2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Anomaly-Aware Change Detection for Oil Refinery Inspection Using a Mobile Robot with Viewpoint-Aligned Novel View Synthesis

Tomohito Hoshii[1], Takuya Igaue[2], Jun Younes Louhi Kasahara[2], Masayoshi Kinoshita[3], Risa Koda[3], Shota Shimizu[3], Shinji Kanda[2], Hajime Asama[4], Qi An[1] and Atsushi Yamashita[1]

Abstract— **Stable operation of oil refineries is essential for ensuring a continuous supply of petroleum products, which underpin the foundation of modern society. However, various anomalies such as leaks can occur even during regular operation, making periodic inspection indispensable. To reduce the burden on human operators, automated visual inspection using mobile robots has been attracting increasing attention as a promising alternative to manual inspections. One promising approach, which we refer to as anomaly-aware change detection, is to compare videos captured during past and current inspections to identify scene changes specifically caused by anomalies. However, the robot cannot perfectly retrace its previous path, resulting in viewpoint misalignment between the two videos, which significantly degrades the performance of naive frame-wise comparison methods. To address this issue, we propose a novel inspection method that reconstructs a 3D model of the refinery from the past inspection video using Structure from Motion and 3D Gaussian Splatting, and then renders novel view images from the same viewpoints as those in the current inspection video. This allows us to obtain geometrically aligned image pairs, enabling anomaly-aware change detection that is robust to viewpoint misalignment. Field experiments conducted in an operational oil refinery achieved an F1-score of 0.806, significantly outperforming conventional methods and demonstrating the effectiveness of our approach.**

I. INTRODUCTION

Petroleum remains a cornerstone of contemporary civilization. Beyond fueling automobiles, planes, and ships, it constitutes a fundamental feedstock for a wide array of petrochemical products, including synthetic fibers and polymer resins. Consequently, securing a stable and reliable oil supply is critical for sustaining global economic and social stability.

Petroleum products are manufactured in oil refineries. A stable supply of petroleum products depends on the continuous operation of each oil refinery. However, aging equipment and machinery inevitably lead to anomalies that reduce operational efficiency and trigger unplanned shutdowns. Regular inspections are therefore essential for the early detection of anomalies and implementing appropriate countermeasures.

[1]T. Hoshii, Q. An and A. Yamashita are with the Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8563, Japan.

[2]T. Igaue, J. Y. Louhi Kasahara and S. Kanda are with the Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.hoshii@robot.t.u-tokyo.ac.jp

[3]M. Kinoshita, R. Koda and S. Shimizu are with ENEOS Corporation, 1-1-2 Otemachi, Chiyoda-ku, Tokyo 100-8162, Japan.

[4]H. Asama is with Tokyo College, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

At present, anomaly detection relies on periodic patrol inspections by oil refinery operators. Each operator walks through the refinery and checks for anomalies using visual, auditory, and tactile cues, typically conducting four to five patrol inspections per day.

However, operator-based inspections face several challenges. Inspection accuracy differs between experienced and inexperienced operators, and the tacit knowledge of experienced operators is difficult to transfer to inexperienced operators [1]. Moreover, the pool of experienced operators is aging, chronic labor shortages persist, and night-time patrol inspections impose a substantial burden on operators.

Automating inspection tasks is essential to overcome these challenges. In particular, this study focuses on the visual inspection of piping and associated mechanical components, which together constitute the majority of infrastructure in oil refineries. These components are responsible for transporting raw materials and products throughout an oil refinery, and failures, especially at pipe joints, valves, and other connection points, are common sources of leaks and other anomalies. Because such anomalies often manifest as visible surface changes, visual inspection is highly effective for their early detection.

A promising strategy for visual anomaly detection is to detect changes by comparing images captured during past and current inspections. AnoDFDNet [2] detects anomalies by highlighting changes between two spatially aligned images taken at different times, but does not address how such image pairs should be selected or aligned. For mobile robot patrol inspections, Shimizu et al. [3] proposed pairing frames from past and current videos using estimated robot poses and feature-based image matching. However, their method does not compensate for viewpoint misalignment between the two videos, which significantly degrades detection performance.

The goal of this study is to develop a change detection method that identifies visual anomalies by comparing videos captured during past and current patrol inspections. We refer to this task as anomaly-aware change detection, which aims not only to detect changes between scenes, but also to determine whether such changes are anomalous or acceptable. A key challenge in this setting is the viewpoint misalignment caused by the mobile robot's imperfect path retracing, which often leads to misaligned frame pairs and degraded detection performance. To address this, we aim to synthesize reference views from the past inspection video to match the viewpoints

978-1-6654-5784-2/26/$31.00 ©2026 IEEE

of the current frames, enabling robust anomaly-aware change detection even under viewpoint misalignment. We evaluate the proposed method through field experiments conducted in an operational oil refinery.

II. RELATED WORKS

A. Sensor-Based and Image-Based Anomaly Detection

A wide range of studies have been conducted to automate inspection tasks for piping systems in industrial plants. Various sensor-based methods have been proposed for detecting pipe anomalies, including ultrasonic sensors for leak detection [4], fiber optic sensors for monitoring temperature changes [5], and both visible lights and infrared cameras [6], [7]. While these methods enable continuous monitoring, they require the installation of a large number of sensors to cover an entire oil refinery, resulting in high installation and maintenance costs.

