applied sciences

MDPI

Article

Robust Localization of Flange Interface for LNG Tanker Loading and Unloading Under Variable Illumination a Fusion Approach of Monocular Vision and LiDAR

Mingqin Liu *, Han Zhang, Jingquan Zhu, Yuming Zhang and Kun Zhu

School of Mechanical Engineering, Jiangsu Ocean University, Lianyungang 222000, China; zhhmoyun@163.com (H.Z.); 2024220537@jou.edu.cn (J.Z.); 2023220530@jou.edu.cn (Y.Z.); 2024210522@jou.edu.cn (K.Z.)

* Correspondence: lmq191@jou.edu.cn

Abstract

The automated localization of the flange interface in LNG tanker loading and unloading imposes stringent requirements for accuracy and illumination robustness. Traditional monocular vision methods are prone to localization failure under extreme illumination conditions, such as intense glare or low light, while LiDAR, despite being unaffected by illumination, suffers from limitations like a lack of texture information. This paper proposes an illumination-robust localization method for LNG tanker flange interfaces by fusing monocular vision and LiDAR, with three scenario-specific innovations beyond generic multi-sensor fusion frameworks. First, an illumination-adaptive fusion framework is designed to dynamically adjust detection parameters via grayscale mean evaluation, addressing extreme illumination (e.g., glare, low light with water film). Second, a multi-constraint flange detection strategy is developed by integrating physical dimension constraints, K-means clustering, and weighted fitting to eliminate background interference and distinguish dual flanges. Third, a customized fusion pipeline (ROI extraction-plane fitting-3D circle center solving) is established to compensate for monocular depth errors and sparse LiDAR point cloud limitations using flange radius prior. High-precision localization is achieved via four key steps: multi-modal data preprocessing, LiDAR-camera spatial projection, fusion-based flange circle detection, and 3D circle center fitting. While basic techniques such as LiDAR-camera spatiotemporal synchronization and K-means clustering are adapted from prior works, their integration with flange-specific constraints and illumination-adaptive design forms the core novelty of this study. Comparative experiments between the proposed fusion method and the monocular vision-only localization method are conducted under four typical illumination scenarios: uniform illumination, local strong illumination, uniform low illumination, and low illumination with water film. The experimental results based on 20 samples per illumination scenario (80 valid data sets in total) show that, compared with the monocular vision method, the proposed fusion method reduces the Mean Absolute Error (MAE) of localization accuracy by 33.08%, 30.57%, and 75.91% in the X, Y, and Z dimensions, respectively, with the overall 3D MAE reduced by 61.69%. Meanwhile, the Root Mean Square Error (RMSE) in the X, Y, and Z dimensions is decreased by 33.65%, 32.71%, and 79.88%, respectively, and the overall 3D RMSE is reduced by 64.79%. The expanded sample size verifies the statistical reliability of the proposed method, which exhibits significantly superior robustness to extreme illumination conditions.

Keywords: LNG; monocular vision; LiDAR; flange; variable illumination; image-based point cloud

Academic Editor: Ephraim Suhir

Received: 16 December 2025
Revised: 14 January 2026
Accepted: 17 January 2026
Published: 22 January 2026

1. Introduction

As a clean energy source, liquefied natural gas (LNG) has been widely adopted due to its advantages of convenience in transportation and high energy density [1]. Onshore LNG transportation is primarily undertaken by tanker trucks, which exhibit high flexibility and cost-effectiveness to meet the gas supply demands of diverse end-users [2]. Compared with pipeline transportation, LNG tanker delivery is more suitable for scenarios characterized by low gas consumption and long delivery distances. It can flexibly transport LNG from receiving terminals to end-points (e.g., power plants and urban gas filling stations), respond rapidly to market demands, and realize efficient energy distribution [3]. However, the current loading and unloading process mainly relies on manual guidance, which gives rise to multiple engineering challenges [4]. The large size of tanker trucks creates visual blind spots for drivers, easily leading to the risk of collision with related equipment due to parking misalignment [5]. Manual operations are time-consuming and inefficient, making it difficult to satisfy high-frequency loading/unloading requirements [6]. In addition, inherent safety risks are also posed due to the unique properties of LNG. Therefore, the precise positioning and docking of the loading and unloading interface have become the key to addressing these problems. Nevertheless, existing research predominantly focuses on autonomous driving or robot navigation, leaving a critical gap in the development of targeted fusion localization methods for industrial scenarios such as the flange interface of LNG tanker trucks.

Light Detection and Ranging (LiDAR) is characterized by sparse yet high-precision depth data, while cameras provide dense but low-precision depth data; the fusion of these two sensors enables the depth value restoration of pixels [7]. Reference [8] accomplishes the fusion of sparse LiDAR depth data and images relying on basic image processing operations; however, it is only applicable to Simultaneous Localization and Mapping (SLAM) navigation scenarios and fails to take into account the fixed-size constraints of industrial components. Reference [9] proposed a fusion method combining RGB-D cameras and LiDAR, which is mainly designed to address the issue of visual tracking failure yet lacks targeted localization optimization for specific targets such as loading and unloading interfaces. Reference [10] employs visual algorithm-based feature points for loop closure detection to enhance the performance of LiDAR, yet it still fails to break through the bottleneck of localization robustness for targets under dynamic lighting conditions. Reference [11] proposed running visual and LiDAR localization algorithms in parallel, and realized data coupling by utilizing the residuals of these two modalities during the optimization phase. Reference [12] proposed an overall framework integrating visual odometry and LiDAR odometry, while simultaneously improving the performance of real-time motion estimation and point cloud registration algorithms. Reference [13] adopted visual odometry to provide an initial position value for the LiDAR localization algorithm, thus achieving real-time and accurate pose estimation. Reference [14] proposed a graph optimization framework based on LiDAR and cameras. Reference [15] integrated vision and ultrasonic sensors for the automated docking of LNG loading arms, achieving millimeter-level positioning accuracy; however, the robustness of such methods under extremely variable lighting conditions (e.g., strong light, low light, and water film reflection) has not been fully verified.

