Change Detection in Pipe Image Pairs Extracted from Inspection Videos by Sequential Filtering

Susumu Shimizu[1], Takuya Igaue[1], Jun Younes Louhi Kasahara[1],
Naoya Yamato[2], Seiji Kasahara[2], Hiroyuki Ito[2],
Taizo Daito[2], Sunao Tamura[2], Akinobu Sasamura[2], Toshiya Kato[2],
Shinji Kanda[3], Keiji Nagatani[1], Hajime Asama[1], Qi An[4] and Atsushi Yamashita[4]

Abstract— **In this study, we propose a method for detecting changes on the outer surface of pipes using inspection videos captured by an inspection robot. It is critical to detect anomalies on the outer surface of pipes during patrol inspections. Anomalies are defined as deviations from the normal state and should be detected as areas that have changed from the normal state. Therefore, for appropriate maintenance of the plants, it is crucial to perform change detection by comparing videos that capture the past normal state with those capturing the current state. The problem with detecting changes from videos is deciding which frame to compare. We therefore propose sequential filtering to determine image pairs based on the position of the images and their similarity. We then apply a deep learning method to perform change detection. An indoor simulated plant environment has been constructed to test the efficacy of the proposed method. Experiments and evaluation results showed that the proposed method outperformed an autoencoder. The proposed method also achieved an F1 score of 0.880 for change detection in the inspection videos by introducing sequential filtering, which prevented mismatching of image pairs and reduced computational costs.**

I. INTRODUCTION

Petroleum refineries and petrochemical plants process raw materials such as crude oil, heavy oil, and petroleum gas to produce a wide range of products, including fuel oil, petroleum products, and petrochemical products such as synthetic fibers and plastics [1]. In these plants, pipes are widely used to transport raw materials and products [2], [3]. These pipes are susceptible to corrosion caused by various factors, such as components in the raw materials, moisture from rainwater, and temperature. This can result in various anomalies, including internal fluid leaks and structural damages to the pipes [4].

Field operators conduct regular patrol inspections at the plant all day and night to detect anomalies early and ensure proper maintenance. Specifically, visual inspections are important in detecting anomalies such as corrosion, leaks, and structural damages. However, human-based inspections often have issues such as a decrease in the number of expert field operators and difficulty in transferring skills to learners. Consequently, camera-based inspection methods have been studied as an alternative to human visual inspections.

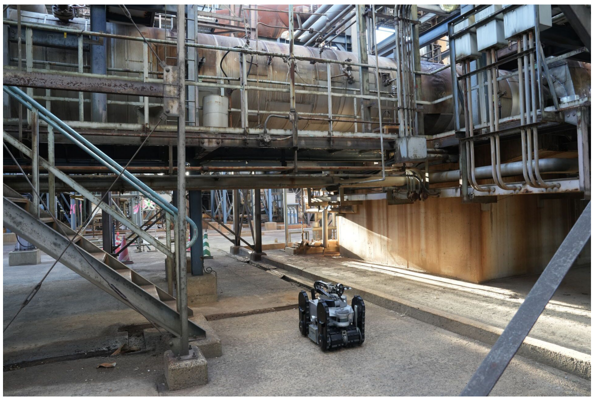


Fig. 1. An inspection robot in a petroleum refinery (ENEOS Corporation).

There have been studies that used multiple fixed cameras to detect gas leaks from pipes [5], and an infrared camera installed in chemical process plants to inspect for pipe leaks [6]. One of the drawbacks of these studies is the requirement of installing fixed cameras within the plant. This means that a substantial number of cameras must be deployed to cover the entire area of large plants. Therefore, as shown in Fig. 1, there has been an increasing trend in recent years to automate patrol inspections using inspection robots equipped with cameras [7], [8]. This approach offers extensive coverage, bypassing the need for fixed cameras and enabling inspections in large plants using just one inspection robot.

Generally, the patrol inspection of plants is considered an anomaly detection task aimed at detecting specific phenomena such as corrosion, cracks, and leaks [9], [10]. Anomalies are defined as deviations from the normal state. From this definition, a comparison with the normal state, that is, posing the anomaly detection problem as a change detection method, is appropriate. Moreover, certain anomalies, such as a sudden emergence of an anomalous object, are difficult to detect from a single image [11]. Consequently, the focus during

[1] S. Shimizu, T. Igaue, J. Y. Louhi Kasahara, K. Nagatani and H. Asama are with the Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8563, Japan. shimizu@robot.t.u-tokyo.ac.jp

[2] N. Yamato, S. Kasahara, H. Ito, T. Daito, S. Tamura, A. Sasamura and T. Kato are with the Engineering & Capital Planning Department, ENEOS Corporation, 1-1-2 Otemachi, Chiyoda-ku, Tokyo 100-8162, Japan.

[3] S. Kanda is with the Research into Artifacts, Center for Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.

