Render-to-Real Image-Based Change Detection of Outdoor Infrastructure Using 3D Gaussian Splatting

Satoko Hattori-Nagao[1], Kazuo Oda[1], Tomoaki Eguchi[1], Takanobu Nagao[1], Satomi Kakuta[1]

[1]Asia Air Survey Co. Ltd., 1-2-2 Manpukuji, Asao-ku, Kawasaki-shi, Kanagawa, Japan
- (stk.hattori, kz.oda, tma.eguchi, tnb.nagao, stm.kakuta)@ajiko.co.jp

Keywords: 3D Gaussian Splatting, Change Detection, Cross-Domain Image Matching, UAV images

Abstract
This study proposes a framework for detecting changes in outdoor civil infrastructure using bi-temporal images and validates its effectiveness through experiments on real-world datasets. The proposed method performs change detection by comparing a 3D Gaussian Splatting (3DGS) model reconstructed from multi-view images acquired before changes occur with a single real image captured from a new observation viewpoint after changes. The processing pipeline consists of: (1) construction of the 3DGS model, (2) generation of an initial rendered image corresponding to the post-change real image, (3) feature matching between the rendered image and the real image followed by camera pose estimation, and (4) change detection. Experiments conducted on a sediment control dam and a bridge dataset demonstrate that the proposed method achieves a maximum Intersection over Union (IoU) of 0.82 for change detection. Furthermore, compared to a baseline method based on bi-temporal real image pairs, the proposed method improves IoU by up to 24 percentage points. The results also indicate that even under limited acquisition conditions after changes, accurate change detection can be achieved when the 3DGS reconstruction quality and pose estimation are sufficiently reliable.

1. Introduction

In the maintenance and management of civil infrastructure, it is essential to rapidly assess site conditions, particularly after natural disasters. Bi-temporal image-based change detection is widely used. However, its effectiveness depends on several factors, including the number of available images, viewpoint differences, camera parameters, and the availability of pre-change 3D information, all of which influence the design of a processing pipeline.

When georeferenced images are available for both epochs—such as satellite or aerial imagery—many change detection methods have been developed with notable success (You et al., 2020). When multi-view imagery is available at both epochs, 3D neural scene representations such as 3DGS-CD (Zhang et al., 2024) enable change detection by comparing two reconstructed 3D models. However, post-disaster surveys often provide only a single image at the second epoch (post-change, t_1), limiting the applicability of multi-view-based approaches. In this study, we consider a setting in which multi-view imagery is available at the first epoch (pre-change, t_0), while only a single image or otherwise limited observations are available at t_1.

In this work, we propose a change detection framework that compares a 3D scene model constructed from multi-view imagery at t_0 with a real image captured from a new observation viewpoint at t_1, by rendering the scene reconstructed at t_0 from the viewpoint of the t_1 image. We also demonstrate the effectiveness of the proposed approach through experiments using real UAV imagery.

2. Render-to-Real Change Detection Framework

An overview of the proposed framework is illustrated in Figure 1, and the detailed processing pipeline is shown in Figure 2. The framework performs change detection by comparing a 3D Gaussian Splatting (3DGS) model reconstructed from multi-view images at t_0 with a single real image captured from a new observation viewpoint at t_1. This framework is designed for rapid post-disaster response and targets object-level changes such as rockfall, debris deposits, and the appearance or disappearance of structural elements.

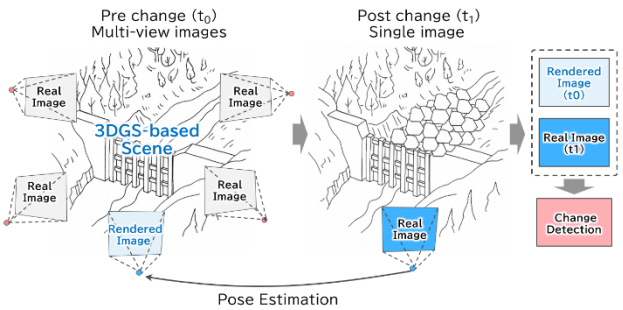

Figure 1. Conceptual overview of bi-temporal change detection using rendered images

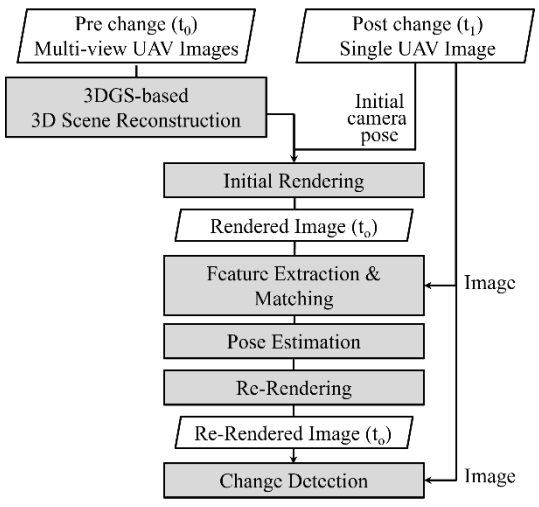

Figure 2. Framework of the proposed render-to-real image-based change detection.