Recent advancements in methods for detecting specific geometric features (e.g., circles/ellipses) are noteworthy. For instance, reference [16] proposed a deep learning-based ellipse detection network (EllipseNet), which can effectively handle viewpoint variations and illumination interference. However, it relies on large-scale labeled data, and its robustness under extreme illumination or sparse point cloud scenarios—precisely the core challenges in LNG flange localization—remains insufficient. In the aspect of LiDAR–camera fusion calibration, reference [17] proposed a targetless continuous-time calibration method

capable of jointly estimating intrinsic and extrinsic parameters even with non-overlapping sensor fields of view and asynchronous multi-sensor data. Yet, it is primarily designed for autonomous driving scenarios and lacks optimization for fixed-size industrial targets like flanges. Furthermore, addressing reflection and water-surface interference, reference [18] introduced a LiDAR feature matching-based dynamic object detection method, enhancing robustness in maritime environments through voxelization and feature descriptor fusion. However, its focus is on object detection rather than high-precision localization of static industrial components. In contrast, the method proposed in this paper specifically targets the fixed geometric dimensions of LNG flange interfaces and extreme illumination conditions. By fusing monocular vision and LiDAR through multi-constraint optimization and an illumination-adaptive mechanism, it achieves high-precision localization in complex scenarios such as intense glare, low light, and water film reflections, demonstrating both methodological novelty and engineering practicality.

To address the aforementioned problem, this paper proposes a light-robust positioning method based on the fusion of monocular vision and lidar. The core solution is to leverage the high-resolution advantage of vision to achieve accurate 2D feature detection of flanges, and compensate for the shortcomings of depth measurement by virtue of the illumination robustness of lidar. Through spatiotemporal synchronization, multi-constraint optimization and 3D fitting, a fusion positioning framework consisting of 2D positioning, depth compensation and 3D solution is established. The specific innovations are as follows:

1. An illumination-adaptive radar-camera fusion framework is proposed. The illumination intensity is evaluated via the gray mean value, and the parameters of edge detection and circle detection are dynamically adjusted to adapt to complex scenarios including uniform illumination, local strong light, and low light with water film.
2. A multi-constraint flange detection and fitting strategy is designed. Background interference is eliminated by combining the physical size constraints of flanges, the left and right flanges are distinguished using K-means clustering, and weighted fitting is adopted to enhance the core features of the outer edge circle, thereby improving the positioning stability.
3. A fusion positioning process featuring Region of Interest (ROI) extraction-plane fitting-3D circle center solving is established. The ROI of LiDAR point cloud is extracted under the constraint of the 2D circle center obtained by vision, valid points are filtered via plane fitting, and the 3D circle center fitting of sparse point cloud is completed by integrating the physical radius constraints of flanges, which accurately compensates for the depth error of monocular vision.

The remainder of this study is organized as follows: Section 2 elaborates on the spatiotemporal synchronization method for LiDAR and cameras. Section 3 presents a detailed introduction to the core algorithms of fusion localization, including flange detection optimization, point cloud processing, and 3D circle center fitting. Section 4 verifies the accuracy and robustness of the proposed method through experiments under multi-illumination scenarios. Section 5 discusses the method in relation to existing works, and Section 6 concludes the study and suggests future directions.

To further clarify the scientific contributions of this work and explicitly distinguish novel designs from adapted techniques, the key distinctions are summarized as follows:

The illumination-adaptive radar-camera fusion framework and multi-constraint flange detection strategy are original innovations, as prior fusion works (e.g., [8,9,15]) lack targeted illumination adaptation and physical constraint integration for fixed-size industrial flanges. The ROI extraction-plane fitting-3D circle center solving pipeline is a scenario-specific improvement: while LiDAR-camera spatial-temporal synchronization and basic clustering algorithms are based on existing techniques, their integration with flange-specific

constraints (fixed radius, dual-flange layout) and monocular vision guidance is newly developed to address sparse point cloud fitting issues of 16-line LiDAR in industrial docking scenarios. The weighted flange center fitting strategy is also a tailored design for dual-circle features (outer edge and central hole) of flanges, which has not been reported in prior flange localization or multi-sensor fusion studies. This work thus goes beyond simple combination of LiDAR and camera data, providing a targeted solution for the practical challenges of LNG tanker automated loading/unloading under variable illumination.

2. Spatiotemporal Synchronization of LiDAR and Cameras

2.1. Spatial Synchronization

The data collected by LiDAR and cameras are in their respective coordinate systems. To fuse these data, it is necessary to transform them into a unified coordinate system [19]. To ensure the spatial consistency of multi-modal data, two types of calibration need to be completed. First, the monocular camera is calibrated using the Zhang Zhengyou calibration method [20] to obtain the camera intrinsic parameters matrix K. The matrix K is defined as a 3×3 upper triangular matrix with the following form:

$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where f_x and f_y represent the effective focal lengths (in pixels) in the horizontal and vertical directions, and u_0 and v_0 denote the horizontal and vertical coordinates of the principal point in the image plane. The final calibration yields the intrinsic parameter matrix, with the specific values given as follows:

$$K = \begin{bmatrix} 1430.53 & 0 & 641.6 \\ 0 & 1430.67 & 346.85 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Subsequently, by collecting multiple sets of synchronized images and point cloud data containing a checkerboard calibration target, we perform joint calibration of the LiDAR and camera using the algorithms of feature matching and reprojection error minimization. Through this process, we solve for the 3×3 rotation matrix R and 3×1 translation vector t between the LiDAR and the camera, thereby achieving spatial alignment of multi-modal data.

The LiDAR coordinate system follows the right-hand rule: the X-axis points forward, the Y-axis points to the left, and the Z-axis points upward, consistent with the ROS (Robot Operating System) standard coordinate convention for mobile robots.

To ensure the reproducibility of the joint calibration, a standard checkerboard calibration target was adopted with the following parameters: the side length of each small square is 86 mm, and the checkerboard size is 7×8 squares. During the acquisition of calibration data, the checkerboard target was placed within the overlapping field of view of the LiDAR and the camera, with a total of 10 different poses set (distance range: 1.0–2.5 m, covering the typical working distance for LNG flange localization). For each pose, 3 consecutive images and corresponding point cloud data were collected. After excluding blurred images and incomplete point clouds, 30 sets of valid images and point clouds were finally obtained. The reprojection error was selected as the core metric for evaluating calibration accuracy (consistent with standard camera-LiDAR calibration practices [16,18]). This error quantifies the deviation between the pixel position of the LiDAR 3D points projected onto the camera image after calibration and the pixel position of the actually detected 2D corner points. The final average reprojection error of the joint calibration is 1.9 pixels, with a maximum error

of 4 pixels and a minimum error of 0.4 pixels, indicating reliable spatial alignment between the two sensors. The final calibrated transformation parameters are as follows:

$$R = \begin{bmatrix} 0.1060 & -0.9941 & 0.0216 \\ -0.0240 & -0.0242 & -0.994 \\ 0.9941 & 0.1054 & -0.0264 \end{bmatrix} \quad (3)$$

$$T = \begin{bmatrix} 0.0563 & -0.1091 & -0.7796 \end{bmatrix} \quad (4)$$

In this study, the Lidar Camera Calibrator tool in MATLAB is adopted to complete the extrinsic parameter calibration of the LiDAR and camera.