[4] Q. An and A. Yamashita are with the Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8563, Japan.

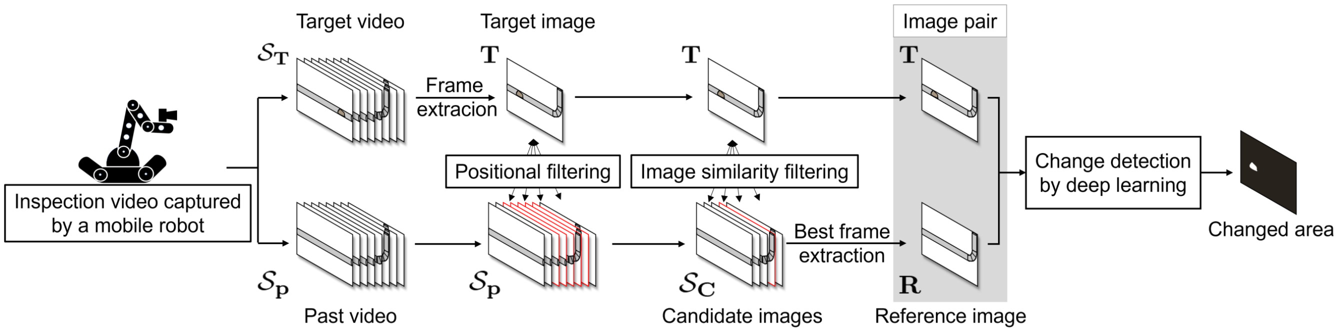


Fig. 2. Overview of the proposed change detection system in inspection videos.

inspections should be on detecting changes between the past normal state and the current state.

The difficulty in detecting changes between the past and current states captured by an inspection robot lies in the variation in shooting conditions during each inspection. The inspection robot follows a predetermined path in the plant and captures videos of the facility. The path taken may slightly differ during each inspection due to errors in self-localization by odometry or simultaneous localization and mapping (SLAM) [12]. Moreover, inspections are conducted at different times of the day, leading to variations in light intensity between inspections. Therefore, change detection in videos captured during robot inspections must be resistant to changes in the robot's position and light intensity.

Image subtraction is one of the most fundamental ideas used in detecting changes [13]. However, this method cannot fully absorb the effects of camera position and brightness changes, leading to false positives and negatives in detecting changes [14]. A more sophisticated method for change detection involves using an autoencoder [15]. Autoencoders are widely used in change detection because of their robustness against changes in camera position. Changes between the past and current images can be obtained by inputting the current image into an autoencoder model that has been trained with past images. However, to conduct change detection using an autoencoder, it is necessary to train a model for each past time point. This requires a significant amount of training data and time, making it impractical for an inspection robot to acquire sufficient data for a large plant.

Wang et al. [16] proposed a change detection method that is adaptable for data acquisition by a robot. In this method, a Siamese network was designed using a pair of images captured in the same region at different times as input. The network then outputs the regions where changes have occurred. However, this method does not address the problem of determining the input image pairs. Consequently, it cannot select appropriate image pairs for comparison from the videos captured during past inspections and the current inspection, which hinders the detection of changes in the inspection videos. Therefore, this study aims to achieve change detection on the outer surface of pipes by identifying proper image pairs.

II. PROPOSED METHOD

A. *Method overview*

A schematic diagram of the proposed method is shown in Fig. 2. We propose sequential filtering for pair determination based on both positional filtering and image similarity filtering. This study aims to achieve change detection in videos captured by an inspection robot. A simple method to accomplish this is by directly inputting each frame of the videos into a change detection network. However, because it involves comparing image pairs that do not capture the same region, this approach reduces the correct change detection rate. Therefore, image similarity filtering based on image features is performed to select image pairs that depict the same region.

Despite this improvement, computational time remains an issue due to the need for image similarity calculations between images acquired at different locations. As images captured at distant locations do not usually capture a common area, there is no need to compute image similarity between these images. Prior to image similarity filtering, positional filtering based on self-localization of the robot is performed to reduce the similarity calculations between images captured at long distances. It is important to note that a pair determination based solely on positional filtering is insufficient due to self-localization errors. The image pairs, determined sequentially using image similarity and positional filtering, are then input into a deep learning model to detect the changes. Thus, change detection can be performed in inspection videos captured during patrol inspections using this method.

B. *Data acquisition by an inspection robot*

Patrol inspection robots operate between machinery in petroleum refineries and petrochemical plants, capturing inspection videos of the outer surface of the pipes with a camera mounted on them. The inspection videos obtained along the same path at different times are divided into the following two types:

1) Past videos:
These videos were captured during previous inspections. They do not contain any anomalies. Each video frame is referred to as a past image.

2) Target videos:
   These videos are captured during the current inspection. They may include changes in the outer surface of the pipes, and it is necessary to determine whether any changes are present. Frames are extracted from these videos at a specific interval and collected as target images.