Mobile-sensor approaches have also been proposed as alternatives to fixed-sensor methods. By enabling sensors to move, these approaches allow wide areas to be inspected with fewer sensors. In particular, robots that perform internal pipe inspections by traveling through the pipe interiors have been widely explored [8]–[11]. While such robots can inspect long stretches of piping using only a few onboard sensors, they require insertion ports for deployment and are physically constrained by pipe diameter, making them unsuitable for small-diameter pipes.

To avoid requiring special infrastructure on the piping, inspection methods using visual sensors mounted on unmanned aerial vehicles [12] or pipe-external mobile robots [13] have been proposed. These methods typically detect anomalies based solely on images during a single inspection, without explicitly comparing them to data from previous inspections. However, in oil refinery inspections, daily patrols are conducted, which makes past inspection data available. Taking advantage of such historical data can enable more robust and reliable anomaly detection.

One approach to utilizing historical inspection data is to compare images of the same scene captured at different times, a process known as change detection [14]. However, not all observed changes correspond to anomalies, as some changes may be caused by lighting variations. Therefore, it is important to distinguish between normal changes and those caused by anomalies. One method designed with this purpose is AnoDFDNet [2], a deep neural network that detects anomalies by comparing two spatially aligned images captured at different times. However, they did not describe how such image pairs should be selected or aligned, leaving the pairing step unresolved. Moreover, applying anomaly-aware change detectors like AnoDFDNet to inspections conducted by camera-equipped mobile robots presents further challenges. The two input images must observe the same object from comparable viewpoints. To address this, Shimizu et al. [3] proposed pairing frames from past and current inspection videos using estimated robot poses and feature-based image matching. Although the robot cannot perfectly retrace its path, their method does not compensate for the resulting viewpoint misalignment between the two videos. This misalignment significantly degrades the performance of the change detector. Therefore, methods that can handle such viewpoint misalignment are desirable, especially in mobile robot inspection settings.

B. 3D Reconstruction-Based Anomaly Detection

Recent advances in 3D reconstruction have introduced methods such as Neural Radiance Fields (NeRF) [15] and 3D Gaussian Splatting (3DGS) [16], which can recover highly detailed 3D scene structures from multiple images. These methods enable the construction of a dense 3D representation from previously captured images or videos, allowing novel view images to be rendered from arbitrary viewpoints.

Studies have also explored applying 3D reconstruction techniques to anomaly detection. Zhou et al. [17] proposed OmniPoseAD, which uses NeRF to build a 3D representation of a scene from training images depicting normal conditions. The method then renders an image from the same viewpoint as a query image and compares the two to detect anomalies. Similarly, Kruse et al. [18] introduced SplatPose & Detect, which uses 3DGS and likewise addresses viewpoint misalignment between training and query images.

Although these methods leverage 3D reconstruction to address viewpoint differences, they have so far been evaluated only on synthetic datasets with relatively simple scenes. Moreover, they typically require optimizing the camera pose of each query image, which can be computationally intensive and prone to failure in complex environments such as oil refineries. These limitations suggest that further development is needed to make 3D reconstruction-based anomaly detection practical for real-world industrial inspections.

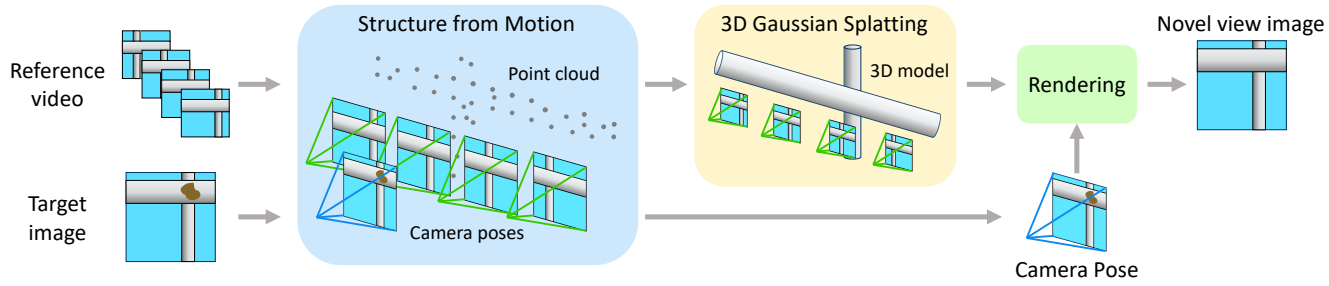
III. PROPOSED METHOD

A. Method Overview

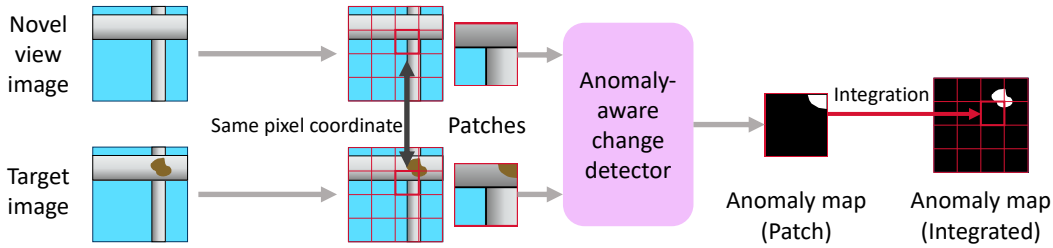
An overview of the proposed method is illustrated in Figure 1. To improve reliability in anomaly detection, our method compares two inspection videos: a reference video recorded during a past patrol inspection and a target video recorded during the current patrol inspection. Assuming that the reference video captures the normal state, anomalies are detected as changes observed in the target video. Both videos are assumed to be captured by a mobile robot patrolling along the same nominal route.