To achieve effective fusion of LiDAR and camera data, this paper establishes a complete coordinate transformation relationship, and the diagram of the coordinate transformation relationship is shown in Figure 1:

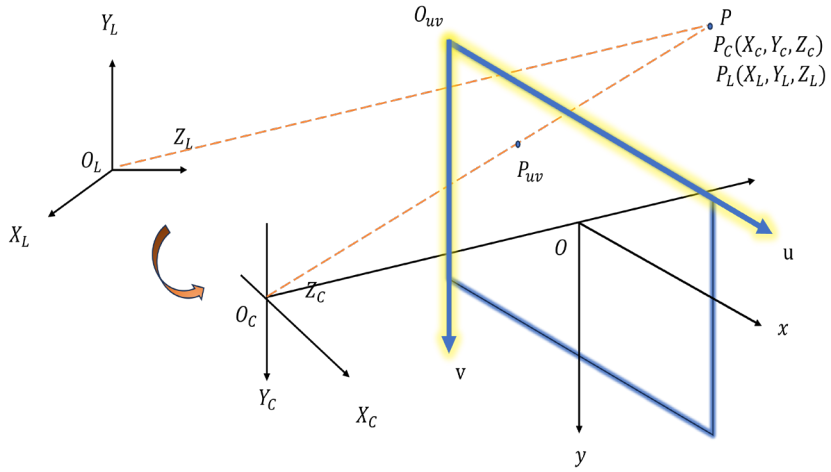

Figure 1. Coordinate Transformation Relationship Diagram.

This diagram mainly involves the LiDAR coordinate system $O_L - X_L Y_L Z_L$, the camera coordinate system $O_C - X_C Y_C Z_C$, the image coordinate system $O - xy$, and the pixel coordinate system $O_{uv} - uv$. The transformation matrix from the LiDAR coordinate system to the camera coordinate system is given by:

$$T_{C \leftarrow L} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad (5)$$

The rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$ in the formula are the parameters obtained from the joint calibration of the LiDAR and camera.

A point P in space is denoted as P_L in the LiDAR coordinate system and P_C in the camera coordinate system. The formula for transforming point P_L from the LiDAR coordinate system to the camera coordinate system is expressed as follows:

$$P_C = RP_L + t = T_{C \leftarrow L} P_L \quad (6)$$

Prior to projecting P_C onto the image plane, lens distortion correction is performed using the calibrated distortion coefficients based on the standard Brown–Conrady distortion model. The correction process involves converting P_C to normalized image coordinates

$(x_n,\ y_n)$ and then compensating for radial and tangential distortions to obtain corrected normalized coordinates $(x_c,\ y_c)$:

$$x_n = \frac{X_C}{Z_C},\ y_n = \frac{Y_C}{Z_C} \tag{7}$$

$$r^2 = x_n^2 + y_n^2 \tag{8}$$

$$\begin{cases} x_c = x_n\left(1 + k_1 r^2 + k_2 r^4\right) + 2p_1 x_n y_n + p_2\left(r^2 + 2x_n^2\right) \\ y_c = y_n\left(1 + k_1 r^2 + k_2 r^4\right) + p_1\left(r^2 + 2y_n^2\right) + p_2 x_n y_n \end{cases} \tag{9}$$

Note: The third-order radial distortion term ($k_3 r^6$) is omitted since k_3 is negligible and not provided by the calibration tool.

The intrinsic parameter matrix $\boldsymbol{K}$ has been obtained via monocular camera calibration as described previously. The formula for projecting the corrected point $(x_c,\ y_c,\ Z_C)$ from the camera coordinate system to the pixel coordinate system is expressed as follows:

$$P_{uv} = \frac{1}{Z_C} K \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \tag{10}$$

where Z denotes the coordinate of point P along the Z_C-axis in the camera coordinate system. Z_C serves as the depth normalization factor, implementing the normalization strategy of perspective projection with depth scaling to map 3D metric coordinates to 2D pixel coordinates.

By combining Equations (6) and (10), the formula for transforming the point P_L in the LiDAR coordinate system to the pixel coordinate system is expressed as follows:

$$P_{uv} = \frac{1}{Z_C} K T_{C \leftarrow L} P_L \tag{11}$$

To ensure the validity of input data for projection and subsequent fusion, the following strategies are implemented for occlusion filtering and invalid point rejection after applying the aforementioned model:

The calculated pixel coordinates (u, v) are constrained within the physical image boundaries ($0 \leq$ u ≤ 1280, $0 \leq$ v ≤ 720). Points falling outside these boundaries are regarded as invalid projections and discarded. Based on the effective working distance defined by the system (see Section 4.1), point clouds with a depth Z_C (in the camera coordinate system) not within the range [1.0, 2.5] meters are directly discarded to eliminate background interference.

A complete projection transformation model from the LiDAR to the image is thus obtained, as illustrated in Figure 2.

2.2. Time Synchronization

The LiDAR and camera are two completely independent operating systems, which operate at different frequencies in data acquisition, and some even suffer from significant time delays. The core of data acquisition is to ensure the temporal consistency of the data from the two sensors. Therefore, time synchronization between the LiDAR and camera is performed prior to data acquisition to ensure that the two sensors capture information at the same moment [21], thus achieving a one-to-one temporal correspondence between images and point clouds. In this study, software synchronization is implemented for the two sensors based on the Ubuntu 20.04 operating system, and the final time error after

synchronization is within 5 ms. The timestamps before and after time synchronization are illustrated in Figures 3 and 4.