2.1 3D Scene Reconstruction at t_0

At t_0, multi-view images are acquired using a UAV to construct a 3D scene model of the target area. The reconstruction is performed using 3DGS. In 3DGS, a scene is represented as a set of 3D Gaussians, each defined by its position, shape, color, and opacity.

During rendering, each Gaussian is projected onto the image plane, and pixel values are computed through weighted blending. This enables faster rendering and higher-quality novel view images compared to conventional NeRF-based methods. In this study, these properties are used to generate high-quality rendered images for change detection.

2.2 Initial Rendering for the t_1 Image

At t_1, the same area is captured from a new observation viewpoint. An initial rendered image corresponding to the new observation viewpoint is generated from the 3DGS model constructed at t_0. To estimate the initial camera position and orientation at t_1, the following three cases are considered:

i) Both position and orientation are known:
Camera position and orientation obtained from onboard GNSS and the camera gimbal are used as the initial parameters.

ii) Only position is known:
The GNSS-derived camera position is used, and multiple candidate orientations (Euler angles ω, ϕ, κ) are generated and evaluated.

iii) Both position and orientation are unknown:
Multiple candidate camera poses are generated based on the direction and elevation of the target structures.

2.3 Feature Matching and Pose Estimation for the New-View Real Image

To estimate geometric correspondences between the initial rendered image at t_0 and the real image at t_1, local feature points and descriptors are extracted from each image using SuperPoint (DeTone et al., 2018). SuperPoint is a convolutional neural network trained using self-supervised learning, which jointly detects keypoints that are robust to geometric transformations and illumination changes, and computes local descriptors for matching within a single network.

Next, the extracted feature descriptors are matched using LightGlue (Lindenberger et al., 2023) to estimate correspondences between the images. LightGlue is a Transformer-based matching method that models relationships between feature points using a self-attention mechanism. It also filters out unreliable matches based on confidence scores, enabling stable matching with reduced mismatches.

Based on the correspondences between the matched feature points and the associated 3D points of the rendered image, the camera pose is estimated by solving the Perspective-n-Point (PnP) problem. PnP is a geometric method that estimates the camera rotation and translation from multiple 2D–3D correspondences given known camera intrinsics parameters. In this study, RANSAC-based PnP is applied to reduce the effect of outliers, and the inlier set is determined based on the reprojection error.

For cases (ii) and (iii), pose estimation is performed for each candidate initial pose using PnP, and the solution with the highest inlier ratio is selected as the final camera pose.

2.4 Bi-Temporal Change Detection

Based on the estimated camera pose, the 3DGS model at t_0 is re-rendered. The resulting rendered image is then compared with the real image to extract change regions. For change detection, we use the Generalizable Scene Change Detection Framework (GeSCF, Kim and Kim, 2025), which is robust to differences in illumination and imaging conditions.

GeSCF consists of three main stages: (i) initial mask generation using the Segment Anything Model (SAM), (ii) feature extraction from the input images, and (iii) computation of change scores based on feature differences. This approach enables the detection of structural changes while reducing the influence of illumination variations and viewpoint differences, which are difficult to handle with simple pixel-wise comparison.

3. Field Tests

3.1 Dataset

In this study, field experiments were conducted at a sediment control dam (Site 1) located in a mountainous area and a bridge (Site 2), both in Japan. The acquisition conditions for each case are summarized in Table 1. At t_0, a 3DGS model was constructed using multi-view images, while at t_1, a limited number of images were treated as single images for change detection.

Site	Target	UAV / Camera	Images	Changes
Site 1*	Sediment control dam	DJI M300 RTK + Zenmuse P1	t_0: 388 t_1: 6	Cardboard box (rock), vehicle
Site 2	Bridge	Skydio 2+	t_0: 609 t_1: 5	Cones, vehicle

Table 1. Overview of the datasets used in the experiments.

* Data provided by the Kii Mountain District Sabo Office, Kinki Regional Development Bureau, Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan.

3.1.1 Site 1: Sediment Control Dam

Site 1 is a sediment control dam located in a mountainous valley in Japan. The area surrounding the dam is densely covered with trees and vegetation. UAV image acquisition was conducted twice, in October 2023 and December 2023, using a DJI M300 RTK equipped with a Zenmuse P1 camera. Each image includes camera position and orientation obtained from onboard GNSS and the camera gimbal.

