Hindawi

Research Article

Visual Localization of an Internal Inspection Robot for the Oil-Immersed Transformer

Yingbin Feng (iD), **Yahui Kou** (iD)**, and Yanju Liu** (iD)

School of Automation and Electrical, Shenyang Ligong University, Shenyang, Liaoning 110159, China

Correspondence should be addressed to Yahui Kou; 314725071@qq.com

Aiming at the problem that the robot is difficult to locate in the oil-immersed transformer, a visual positioning of the robot is proposed for internal inspection. First, in order to solve the problem of blur, distortion, and low contrast of the image obtained by the camera in the deteriorated and discolored transformer oil, an image enhancement algorithm based on multiscale fusion is developed to provide a reliable data source for robot localization. Then, the FAST key points are extracted and the BRIEF descriptors are calculated from the enhanced images, and the pose transformation of the robot between image frames is calculated by using polar constraint and EPnP method. A pose optimization model of the robot is designed to improve the positioning accuracy. Finally, to verify the effectiveness of the proposed methods, function tests are carried out by using the real continuous image sequence acquired by the robot in Mitsubishi transformer. The experimental results show that the trajectory of the robot in the transformer can be accurately drawn, the position data of the robot can be efficiently obtained, and autonomous positioning of the robot in the transformer can be well achieved.

1. Introduction

Oil-immersed transformers are key equipment of the national power grid and have the advantages of excellent performance and low price. Because transformer oil has the characteristics of low viscosity, good insulation, and good heat transfer performance, it can protect transformer cores and windings well. However, with increasing transformer use time, failure will occur more frequently [1]. To solve the problems of low efficiency and high cost of manually inspecting internal faults of transformers, scholars have designed internal fault detection robots for oil-immersed transformers, which are based on vision [2–4]. The robots make use of their vision system to obtain the internal image of the transformer for judging the fault type. However, the internal structure of transformers is complicated, and the location information of the faults cannot be accurately obtained only through the image. In addition, these robots do not have autonomous localization ability and cannot realize autonomous operation. Therefore, transformer fault detection can only be achieved by manual manipulation, which is inefficient.

The oil-immersed transformer has a closed shell, compact interior structure, and full transformer oil. The complex transformer structure brings difficulties to robot positioning. The research on submersible transformer inspection robots is still limited, and there are few research results on the positioning method of robots under transformer oil. Considering that both transformer oil and water belong to the fluid medium and that the motion characteristics of robots in fluid are similar, this paper mainly refers to the positioning methods of underwater vehicles [5, 6].

According to the different measuring principles of sensors, underwater vehicle positioning technology is divided into underwater acoustic positioning methods, dead reckoning methods, and vision-based positioning methods [7–12]. In underwater acoustic localization, the position of the underwater vehicle can be calculated by the transmission time and phase difference of the acoustic wave between the underwater robot and acoustic beacon. It can be divided into long baseline, short baseline, and ultrashort baseline based on the length of the baseline. In the dead reckoning method, the position of the underwater vehicle can be obtained by

integrating the time according to the pose and speed of the underwater vehicle measured by the inertial navigation and Doppler velocity log (DVL), but accumulated errors exist in this method, which must be eliminated by the underwater acoustic positioning method. An underwater location method based on a particle filter was proposed by Martínez-Barberá et al. [13]. This method processes and fuzes the data of several underwater sensors, which improves the positioning accuracy of underwater robots. An ultrashort baseline positioning method based on a Kalman filter was proposed by Luo et al. [14], which effectively improved the positioning accuracy.

The underwater acoustic positioning method needs to arrange the acoustic beacon array in advance. However, the structure of the oil-immersed transformer is closed, which does not have the conditions to arrange the array. There are many obstacles in the transformer, and acoustic communication cannot be realized, so underwater acoustic positioning cannot be applied in oil-immersed transformers. In addition, the dead reckoning method requires high-precision inertial navigation and DVL. Considering the cost and volume of the transformer internal inspection robot, the sensors mentioned above are not suitable for a transformer internal inspection robot.

In underwater location methods based on vision, underwater vehicles can be located with image information [15–18], which is obtained by inexpensive vision sensors. An underwater vision positioning technology based on refraction correction was provided by Suresh et al. [19], which uses refraction correction and triangulation to locate underwater vision by observing aerial landmarks in the water. An underwater localization method based on binocular vision was proposed by Du et al. [20]. The method takes advantage of the combination of feature extraction and back-end global optimization to verify the feasibility of binocular vision positioning in underwater environments. An improved front-end method of underwater visual SLAM based on binocular cameras was given by Wang et al. [21]. The histogram equalization algorithm was used to improve image sharpness for feature point extraction and tracking, and the Rtab-map algorithm was used to optimize its tracking and positioning process. The feasibility of achieving accurate positioning by the visual sensor in a small underwater area has been verified by the above methods. However, most underwater positioning methods based on vision use binocular vision sensors. Considering the size of the robot and the cost of fault detection, binocular cameras are not suitable for submersible transformer inspection robots. In addition, the environment in the transformer oil is different from the underwater environment, and the transformer oil has the problems of deterioration and discoloration when being used high pressure and high-temperature environments for a long time. Therefore, the internal image of the transformer obtained by the robot has the problem of low contrast and color distortion, which brings new challenges to the visual positioning method of the submersible transformer inspection robot.

In recent years, scholars have conducted relevant research on the positioning of inspection robots inside oil-immersed transformers. A high-precision detection model for transformer

components based on the Fast R-CNN structure was provided by Liu et al. [22]. This model employs the dual feature mapping method, enabling automatic detection of the category and location of various transformer components in the detection image. However, this position information is only relative between components and cannot provide global positioning within the transformer. In addition, the R-CNN model-based approach demands significant computational power for transformer inspection robots, making it unsuitable for real-time detection and localization. Zhang et al. [23] implemented robot autonomous navigation through path planning based on the open-source Cartographer algorithm. Oil levels were acquired using a depth camera, and YOLOv4 was employed to train and learn oil level states under various weather conditions. This method is suitable for single inspection tasks such as oil level detection but is not applicable to fault detection and localization tasks that require more visual information. Moreover, significant measurement errors are introduced in the measurement of infrared light emitted by the depth camera in transformer oil. In the face of various complex environmental stresses, Pan et al. [24] conducted an analysis of the performance failure mechanisms of internal inspection robots. They completed a study on improving the reliability of internal inspection robots and their control systems, effectively enhancing the reliability of the internal inspection robotic arm. However, no analysis was performed on the positioning of internal inspection robots within transformers. A positioning method for an oil-immersed transformer internal inspection robot was proposed by Feng et al. [25]. In this method, the robot's position within the transformer is measured and analyzed using a detection instrument and laser radar carried by the inspection robot. However, high-precision requirements are imposed on the accuracy of laser radar measurements in this method, and there is a lack of abundant visual information.

In summary, the visual-based positioning method for the transformer internal inspection robot proposed in this paper does not require expensive high-precision radar. After image enhancement, it is capable of providing abundant visual information. Achieving visual positioning and utilizing images for fault detection simultaneously, waste of robot computational resources is avoided, and real-time positioning of the internal inspection robot is achieved.

The specific implementation process of the algorithm in this paper is as follows: first, the image enhancement method based on multiscale fusion is used to enhance the image acquired by a monocular camera to improve the brightness and contrast of the image in the transformer oil. Second, the enhanced image is converted into grayscale to improve the efficiency of extracting features from accelerated segment test (FAST) key points to ensure that each image can be tracked quickly. Then, the position and pose transformations of different camera positions were solved by the epipolar geometry constraint and efficient perspective-n-point (EPnP) algorithm. Finally, the nonlinear optimization method is applied to optimize the position and attitude of the robot to obtain a more accurate trajectory. The experimental results show that the autonomous positioning of the submersible transformer

FIGURE 1: The structure of the oil-immersed transformer.

FIGURE 2: The transformer internal inspection robot.

inspection robot can be realized, and the problem of robot position initialization failure and tracking loss caused by the lack of image information can be solved.