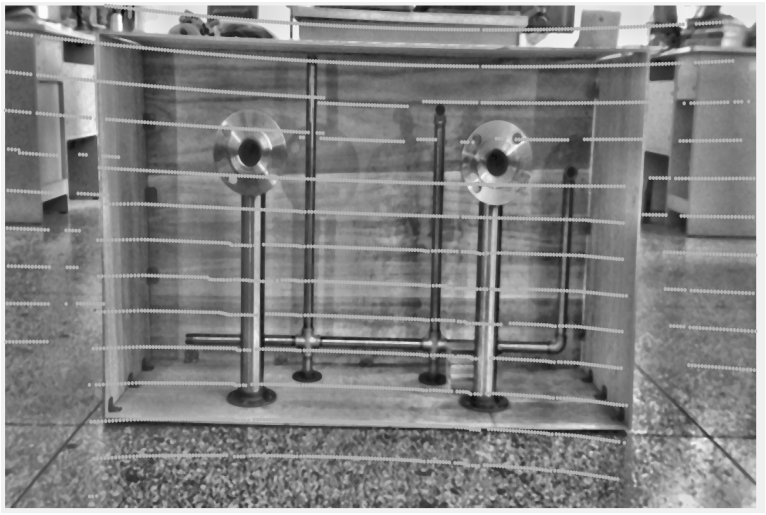

Figure 2. Projection of the LiDAR Point Cloud onto Images.

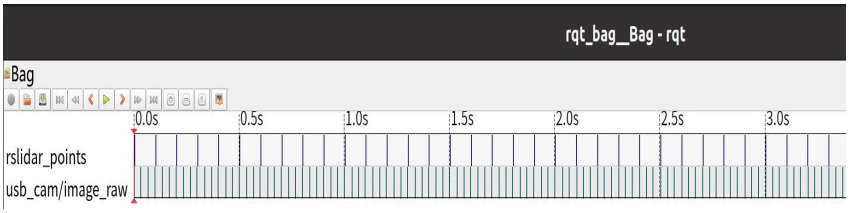

Figure 3. Timestamps Before Time Synchronization.

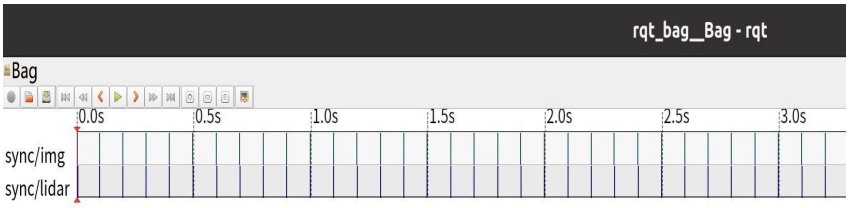

Figure 4. Timestamps After Time Synchronization.

3. Research Methodology

The overall framework of the fusion localization system proposed in this paper is illustrated in Figure 5, which mainly comprises five stages: sensor joint calibration, synchronized data acquisition, visual flange detection and optimization, visual-point cloud fusion localization, and 3D circle center fitting. This framework fully exploits the high-resolution advantage of monocular vision in 2D feature localization and the illumination robustness of LiDAR in depth measurement, thereby jointly addressing the localization challenges under variable illumination conditions.

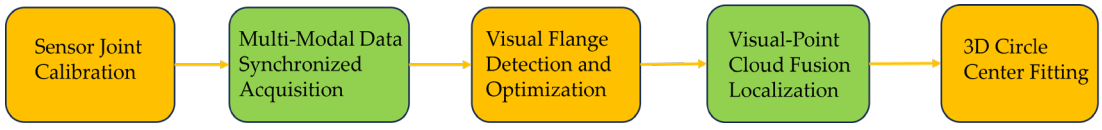

Figure 5. Overall Framework Diagram of the Fusion Localization System.

3.1. Basic Principles of Circle Detection

Prior to circular detection on the image, image preprocessing is first implemented. Given the variations in image quality under different lighting conditions, the preprocessing is designed to enhance edges and suppress interference. Specifically, it sequentially involves image grayscale conversion to eliminate the interference of color information, median filtering to suppress salt-and-pepper noise [22], histogram equalization to improve the contrast of regions with uneven illumination, and Canny edge detection [23] to extract the edge contour of the flange, thereby providing high-quality input for subsequent flange circle detection. In this study, a gradient voting-based circular detection method [24] was adopted, whose core principle consists of two steps: edge gradient calculation and circular feature voting.

First, the Canny edge detector was employed to extract image edges, followed by the calculation of the gradient direction of edge points, which theoretically points to the circle center. The Canny detector suppresses noise and preserves continuous edges via double thresholds, and the gradient vector of an edge point (u, v) is expressed as follows:

$$\nabla I(u, v) = \left[\frac{\partial I}{\partial u}, \frac{\partial I}{\partial v}\right]^T \tag{12}$$

where $I(u, v)$ denotes the grayscale value of the pixel (u, v), and $\frac{\partial I}{\partial u}$ and $\frac{\partial I}{\partial v}$ represent the gradient components in the horizontal and vertical directions, respectively.

Subsequently, within a preset radius range, voting is performed for each edge point along the gradient direction, and the circle center-radius combination with the highest number of votes is defined as the candidate circle. For a candidate circle with radius r, its center (a, b) satisfies the following condition:

$$(u - a)^2 + (v - b)^2 = r^2 \tag{13}$$

where (u, v) denotes the pixel coordinates of the pixel point. This method achieves circle feature localization through the intersection of gradient directions of multiple edges, thus avoiding interference from noise of individual edge points.

3.2. Rationale for Circle Detection Method Selection and Co-Design

In industrial scenarios with variable illumination, circle detection methods must balance accuracy, robustness, computational efficiency, and integrability with prior knowledge. This study selects the gradient-based voting circle detection method [24] as the foundational component for visual perception, primarily based on the following three considerations:

(1) Inherent Robustness to Illumination-Induced Edge Defects: The core challenge in this scenario is extreme illumination (intense glare, low light, water film), causing target edge blurring, fragmentation, or generation of false edges. Classical circle detection methods (e.g., the standard Hough transform) heavily rely on complete and continuous edges, leading to significant performance degradation when edges are broken [25]. In contrast, the gradient voting method operates on the principle that edge point gradient directions converge toward the circle center for voting, demonstrating stronger tolerance to edge discontinuities [24]. This aligns better with the interference patterns caused by illumination in our specific scenario.