No natural changes such as slope failure or rockfall occurred between the two epochs; therefore, cardboard boxes simulating rockfall were placed as artificial changes. At t_0, nadir images were captured with an overlap of 80% (forward) and 60% (side). At t_1, oblique images were captured.

3.1.2 Site 2: Bridge

Site 2 is a bridge located at the Fukushima Robot Test Field in Japan. This facility is a large-scale outdoor experimental site for robotics, where various infrastructure structures are installed. UAV image acquisition was conducted twice, in June 2023 and May 2024, using a Skydio 2+. Each image contains only camera position obtained from GNSS.

Since no actual changes occurred between the two epochs, objects such as traffic cones and vehicles were treated as changes. Oblique images were captured at both epochs.

3.2 Baseline Method

In this study, experiments were conducted using both the baseline method and the proposed method. The baseline method performs change detection directly between real images acquired at two epochs.

First, from the set of images at t_0, the image with the closest camera position and orientation to the image at t_1 is selected. For the selected image pair, local feature points and descriptors are extracted using SuperPoint, and correspondences are estimated using LightGlue. Next, RANSAC is applied to remove outliers from the matched correspondences. The image at t_0 is then aligned to the image at t_1 using an affine transformation. Finally, change regions are extracted using GeSCF.

3.3 Evaluation Metrics

Evaluation was conducted at each stage of the processing pipeline. The reconstruction quality of the 3DGS model at t_0 was quantitatively evaluated on the validation dataset using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM, Wang et al., 2004), and Learned Perceptual Image Patch Similarity (LPIPS, Zhang et al., 2018).
For feature matching, the number of matched feature points was used as the evaluation metric. For pose estimation from a new observation viewpoint, geometric consistency was evaluated using the number of inliers and the inlier ratio obtained from PnP with RANSAC (Fischler and Bolles, 1981):

$$r_{inlier} = \frac{N_{inlier}}{N_{all}} \quad (1)$$

where N_{inlier} = number of inlier correspondences
N_{all} = total number of correspondences

To evaluate projection consistency, the reprojection error was used. The reprojection error is defined as the Euclidean distance between the observed 2D point and the projected point of a 3D point onto the image plane using the estimated camera parameters:

$$u_i = \|u_i - \pi(X_i\ ;\ R, t, K)\|_2 \quad (2)$$

where X_i = 3D point
u_i = observed 2D image point
R, t = camera rotation and translation
K = camera intrinsic matrix
$\pi(\cdot)$ = projection function

In this study, both the mean reprojection error over all correspondences and the mean reprojection error over the inlier set I selected by RANSAC were used as evaluation metrics:

$$E_{all} = \frac{1}{N}\sum_{i=1}^{N} e_i \quad (3)$$

$$E_{inlier} = \frac{1}{|I|}\sum_{i\in I} e_i \quad (4)$$

where N = total number of correspondences
I = set of inlier correspondences selected by RANSAC

Finally, the change detection results between the two epochs were evaluated using the Intersection over Union (IoU) by comparing the detected change regions with manually annotated ground-truth masks.

3.4 Implementation Details

Experiments were conducted on a workstation equipped with an Intel Xeon Silver 4214R CPU × 2 (24 cores / 48 threads), 384 GB of memory, and an NVIDIA Quadro RTX 5000 GPU (16 GB VRAM). For 3DGS training, Splatfacto implemented in Nerfstudio v1.1.5 was used (Nerfstudio Development Team, 2023). The implementation environment consisted of Python 3.10, PyTorch 2.1.2, and torchvision 0.16.2.

4. Results and Discussion

4.1 3D Scene Reconstruction at t_0

3DGS was trained for the scene at t_0. Examples of rendered images on the validation dataset are shown in Figure 3, and the quantitative evaluation results are presented in Table 2.

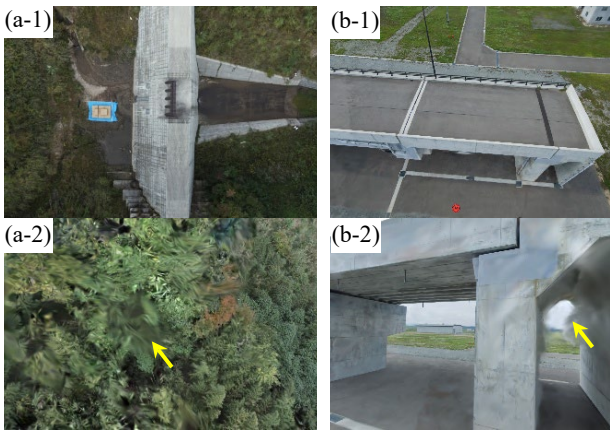

Figure 3. Rendered images for validation data: (a) Site 1, (b) Site 2. Yellow arrows indicate regions with low reconstruction quality.