Pair determination is performed on each of these target images. The pair determination is accomplished by comparing the similarity between a target image and past images captured in nearby locations, as described in the following sections.

C. Positional filtering considering the self-localization error of the robot

When determining image pairs, they must capture the same region. The closer the locations where the images are captured, the higher the probability that they capture the same region. Therefore, we propose a positional filtering approach based on the location where the images are captured.

We use the results of self-localization by an inspection robot to obtain the image shooting locations. It is possible to estimate the shooting position of each frame of the video by synchronizing the video and the self-localization data. In the context of plant patrolling using robots, the robot estimates its position using odometry or SLAM. However, odometry can be prone to significant errors caused by crawler slippage. Moreover, SLAM has an estimated self-position error of approximately 0.1 m due to point cloud misalignment [17]. Hence, when determining pairs of images in the neighborhood, it is necessary to consider the self-localization error. Positional filtering is accomplished by the following equation:

$$\mathcal{S}_{\mathbf{C}} = \{\mathbf{P} \in \mathcal{S}_{\mathbf{P}} \mid \|\mathbf{x_T} - \mathbf{x_P}\| \leq T\}, \tag{1}$$

where $\mathbf{P}$ represents a past image, $\mathbf{C}$ represents a candidate image, and $\mathbf{T}$ represents a target image. $\mathcal{S}_{\mathbf{C}}$ and $\mathcal{S}_{\mathbf{P}}$ refer to sets of candidate images and past images, respectively. $\mathbf{x_T} \in \mathbb{R}^{2\times 1}$ and $\mathbf{x_P} \in \mathbb{R}^{2\times 1}$ denote estimated 2D positions of target and past images, respectively. $T \in \mathbb{R}$ and represents a threshold of the distance between the target and past images. This filtering prevents the selection of pairs of images captured at clearly different locations.

D. Image similarity filtering for reference image determination

The positional filtering described in the previous section, which is based on the shooting distance between images, is insufficient for pair determination because errors can occur in estimating the positions of the images. This means that even if two images are identified as being the closest in distance, the self-localization error may prevent them from accurately capturing the same area. Therefore, it is necessary to compare the images themselves to examine whether they captured the same region.

In this study, we employ image similarity filtering between two images based on feature matching to determine whether the images captured the same region. The image similarity is determined by evaluating the number of matched feature points obtained through feature matching. The input image pair is selected as the candidate image with the highest similarity to the target image, referred to as the reference image, along with the target image. As petroleum refineries and petrochemical plants are often situated outdoors where lighting conditions vary considerably, robust feature extraction is achieved by employing the brightness-robust Accelerated-KAZE (AKAZE) feature descriptor [18]. The feature point with the smallest Hamming distance among the feature point sets of the two images is chosen for feature matching [19]. The reference image is selected as follows:

$$\mathbf{R} = \arg\max_{\mathbf{C}\in\mathcal{S}_{\mathbf{C}}} f(\mathbf{T}, \mathbf{C}), \tag{2}$$

where $\mathbf{R}$ represents a reference image. $f(\mathbf{T}, \mathbf{C})$ denotes the number of matched feature points between the target image $\mathbf{T}$ and the candidate image $\mathbf{C}$. This method determines the image pair with the highest similarity based on the number of matched feature points as the input image pair.

E. Change detection by deep learning

When detecting changes on the outer surfaces of pipes, it is essential to compare images captured in the past with those captured currently. However, images captured at different times have varying camera positions and lighting conditions. Hence, detecting significant changes without being influenced by varying conditions is crucial. To achieve this, a Siamese change detection network is introduced. It is possible to output the changed area between two images by inputting them as a pair into this network. In this way, change detection in inspection videos is achieved.

III. EXPERIMENTS

A. Conditions

To evaluate the effectiveness of the proposed method, experiments were conducted in an indoor environment that simulated a petroleum refinery and petrochemical plant, as shown in Figs. 3 and 4. PVC pipes were installed to reproduce the plant environment, and changes on the pipe surfaces were replicated by attaching black vinyl tape. Fluorescent lamps provided the primary lighting in the room, while halogen lamps were used to change the brightness of the environment.