2. Framework of the Visual Localization for the Oil-Immersed Transformer

2.1. Oil-Immersed Transformer Structure. The structure of the oil-immersed transformer is shown in Figure 1. The structure of the transformer (510 cm × 230 cm × 350 cm) is mainly composed of an oil tank, core, winding, bushing, cooler, manhole, and conservator. The core and winding are located in the center of the transformer. There are a large number of locking nuts, screws, and cables in the transformer, which may interfere with robot motion. The tank is filled with 25# transformer oil, which is mainly used for insulation, heat dissipation, and arc extinction. The internal structure of the transformer is compact, and the space is narrow, so it is difficult to realize robot positioning by arranging external sensors.

According to the structural characteristics of the transformer, our research group has designed a submersible transformer inspection robot, as shown in Figure 2. The diameter of the robot is only 19 cm, and the propulsion system, wireless communication system, control system, sensing and detection system, and power supply system are integrated inside. The robot is propelled by the propulsion system, which has six oil-jet thrusters. The sensor detection system mainly includes single-line laser radar, monocular camera, electronic compasses, and other sensors. The laser radar can only detect the distance between the robot and the obstacle, and it is influenced by the transformer oil and Lidar hull, so the measurement error is large. The electronic compass is affected by the oil-jet thruster and the closed structure of the transformer, so the measurement data are not accurate. Therefore, the robot can only rely on the images obtained by monocular vision to realize autonomous positioning.

2.2. The Proposed Framework. According to the above analysis, constrained by the internal structure of the transformer and the size of the robot, this paper uses a monocular camera to achieve robot positioning. Compared with the depth camera, which can obtain the distance information directly, the monocular camera can only estimate the depth by feature matching from the adjacent images obtained by the camera position change. For this reason, image feature extraction is very important in the positioning process. However, with
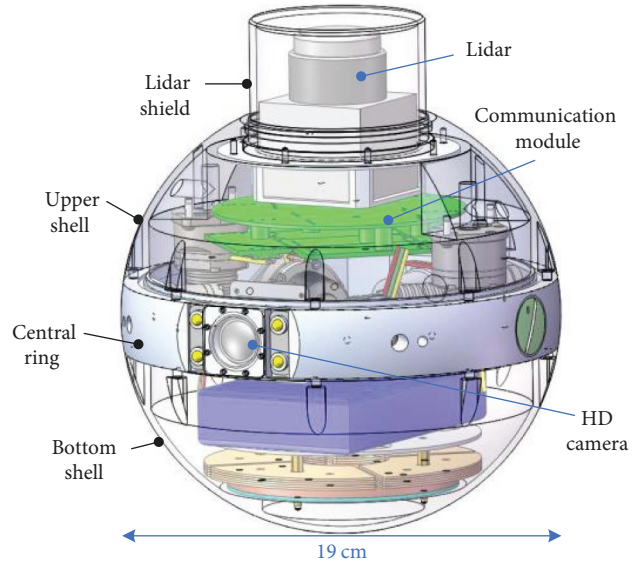
increasing transformer working time, the transformer oil color gradually changes from pure transparent and light yellow to light brown. The image obtained by the camera in the degraded and discolored oil exhibits distortion, discoloration and low contrast, which increases the difficulty of image feature point extraction and registration. Therefore, a visual localization method of the submersible transformer inspection robot based on image enhancement is proposed. The flowchart of this method is shown in Figure 3 and is elaborated as follows:

(1) The image enhancement method based on multiscale fusion is applied to process the original image obtained by the robot to improve the brightness and contrast of the image to increase the number and speed of image feature point extraction.

(2) To improve the extraction efficiency of FAST key points, the enhanced image is converted to a grayscale image.

(3) The binary robust independent elementary features (BRIEF) descriptors of the feature points are calculated to compare the similarity of the features of the two frames for feature matching.

(4) The epipolar geometry constraint method is used to initialize the motion process of the robot, and the EPnP algorithm is applied to solve the pose transformation of the robot between image frames.

(5) To solve the pose between image frames, the tracking process of the robot is optimized by a nonlinear optimization method to realize more accurate tracking and positioning.

3. The Proposed Algorithms

3.1. Image Enhancement in Transformer Oil. Because the transmittance of artificial light in transformer oil of different
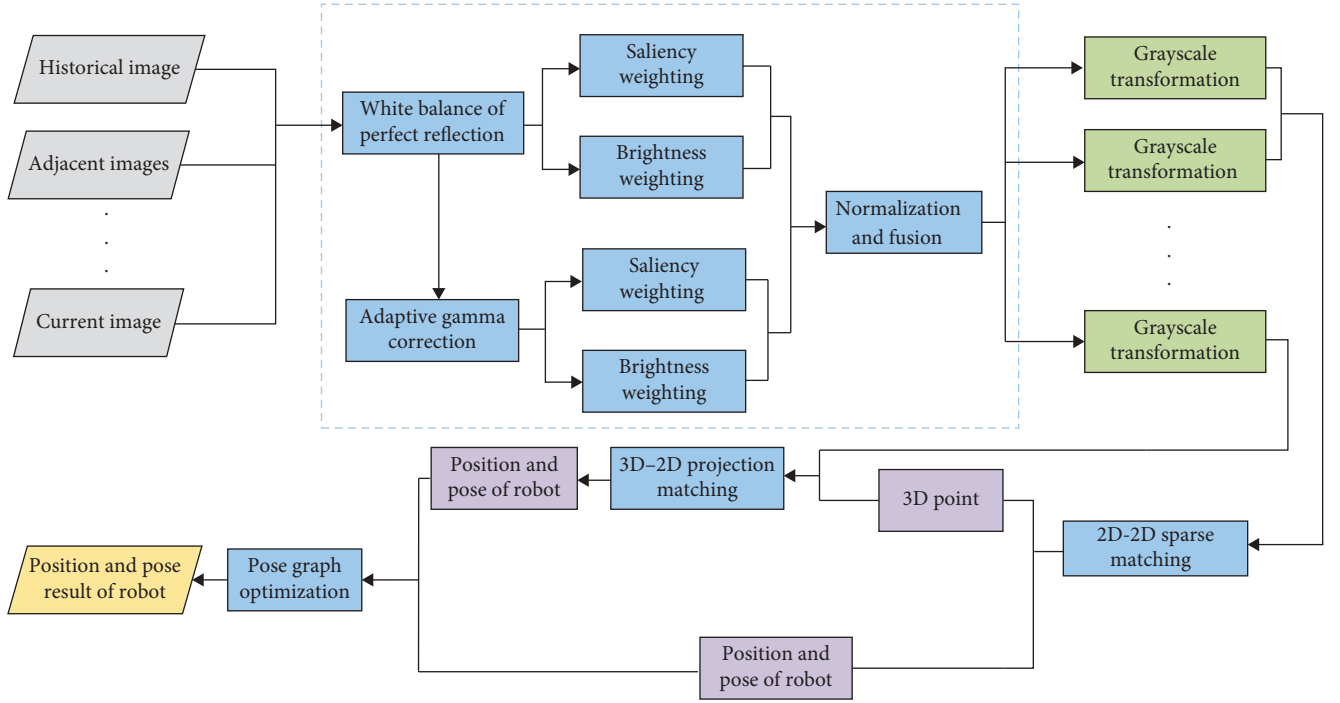
FIGURE 3: The proposed framework the visual localization for the oil-immersed transformer.

years is different, the image quality under transformer oil cannot be improved by restoring the optical imaging model under transformer oil. Consequently, the image enhancement method under transformer oil based on a multiscale fusion strategy is adopted.

To solve the problem of color distortion in the transformer oil image [26], a color correction method based on a perfect reflection algorithm is adopted to correct the image color channel to balance the image color. The correction formula can be expressed as follows:

$$
\begin{cases}
K = \mathrm{Max}(R_{\max}, G_{\max}, B_{\max}) \\
R' = \dfrac{K}{\overline{R}} \times R \\
G' = \dfrac{K}{\overline{G}} \times G \\
B' = \dfrac{K}{\overline{B}} \times B
\end{cases}, \tag{1}
$$

where R, G, and B indicate the values of each channel of the original picture pixel. $R_{\max}, G_{\max}$, and $B_{\max}$ are the maximum values of each of the three channels R, G, and B in the image, respectively. The histogram of the sum of three channels for each pixel R, G, and B in the image is traversed backward, and the pixel with values in the top 10% is found as the white point to find the threshold T of the white point. Where $\overline{R}, \overline{G}$, and $\overline{B}$ indicate the average values that the sums of pixel points R, G, and B are greater than T. R', G', and B' are the correction results of each channel value of the pixel in the picture.