(2) Deep Synergy with the Multi-Stage, Strongly Constrained Optimization Pipeline: The core innovation of our localization framework lies not in a single detection module, but in a multi-stage optimization pipeline. The gradient voting method can output all possible candidate circles in the image (including the inner/outer flange circles and background interference circles). This provides an ideal input interface for the

subsequent series of customized optimization steps:
Physical Dimension Constraint Filtering: Quickly eliminates interferences that do not conform to the known flange dimensions from the candidate set.
K-means Spatial Clustering: Utilizes the center coordinates of all candidate circles for clustering to reliably distinguish the left/right flange layout.
Weighted Center Fitting: Comprehensively utilizes the more reliable outer edge circle and the auxiliary central hole circle to achieve robust center estimation.
Employing an end-to-end deep learning approach would make it difficult to transparently and flexibly embed the above optimization logic based on strong physical priors due to its "black-box" nature. Conversely, methods like direct ellipse fitting cannot provide a rich set of candidates for screening and fusion.

(3) Determinism and Debuggability in Engineering Deployment: Given the characteristics of this task—fixed target size and clear geometric rules—a geometric method with high interpretability offers greater engineering advantages over a data-driven deep learning solution. The former possesses the merits of relatively controllable computational load, high output determinism, and logical transparency, facilitating implementation, debugging, and integration with control algorithms in industrial embedded systems. Furthermore, its modular pipeline (edge detection → gradient calculation → voting) allows our proposed illumination-adaptive mechanism (Section 3.3.2)—which dynamically adjusts Canny detection parameters via grayscale evaluation—to be directly and efficiently integrated at the front end, forming a "perception-evaluation-adjustment" closed loop.

Therefore, due to its tolerance to incomplete edges, its characteristic as an ideal front-end module for the multi-constraint optimization pipeline, and its good support for the adaptive mechanism, the gradient voting method becomes the optimal technical choice for connecting visual perception with subsequent fusion localization in this study. The core contribution of this work lies in combining this foundational detector with scenario-customized constraints and fusion strategies, constructing a specialized solution that surpasses generic detectors.

3.3. Optimization of Flange Interface Detection

Combined with the physical characteristics of the flange with fixed dimensions (an outer diameter of 140 mm and a central hole diameter of 45 mm) and the characteristics of scene interference, this study carries out optimization from three dimensions: dimension constraint, illumination adaptability, and K-means clustering.

3.3.1. Geometric Dimension Constraint

The physical dimensions of the flange are the most reliable prior information. By converting the 3D physical dimensions into 2D image pixel dimensions using the millimeter-to-pixel conversion coefficient, the radius range for circle detection is constrained, thereby eliminating background interference circles from the source.

First, mm2pixel is defined as the number of pixels corresponding to a physical length of 1 mm, and its calculation formula is given as follows:

$$\text{mm2pixel} = \frac{D_{\text{pixel}}}{D_{\text{mm}}} \tag{14}$$

where $D_{\text{mm}} = 140$ mm is the standard physical diameter of the flange outer edge in this experiment, and D_{pixel} is the pixel length of this diameter in the image. In this study, the mm2pixel coefficient was determined through calibration at the standard working distance (1.5 m) under a frontal view. Multiple images of the flange with its known physical

diameter (140 mm) were captured. The projected pixel diameter D_{pixel} was obtained via image processing, and the ratio to the physical dimension was calculated. The average value from multiple measurements yielded mm2pixel = 0.91.

The rationale and stability of this coefficient are grounded in the following design premises of our system:

1. Sensor Invariance: The experiment employs a fixed-focus camera (Logitech C270) with no optical zoom. Therefore, its intrinsic matrix remains constant during operation, providing the physical basis for scale stability.
2. Constrained Working Range: As defined in Section 4.1, the system's effective working distance is limited to 1.0–2.5 m. Within this short-to-medium range, the nonlinear scale variation caused by perspective projection is minimal, allowing mm2pixel to be treated as approximately linear.
3. Algorithmic Tolerance Design: As shown in Equations (16) and (17), we introduced a $\pm 20\%$ tolerance band for radius detection. This tolerance band adequately accommodates potential deviations in pixel radius estimation arising from minor distance variations, residual camera calibration errors, and slight fluctuations in the mm2pixel coefficient itself. Consequently, the system is not sensitive to the precise value of mm2pixel.
4. Robustness of Subsequent Processing: Even if some candidate circles are incorrectly included due to scale estimation deviation, the subsequent K-means spatial clustering (Section 3.3.3) and weighted center fitting (Section 3.4) steps can effectively suppress outliers, ensuring the stability of the final localization results.

In summary, mm2pixel = 0.91 is a stable reference value calibrated under the specific hardware and working conditions of our system. The system's performance does not rely on the absolute accuracy of this coefficient but is ensured by the generous geometric constraints and multi-stage decision fusion. If fundamental changes occur in future application scenarios (e.g., camera replacement, significant alteration of working distance), recalibration of this coefficient would be necessary.

Then, according to the inner and outer circle dimensions of the flange, combined with the theoretical pixel radius derived from the conversion coefficient, an error range of $\pm 20\%$ is set, and the constraint formula is given as follows:

$$r_{\text{pixel}} = \frac{D_{\text{mm}}}{2} \times \text{mm2pixel} \tag{15}$$

$$r_{\min} = r_{\text{theo}} \times (1 - \delta) \tag{16}$$

$$r_{\max} = r_{\text{theo}} \times (1 + \delta) \tag{17}$$

where r_{pixel} is the theoretical pixel radius of the flange circle, and $\delta = 0.2$ is the error coefficient.

The theoretical radius detection range of the outer edge circle is calculated to be 51–76 pixels, and that of the central hole is 16–25 pixels. Finally, the candidate circles whose radii fall within the valid ranges are filtered out.

3.3.2. Adaptive Illumination Adjustment

Low illumination causes blurred gradients of the flange interface edges, while high illumination may lead to specular reflections on the flange surface, which in turn form false edges; both of these issues degrade the performance of Canny edge detection. In this study, a closed-loop mechanism consisting of illumination intensity evaluation and dynamic parameter adjustment is adopted to achieve optimized edge extraction under varying illumination conditions.

1. Illumination Evaluation Index and Threshold Setting Basis

The average grayscale value of the grayscale image is taken as the illumination evaluation index and normalized to the range of $[0, 1]$, with the formula given as follows:

$$L = \frac{1}{M \times N} \sum_{u=1}^{M} \sum_{v=1}^{N} \frac{I(u, v)}{255} \tag{18}$$

where $M \times N$ denotes the image size, and $\frac{I(u,v)}{255}$ represents the normalized result of the pixel grayscale value.