A 5-axis manipulator for S2-P1-F4 and an experimental rescue robot system S2-P1 module manufactured by Topy Industries were used as the inspection robot. The robot moved along the pipes in a straight line for approximately 3 m at a speed of 0.014 m/s, capturing videos of the pipes. In each trial, the robot's starting position was visually set to the same location. The self-localization of the robot was performed by calculating odometry using the encoder values

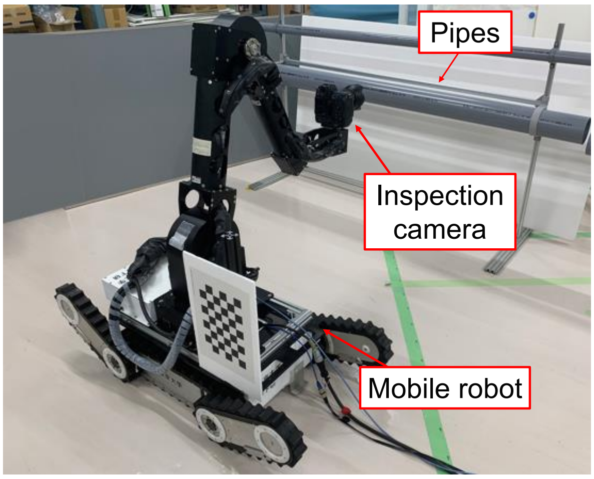


Fig. 3. Mobile robot, inspection camera, and pipes used for experiments.

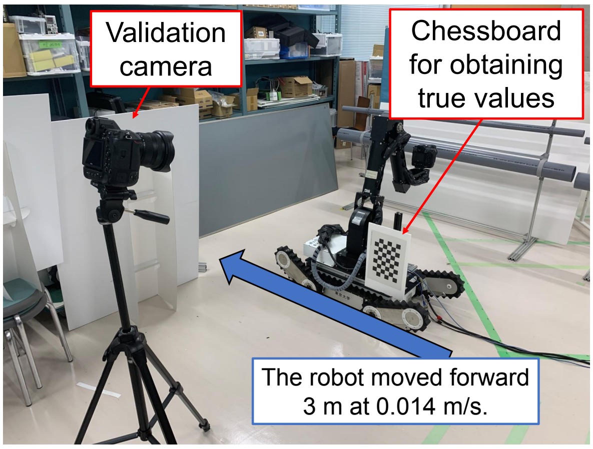

Validation camera
Chessboard for obtaining true values
The robot moved forward 3 m at 0.014 m/s.

Fig. 4. Validation camera and chessboard used to obtain the true values of the robot's movement.

of the crawler. In the proposed method, the threshold T was set to 0.30 m.

To capture videos of the pipes, the mobile robot was equipped with a Nikon Z9 camera mounted with a NIKKOR Z 24-70 mm f/4 S lens (inspection camera). The video resolution of the inspection camera was set to 3840 × 2160 pixels, frame rate to 30 fps, and focal length to 24 mm. The distance between the inspection camera and the surface of the pipe was approximately 0.6 m.

To acquire the true position of the robot, a chessboard pattern was attached to the robot, and the coordinates of the chessboard were obtained using a pre-calibrated Nikon Z9 camera equipped with a NIKKOR Z 14-24 mm f/2.8 S lens (validation camera). As the relative position between the inspection camera and the chessboard was fixed, there was no need for calibration between the inspection camera and the chessboard. The video resolution of the validation camera was set to 3840 × 2160 pixels, frame rate to 30 fps, and focal length to 24 mm.

We employed AnoDFDNet, a deep learning model proposed by Wang et al. [16], for change detection. This network is based on convolutional neural networks and vision transformers. By training this network with the features of the target changes to be detected, it can achieve change detection that is tolerant to non-target changes such as camera position and lighting conditions.

An autoencoder model [20] trained on the same past images as the proposed method was used to evaluate the performance of the proposed method. The autoencoder was based on VGG16 architecture [21]. In this approach, the threshold of the autoencoder was determined based on the reconstruction error of past images. Specifically, kernel density estimation was performed on the histogram of reconstruction errors, and the threshold value was set as the reconstruction error at which the area of the probability distribution function reached 90%. When the reconstruction error of a target image exceeded this threshold value, it was interpreted as indicating a changed image. When comparing the proposed method with the autoencoder, any image where changes were detected, even in a single pixel, was classified as an image with changes. Three evaluation metrics were used to evaluate model performance: precision, recall, and F1 score.

In addition, to evaluate the effectiveness of sequential filtering for pair determination, we compared the proposed method against the following two measures:

Similarity measure:

In this measure, image pair determination relies solely on image similarity filtering. The reference image is determined as follows:

$$\mathbf{R} = \arg\max_{\mathbf{P} \in \mathcal{S}_\mathbf{P}} f(\mathbf{T}, \mathbf{P}). \tag{3}$$

Position measure:

In this measure, image pair determination relies solely on the estimated self-position of the robot. The reference image is determined as follows:

$$\mathbf{R} = \arg\min_{\mathbf{P} \in \mathcal{S}_\mathbf{P}} \|\mathbf{x}_\mathbf{T} - \mathbf{x}_\mathbf{P}\|. \tag{4}$$

When comparing the proposed method against these two measures, the effectiveness of the proposed sequential filtering for pair determination was evaluated in two aspects: robustness against self-localization errors and computational cost. To assess robustness, pair determination and change detection were conducted by programmatically introducing constant noise to the robot's estimated self-position. In this assessment, the true shooting distance of image pairs, F1 score, and area under the curve (AUC) of the precision-recall (PR) curve were used. The PR curve was utilized due to the imbalanced nature of the data acquired in this experiment, which contained few changed areas. The execution time required for each method to determine a single image pair was measured to compare the computational cost. In these evaluations, change detection results were evaluated pixel-by-pixel.