To solve the problem of low-image contrast caused by different attenuation speeds of different colors of light in transformer oil, the adaptive gamma correction method is devoted to improving the image contrast. The gamma correction can be expressed as follows:

$$
G(I) = I^{\gamma}, \tag{2}
$$

where γ indicates the adjustment parameter of gamma correction. I is the input image. To avoid the same intensity change after correction, the weighted modified probability density and distribution statistics are constructed. The adaptive gamma correction method is obtained as follows:

$$
\begin{cases}
G(s) = S_{\max} \times \left(\dfrac{s}{S_{\max}}\right)^{\gamma} \\
\gamma = 1 - \sum_{s=0}^{S_{\max}} \dfrac{G_{\mathrm{pdm}}(s)}{\sum_{s=0}^{S_{\max}} G_{\mathrm{pdm}}(s)}
\end{cases}, \tag{3}
$$

where $G(s)$ is the constructed adaptive gamma calibration, s indicates the image intensity, and $S_{\max}$ is the maximum intensity.

To further improve the image quality, the color-corrected images and contrast-improved images are normalized by saliency weighting and brightness weighting. The process can be expressed as follows:

$$
\overline{W}_m(x, y) = \frac{\|p_\mu - p_\varepsilon(x, y)\|}{\sum_m \|p_\mu - p_\varepsilon(x, y)\|} = \frac{w_m(x, y)}{\sum_m w_m(x, y)}, \tag{4}
$$

$$W_m(x,y) = \left\| p_\mu - p_\varepsilon(x,y) \right\| + \exp\left(-\frac{\left(p_g^m(x,y) - 0.5 \right)^2}{2\sigma^2} \right)$$
$$= W_s(x,y) + W_E(x,y), \tag{5}$$

where p_μ indicates the eigenvector of the image and $p_\varepsilon(x,y)$ is the corresponding pixel vector value of the original image after Gaussian blur processing. $p_g^m(x,y)$ describes the pixel position of the input image p^m to grayscale image p_g^m, and σ is the standard deviation. $\overline{W}_m(x,y)$ is the required normalized weight and $W_m(x,y)$ is the sum of saliency weighting and brightness weighting.

Finally, the images after the white balance of perfect reflection correction and adaptive gamma correction are processed by the multiscale fusion method. The fusion formula is represented as follows:

$$F_k(x,y) = \sum_m L_k[p_m(x,y)]G_k\left[\overline{W}_m(x,y)\right], \tag{6}$$

where $F_k(x,y)$ indicates the final output image pyramid, k is the number of pyramid layers, m is the number of images to be fuzed, $L_k[p_m(x,y)]$ indicates that the input image is decomposed by the Laplacian pyramid, and $G_k[\overline{W}_m(x,y)]$ indicates that the standardized weight image is decomposed by the Gaussian pyramid. The fuzed pyramid is upsampled to enhance the transformer oil image.

3.2. Feature Extraction and Matching. The submersible transformer inspection robot has a small size and limited computing power. To reduce the computation time and improve the efficiency of key point detection, the FAST key points and BRIEF descriptors are selected as the feature points in the feature extraction of the transformer oil image [27]. The processes of extracting FAST key points are as follows: a certain pixel point p is selected in the image, its brightness is set to I_p, the threshold is set to $T = 0.2I_p$, and 16 pixel points on a circle with point p as the center and three pixels as the radius are selected. If there are 12 consecutive points in these 16 pixels and their brightness is in the range of $I_p \pm T$, then the point p is a key point. In this method, only the brightness difference between pixels is compared, which is efficient.

The moving direction of the submersible transformer inspection robot is variable, and the FAST key points do not have direction information, so the gray centroid method is used to describe the rotation direction of the features. The moment of the area image is defined as follows:

$$m_{\text{pq}} = \sum_{x,y} x^p y^q I(x,y), p, q = \{0, 1\}, \tag{7}$$

where $I(x,y)$ indicates the gray value at pixel (x,y), and the centroid of the image block is represented as follows:

$$C = (c_x, c_y) = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right). \tag{8}$$

The principal direction of the key point can be expressed as the direction vector $\overrightarrow{OC}$, which is from the centroid O of the circular image to the centroid C. Then, the direction of the feature point is defined as follows:

$$\theta = \arctan(m_{01}/m_{10}). \tag{9}$$

The intensity centroid method makes the FAST key points have rotation information, and the accuracy of feature matching is improved when the robot rotates in the transformer.

In addition, the position of the robot changes every moment in the running process, which causes the size of the internal components of the transformer photographed by the monocular camera to be different in the picture. The circle with a radius of 3 is selected for the FAST key point, so there is a scale problem. Therefore, the paper constructs the image pyramid and detects the key points at each layer of the pyramid to solve the above problems [28]. In this work, the number of layers of the image pyramid is 8, and the scaling factor is 1.2.

Key points are the positions of feature points in the image, and descriptors describe the information of pixels around key points. The similarity of key points descriptors has been compared to achieve feature matching between images. Considering the computing power and real-time positioning of the submersible transformer inspection robot, the binary BRIEF descriptor with convenient storage and fast calculation is selected. Its description vector is composed of 0 and 1, and only the number of different digits in the binary string needs to be compared in the process of feature point matching, which is suitable for real-time matching of feature points in transformer oil images. We choose to utilize a 256-bit binary descriptor in this paper. When matching feature points between different images, the similarity between two BRIEF descriptors is assessed by calculating their Hamming distance. A smaller Hamming distance indicates a higher degree of similarity between the two descriptors, thus implying a higher level of matching. To determine whether two descriptors match, a threshold is typically set. If the Hamming distance between two descriptors is less than the threshold, they are considered to be a match. The steps to set the threshold are as follows: first, the minimum Hamming distance is calculated. Second, 30 is set as the lower limit for the distance. Third, if the distance between descriptors is greater than 30 and less than twice the minimum distance, then the two feature points are considered to be a match.

3.3. Robot Pose Estimation. The images captured by the robot's monocular camera have no depth information, so it is necessary to estimate the pose and depth of the robot at the beginning of its movement. In this paper, the epipolar geometry constraint method is used to estimate at the same time, and the method with smaller error is selected as the result of motion estimation [29]. After obtaining some 3D spatial
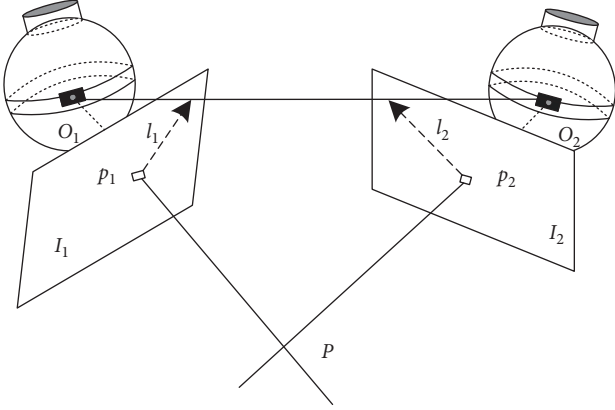
FIGURE 4: Principle of Epipolar Constraint.

points and their projections on 2D images, the relative motion of the camera can be estimated based on the coordinates of n known 3D points and the coordinates of these 3D points projected onto the image as 2D points in the pixel coordinate system. Various solution methods can be employed for such problems, including direct linear transform (DLT), perspective-three-point (P3P), efficient perspective-n-point (EPnP) [30], and others. The issues that need to be considered include the complexity of capturing images inside oil-immersed transformers, significant noise levels, and limitations in the computational power of the internal inspection robot. At least six pairs of points are required for the DLT method, and it is highly sensitive to noise and errors. The P3P method requires four pairs of points, with three pairs used for solving and one pair used for algorithm validation. However, P3P struggles to utilize information from additional matching point pairs, and it becomes ineffective in the presence of noise or mismatches. The EPnP algorithm transforms through control points, exhibiting better performance with noisy feature points. The solution obtained is a closed-form solution, eliminating the need for iteration and initial estimates, resulting in exceptionally high-computational efficiency. Therefore, the EPnP algorithm is chosen in this paper for camera pose estimation. It consists of the following steps:

(1) The motion model between two images acquired by a monocular camera is shown in Figure 4, where O_1 and O_2 are camera optical centers and l_1 and l_2 are polar lines. If the feature point p_1 and the feature point p_2 are correctly matched, the position of the spatial point P can be calculated. According to the pinhole camera model and the epipolar geometry, the following equation is true:

$$\mathbf{p}_2^T\mathbf{K}^{(-T)}\mathbf{t}^\Lambda\mathbf{R}\mathbf{K}^{(-1)}\mathbf{p}_1 = 0, \quad (10)$$

where $\mathbf{K}$ indicates the internal reference matrix of the camera, $\mathbf{R}$ is the camera rotation matrix between two coordinate systems, and $\mathbf{t}$ is the translation vector of the camera.