Site	PSNR ↑	SSIM ↑	LPIPS ↓
Site 1	23.868 ± 5.213	0.664 ± 0.203	0.350 ± 0.222
Site 2	26.800 ± 3.579	0.765 ± 0.064	0.547 ± 0.111

Table 2. Quantitative evaluation of rendered images on the validation dataset (mean ± standard deviation).

For Site 1, the rendered results on the validation dataset show high overall fidelity. In particular, the geometry and texture of the sediment control dam are clearly reconstructed (Fig. 3a-1). In contrast, blur and artifacts are observed in vegetated areas and near the boundaries of the scene (Fig. 3a-2).

For Site 2, the overall structure of the scene is well reconstructed from a top-down viewpoint (Fig. 3b-1). However, the reconstruction quality degrades in the lower parts of the structure and in regions affected by occlusion (Fig. 3b-2).

Quantitative evaluation was performed using PSNR, SSIM, and LPIPS. Typical ranges of evaluation metrics for outdoor scenes have been reported in prior studies (Barron et al., 2022). Compared with these ranges, the PSNR, SSIM, and LPIPS obtained in this study are generally within or slightly below the typical values. This can be attributed to the effects of vegetation and occlusions, as well as viewpoint differences and variations in acquisition conditions.

4.2 Initial Rendering for the t_1 Image

Initial rendered images at t_0 corresponding to the new observation viewpoint of the real images at t_1 were generated from the 3DGS model constructed at t_0.

For Site 1, the camera position and orientation at t_1 are known; therefore, initial rendered images were generated using the camera pose obtained from onboard GNSS and the camera gimbal (Fig. 4). Rendered images from viewpoints close to the nadir direction show clear reconstruction with fine details. In contrast, rendered images from oblique viewpoints are less clear, making it difficult to recognize structures such as steel slits (Fig. 4a). This is because the 3DGS model at t_0 was trained using only nadir images, resulting in insufficient information for reconstructing the front and back surfaces of the dam. Furthermore, when the viewpoint is higher than the camera positions at t_0, ghosting artifacts appear in the rendered images due to artifacts near the boundaries of the scene (Fig. 4b).

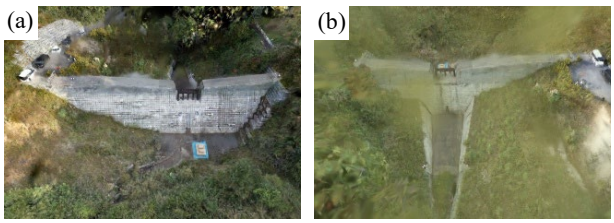

Figure 4. Initial rendered images for Site 1: (a) oblique viewpoint, (b) viewpoint affected by ghosting artifacts.

For Site 2, only the camera position at t_1 is known; therefore, multiple candidate camera orientations were defined based on the GNSS-derived position (Table 3). First, an approximate orientation was computed from the relative relationship between the bridge center and the camera position, and this was used as the initial orientation. Additional candidate positions were generated by shifting the camera position by 0 m, 3 m, and 5 m in the direction opposite to the viewing direction toward the center of the structure. This was intended to widen the field of view and improve the stability of feature matching. Candidate heights of -5 m, -3 m, and 0 m were also considered. The camera orientation was represented using Euler angles (ω,ϕ,κ), and candidate values of -5, 0, 5 degree were set for ϕ and κ relative to the initial orientation.

As a result, a total of 81 initial rendered images were generated for each new observation viewpoint at t_1 (an example is shown in Fig. 5). In addition, depth images for PnP-based pose estimation were generated for each rendered image. Since both epochs at Site 2 were captured from oblique viewpoints, the quality of the initial rendered images was comparable to that of the validation data.

Parameter	Description	Values
d	Distance from structure center	0, 3, 5 (m)
z	Height offset	-5, -3, 0 (m)
ω	Rotation about X-axis	0 (deg)
ϕ	Rotation about Y-axis	-5, 0, +5 (deg)
κ	Rotation about Z-axis	-5, 0, +5 (deg)

Table 3. Candidate settings of camera position and orientation for Site 2.

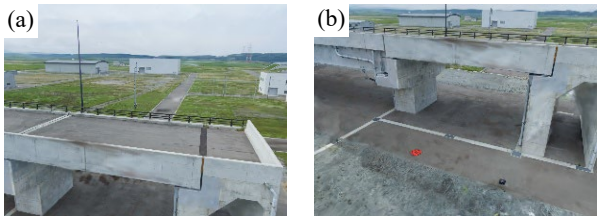

Figure 5. Initial rendered images for Site 2 under different camera parameter settings : (a) case with distance 0 m, height offset 0 m, φ +5°, κ +5°; (b) case with distance 5 m, height offset -5 m, φ -5°, κ -5°.