In the experiment, the definitions are as follows: $L < 0.3$ indicates a low-illumination scenario, $L > 0.7$ indicates a high-illumination scenario, and the rest are classified as normal illumination.

The threshold division (L < 0.3 for low illumination, L > 0.7 for high illumination, and $0.3 \leq$ L ≤ 0.7 for normal illumination) is determined based on the qualitative analysis of image features and experimental iterative optimization under the four typical illumination scenarios described in Section 4, without the need for additional data collection:

- Low illumination threshold (L < 0.3): In the scenarios of uniform low illumination and low illumination with water film, the overall grayscale of the image is relatively low, the contrast between the target and the background is weakened, and the default parameters of Canny are prone to edge breakage or missed detection, so it is necessary to reduce the threshold to retain weak edges.
- High illumination threshold (L > 0.7): In the scenario of local strong illumination, the specular reflection on the metal surface of the flange will introduce gradient noise and false edges, so it is necessary to increase the threshold and enhance filtering to suppress interference.
- Normal illumination ($0.3 \leq$ L ≤ 0.7): The illumination condition is ideal, the edge features are clear, and the use of compromised parameters can achieve stable detection.

2. Dynamic Parameter Adjustment Strategy and Its Performance Impact

The system automatically matches parameter sets based on the real-time calculated L value, with the core logic: improve detection sensitivity in low-illumination scenarios (where edges are prone to loss) and enhance detection specificity in high-illumination scenarios (where edges are prone to overload). The specific parameter settings are as follows:

- Low illumination (L < 0.3): 2×2 median filtering + Canny thresholds [0.1, 0.2] + circle detection sensitivity 0.92.
- High illumination (L > 0.7): 5×5 median filtering + Canny thresholds [0.25, 0.4] + circle detection sensitivity 0.8.
- Normal illumination ($0.3 \leq$ L ≤ 0.7): 3×3 median filtering + Canny thresholds [0.15, 0.3] + circle detection sensitivity 0.86.

The effectiveness of this mechanism has been verified by the experimental results in Section 4.3. This indicates that the dynamic parameter adjustment strategy effectively mitigates the impact of illumination interference on edge detection, laying a solid foundation for the subsequent improvement of 3D localization accuracy and robustness.

3.3.3. K-Means Clustering Algorithm

The experimental objects in this study include two flanges (left and right) to simulate the left and right flanges of the LNG tank truck loading/unloading port. Given that the spatial distribution of flange images features left-right separation, the K-means clustering algorithm [26] is employed to distinguish the left and right flange circles, thereby deriving the 2D parameters of the target flanges. Taking the center coordinates (u, v) of candidate circles as features, the K-means algorithm is adopted to classify the circles into two cate-

gories corresponding to the left and right flanges. The clustering objective function is to minimize the sum of squared errors within clusters, expressed as follows:

$$J = \sum_{k=1}^{2} \sum_{(u,v) \in C_K} \| (u,v) - \mu_k \|^2 \tag{19}$$

where J denotes the value of the objective function, with a smaller value indicating a better clustering effect; K = 1, 2 represents the number of clustering categories; C_K denotes the geometry of all candidate circles in the k-th category; $(u,v) \in C_K$ indicates traversing the center coordinates of each candidate circle in the k-th category; $\mu_k = \left(\mu_u^k, \mu_v^k\right)$ represents the clustering center of the k-th category, where $\mu_u^k = \frac{1}{|C_K|}\sum_{(u,v)\in C_K} u$ and $\mu_v^k = \frac{1}{|C_K|}\sum_{(u,v)\in C_K} v$, i.e., the average values of the u and v coordinates of all circle centers in this category; $\| (u,v) - \mu_k \|^2$ denotes the square of the Euclidean distance from a single circle center (u,v) to the clustering center μ_k of its affiliated category. The implementation process is shown in Figure 6:

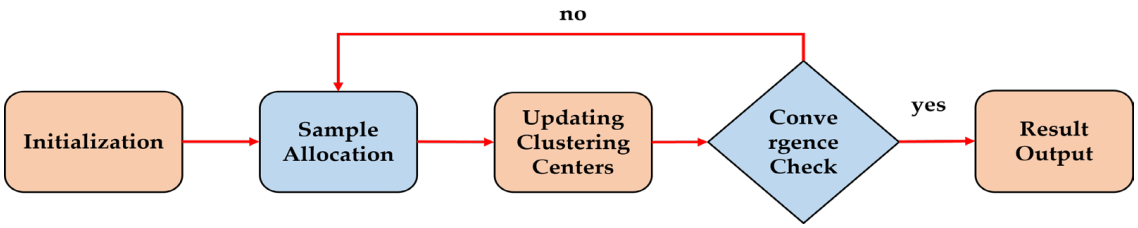

Figure 6. Flow Chart of K-Means Clustering Iterative Optimization.

Among them, initialization involves randomly selecting the centers of two candidate circles as the initial clustering centers μ_1 and μ_2; sample assignment refers to calculating the distance from each candidate circle to μ_1 and μ_2 and assigning it to the category with the closer distance; updating the clustering centers means recalculating the mean centers μ_1 and μ_2 of the two categories. Steps 2 and 3 are repeated until the clustering centers no longer change or the maximum number of iterations is reached, at which point the convergent minimum value of J is output. In this study, the number of clustering repetitions is set to 10 and the maximum number of iterations to 200. After clustering, the left and right flanges are distinguished by the x-coordinates of the clustering centers.

3.4. Weighted Fitting of Flange Circle Centers

In flange detection, a single flange is often detected with two types of candidate circles simultaneously: the outer edge circle and the central hole circle. The outer edge directly reflects the physical contour of the flange and thus has higher positioning reliability, while the central hole serves as an auxiliary feature; there are differences in the detection accuracy between the two. This study proposes a weighted fitting strategy, which strengthens the role of core features through differential weight assignments to achieve robust estimation of the circle center coordinates. The weighted fitting formula is given as follows:

$$\begin{cases} u_0 = \frac{\sum_{i=1}^{n} w_i u_i}{\sum_{i=1}^{n} w_i} \\ v_0 = \frac{\sum_{i=1}^{n} w_i v_i}{\sum_{i=1}^{n} w_i} \end{cases} \tag{20}$$

where u_0, v_0 are the fitted image pixel coordinates of the flange circle center; u_i, v_i are the pixel coordinates of the i-th candidate circle center; w_i is the weight coefficient, which is assigned a value of 1 for the outer edge circle and 0.6 for the central hole circle.