Suppose the fundamental matrix from pixel $\mathbf{p}_2$ to $\mathbf{p}_1$ is $\mathbf{F}$. Equation (10) can be written as follows:

$$\mathbf{p}_2^T\mathbf{F}\mathbf{p}_1 = 0. \quad (11)$$

Equation (11) is written in matrix form, which can be expressed as follows:

$$\begin{bmatrix} u_2 & v_2 & 1 \end{bmatrix}\begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0, \quad (12)$$

where the pixel point $\mathbf{p}$ is represented by homogeneous coordinates. Considering the scale equivalence, Equation (12) shows that the $\mathbf{F}$ matrix can be solved by constructing 8 constraint equations through 8 pairs of matching points.

(1) If the robot only moves in translation, it can be described by homography matrix $\mathbf{H}$. The matched pixels on the two images I_1 and I_2 can be represented by the following equation:

$$\mathbf{p}_2 = \mathbf{H}\mathbf{p}_1. \quad (13)$$

Equation (13) is written in matrix form as a homogeneous equation,

$$\begin{bmatrix} 0 & -1 & v_2 \\ 1 & 0 & -u_2 \\ -v_2 & u_2 & 0 \end{bmatrix}\begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0. \quad (14)$$

Equation (14) shows that the H-matrix can be obtained by constructing 8 constraint equations through 4 pairs of matching points. The rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ can be obtained by singular value decomposition of the homography matrix $\mathbf{H}$ [31].

(3) The depth information of the space point P is estimated by the triangulation method, and the depth estimation model is shown in Figure 5. The geometric relation of the position of feature points in three-dimensional space can be expressed as

$$s_2\mathbf{x}_2 = s_1\mathbf{R}\mathbf{x}_1 + \mathbf{t}, \quad (15)$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ indicate the normalized coordinates of the feature points after successful matching in the image and s_1 and s_2 are the depths of the two feature points.

(4) To reduce the computation load and improve the pose accuracy of the robot, the EPnP algorithm is
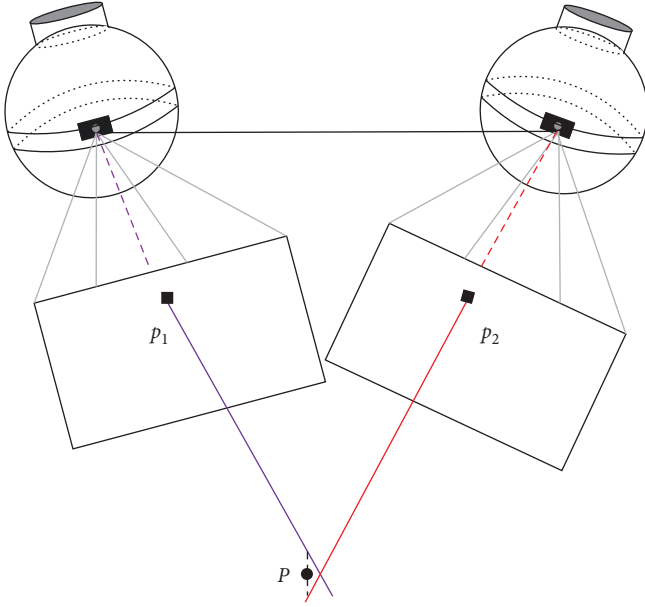
FIGURE 5: Depth estimation.

$$\forall i, \mathbf{p}_i^w = \mathbf{R}\mathbf{p}_i^c + \mathbf{t}. \tag{18}$$

To solve matrix $\mathbf{R}$ and vector $\mathbf{t}$, the error Equation (19) can be defined as follows:

$$\min_{\mathbf{R},\mathbf{t}} \mathbf{J} = \frac{1}{2} \sum_{i=1}^{n} \|(\mathbf{p}_i^w - (\mathbf{R}\mathbf{p}_i^c + \mathbf{t}))\|_2^2. \tag{19}$$

The centroids of the 3D point $\mathbf{p}_i^w$ in the world coordinate system and the 3D point $\mathbf{p}_i^c$ in the camera coordinate system are expressed as follows:

$$\mathbf{p}^w = \frac{1}{n} \sum_{i=1}^{n} \mathbf{p}_i^w, \tag{20}$$

$$\mathbf{p}^c = \frac{1}{n} \sum_{i=1}^{n} \mathbf{p}_i^c. \tag{21}$$

The error equation can be expressed as follows:

$$\min_{\mathbf{R},\mathbf{t}} \mathbf{J} = \frac{1}{2} \sum_{i=1}^{n} \|(\mathbf{p}_i^w - \mathbf{p}^w - (\mathbf{R}\mathbf{p}_i^c - \mathbf{p}^c))\|^2 + \|\mathbf{p}^w - \mathbf{R}\mathbf{p}^c - \mathbf{t}\|^2. \tag{22}$$

The rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$ of the robot can be obtained by the least square method.

3.4. Robot Pose Optimization. The relative camera poses $\Delta \mathbf{T}_{k,k-1}$ of two consecutive pictures can be calculated by the above steps. Due to the existence of noise, accumulated errors will inevitably occur with increasing running time. In addition, the motion of the robot is nonlinear. To obtain a relatively accurate pose of the robot, it is necessary to optimize the pose of the robot.

Considering the computing power of the robot, on the premise that the computing performance and the optimization accuracy are balanced, a bundle adjustment (BA) optimization method with a small computing scale is used in this paper [32]. On the premise of having no error, the camera pose calculated by projection and the camera pose observed by feature matching should satisfy the Lie group transformation,

$$\mathbf{T}_{i,j} = \mathbf{T}_i^{-1} \mathbf{T}_j, \tag{23}$$

where $\mathbf{T}_{i,j}$ indicates the calculated pose, and $\mathbf{T}_i^{-1} \mathbf{T}_j$ is the pose obtained by sparse matching. To optimize the pose of the robot, the optimization objective function can be constructed based on Equation (23). The optimization objective equation can be written as follows:

$$e_{ij} = \ln\left(\mathbf{T}_{ij}^{-1} \mathbf{T}_i^{-1} \mathbf{T}_j\right)^\vee. \tag{24}$$

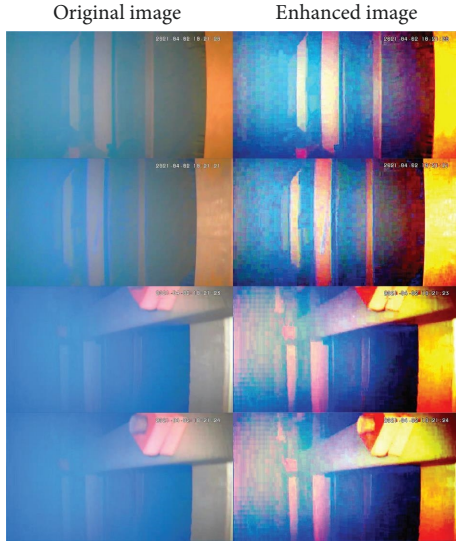The optimization function represents the error between the robot motion model and the observation model. In this

used to estimate the relative motion of the robot. The EPnP algorithm only needs 4 pairs of matching points to solve the camera pose, as described below.