4.3 Feature Matching and Pose Estimation for the New-View Real Image

Feature extraction and matching were performed between the initial rendered image at t_0 and the real image at t_1, and the camera pose was estimated using PnP. The quantitative evaluation results are summarized in Table 4.

Site	Metric	Mean	Median	Max	Min
Site 1 (N = 6)	Matches	1398	1368	1623	1241
	Inliers	338	180	882	6
	Inlier ratio	0.22	0.13	0.54	0.00
	Reprojection error (all)	628.18	589.69	1095.14	364.82
	Reprojection error (inliers)	4.21	3.00	7.93	1.72
Site 2 (N = 5)	Matches	1363	1386	1488	1233
	Inliers	336	361	471	100
	Inlier ratio	0.24	0.27	0.32	0.08
	Reprojection error (all)	464.81	413.06	678.78	367.84
	Reprojection error (inliers)	4.28	4.92	5.14	3.04

Table 4. Quantitative evaluation results of feature matching and pose estimation. Mean, median, maximum, and minimum values are reported for each metric.

For Site 1, multiple images at t_1 were processed, and pose estimation was successful for 3 out of 6 image pairs, while the remaining 3 pairs failed. The successful pairs were all located near the center of the 3D scene, where the rendered image quality was relatively high. Although the maximum inlier ratio was 0.54, which is not particularly high, the inlier points were widely distributed across the image. Figure 6c-1 shows an example of successful feature matching at Site 1, where a sufficient number of correspondences were obtained between the rendered and real images. In contrast, for the failed pairs, the number of inliers was significantly reduced due to blur and ghosting artifacts near the boundaries of the scene (Fig. 6a-2). For these pairs, the number of inliers was fewer than 10, and almost no inliers were obtained in regions affected by blur or ghosting.

For Site 2, 81 image pairs were generated for each viewpoint, and the pair with the highest inlier ratio was selected. The maximum number of inliers was 471, and the maximum inlier ratio was 0.32, both of which were lower than those of Site 1. The inlier distribution was also concentrated in a limited region of the structure, showing a clear spatial bias (Fig. 6a-4). In the rendered images, very few feature points were detected on smooth concrete surfaces, and consequently, inlier points were rarely observed in these regions. A similar tendency was observed in distant regions, where inliers were also scarce. These trends were consistently observed across all five pairs.