B. Dataset generation and training

The robot captured three target videos and three past videos for approximately 200 seconds for testing purposes,

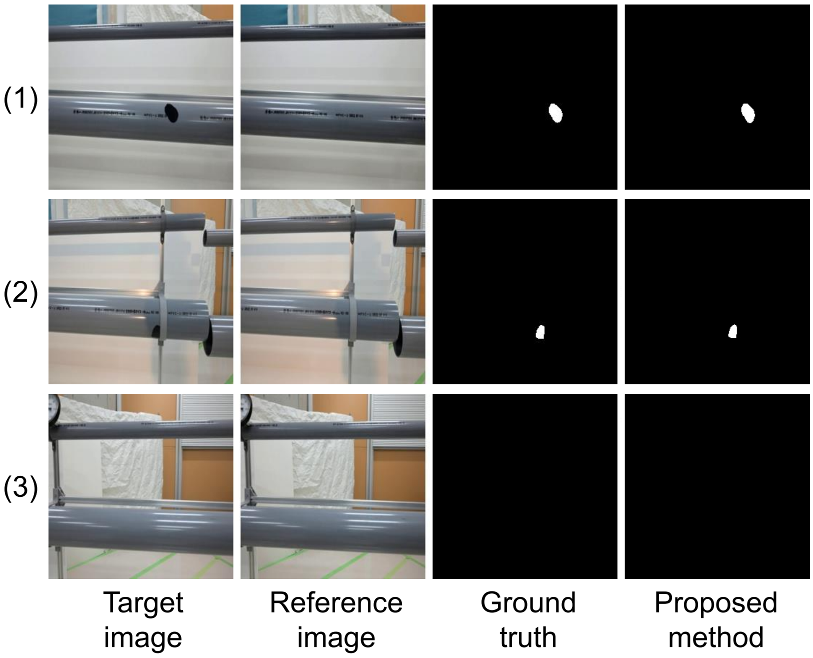


Fig. 5. Result of pair determination and change detection using the proposed method. Columns from left to right show the target image, reference image, ground truth for the changed area, and result of change detection by the proposed method, respectively. Row (1) shows an example of a relatively small change in camera position, row (2) shows an example of different lighting conditions with a halogen lamp irradiating the area, and row (3) shows an example of no change on the outer surface of the pipes.

TABLE I

EVALUATION OF THE PROPOSED METHOD AND AUTOENCODER.

	Autoencoder	Proposed method
Precision	1.00	0.994
Recall	0.260	0.983
F1 score	0.413	0.989

resulting in a total of 209 target images by extracting frames from these target videos every three seconds. Subsequently, sequential filtering was performed to determine the image pairs. As a result, these target images were paired with their corresponding reference images, generating 209 pairs of testing data.

AnoDFDNet was pre-trained using a self-made dataset captured in a similar environment. Data for training and validation were collected such that the ratio of training and validation data to testing data was about 8:2. A total of 727 pairs of data were obtained and split in an approximately 8:2 ratio for training and validation data, resulting in 559 pairs for training and 168 pairs for validation. The validation data were used to adjust the hyperparameters of the model. For the training and validation data, input image pairs were manually selected to ensure that the past images and the reference images captured the same region.

AnoDFDNet was trained using an Nvidia GeForce RTX3070 Ti GPU and Intel core i9-11900K CPU. We used the Adam optimizer [22] in batches of four, with a learning rate of 0.01, and completed 20 training epochs. All images were resized to 256× 256 pixels before being input into the network.