In the world coordinate system, 4 control points are selected to describe the distribution of all 3D points in the space. Each 3D point $\mathbf{p}_i^w$ can be expressed as

$$\mathbf{p}_i^w = \sum_{j=1}^{4} \alpha_{ij} \mathbf{c}_j^w, \quad \sum_{j=1}^{4} \alpha_{ij} = 1, \tag{16}$$

where $\mathbf{c}_i^c$ indicates the control point and α_{ij} represents the weight of the four control points. Points in the world coordinate system are transformed to the camera coordinate system by rotation and translation, and the conversion equation can be expressed as

$$\mathbf{p}_i^c = \mathbf{R}_{cw} \mathbf{p}_i^w + \mathbf{t} = \mathbf{R}_{cw}\left(\sum_{j=1}^{4} \alpha_{ij} \mathbf{c}_j^w\right) + \mathbf{t} = \sum_{j=1}^{4} \alpha_{ij}\left(\left(\mathbf{R}_{cw} \mathbf{c}_j^w\right) + \mathbf{t}\right)$$

$$= \sum_{j=1}^{4} \alpha_{ij} \mathbf{c}_j^c, \tag{17}$$

where $\mathbf{c}_i^c$ indicates the coordinate of the control point in the camera coordinate system. Because α_{ij} is invariant in the camera coordinate system, the weight of α_{ij} in the world coordinate system is known. All points in the world coordinate system can obtain their coordinates $\mathbf{p}_i^c$ in the camera coordinate system.

Assuming that the coordinate of a 3D point in the camera coordinate system is $\mathbf{p}_i^c$ and the coordinate in the world coordinate system is $\mathbf{p}_i^w$, there are rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$ that can satisfy the following equation:

FIGURE 6: The comparison before and after image enhancement.

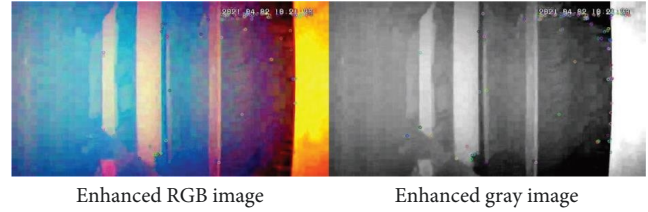FIGURE 7: Feature extraction of the original image.

FIGURE 8: Feature extraction of the enhanced image.

TABLE 1: Results at each layer of the 8-layer image pyramid.

Layer	0	1	2	3	4	5	6	7
Number	213	118	70	48	37	20	7	7

paper, the Levenberg–Marquardt gradient descent strategy is used to optimize the objective function, and convergence can be realized after 10 iterations. To this point, the posture of the robot has been optimized.

4. Experimental Results and Discussion

The robot was tested in the Mitsubishi transformer of Shuibei Power Supply Station in Shenzhen. To evaluate the positioning performance of this algorithm on the transformer, an experiment was conducted with the image frame sequence that was obtained during the experiment. The internal space of the transformer is small and compact, and it has been working in a high temperature and high-pressure environment for more than 20 years. The discoloration of the transformer oil is serious.

4.1. Comparative Experiment on Feature Point Extraction. To verify the improvement of the image enhancement algorithm on feature point extraction, a comparative experiment of feature point extraction was carried out between the original image obtained by the robot and the enhanced image. The comparison before and after image enhancement obtained from Section 3.1 is shown in Figure 6. The results of the comparative experiment of feature extraction are shown in Figures 7 and 8. As shown in Figure 7, the original image is distorted and blurred by the metamorphic and discolored transformer oil. In the original image and grayscale image, only a few feature points can be extracted in the upper right corner of the image, so the pose of the robot cannot be solved by using the original image directly.

As shown in Figure 8, more feature points can be extracted where the pixel brightness contrast is strong in the enhanced picture. The experimental results of feature extraction using an 8-layer pyramid are presented in Table 1 and illustrated in Figure 9.

The algorithm's robustness and stability can be improved, and effective feature detection and matching across multiple scales can be allowed by using an image pyramid for feature point extraction.

To verify the robustness of the image enhancement algorithm, we selected 10 images for comparative experiments, and the results are shown in Table 2.

As shown in Table 2, the average number of feature points extracted from 10 images without enhancement is 33.5. After the image is enhanced, the average number of feature points is approximately 333. The results show that the number of extracted feature points was greatly increased.

4.2. Comparison of Feature Matching between Images. To verify the performance of the image enhancement algorithm on interframe matching results, a comparative experiment of interframe feature matching was carried out. First, the feature points of two consecutive original images are extracted for feature matching, and mismatching points are deleted. Then, the same is done for two consecutive gray images. The matching results are shown in Figure 10, where only 14 pairs of points are successfully matched.

The feature matching experiments are carried out on the above two consecutive images after image enhancement, and the mismatching points are deleted. The matching results are shown in Figure 11. The number of matching pairs is greatly increased, reaching 72 pairs, which is beneficial for solving and optimizing the robot's pose.

To realize the tracking and positioning of the robot, it is necessary to match successive multiframe images. Therefore, the matching experiment was carried out on the 10 consecutive enhanced images, and the feature matching results before and after image enhancement are shown in Table 3.
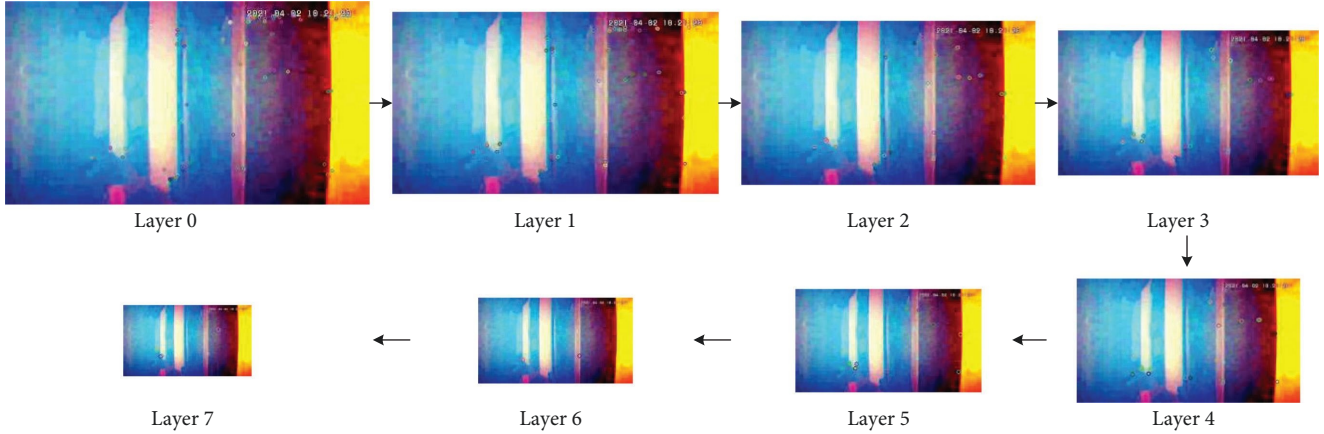
FIGURE 9: Results at each layer of the 8-layer image pyramid.

TABLE 2: Comparison results of feature extraction quantity.

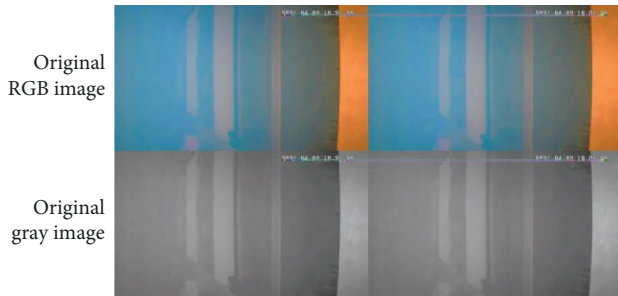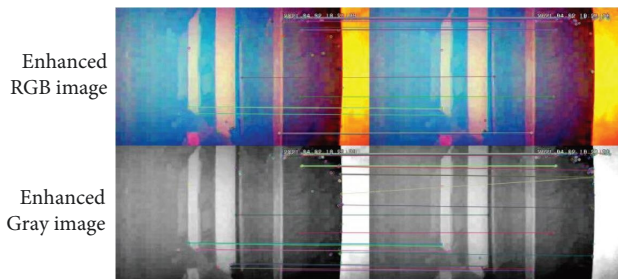| Number	Original image	Enhanced image
Image 1	33	339
Image 2	32	334
Image 3	33	302
Image 4	32	305
Image 5	36	320
Image 6	35	323
Image 7	36	328
Image 8	32	340
Image 9	33	395
Image 10	33	345

FIGURE 10: Feature matching of the original image.