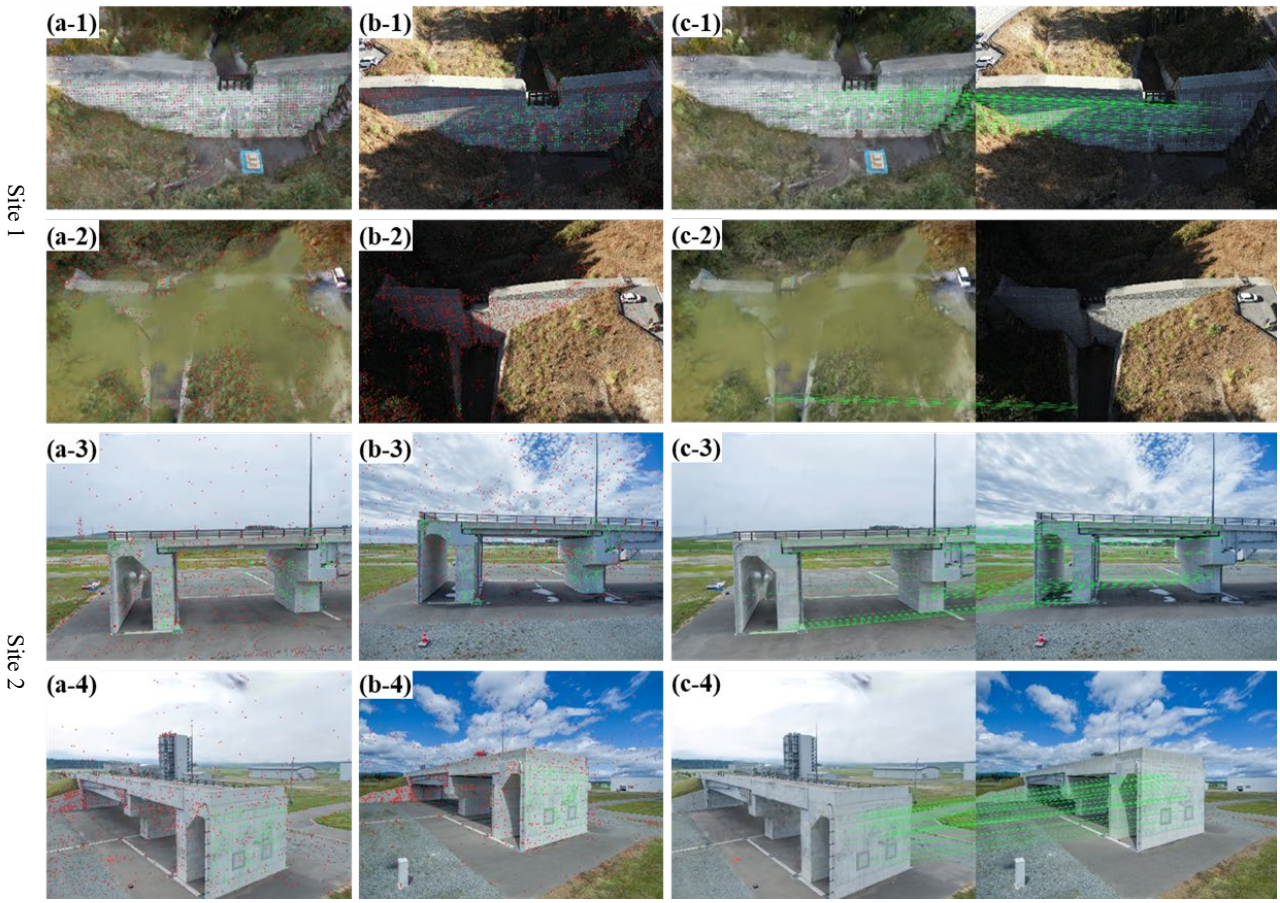

Figure 6. Feature extraction and matching results: (a) Detected feature points in the rendered image at t_0, (b) detected feature points in the real image at t_1, (c) matched feature correspondences between the rendered and real images. Green points indicate inliers, red points indicate outliers, and green lines represent matched correspondences.

4.4 Bi-Temporal Change Detection

Based on the estimated poses in Section 4.3, the 3DGS model at t_0 was re-rendered. The resulting rendered images were compared with the real images at t_1, and change regions were extracted using GeSCF. The change detection results were quantitatively evaluated using Intersection over Union (IoU) (Table 5).

Site	Method	IoU Mean	IoU Max	IoU Min
Site 1	Baseline (N = 0)	–	–	–
	Proposed (N = 3)	0.64	0.82	0.48
Site 2	Baseline (N = 5)	0.01	0.02	0.00
	Proposed (N = 5)	0.08	0.27	0.01

Table 5. Quantitative evaluation of change detection using IoU. Baseline results are not available for Site 1 due to failure in image matching and alignment.

The change detection results using the proposed method are shown in Figure 7. For Site 1, objects simulating rockfall (cardboard boxes) and vehicles (indicated by arrows in Fig. 7d-1 and 7d-2) were generally detected correctly. Although the images at t_1 contain large shadowed regions, change detection was successful, achieving a maximum IoU of 0.82 (Table 5). Relatively small changes were also detected. This can be attributed to the fact that inlier points were widely distributed across the image, leading to stable feature matching and reduced geometric misalignment between the image pairs. On the other hand, some vehicles were not detected (Fig. 7d-2, upper left). This is likely because GeSCF detects changes based on feature differences at the level of segmented regions obtained by SAM, rather than pixel-wise intensity differences. As a result, objects with similar position and shape are less likely to be identified as changes.

For Site 2, changes such as traffic cones and construction materials were successfully detected in image pairs with relatively accurate pose estimation (Fig. 7d-3). In contrast, for pairs with uneven inlier distributions, over-detection occurred across large regions of the image (Fig. 7d-4). This is because inlier points were concentrated on foreground structures, causing geometric misalignment in other regions and increasing false detections. The mean IoU was 0.08, indicating overall low change detection performance.

These results indicate that the accuracy of pose estimation is the most critical factor affecting change detection performance. In particular, when inlier points are evenly distributed across the image, geometric misalignment between image pairs is reduced, leading to more accurate change detection.