To enable robust comparison under viewpoint misalignment, we reconstruct geometrically aligned reference views using 3D reconstruction techniques. In contrast to methods that require per-frame pose optimization, we estimate camera poses using SfM, which improves scalability and robustness in real-world environments. Furthermore, by leveraging the temporal continuity of inspection videos, our method integrates detection results across frames, enhancing reliability through spatial consistency.

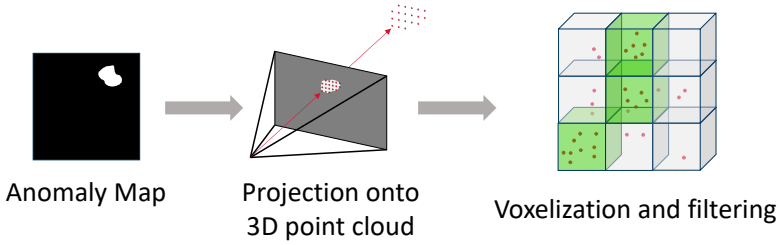
First, frames are extracted from both reference and target videos. The extracted frames are referred to as reference and target images, respectively. SfM [19] is then used to estimate the camera poses for each reference and target image.



(a) 3D structure estimation and novel view synthesis.



(b) 2D anomaly map generation.



(c) 3D anomaly localization.

Fig. 1: Overview of the proposed method.

Next, using the reference images as input, we reconstruct the 3D structure of the refinery at the time of the reference video capture by applying 3DGS [16], which enables fast and high-quality 3D reconstruction and rendering. Based on the reconstructed model, we synthesize novel view images from the viewpoints of the target images. These synthesized images and the target images can be regarded as having been captured from the same camera poses at different times.

Because anomalous regions typically occupy only a small portion of the entire image, we crop small patches from each image to increase the likelihood that an anomaly, if present, dominates the cropped patch. This improves the effectiveness of the anomaly-aware change detector. Due to viewpoint misalignment between the reference and target videos, including differences in camera position and field of view, some patches may not be reliably reconstructed in the synthesized images. We therefore exclude such patches from the detection process. We integrate the outputs of the anomaly-aware change detector for all remaining patches to obtain the final detection results for the target images.

Finally, these per-image detection results are integrated in 3D space. By considering only those 3D regions that are consistently flagged as anomalous across multiple images, we reduce false positives and enhance the reliability of the results, while also making the results easier to understand through spatial visualization.

B. Data Acquisition and Preprocessing

Detecting anomalies from a single image is challenging, especially in complex environments like oil refineries. To address this, we compare inspection videos captured from two different time points, enabling more reliable anomaly detection. We use a reference video recorded during a past patrol inspection and a target video recorded during the current patrol inspection. Frames are extracted from each video and referred to as reference images and target images, respectively. Each reference and target image is then used as input to SfM to estimate the camera pose at the time of capture. SfM additionally provides 3D point clouds, along with the estimated intrinsic and extrinsic parameters for all reference and target images. SfM is used because it provides globally consistent camera poses and 3D structure estimation, allowing us to align reference and target views without per-frame pose optimization.

C. 3D Structure Estimation and Novel View Synthesis

To enable accurate anomaly detection, we aim to compare each target image with a corresponding view of the refinery under normal conditions, as captured in the reference video. However, since the reference and target videos are recorded at different times and from slightly different viewpoints, directly comparing frames from the two videos is unreliable due to viewpoint misalignment. To address this, we reconstruct a 3D model of the refinery from the reference video and use it to synthesize novel view images from the same viewpoints as the target images. This ensures that both images are viewed from the same camera pose, enabling a reliable comparison between the normal and current states of the refinery.

In this study, we adopt 3DGS, which enables fast and high-quality image synthesis. 3DGS is a technique that places a set of 3D Gaussian distributions in 3D space, where each distribution represents color and shape information, based on images captured from multiple viewpoints. These Gaussians are then used to render images from arbitrary viewpoints.

As inputs to 3DGS, we use the reference images and the outputs of SfM, which include the intrinsic and extrinsic camera parameters as well as the 3D point cloud. These inputs are used to reconstruct a 3D model of the refinery as it appeared at the time of the reference video capture. At this stage, no information from the target images is used to generate the 3D model. However, the target images will later be used independently in another 3DGS process for estimating depth, separate from the 3D reconstruction based on the reference images.