FIGURE 11: Feature matching of the enhanced image.

4.3. Algorithm Comparison Simulation Location Experiment.

The real trajectory of the robot cannot be obtained inside the transformer. To verify the error between the trajectory obtained by the robot positioning method proposed in this paper and the real trajectory of the robot, the ROS and Gazebo environments [33] are used to build the simulation environment of the robot movement in transformer oil. The environment is shown in Figure 12, including the robot and the observation target.

The robot is equipped with a monocular camera. Gaussian noise is added to the robot monocular camera in the simulation process, and the images obtained before and after increasing the noise are shown in Figure 13.

As shown in Figure 13, the blurred image obtained by the robot after adding noise increases the difficulty of robot positioning. Image enhancement and feature matching experiments are performed on the data with added noise. The matching results are shown in Figure 14. After matching and screening, there are 46 pairs of matching points in the image with noise, and the number of matched feature points in the enhanced image reaches 93 pairs. The experimental results show that the image enhancement algorithm can effectively improve the matching number of image feature points.

To verify the positioning accuracy of the algorithm, robot motion observation and control experiments are outperformed. Conduct comparative experiments with the open-source monocular visual odometry algorithm direct sparse odometry (DSO). DSO is a monocular vision-based localization algorithm released by Dr. Jakob Engel from the Computer Vision Laboratory at the Technical University of Munich (TUM) in 2016. DSO falls under the category of sparse direct methods. The 3D trajectory of the robot in the simulation environment is shown in Figure 15, in which the blue trajectory represents the trajectory obtained by the localization algorithm, green represents the real trajectory of the robot, and pink indicates the trajectory obtained by the DSO algorithm based on a monocular visual odometer. As shown in Figure 15, the motion trajectory obtained by the algorithm in this paper is basically consistent with the real

TABLE 3: Comparison of the number of matching features before and after image enhancement.

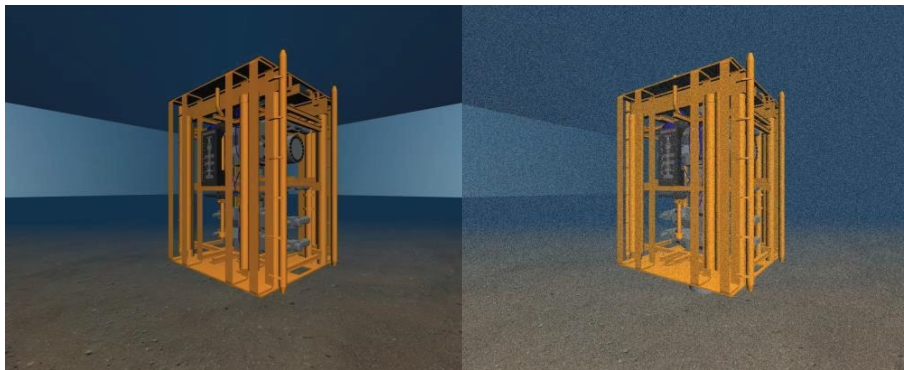| Number	Original image	Enhanced image	Increased (%)
Image 1 and Image 2	14	72	414
Image 2 and Image 3	16	34	113
Image 3 and Image 4	22	86	290
Image 4 and Image 5	24	79	229
Image 5 and Image 6	26	119	358
Image 6 and Image 7	27	100	270
Image 7 and Image 8	22	85	286
Image 8 and Image 9	23	58	152
Image 9 and Image 10	25	88	252
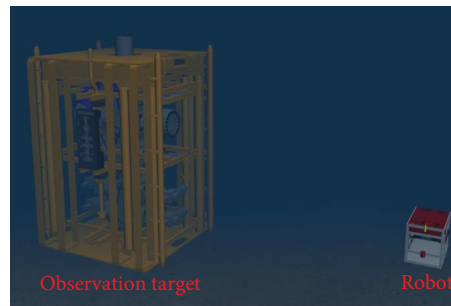
FIGURE 12: The simulation environment.

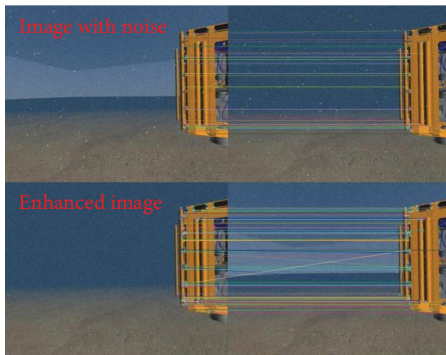FIGURE 13: Comparison of images before and after adding noise.

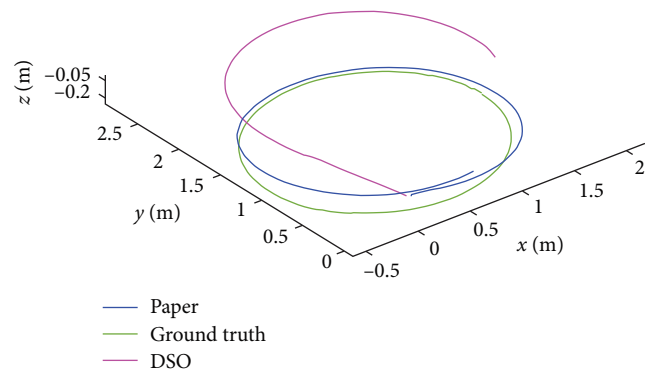FIGURE 14: Feature matching comparison before and after image.

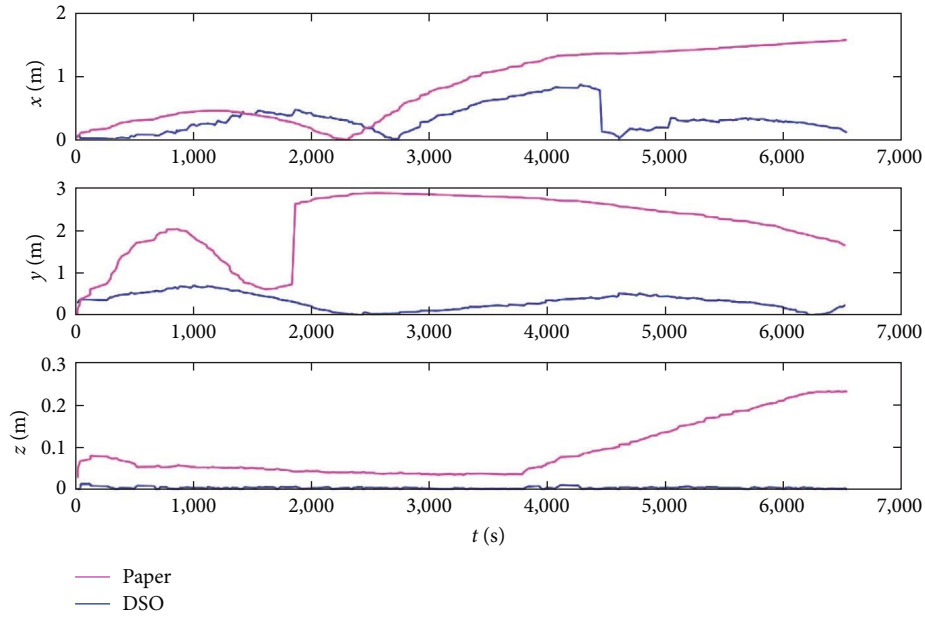FIGURE 15: The results of the 3D trajectory estimation of each algorithm.

FIGURE 16: The position error of each algorithm on each coordinate axis.

TABLE 4: The comparative results of the root mean square error.