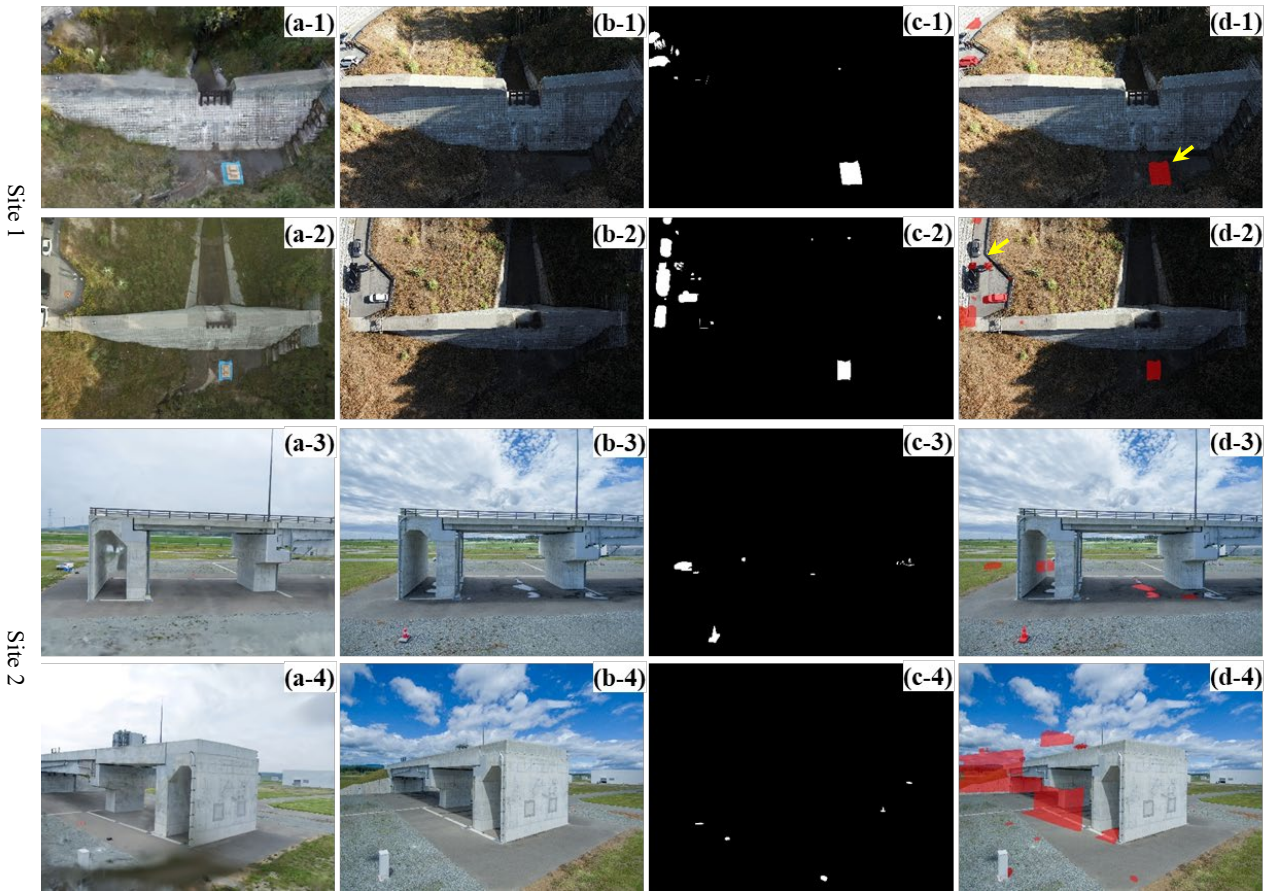

Figure 7. Change detection results using the proposed method : (a) Rendered image at t_0, (b) Real image at t_1, (c) Ground truth, (d) Overlay of detected changes on the t_1 image (red = detected changes)

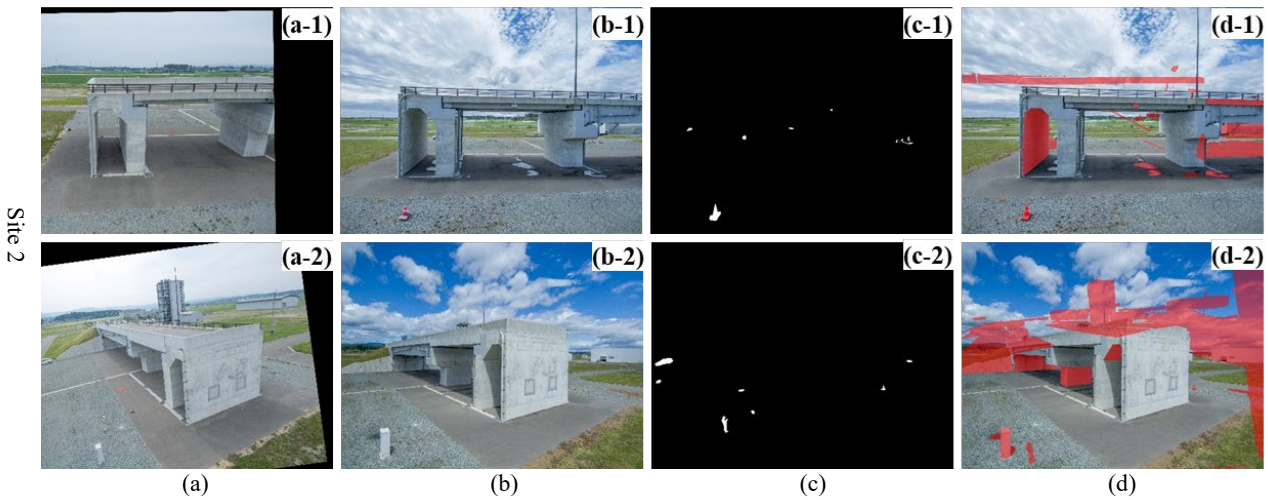

Figure 8. Change detection results using the baseline method : (a) Transformed real image at t_0, (b) Real image at t_1, (c) Ground truth, (d) Overlay of detected changes on the t_1 image (red = detected changes)

Furthermore, to achieve stable pose estimation, it is important to reduce blur and ghosting artifacts in the 3D scene and to generate rendered images from which feature points can be reliably extracted. In addition, under imaging conditions where the viewing direction is close to horizontal and distant structures are observed, feature points tend to concentrate in the foreground, resulting in reduced global consistency. In such cases, regions with a high density of inlier points should be treated as high-confidence areas, while regions with few or no inliers should be considered less reliable in the change detection results.