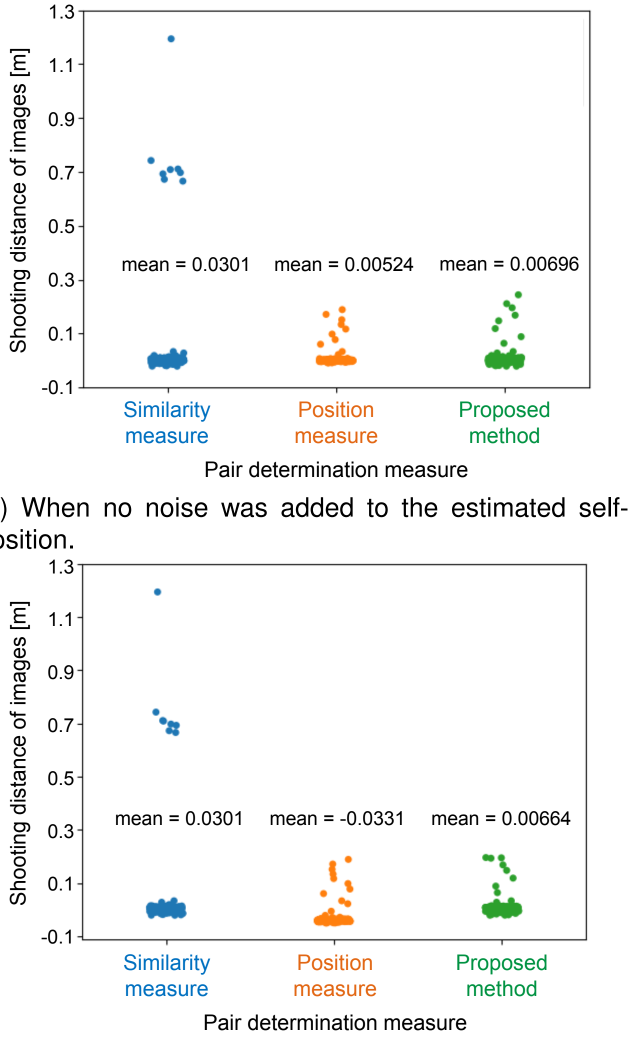
IV. RESULTS AND DISCUSSION

A. Results

The change detection results using the proposed method are shown in Fig. 5. For the example shown in the top row (row(1)), the proposed method appropriately selected the reference image for the target image and accurately detected only the changed regions. The example in the middle row (row(2)) confirmed the robustness of the proposed method to changes in the brightness of the environment. The example in the bottom row (row(3)) demonstrated the ability of the proposed method to accurately determine the absence of change when there is no change.

Table I lists the values of the three performance metrics of the proposed model and the autonecoder. The recall of the autoencoder was lower than that of the proposed method, while the precision of the autoencoder was 1.00. This implied that the autoencoder was prone to making false negative and true negative decisions. In particular, a low recall means that changes were overlooked, leading to safety risks and failures.

The true shooting distances between the matched pair of images are shown in Fig. 6. The similarity measure determined image pairs without being affected by noise because it does not use the robot's positional information. However, there were mismatches with images at distances as far apart as 0.7 m and 1.2 m. The position measure showed fewer matches with images at a distance. However, the mean value in Fig. 6a changed significantly from that in Fig. 6b, indicating that it was influenced by noise. The proposed method prevented matching at a distance by performing



(a) When no noise was added to the estimated self-position.

(b) When noise of 0.050 m was added to the estimated self-position.

Fig. 6. True shooting distance of paired images with and without noise in the estimated self-position: (a) no noise and (b) noise of 0.050 m. The horizontal axis represents each pair determination measure, and the vertical axis represents the shooting distance between matched images. The true shooting distances are calculated based on the true values of the robot acquired by the validation camera. The results of 209 samples are plotted. As the similarity measure is not affected by noise, the results of the similarity measure in these figures are the same.

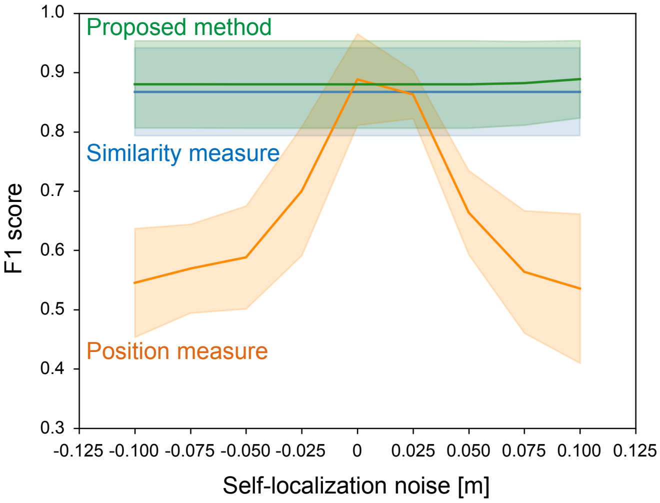


Fig. 7. Comparison of F1 score in the presence of noise in self-localization. The horizontal axis represents the noise added to the estimated self-position of the robot, and the vertical axis represents the F1 score. The red line represents the case where the proposed method is performed to determine pairs, the blue line represents the case where pairs are determined by the similarity measure, and the green line represents the case where pairs are determined by the position measure. The average over 3 target videos are plotted.

positional filtering and achieved robust matching against noise through image similarity filtering.