Next, we generate novel view images from the viewpoints of the target images using the reconstructed 3D model from the reference images. For this process, we use the intrinsic and extrinsic camera parameters of the target images, as estimated by SfM. Each synthesized image represents a projection of the 3D model of the refinery, as it appeared at the time of the reference video capture, from the viewpoint of the corresponding target image.

D. Patch Extraction and Filtering

We perform change detection on each pair consisting of a synthesized novel view image and its corresponding target image. To enhance the relative prominence of anomalous regions within the images, which typically occupy only a small portion of the entire image, we crop small patches from each image. Following the approach of Shimizu et al. [3], we select patches at the same pixel coordinates in both images. In their method, differences in camera positions often led to inconsistencies between the paired patches, which degraded detection performance. In contrast, our method ensures alignment between the synthesized novel view and the target image, as they share the same camera pose and field of view, thereby improving change detection performance.

However, synthesized novel view images may contain incomplete regions. When the camera positions differ between the reference and target videos, some objects in the target images may not be sufficiently reconstructed in the synthesized images. Even when the fields of view are similar, regions near the edges of the reference video may still lack reliable 3D reconstruction due to insufficient coverage. To address these issues, we assess whether each cropped patch is sufficiently reconstructed. Patches determined to be unreconstructed are excluded from the detection process and do not contribute to the final detection results.

To determine whether a cropped region in a target image is sufficiently reconstructed, we check whether the region is visible in multiple reference images. As illustrated in Figure 2, we project the four corners of the cropped region from the target image onto each reference image using the intrinsic and extrinsic camera parameters estimated by SfM. To perform this projection, we require depth information for each target image, which we obtain by computing the median of the 3D feature points reconstructed by SfM that lie within the cropped patch. Let $(u_\mathrm{t}, v_\mathrm{t})$ be a pixel coordinate of a corner of the cropped patch in the target image, and let $\mathbf{K}_\mathrm{t} \in \mathbb{R}^{3\times3}$ be the intrinsic camera matrix, and $\mathbf{R}_\mathrm{t} \in \mathbb{R}^{3\times3}$ and $\mathbf{t}_\mathrm{t} \in \mathbb{R}^3$ be the extrinsic parameters (rotation and translation) of the target image as estimated by SfM. We define $d \in \mathbb{R}$ as the approximate depth at $(u_\mathrm{t}, v_\mathrm{t})$, estimated from the median depth of SfM-reconstructed points within the patch. Then, the 3D point in world coordinates corresponding to $(u_\mathrm{t}, v_\mathrm{t})$, denoted by $(x_\mathrm{w}, y_\mathrm{w}, z_\mathrm{w})$, is given by:

$$\begin{pmatrix} x_\mathrm{w} \\ y_\mathrm{w} \\ z_\mathrm{w} \end{pmatrix} = \mathbf{R}_\mathrm{t}^{-1} \left\{ \mathbf{K}_\mathrm{t}^{-1} d \begin{pmatrix} u_\mathrm{t} \\ v_\mathrm{t} \\ 1 \end{pmatrix} - \mathbf{t}_\mathrm{t} \right\}. \tag{1}$$

To determine whether this 3D point is visible in a reference image, we project it using the intrinsic matrix $\mathbf{K}_\mathrm{r} \in \mathbb{R}^{3\times3}$ and extrinsic parameters $\mathbf{R}_\mathrm{r} \in \mathbb{R}^{3\times3}$ and $\mathbf{t}_\mathrm{r} \in \mathbb{R}^3$ of the reference image:

$$\begin{pmatrix} u_\mathrm{r} \\ v_\mathrm{r} \\ 1 \end{pmatrix} \propto \mathbf{K}_\mathrm{r} \left\{ \mathbf{R}_\mathrm{r} \begin{pmatrix} x_\mathrm{w} \\ y_\mathrm{w} \\ z_\mathrm{w} \end{pmatrix} + \mathbf{t}_\mathrm{r} \right\}, \tag{2}$$

where $(u_\mathrm{r}, v_\mathrm{r})$ is the projected pixel coordinate in the reference image.

We then count how many reference images contain all four projected points within their image boundaries. If the number of such reference images falls below a predefined threshold, the region is considered unreconstructed and is excluded from the anomaly detection process.

E. Anomaly-Aware Change Detection

Each pair of cropped patches from the synthesized novel view image and the corresponding target image is input to an anomaly-aware change detector, such as AnoDFDNet [2]. Our method does not depend on a specific model and can incorporate any method designed to detect anomalies based on differences between the two images. Since anomalies are typically localized and small, patch-based comparison helps the detector focus on fine-grained differences between the normal and current states. The detection results obtained for each patch pair are then integrated based on the pixel

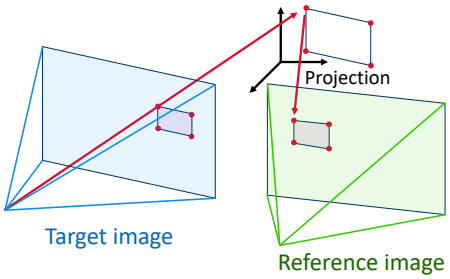


Fig. 2: Projection of the corners of a cropped region from the target image to the reference image.

