shows the results obtained by applying Standard Mode to identify the overall defect distribution across the structure. In contrast, Figure 9 (b) presents the results of applying Crop Mode of the proposed framework to the region requiring local-area inspection. This comparison was conducted with all processing steps identical except for the detection scale. Based on YOLOv11 model inference on cropped images, defects that were not detected in Standard Mode were additionally detected.

These results demonstrate that the proposed framework can accommodate both wide-area inspection and local-area inspection within a unified framework, as required in practical inspection environments.



No.	Type	Length (mm)	Width (mm)	Area (mm^2)
5	Crack	1145.6	0.39	-
15	Crack	650.0	0.44	-
23	Detachment	-	-	2515.3
26	Spalling	-	-	888.7

(d)

Figure 8. (a) point cloud from SfM algorithm, (b) all defect points visualization, (c) 2D inspection map, (d) example of defect quantification results

(a) (b)

Figure 9. Multi-scale inspection comparison at diaphragm site; (a) wide-area inspection using Standard Mode, (b) local-area inspection using Crop Mode

4 Discussion

The experimental results demonstrated that the proposed framework operates as a robotic inspection system for real concrete bridge surfaces. In inspection environments using multi-view images, issues such as duplicate detections, fragmented observations, and false positives frequently occur. The proposed framework mitigates these issues in 3D space and subsequently performs defect quantification. As shown in Section 3.3.1, the quantitative evaluation confirmed that the multi-view voting mechanism substantially reduces false positives while maintaining comparable Recall. However, direct comparison with the manual inspection map remains limited because the manual inspection was conducted at a different time than the experimental data acquisition.

Meanwhile, 3D reconstruction through SfM was advantageous in establishing explicit 2D-3D correspondence but is limited by the scarcity of extractable features in the acquired images. Sparse point clouds were generated in feature-poor regions, such as areas covered with repair tape on the interior diaphragm of the box girder, due to the lack of distinctive visual features or motion blur. To overcome these limitations, future work could explore RGB-D Visual SLAM or fusion with LiDAR sensors. Additionally, while the diaphragm site experiments demonstrated the effectiveness of multi-scale inspection, the use of handheld camera data can also be considered a limitation.

While the detection performance has been quantitatively evaluated in Section 3.3.1, further validation of the quantification process remains necessary. Ground truth measurements of defect dimensions should be obtained through manual inspection conducted concurrently with robotic data acquisition to enable direct comparison with the framework's quantification outputs.

5 Conclusion

This paper proposed an integrated framework that unifies defect detection, quantification, and report generation into a continuous workflow for accessibility-constrained concrete bridge surfaces using multi-view RGB-D image sequences. The framework addresses multi-scale inspection by supporting wide-area and local-area inspection while systematically mitigating duplicate detections and false positives in 3D space.

Experiments at the exterior abutment and interior diaphragm of a PSM box-girder bridge confirmed reliable operation in real environments, generating quantification results and 2D inspection maps for various defect types. This work demonstrates the feasibility of advancing robotic bridge inspection from simple data acquisition to an integrated system encompassing multi-view analysis and quantification.

Future research should address the limitations identified in Section 4. More robust reconstruction approaches must be developed to ensure stable 3D reconstruction. Rigorous validation of the quantification process is required to demonstrate that the proposed framework provides reliable results.

Acknowledgment

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. 2018R1A6A1A08025348) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00457308). ※ MSIT: Ministry of Science and ICT. This research was conducted with the support of i-thetto Co., Ltd. and Korea Expressway Corporation.

References

[1] CERIK(Construction & Economy Research Institute of Korea). Construction Trends Briefing No.941 for 2024. On-line : https://www.cerik.re.kr/report/briefing/2821, Accessed: 15/01/2026.

[2] Abdallah, A. M., Atadero, R. A., & Ozbek, M. E. (2022). A state-of-the-art review of bridge inspection planning: Current situation and future needs. *Journal of Bridge Engineering*, *27*(2), 03121001.

[3] Zhang, Y., & Yuen, K. V. (2022). Review of artificial intelligence-based bridge damage detection. *Advances in Mechanical Engineering*, *14*(9), 16878132221122770.

[4] Kang, D., Benipal, S. S., Gopal, D. L., & Cha, Y. J. (2020). Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning.

[5] Moore, Mark, et al. *Reliability of visual inspection for highway bridges, volume I*. No. FHWA-RD-01-105. Turner-Fairbank Highway Research Center, 2001.

[6] Yang, K., Bao, Y., Li, J., Fan, T., & Tang, C. (2024). Deep learning-based YOLO for crack segmentation and measurement in metro tunnels. *Automation in Construction*, *168*, 105818.

[7] Lin, C., Chen, Y., Itakura, K., Maharjan, S., & Chun, P. J. (2025). Bridge inspection using image–point cloud fusion with image filtering, damage detection and 3D registration. *Automation in Construction*, *180*, 106538.

[8] Torok, M. M., Golparvar-Fard, M., & Kochersberger, K. B. (2014). Image-based automated 3D crack detection for post-disaster building assessment. *Journal of Computing in Civil Engineering*, *28*(5), A4014004.

[9] Deng, L., Sun, T., Yang, L., & Cao, R. (2023). Binocular video-based 3D reconstruction and length quantification of cracks in concrete structures. Automation in Construction, 148, 104743.

[10] Feng, C. Q., Li, B. L., Liu, Y. F., Zhang, F., Yue, Y., & Fan, J. S. (2023). Crack assessment using multi-sensor fusion simultaneous localization and mapping (SLAM) and image super-resolution for bridge inspection. Automation in Construction, 155, 105047.

[11] Li, X., Liu, C., Yang, X., & Yan, H. (2025). 3D surface defect detection and measurement using point cloud unwrapping via dimension reduction. *Automation in Construction*, *177*, 106349.

[12] Lattanzi, D., & Miller, G. (2017). Review of robotic infrastructure inspection systems. *Journal of Infrastructure Systems*, *23*(3), 04017004.

[13] Halder, S., Afsari, K., Chiou, E., Patrick, R., & Hamed, K. A. (2023). Construction inspection & monitoring with quadruped robots in future human-robot teaming: A preliminary study. *Journal of Building Engineering*, *65*, 105814.

[14] *Enforcement Decree of the Special Act on the Safety Control and Maintenance of Establishments* (Presidential Decree No. 35948, Jan. 2, 2026, Korea).

[15] Yan, T. YOLOv11 by Ultralytics. On-line : https://github.com/yt7589/yolov11, Accessed: 15/01/2026

[16] AI-Hub. (2023). *The SOC facility crack-pattern image dataset.* On-line : https://www.aihub.or.kr, Accessed: 15/01/2026