Figs. 7 and 8 show the evaluation results when noise was added to the estimated self-position. In Fig. 7, F1 scores are plotted. When no noise was added to the estimated self-position, the F1 scores of the proposed method, similarity measure, and position measure were 0.880, 0.861, and 0.892, respectively. This confirmed that the proposed method was more robust to errors in robot self-localization than the position measure and could detect changes with an F1 score of 0.880. Moreover, the proposed method demonstrated a higher F1 score in change detection compared to the similarity measure. This can be attributed to the proposed method's ability to suppress false matches with images captured at distant locations.

In Fig. 8, PR curves are shown when a noise of 0.050 m was added to the self-localization. The proposed method had an AUC higher than that of the position measure and comparable to that of the similarity measure.

In terms of computational time, the time required to determine a pair is shown in Table II. The position measure was the fastest in determining pairs, and the proposed method was the second fastest and could perform pair determination in approximately one-fifth of the time required by the similarity measure.

B. *Discussion*

The proposed method outperformed the autoencoder. This superior performance of the proposed method could be attributed to the inadequacy of the data obtained from the inspection videos captured by the mobile robot for effectively training the autoencoder. As a result, the proposed method proved to be better suited for inspection tasks utilizing a mobile robot than the autoencoder.

The proposed method demonstrated superior change detection capability compared to the position and similarity measures. Although the position measure had the lowest computational cost, its change detection performance was significantly degraded if errors occurred in self-localization. The similarity measure was not affected by self-localization errors and demonstrated high change detection capability. However, the impact of mismatching and high computational cost indicated that operating the similarity measure in a large plant could be difficult. The proposed method achieved high reliability in change detection while addressing self-localization errors, preventing mismatches, and reducing computational costs. In other words, positional filtering limits

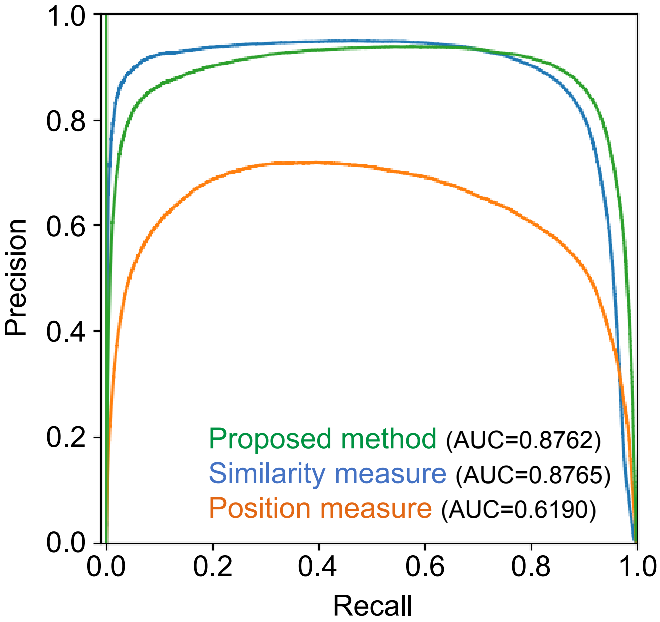


Fig. 8. Precision-recall (PR) curves when the self-location noise is 0.050 m, where AUC denotes the area under the curve. The horizontal axis represents the recall and the vertical axis represents the precision. The green line represents the case where the proposed method is performed to determine pairs, the blue line represents the case where pairs are determined by the similarity measure, and the orange line represents the case where pairs are determined by the position measure.

the number of images to be processed, regardless of the inspection video length. Image similarity filtering identifies pairs that capture the same region.

While applying the proposed method to real plant environments, two issues need to be addressed. First, inspection robots are expected to perform various motions, such as translations, rotations, and manipulator movements in large and complex plants, whereas in this experiment, the robot moved in a straight line with its manipulator fixed. Therefore, higher degrees of freedom of camera motion need to be implemented for the inspection robot in real plants. Second, a broad range of anomaly training data is required to detect anomalies that occur in real pipes. Collecting such diverse training data is time-consuming and beyond the scope of this study. Efficient data collection will be a future work.

The reliability of the proposed method depends on the accuracy of self-localization. As shown in Fig. 7, when self-localization was highly accurate, the position measure exhibited a higher F1 score than the proposed method. In addition, the accuracy of self-localization affects how threshold T is determined, and T is related to the performance of the proposed method. As T is determined heuristically in this paper, self-localization performance should be carefully checked during real-world operations.

In summary, the similarity measure showed robustness to self-localization errors, and the position measure had the lowest computational cost. Although the proposed method depended on the performance of self-localization, it exhibited both robustness and low computational cost, rendering it the most suitable model for determining pairs and detecting changes among the three measures.