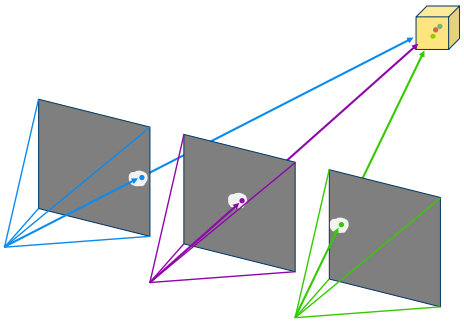


Fig. 3: Example of a voxel where projected anomalies from multiple target images overlap.

coordinates in the corresponding target image. This allows us to obtain a complete detection result for each target image.

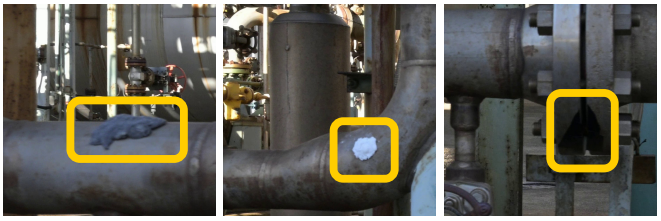
F. Integration of Detection Results into 3D Space

The overall procedure for integrating the detection results into 3D space is illustrated in Figure 1 (c). For each target image, the corresponding detection results are projected into 3D space to generate a set of 3D points. To obtain the depth information for this projection, we use the 3D model reconstructed by 3DGS with the target image as input, allowing us to estimate the depth at each pixel in the target image. Then, we divide the 3D space into voxels and evaluate each voxel based on how many points from different target images fall within it. As illustrated in Figure 3, when a voxel contains points originating from multiple target images, it is more likely that a true anomaly exists within that voxel. Therefore, we define a threshold and consider a voxel to contain an anomaly only if the number of distinct target images contributing points to the voxel exceeds the threshold. The threshold is determined based on how many times the same spatial region appears across the target images. This aggregation process enhances the reliability of anomaly detection in each voxel.

IV. EXPERIMENTS

A. Condition

To validate the effectiveness of the proposed method, we conducted experiments using four videos obtained at an oil refinery operated by ENEOS Corporation. A straight path approximately 9 meters long was defined within the refinery,



(a) Gray clay used to simulate an anomaly. (b) White clay used to simulate an anomaly. (c) Vinyl tape used to simulate an anomaly.

Fig. 4: Example of simulated anomalies.

and all videos were recorded along this path. The camera used for recording was a Nikon Z9 equipped with a NIKKOR Z 24–70mm f/4 S lens. The focal length was fixed at 24 millimeters, the frame rate was 30 frames per second, and the resolution was $7,680 \times 4,320$ pixels. Each video was approximately 45 seconds in duration and was recorded by manually moving a cart equipped with the camera, which we regarded as a proxy for a mobile robot. For safety reasons, as the experimental equipment was not explosion-proof, the experiments were conducted under the supervision of refinery personnel carrying a gas detector.

Two of the four videos contain simulated anomalies. Clay and vinyl tape were used to simulate these anomalies. Specifically, as shown in Figure 4, gray and white clay was used to represent deposits that may adhere to pipelines due to leakage, while vinyl tape was used to simulate colored fluid leaks. The simulated anomalies in the two anomaly-containing videos were designed to have different shapes.

From these videos, we constructed two reference-target pairs, where one anomaly-free video was used as the reference, and the other anomaly-containing video was used as the target. We used AnoDFDNet by Wang et al. [2] as the anomaly-aware change detector, as it is a high-performing method well-suited to our problem setting. One video pair was used as the training dataset, and the other was used as the validation dataset.

B. Data Processing

We extracted frames from the target video at 0.5 second intervals and from the reference video at 0.2 second intervals. From the target video used for evaluation, 76 frames were extracted, of which 42 frames contained simulated anomalies. For each extracted frame, we performed SfM using COLMAP [19] to estimate the intrinsic and extrinsic camera parameters at the time of capture, along with a sparse 3D point cloud. Specifically, our process utilized the `feature_extractor` module, specifying the camera model as `SIMPLE_PINHOLE`, and the `exhaustive_matcher` module for feature matching. The sparse reconstruction (camera poses and 3D points) was then obtained using the `mapper` module.

In addition to the proposed method, we compared the methods of Shimizu et al. [3], which uses image pairs from inspection videos without compensating for viewpoint differences. For both methods, we defined a 960×960-

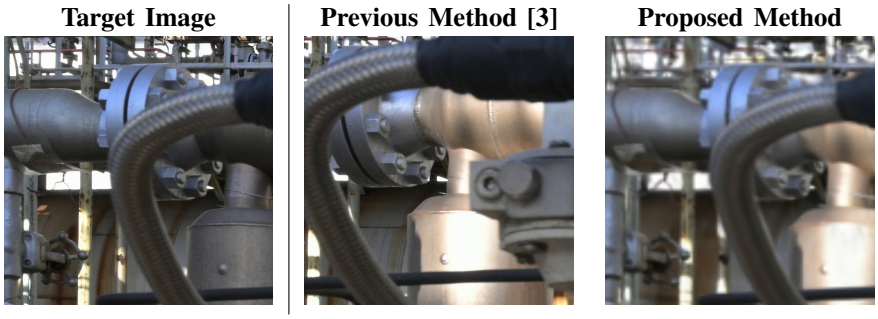


Fig. 5: Example of extracted regions for the proposed method and comparison method.