Through the aforementioned methods, the left and right flanges are finally detected and distinguished, and the pixel coordinates of the flange circle centers are fitted, which provides a foundation for subsequent fusion with LiDAR. The results are shown in Figure 7.

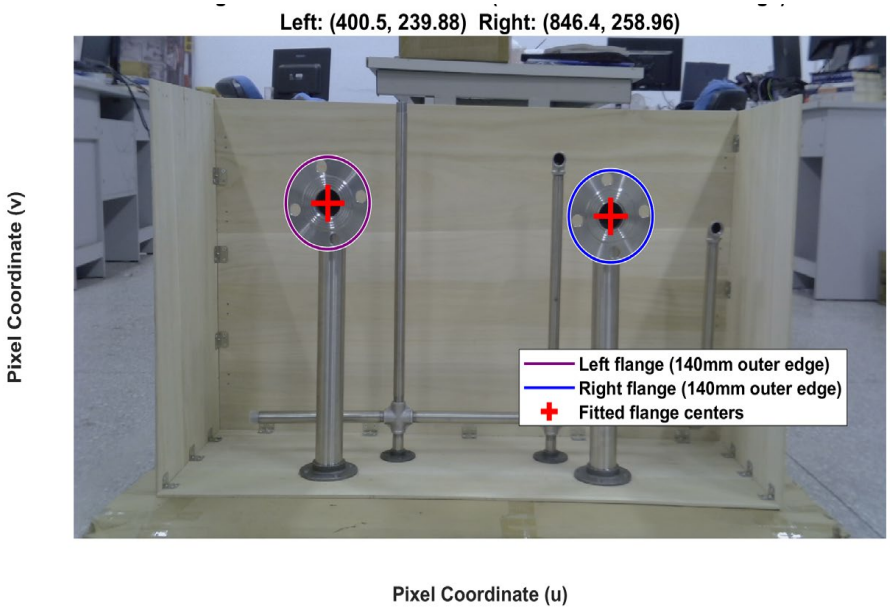

Figure 7. Effect Diagram of Flange Circle Detection.

3.5. Fusion Positioning

In general, industrial sites are subject to various interference factors such as variable illumination and water films, which tend to cause fusion misalignment as well as fitting errors induced by the sparse point clouds of 16-channel LiDAR. To address these issues, this study proposes a method that takes the 2D pixel coordinates of the circle centers obtained via vision as constraints, combines the depth information from LiDAR to achieve accurate extraction of ROI point clouds, and performs circle fitting with the physical radius of the flange as a constraint. This method specifically resolves the fitting deviation caused by sparse point clouds, thus enabling high-precision 3D positioning of the flange circle centers. The proposed method fully leverages the advantages of LiDAR (i.e., immunity to illumination interference in distance measurement) and vision (i.e., high-precision 2D positioning), resulting in a significant improvement in positioning accuracy under complex scenarios.

3.5.1. Point Cloud Preprocessing

First, the removeInvalidPoints function is used to filter out invalid points in the point cloud to ensure data validity. Then, the pcdenoise function is employed to perform statistical outlier detection: by calculating the distance distribution between each point and its neighboring points, noise points are identified and removed. Finally, based on the height distribution characteristics of the point cloud, the 5th percentile of the Z-axis coordinates is calculated, and a buffer of 0.1 m is added as the height threshold; the point cloud with Z-values higher than the threshold is filtered out to separate ground points from non-ground points. Ultimately, only the target point cloud covering the experimental area is retained, which improves the calculation speed.

3.5.2. Point Cloud ROI Extraction

To clarify the basis for setting the core thresholds of ROI extraction, this section first formalizes the filtering thresholds and then verifies their rationality through parameter ablation experiments, ensuring the scientific rigor and scenario adaptability of the threshold selection.

The core of ROI extraction is to accurately screen flange point clouds and exclude background interference through dual constraints of pixel range and depth range. The thresholds are defined as follows:

- Pixel Range Threshold (R_{pixel} = 60): Based on the pixel coordinates (u_0, v_0) of the flange center obtained through visual detection, only LiDAR projected points (u_p, v_p) that satisfy the Euclidean distance constraint from this center are retained as ROI candidates. The distance calculation is shown in Equation (21):

$$d_{2D} = \sqrt{(u_p - u_0)^2 + (v_p - v_0)^2} \tag{21}$$

where d_{2D} is the Euclidean distance in the pixel coordinate system, and the constraint condition is $d_{2D} \leq 60$ pixels. The initial setting of this threshold is derived from the calculation of the flange's pixel size in Section 3.3.1 (the theoretical pixel radius of the flange's outer edge is 51–76 pixels), ensuring coverage of the effective contour of the flange while reserving redundancy for perspective distortion.

- Depth Range Threshold ($Z \in [1.0, 2.5]$ m): Combined with the effective working distance of the system defined in Section 4.1, point clouds with depths in the range of 1.0–2.5 m in the camera coordinate system are filtered. The lower limit of 1.0 m avoids near-field blind spot data of the LiDAR, and the upper limit of 2.5 m ensures a point cloud density, meeting the requirements for subsequent plane and circle fitting.

To verify the optimality of the aforementioned thresholds (especially R_{pixel} = 60), a parameter ablation experiment was conducted under four typical illumination scenarios (basic uniform illumination, local strong illumination, uniform weak illumination, weak illumination with water film). The experimental design is as follows:

- Variable Setting: The pixel range threshold R_{pixel} was set as the only variable, taking values of 30, 40, 50, 60, 70, 80, 90, and 100 pixels (with a step size of 10), while the depth range threshold was fixed at [1.0, 2.5] m.
- Evaluation Metrics: Positioning accuracy (3D error), ROI point cloud quantity, and algorithm success rate (number of valid plane fittings/total number of tests).
- Experimental Process: Each R_{pixel} value was tested repeatedly 10 times, and the statistical average value was taken as the final result. The experimental data and visualization results are shown in Figures 8 and 9.