Algorithm	Error of X axis (m)	Error of Y axis (m)	Error of Z axis (m)
DSO	1.016	2.278	0.115
Paper	0.377	0.370	0.003

trajectory, with a certain deviation. The trajectory obtained by the DSO algorithm is quite different from the real trajectory.

It is not easy to evaluate the positioning effect using a 3D trajectory, so the positioning errors of the trajectory on the X axis, Y axis, and Z axis are plotted. The results are shown in Figure 16, in which the pink curve represents the positioning error of the DSO algorithm, and the blue curve represents the positioning error of the algorithm in this paper. As shown in Figure 16, the positioning error of the proposed algorithm is smaller than that of the DSO algorithm in the X-axis, Y-axis, and Z-axis directions. The maximum positioning error of the proposed algorithm is only 1 m, while the maximum positioning error of the DSO algorithm is close to 3 m. In addition, the root-mean-square errors of the robot position on the X-axis, Y-axis, and Z-axis are calculated. As shown in Table 4, the root-mean-square errors of this algorithm are obviously smaller than those of the DSO algorithm.

The simulation results show that the initial trajectory of the DSO location algorithm is far from the real trajectory, and the tracking fails. After image enhancement, the algorithm in this paper effectively realizes feature extraction and accurately restores the running trajectory of the robot under oil. From the error comparison experimental data, it can be seen that the positioning accuracy of the proposed method is better than that of the DSO positioning algorithm, thus verifying the effectiveness of the proposed method in the transformer oil environment.
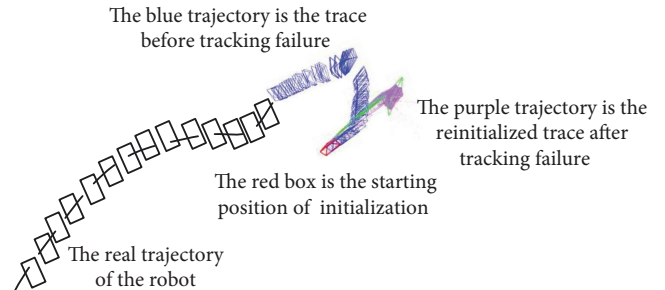

FIGURE 17: The failed trajectory of using raw data.

4.4. Positioning Experiment of the Robot in Transformer Oil.
To verify the effectiveness of the positioning method of the submersible transformer inspection robot, the positioning method is tested and analyzed by the data collected by the robot in the Mitsubishi transformer. The time length of the experimental data is 57 s, and the frame rate is 25 fps. The video data are extracted into a picture sequence, and the positioning performance test is carried out.

The robot tracking and positioning experiment is carried out by the original data, as shown in Figure 17, which mainly includes a blue trajectory, a purple trajectory and a red starting position. After the robot is successfully initialized in the red position, the blue track indicates the robot trajectory drawn from the original image. At the end of the blue track,

The blue box is the
historical position of
the robot

The red box is the
starting position
of the robot

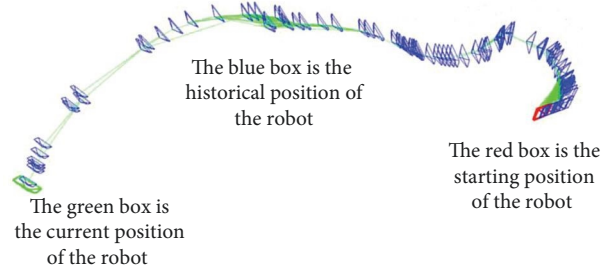The green box is
the current position
of the robot

FIGURE 18: The robot trajectory based on the proposed method.

TABLE 5: Robot trajectory coordinates.

Time	Translation in the world coordinate system			Rotation of camera to world coordinate system			
	t_x	t_y	t_z	q_x	q_y	q_z	q_w
189.000000	−0.1369671	−0.1500254	−0.5803908	0.0149806	−0.5032533	0.1879581	0.8433169
191.000000	−0.1364136	−0.1501418	−0.5806609	0.0155979	−0.5164054	0.1937663	0.8339885
192.000000	−0.1364136	−0.1505871	−0.5823005	0.0154008	−0.5224573	0.1981140	0.8291876
193.000000	−0.1349777	−0.1504681	−0.5819005	0.0158103	−0.5293952	0.2010548	0.8240557
196.000000	−0.1217337	−0.1564949	−0.5938352	0.0128108	−0.5457574	0.2140372	0.8100450
199.000000	−0.1082227	−0.1607452	−0.6106782	0.0127889	−0.5609802	0.2266808	0.7960864
200.000000	−0.1013026	−0.1660932	−0.6151043	0.0095857	−0.5662362	0.2314494	0.7910221
204.000000	−0.0721812	−0.1746682	−0.6406906	0.0094067	−0.5877521	0.2468827	0.7703946
207.000000	−0.0239709	−0.1925754	−0.6715866	0.0039448	−0.5989393	0.2570543	0.7584057
208.000000	−0.0158377	−0.1951312	−0.6766809	0.0024216	−0.6025855	0.2606056	0.7543007
209.000000	0.0074293	−0.2035241	−0.6885141	−0.0004768	−0.6057649	0.2638864	0.7506082
210.000000	0.0324675	−0.2134237	−0.7010788	−0.0042803	−0.6082725	0.2677946	0.7471762
211.000000	0.0523740	−0.2196917	−0.7121715	−0.0064705	−0.6113610	0.2710429	0.7434592
212.000000	0.0739813	−0.2284881	−0.7206520	−0.0101452	−0.6147486	0.2753150	0.7390419
214.000000	0.1247762	−0.2471465	−0.7436563	−0.0156883	−0.6214890	0.2820806	0.7307091

the robot failed to track because the image obtained by the robot was not clear. Then, the robot localization algorithm is reinitialized. After reinitialization, according to the image currently acquired by the robot, the motion trajectory is drawn, which is shown in purple. From the results, we can see that the robot trajectory can be drawn according to the acquired visual information. However, if the image quality obtained by the robot is poor, it will lead to the failure of robot tracking, the repeated reinitialization of the positioning algorithm, and the discontinuous trajectory of the robot.

To verify the effectiveness of the proposed algorithm in the transformer, the image sequence is processed by multi-scale fusion enhancement, and then the robot tracking and positioning experiment is carried out. The results are shown in Figure 18. It includes a red box, some blue boxes, and a green box. The red box indicates the initial position of the robot, the blue boxes indicate the historical trajectory of the robot, and the green box indicates the current position of the robot.

From the result in Figure 18, the robot trajectory can be displayed intuitively and clearly by the positioning method in this paper, and relatively accurate position information

can be obtained. In addition, tracking failure and the phenomenon of reinitialization do not appear. The pose information of each frame in the image sequence can be provided by the positioning method. Part of the data are shown in Table 5.

On the basis of the coordinate information of each image obtained by the robot, the 3D motion trajectory of the robot inside the transformer is drawn by MATLAB, which is shown in Figure 19. The red trajectory represents the robot trajectory obtained by the original image, and the blue trajectory represents the robot trajectory obtained after image enhancement. It can be seen from the results that maximum errors of 0.03 m along the X-axis, 0.008 m along the Y-axis, and 0.005 m along the Z-axis within the actual transformer space. Root-mean-square errors are calculated at 0.0158 m along the X-axis, 0.0026 m along the Y-axis, and 0.0028 m along the Z-axis. Errors in positioning along each axis inside the transformer by the robot are found to be within permissible limits. And the robot's trajectory in the transformer can be drawn by monocular vision in this algorithm, and the algorithm solves that the difficulty of locating the robot in the transformer.
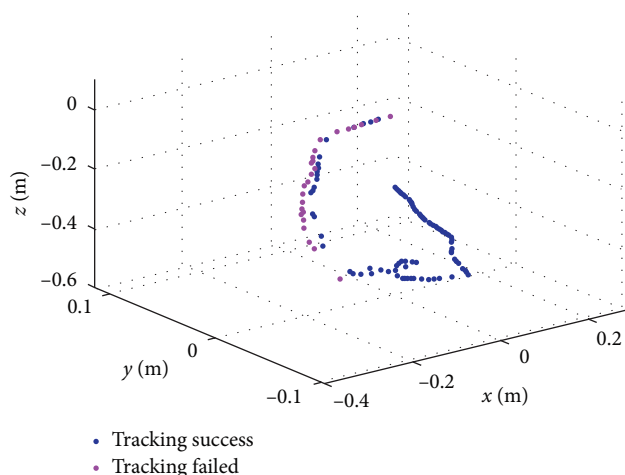
FIGURE 19: Trajectory comparison based on recovery coordinate.