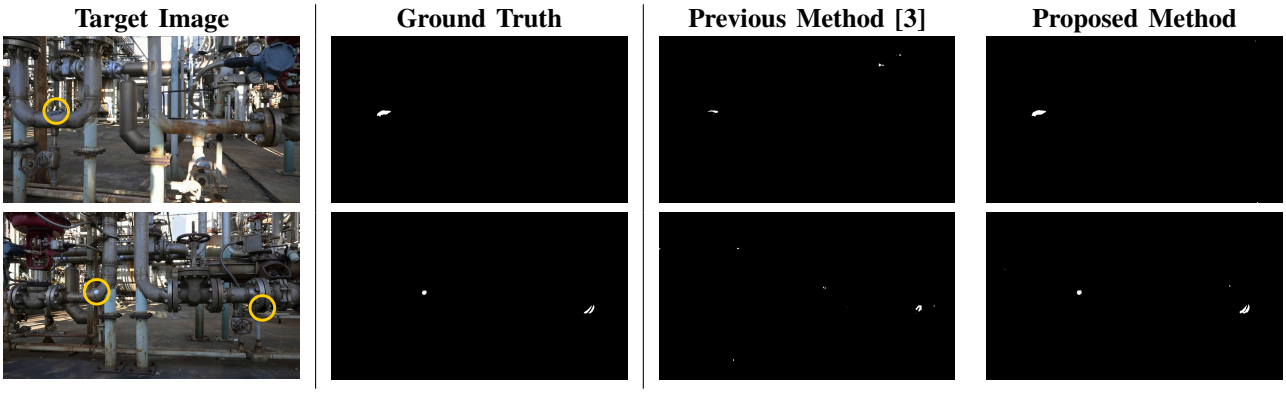


Fig. 6: Example of anomaly detection results for the proposed method and comparison method. Black and white indicate normal and anomalous regions, respectively. Yellow circles in the target images indicate the locations of simulated anomalies.

pixel window to specify the regions to be cropped. The window was moved with a stride of 480 pixels in both the horizontal and vertical directions. To increase the proportion of anomalous image pairs in the training data, we applied undersampling and used 500 pairs for training and 100 pairs for validation. The number of training epochs was set to 100, and the threshold for anomaly detection was set to 0.5. If a pixel was included in multiple patches, it was regarded as anomalous if any of the corresponding patch-wise detections indicated an anomaly. In the proposed method, cropped regions that were determined to be observed from four or fewer reference images were considered unreconstructed and excluded from the detection results. To reconstruct the 3D model with 3DGS, we utilized the official implementation provided by the authors [16] and trained separate models using the reference and target videos, each for 30,000 iterations.

C. Results

Figure 5 shows extracted regions obtained by the proposed method and the comparison method. In the proposed method, it can be observed that objects from the foreground to the background appear at consistent positions within the image, whereas the comparison method misaligned the regions due to differences in viewpoint.

Figure 6 shows examples of the target images, ground truth labels, and detection results by the proposed method and the comparison method. Compared to the comparison method, the proposed method more accurately captures the shapes of the anomaly regions and results in fewer false positives.

Table I shows the detection results for the proposed method and the comparison method. The proposed method outperformed the comparison method in all metrics, including accuracy, precision, recall, F1 score, and Intersection over Union (IoU), demonstrating its effectiveness.

As part of our proposed method, we integrated the detection results into 3D space and applied spatial filtering, both to represent anomalies in 3D space and to enhance reliability. The 3D space was divided into voxels of approximately 0.02 meters per side, and voxels that were associated with anomaly detections from at least 8 different target images were retained. The 3D model and the filtered results are shown in Figure 7, where voxels identified as containing anomalies are highlighted in green. The visualization was performed using CloudCompare [20], a 3D point cloud processing software. For ease of visualization, the 3D model of the refinery was reconstructed using OpenMVS [21], a type of Multi-View Stereo. The locations of the simulated anomalies are indicated with yellow circles in the figure. The aggregation process helped reduce false positives in both methods by suppressing inconsistent detections. However, due to its lower per-image detection precision, the comparison method still resulted in several false positive voxels. In contrast, the proposed method yielded three connected components, all of which corresponded to true anomaly locations, with neither false positives nor false negatives.

TABLE I: Detection results for the proposed method and comparison method.