TABLE II

TIME TO DETERMINE A REFERENCE IMAGE FOR ONE TARGET IMAGE FOR EACH METHOD. THE MEANS AND SDS OVER 10 SAMPLES OF THE TIME EACH METHOD TOOK TO DETERMINE THE REFERENCE IMAGE ARE DISPLAYED.

	mean [s]	SD [s]
Similarity measure	138.494	0.424
Position measure	0.344	0.073
Proposed method	28.231	0.566

V. CONCLUSION

In this study, we proposed a method for detecting changes on the outer surface of pipes using inspection videos captured by an inspection robot. To achieve this objective, we introduced sequential filtering for determining image pairs that utilizes positional filtering and image similarity filtering. According to the experimental results, the proposed method achieved a higher F1 score than that of the autoencoder. The comparison of the proposed method with the two measures revealed that the similarity measure demonstrated robustness to self-localization errors, whereas the position measure required the least computational time. In contrast, the proposed method achieved both robustness against self-localization errors and time efficiency in determining image pairs.

Future works include conducting outdoor experiments in actual petroleum refineries and petrochemical plants. It is necessary to carefully consider the self-localization error that can occur in a real plant and the computational time for real-world operations. In addition, addressing the wide range of changes that can occur in a plant environment is essential. This can be accomplished by efficient data collection, such as data generation using a 3D model or by employing a pre-trained model. Pair determination methods that correspond more effectively to various camera motions should be considered. It is also important to consider other pair determination methods, such as deep learning, to overcome the dependence on the self-localization performance.

REFERENCES

[1] S. Jafarinejad and S. C. Jiang, "Current technologies and future directions for treating petroleum refineries and petrochemical plants (prpp) wastewaters," *Journal of Environmental Chemical Engineering*, vol. 7, no. 5, p. 103326, 2019.

[2] C. Spandonidis, P. Theodoropoulos, and F. Giannopoulos, "A combined semi-supervised deep learning method for oil leak detection in pipelines using iiot at the edge," *Sensors*, vol. 22, no. 11, p. 4105, 2022.

[3] P. Lin, X. Li, Z. Long, P. Ma, G. Zhu, J. Wei, and Z. Fan, "Pipeline leak detection, location and repair," in *Proceedings of the 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2021, pp. 102–108.

[4] 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, p. 1588, 2022.

[5] 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.

[6] 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.

[7] K. Shukutani, K. Onishi, N. Onishi, H. Okazaki, H. Kojima, and S. Kobori, “Development of explosion-proof autonomous plant operation robot for petrochemical plants,” *Mitsubishi Heavy Industries Technical Review*, vol. 55, no. 4, pp. 1–6, 2018.

[8] M. S. Saeed and N. Alim, “Design and implementation of a dual mode autonomous gas leakage detecting robot,” in *Proceedings of the 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2019, pp. 79–84.

[9] P. Daogang, G. Ming, W. Danhao, and H. Jie, “Anomaly identification of critical power plant facilities based on yolox-cbam,” in *Proceedings of the 2022 Power System and Green Energy Conference (PSGEC)*, 2022, pp. 649–653.

[10] L. Yu, E. Yang, P. Ren, C. Luo, G. Dobie, D. Gu, and X. Yan, “Inspection robots in oil and gas industry: a review of current solutions and future trends,” in *Proceedings of the 25th International Conference on Automation and Computing (ICAC)*, 2019, pp. 1–6.

[11] Z. Wang, J. Peng, W. Song, X. Gao, Y. Zhang, X. Zhang, L. Xiao, and L. Ma, “A convolutional neural network-based classification and decision-making model for visible defect identification of high-speed train images,” *Journal of Sensors*, vol. 2021, pp. 1–17, 2021.

[12] A. R. Khairuddin, M. S. Talib, and H. Haron, “Review on simultaneous localization and mapping (slam),” in *Proceedings of the 2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2015, pp. 85–90.

[13] 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.

[14] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, “Image change detection algorithms: a systematic survey,” *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, 2005.

[15] D.-M. Tsai and P.-H. Jen, “Autoencoder-based anomaly detection for surface defect inspection,” *Advanced Engineering Informatics*, vol. 48, p. 101272, 2021.

[16] Z. Wang, Y. Zhang, L. Luo, and N. Wang, “Anodfdnet: A deep feature difference network for anomaly detection,” *Journal of Sensors*, vol. 2022, pp. 1–14, 2022.

[17] K. Kitajima, Y. Kinouchi, and K. Sugimoto, “Robust and accurate self-localization method under varying lighting conditions,” *Mitsubishi Heavy Industries Technical Review*, vol. 58, no. 1, p. 1, 2021.

[18] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, “Fast explicit diffusion for accelerated features in nonlinear scale spaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2011.

[19] S. A. K. Tareen and Z. Saleem, “A comparative analysis of sift, surf, kaze, akaze, orb, and brisk,” in *Proceedings of the 2018 International conference on computing, mathematics and engineering technologies (iCoMET)*, 2018, pp. 1–10.

[20] Horizon2333, “imagenet-autoencoder,” 2022. [Online]. Available: https://github.com/Horizon2333/imagenet-autoencoder

[21] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[22] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.