5. Conclusion

To realize the autonomous positioning of the robot in the deteriorated and discolored transformer oil, a visual positioning method of the submersible transformer inspection robot is proposed in this paper. Aiming at the problem of transformer oil deterioration and discoloration, an image enhancement algorithm based on multiscale fusion strategy is proposed, to improve the image quality, brightness, and contrast. Aiming at accurately locating the robot pose in the transformer, the rotation and scale information of the robot are added based on the feature extraction of FAST key points and BRIEF descriptors. The interframe pose is solved by epipolar constraint and EPnP, and the robot pose optimization model is designed to further improve the positioning accuracy of the robot. The performance of the algorithm proposed in this paper is evaluated by using the monocular image obtained by the robot inside the transformer. The experimental results show that the enhanced images can satisfy feature point extraction, the trajectory of robot in transformer the algorithm can be plotted.

Our future work involves two aspects: first, investigating how to test the positioning accuracy of the algorithm within the transformer; second, considering ways to enhance positioning accuracy in larger and more complex scenarios. Building upon the existing positioning methods, incorporating multiple sensors such as inertial navigation or radar is contemplated to improve the robot's positioning accuracy. Additionally, the addition of loop detection to the system to mitigate the accumulation of errors over time is seen as an important direction for our next research steps.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this article.

References

[1] S. Tenbohlen, S. Coenen, M. Djamali, A. Müller, M. H. Samimi, and M. Siegel, "Diagnostic measurements for power transformers," *Energies*, vol. 9, no. 5, Article ID 347, 2016.

[2] F. Yingbin, Y. Yang, G. Hongwei, and L. Zhigang, "Floating robot design method for inspecting the inside of transformer," *Journal of Mechanical Engineering*, vol. 56, no. 7, pp. 52–59, 2020.

[3] Y. B. Feng, X. H. Zhao, Z. He, Z. G. Li, and Y. B. Wang, "Depth hover control of spherical robot for internal detection of oil-immersed transformer," *Control and Decision*, vol. 35, no. 2, pp. 375–381, 2020.

[4] Q. Zhan and W. Li, "Research progress and development trend of spherical mobile robots," *Journal of Mechanical Engineering*, vol. 55, no. 9, pp. 1–17, 2019.

[5] T. Zhang, M. D. Xia, J. Y. Zhang, Y. Y. Zhu, and J. W. Tong, "Overview of underwater navigation and positioning technology," *GNSS World of China*, vol. 47, pp. 1–16, 2022.

[6] W. W. Yin and S. L. Guo, "Survey on non-satellite underwater navigation and positioning technology," *Ship Electronic Engineering*, vol. 37, pp. 8–11, 2017.

[7] C. Y. Li, D. S. Zhang, Y. S. Zhang, Q. Zhou, and G. J. Qin, "The underwater navigation and location method based on acoustic positioning and dead reckoning," *Ship Engineering*, vol. 37, pp. 57–60, 2015.

[8] D. J. Sun, C. E. Zheng, J. C. Zhang, Y. F. Han, and H. Y. Cui, "Development and prospect of underwater acoustic positioning and navigation technology," *Bulletin of the Chinese Academy of Sciences*, vol. 34, pp. 331–338, 2019.

[9] B. W. Wei, W. H. Lv, X. J. Fan, Y. K. Zhu, and Y. J. Guo, "Present situation and prospect of autonomous undersea vehicle navigation technologies," *Journal of Unmanned Undersea Systems*, vol. 27, pp. 1–9, 2019.

[10] L. Paull, S. Saeedi, M. Seto, and H. Li, "AUV navigation and localization: a review," *IEEE Journal of Oceanic Engineering*, vol. 39, no. 1, pp. 131–149, 2014.

[11] J. Z. Yu, S. H. Kong, and Y. Meng, "Methods and technologies for visual perception of underwater environment," *Robot*, vol. 44, pp. 224–235, 2022.

[12] S. R. Zekavat, R. M. Buehrer, G. D. Durgin et al., "An overview on position location: past, present, future," *International Journal of Wireless Information Networks*, vol. 28, pp. 45–76, 2021.

[13] H. Martínez-Barberá, P. Bernal-Polo, and D. Herrero-Pérez, "Sensor modeling for underwater localization using a particle filter," *Sensors*, vol. 21, no. 4, pp. 1549–1572, 2021.

[14] Q. Luo, X. Yan, C. Ju, Y. Chen, and Z. Luo, "An ultra-short baseline underwater positioning system with Kalman filtering," *Sensors*, vol. 21, no. 1, pp. 143–162, 2021.

[15] J. Li, *Research of underwater spherical robot vision system*, M.S. thesis, Beijing University of Posts and Telecommunications, China, 2019.

[16] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: real-time single camera SLAM," *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 1052–1067, 2007.

[17] G. Simon, G. Zachár, and G. Vakulya, "Lookup: robust and accurate indoor localization using visible light communication," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2337–2348, 2017.

[18] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: large-scale direct monocular SLAM," in *Computer Vision—ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., pp. 834–849, vol. 8690, Springer, Cham, 2014.

[19] S. Suresh, E. Westman, and M. Kaess, "Through-water stereo SLAM with refraction correction for AUV localization," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 692–699, 2019.

[20] P. F. Du, J. Han, J. Y. Wang et al., "View-based underwater SLAM using a stereo camera," in *Oceans 2017—Aberdeen*, pp. 1–6, IEEE, Aberdeen, UK, 2017.

[21] F. M. Wang, Y. Zhang, X. Zhang, and J. Z. Li, "Improvement of underwater visual SLAM front-end based on stereo camera," *Machine Tool & Hydraulics*, vol. 48, pp. 35–39, 2020.

[22] Z. Liu and H. Wang, "Automatic detection of transformer components in inspection images based on improved faster R-CNN," *Energies*, vol. 11, no. 12, Article ID 3496, 2018.

[23] H. Zhang, J. Yang, and Z. Sui, "Research on transformer oil level monitoring system based on inspection robot," *Journal of Physics: Conference Series*, vol. 2378, Article ID 012085, 2022.

[24] Y. H. Pan, L. Wang, X. D. Wang, F. Zhao, and H. Tao, "A reliability improvement technology of transformer internal inspection robot and its control system," *Journal of Physics: Conference Series*, vol. 2399, Article ID 012033, 2022.

[25] Y. Feng, Y. Kou, Y. Zhao, and H. Yang, "Locating method of oil-filled transformer inspection robot," in *2nd International Conference on Machine Learning and Computer Application*, pp. 1–5, IEEE Xplore, Shenyang, China, 2021.

[26] Y. X. Wang, M. Diao, and C. Han, "Underwater image restoration algorithm based on least squares estimation," *Journal of Computer-Aided Design and Computer Graphics*, vol. 30, no. 11, pp. 2125–2133, 2018.

[27] L.-C. Chiu, T.-S. Chang, J.-Y. Chen, and N. Y.-C. Chang, "Fast SIFT design for real-time visual feature extraction," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3158–3167, 2013.

[28] T. Lindeberg, *Scale-Space Theory in Computer Vision*, pp. 33–38, Springer, New York, NY, 2013.

[29] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, England, 2003.

[30] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: an accurate O(n) solution to the PnP problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166, 2009.

[31] D. Kalman, "A singularly valuable decomposition: the SVD of a matrix," *The College Mathematics Journal*, vol. 27, no. 1, pp. 2–23, 2018.

[32] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert, "Initialization techniques for 3D SLAM: a survey on rotation estimation and its use in pose graph optimization," in *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 4597–4604, IEEE, Seattle, USA, 2015.

[33] C. X. Hu, *Development Practice of ROS Robot*, pp. 78–81, Machine Press, Beijing, China, 2018.