Method	Accuracy ↑	Precision ↑	Recall ↑	F1 Score ↑	IoU ↑
Previous Method [3]	0.999	0.099	0.154	0.121	0.064
Proposed Method	**1.000**	**0.741**	**0.883**	**0.806**	**0.675**

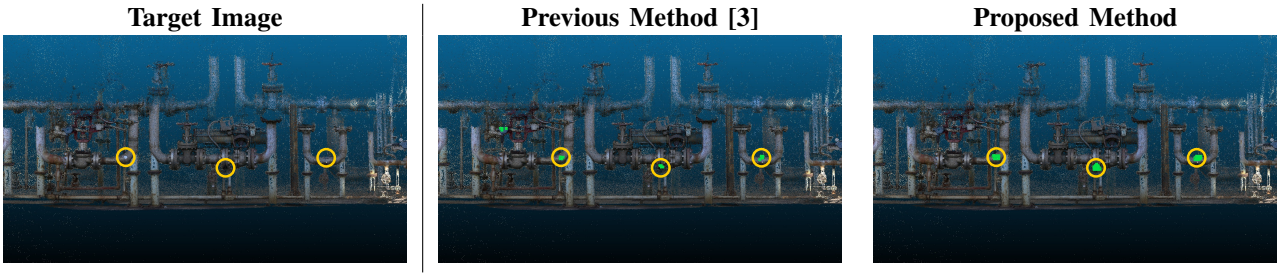


Fig. 7: 3D visualization of anomaly detection results. The yellow circles indicate the locations of simulated anomalies, and the green voxels indicate detected anomalies after filtering.

The processing time of each stage was measured on a desktop PC equipped with an Intel Core i7-14700F CPU, an NVIDIA RTX 4070 Ti SUPER GPU, and 128 GiB of RAM. SfM, applied to 309 high-resolution images (7680 × 4320), took approximately 17 minutes. 3DGS training, using 233 resized images (960 × 540) due to memory constraints, took approximately 25 minutes. Rendering of novel view images was nearly instantaneous and negligible in the overall runtime. The anomaly detector (AnoDFDNet) was trained in 27 minutes, and inference on 76 target images took 9 minutes in total. Each target image was divided into multiple cropped patches before inference, resulting in a total of approximately 7,400 processed patches across all 76 images.

V. DISCUSSION

While anomaly detection might be possible in feature space even with different viewpoints, we posit that explicitly aligning the images as our method does by synthesizing novel view images is a more robust and effective approach for this task. This alignment was especially beneficial when combined with AnoDFDNet, the anomaly-aware change detector used in this study. Since AnoDFDNet is based on convolutional neural networks, bringing semantically corresponding regions into spatial proximity is critical. When the viewpoints differ significantly, low-level feature differences become meaningless, leading to poor detection performance. Aligning viewpoints mitigates this problem and allows the detector to better capture meaningful changes. Additionally, for fair evaluation, each detector was trained on image pairs produced by its corresponding method. The better-aligned pairs in our approach likely enabled more effective learning of both normal and anomalous features.

3DGS provides fast and high-quality reconstruction, and in real-world inspection scenarios, the 3D model can be built during non-operating hours, making its computational cost negligible. However, since SfM is also used to estimate the camera poses of the target images, which must be computed for each new inspection and can become a bottleneck. While real-time processing is ideal, we note that our method is primarily intended for offline detailed inspection; if the analysis completes within a few hours after data collection, it is sufficient for most inspection workflows. Incorporating faster pose estimation methods, such as SLAM-based approaches or localization techniques against the existing 3D model, could further reduce this processing time and enhance the practicality of the proposed method.

Challenges in change detection arise from both environmental disturbances, such as lighting variations, and path misalignments. Our method primarily focuses on the latter, as path retracing is particularly difficult to control perfectly. However, the proposed method also shows robustness to changes in lighting conditions, such as differences in sunlight, which are often encountered in outdoor environments. This is attributed to the anomaly detector being trained to distinguish between changes caused by actual anomalies and those caused by non-anomalous factors such as lighting conditions. Through supervised learning, the model learns to recognize meaningful differences while ignoring irrelevant differences. However, this capability relies on having labeled data for both normal and anomalies. In practice, collecting a diverse and sufficient set of anomaly examples is often challenging. Therefore, developing methods that require fewer or no anomalous labels remains an important direction for future research. Furthermore, a more rigorous validation that decouples these environmental factors from path misalignment remains an important future direction.

While the detection algorithm assumes that a mobile robot patrols along a predefined path, the path planning itself is not yet addressed in this study. Especially in complex refinery environments where objects such as pipes may occlude each other, determining how the robot navigates around such obstacles (e.g., behind piping) remains an open problem. Incorporating motion planning and exploration strategies is essential for practical deployment.

Furthermore, while this study focuses on oil refinery inspections and uses a ground-based mobile robot, the proposed approach is not limited to this platform. It is also applicable to aerial inspection scenarios using unmanned aerial

vehicles. Exploring such extensions could further broaden the applicability and versatility of the proposed method.

VI. CONCLUSION

In this study, we proposed an anomaly-aware change detection method for oil refinery inspection using a mobile robot equipped with a camera. To perform anomaly detection based on image-based change detection, two images capturing the same object at different times are required. However, differences in camera positions and fields of view between videos often cause misalignment between images, making accurate comparison challenging. To address this issue, we proposed a method that utilizes 3DGS to synthesize novel view images. This allows us to generate a pair of images without misalignment, thereby improving the performance of anomaly detection.

Experimental results on real-world data collected at an oil refinery demonstrated that the proposed method outperformed the comparison method in terms of accuracy, precision, recall, F1 score, and IoU. Furthermore, by integrating the detection results into 3D space and applying spatial filtering, the method successfully suppressed false positives and localized anomalies accurately.