4.5 Comparison with the Baseline Method

In the baseline method, change detection was performed by applying feature matching followed by image alignment using an affine transformation.

For Site 1, due to large viewpoint differences, image retrieval failed for 4 out of 6 viewpoints. Even for the two successful pairs, stable feature matching and geometric transformation could not be achieved because of differences in viewing direction and field of view. As a result, change detection could not be performed for Site 1.

For Site 2, feature matching and affine transformation were successfully applied to all five pairs. However, the number of inliers remained low, ranging from 7 to 57, and the maximum inlier ratio was only 0.07. These results indicate that geometric consistency was insufficient (Table 6).

Site	Method	Metric	Mean	Max	Min
Site 1 (N = 0)	Baseline	Matches	–	–	–
		Inliers	–	–	–
		Inlier ratio	–	–	–
Site 2 (N = 5)	Baseline	Matches	666	1049	382
		Inliers	23	57	7
		Inlier ratio	0.03	0.07	0.01

Table 6. Matching performance of the baseline method.

Figure 8 shows an example of a pair with a relatively high inlier ratio in the baseline method. However, the change detection results reveal a large number of false detections and over-detection caused by image distortion and misalignment. These results suggest that image alignment based on a simple affine transformation is not sufficient for stable change detection under significant viewpoint differences.

In contrast, the proposed method improved the IoU by up to 24 percentage points compared to the baseline. By using rendered images based on a 3D scene, the proposed method can generate geometrically consistent image pairs even under large viewpoint differences. As a result, it is less affected by misalignment and enables more stable change detection.

5. Conclusion and Future Work

In this study, we proposed a change detection framework that compares a 3DGS model at t_0 with a single-view image at t_1, and validated its effectiveness using real-world data. The results demonstrate that even under limited acquisition conditions at t_1, major changes can be successfully detected when the reconstruction quality of 3DGS and the accuracy of pose estimation are sufficiently high.

On the other hand, it was confirmed that ghosting artifacts and distortions in 3DGS degrade the quality of rendered images, which in turn affects the accuracy of pose estimation and change detection. In addition, when targeting structures that extend significantly in the depth direction, correspondences in distant regions are less likely to be identified as inliers. This leads to reduced accuracy in both pose estimation and change detection.

Experiments were conducted on two outdoor field sites in this study. Although evaluation using open datasets was considered, datasets that fully satisfy the conditions of bi-temporal outdoor infrastructure are limited, and thus comprehensive validation could not be performed. Since the proposed method addresses a problem setting not covered by existing datasets, further validation with additional case studies is necessary.

For future work, we plan to evaluate the proposed method under more diverse scenes and imaging conditions. In addition, improvements in rendering quality and feature matching are expected to enable more stable change detection in outdoor environments.

Acknowledgements

The dataset of sediment control dams used in this study was provided by the Kii Mountain District Sabo Office, Kinki Regional Development Bureau, Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan. The authors also appreciate the valuable advice and support provided during this study.

References

Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P., 2022: Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, pp. 5470–5479.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. SuperPoint: Self-Supervised Interest Point Detection and Description. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2018).* arXiv: 1712.07629.

Fischler, M. A., Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6), pp. 381–395.

Kim, Y., & Kim, G., 2025. Towards Generalizable Scene Change Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025).* arXiv:2409.06214.

Lindenberger, J., Sarlin, P.-E., Pollefeys, M., 2023. LightGlue: Local Feature Matching at Light Speed. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2023).* arXiv: 2306.13643.

Nerfstudio Development Team, 2023. Nerfstudio. Open Source Project. https://nerf.studio (accessed 20 March 2026).

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004: Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process*., 13(4), 600-612.

You, Y., Cao, Q., Zhang, Y., 2020. A Survey of Change Detection Methods Based on Remote Sensing Images for Multi-Source and Multi-Objective Scenarios. *Remote Sensing*, 12(15), 2460.

Zhang, H., Chen, Y., Wang, Z., et al., 2024. 3DGS-CD: Change Detection Using 3D Gaussian Splatting Representations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2024).* arXiv: 2411.03706.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., Wang, O., 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, pp. 586–595.