Future work includes further validation of the proposed method under more complex conditions (e.g., dense piping, occlusions, or moving personnel), characterizing the type of detected anomalies (e.g., crack, corrosion, or fluid leaks), a detailed ablation study to analyze the contribution of each pipeline component, broader experimental comparisons with other 3D reconstruction-based methods, real-world implementation in oil refinery environments, improvement of pose estimation speed, and adaptation to other platforms such as unmanned aerial vehicles for broader applicability.

REFERENCES

[1] R. Takamido, S. Kurihara, Y. Umeda, H. Asama, S. Kasahara, Y. Tanaka, S. Fukumoto, T. Kato, M. Korenaga, M. Hoshi, and J. Ota, "Evaluation of expert skills in refinery patrol inspection: Visual attention and head positioning behavior," *Heliyon*, vol. 8, no. 12, p. e12117, 2022.

[2] Z. Wang, Y. Zhang, L. Luo, and N. Wang, "Anodfdnet: A deep feature difference network for anomaly detection," *Journal of Sensors*, vol. 2022, no. 1, p. 3538541, 2022.

[3] S. Shimizu, T. Igaue, J. Y. Louhi Kasahara, N. Yamato, S. Kasahara, H. Ito, T. Daito, S. Tamura, A. Sasamura, T. Kato, F. Nonaka, S. Kanda, K. Nagatani, H. Asama, Q. An, and A. Yamashita, "Change detection in image pairs for plant inspection using mobile robot," *International Journal of Automation Technology*, vol. 19, no. 3, pp. 178–191, 2025.

[4] X. Diao, Z. Chi, J. Jiang, A. Mebarki, N. Lei, Z. Wang, and Y. Hao, "Leak detection and location of flanged pipes: An integrated approach of principle component analysis and guided wave mode," *Safety Science*, vol. 129, p. 104809, September 2020.

[5] I. Walker and D. Carr, "Fibre optic leak detection," in *Proceedings of the OTC Offshore Technology Conference*, May 2003, pp. OTC–15 360–MS.

[6] S.-O. Kim, J.-S. Park, and J. W. Park, "A leak detection and 3d source localization method on a plant piping system by using multiple cameras," *Nuclear Engineering and Technology*, vol. 51, no. 1, pp. 155–162, 2019.

[7] M. Fahimipirehgalin, E. Trunzer, M. Odenweller, and B. Vogel-Heuser, "Automatic visual leakage detection and localization from pipelines in chemical process plants using machine vision techniques," *Engineering*, vol. 7, no. 6, pp. 758–776, 2021.

[8] J. Elliott, R. Fletcher, and M. Wrigglesworth, "Seeking the hidden threat: Applications of a new approach in pipeline leak detection," in *Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference*, November 2008, pp. SPE–118 070–MS.

[9] S. Berjaoui, R. Alkhatib, A. Elshiekh, M. Morad, and M. O. Diab, "Free flowing robot for automatic pipeline leak detection using piezoelectric film sensors," in *Proceedings of the 2015 International Mediterranean Gas and Oil Conference*, 2015, pp. 1–3.

[10] D. Waleed, S. H. Mustafa, S. Mukhopadhyay, M. F. Abdel-Hafez, M. A. K. Jaradat, K. R. Dias, F. Arif, and J. I. Ahmed, "An in-pipe leak detection robot with a neural-network-based leak verification system," *IEEE Sensors Journal*, vol. 19, no. 3, pp. 1153–1165, 2019.

[11] Y. Zhao, Z. Su, H. Zhou, and J. Lin, "Video detection of small leaks in buried gas pipelines," *IEEE Access*, vol. 11, pp. 109 708–109 721, 2023.

[12] S. Roos-Hoefgeest, J. Cacace, V. Scognamiglio, I. Álvarez, R. C. González, F. Ruggiero, and V. Lippiello, "A vision-based approach for unmanned aerial vehicles to track industrial pipes for inspection tasks," in *2023 International Conference on Unmanned Aircraft Systems*, 2023, pp. 1183–1190.

[13] N. Bao, Y. Fan, Z. Ye, and A. Simeone, "A machine vision—based pipe leakage detection system for automated power plant maintenance," *Sensors*, vol. 22, no. 4, 2022.

[14] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *International Journal of Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[15] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[16] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, July 2023.

[17] Q. Zhou, W. Li, L. Jiang, G. Wang, G. Zhou, S. Zhang, and H. Zhao, "Pad: A dataset and benchmark for pose-agnostic anomaly detection," *arXiv preprint arXiv:2310.07716*, 2023.

[18] M. Kruse, M. Rudolph, D. Woiwode, and B. Rosenhahn, "Splatpose & detect: Pose-agnostic 3d anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 2024, pp. 3950–3960.

[19] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4104–4113.

[20] CloudCompare (version 2.11.3) [GPL software]. [Online]. Available: https://www.cloudcompare.org

[21] D. Cernea, "OpenMVS: Multi-view stereo reconstruction library," 2020. [Online]. Available: https://cdcseacave.github.io/openMVS