Experimental Result Analysis:

- Performance Defects of $R_{pixel} < 60$: When R_{pixel} = 30–40, the number of ROI points is $\leq$9.5, leading to complete failure of plane fitting due to insufficient valid points (success rate = 0%); when R_{pixel} = 50, although the success rate reaches 100%, the number of ROI points is only 15, the positioning error (0.0604 m) is higher than that of R_{pixel} = 60, and the error stability is poor (coefficient of variation = 0.50%), making it difficult to meet the fitting requirements under complex illumination.
- Optimality Verification of R_{pixel} = 60: Under this threshold, the number of ROI points stabilizes at 19.5, which not only meets the minimum number of points required for plane fitting but also does not introduce redundant background points; the positioning error reaches the minimum value (0.0592 m) with a coefficient of variation in only 0.34%, showing optimal stability. The ROI point clouds extracted under this threshold

can accurately separate flange targets from background interference such as pipes and brackets, as shown in Figure 10, providing high-quality input for subsequent fitting.

- Performance Degradation of $R_{pixel} > 60$: When $R_{pixel} \geq 70$, although the number of ROI points continues to increase, the expanded pixel range incorporates a large amount of background noise, leading to a significant rise in positioning error (reaching 0.1243 m at 80 pixels, 2.1 times that at 60 pixels). Moreover, the error continues to deteriorate as R_{pixel} increases, failing to meet the accuracy requirements for industrial docking.
- Cross-Scenario Robustness: $R_{pixel} = 60$ maintains optimal accuracy and stability in all illumination scenarios: the error is 0.0586 m under basic uniform illumination, 0.0632 m under local strong illumination, and 0.0592 m under weak illumination with water film, verifying the strong adaptability of this threshold to illumination changes.

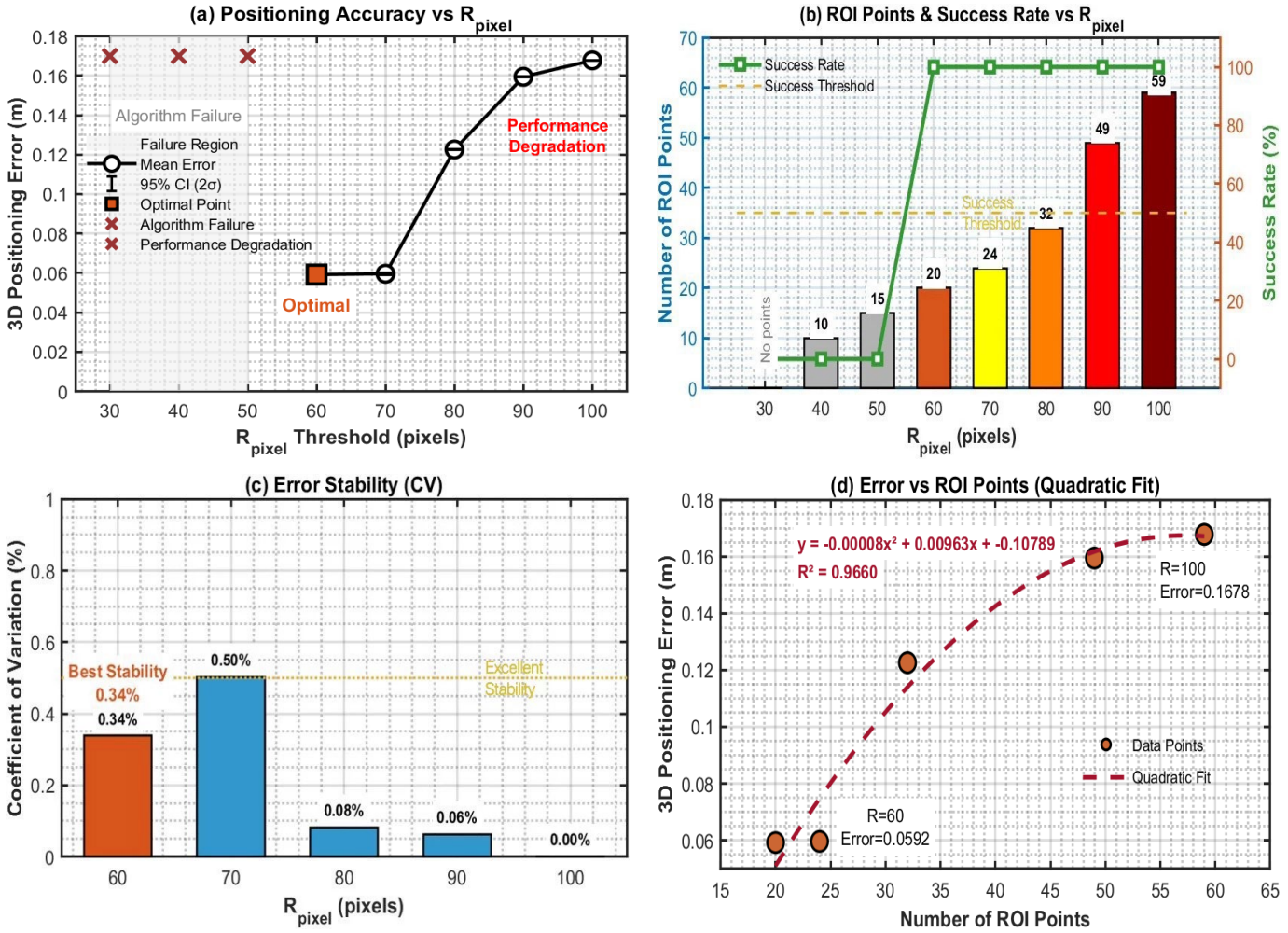

Figure 8. Comprehensive Ablation Results of R_{pixel} Threshold. (**a**) Variation trend of positioning error with R_{pixel}; (**b**) Distribution of ROI points and success rate; (**c**) Error stability analysis based on coefficient of variation; (**d**) Quadratic fitting relationship between ROI points and positioning error.

Based on the comprehensive analysis of the ablation experiment results, the optimal thresholds for ROI filtering are determined as $R_{pixel} = 60$ and $Z \in [1.0, 2.5]$ m. This setting not only ensures the integrity and purity of the flange point cloud through pixel range constraints but also excludes invalid background data through depth range constraints, achieving an optimal balance between positioning accuracy, fitting stability, and scenario adaptability, and laying a foundation for subsequent plane fitting and 3D circle center solving.

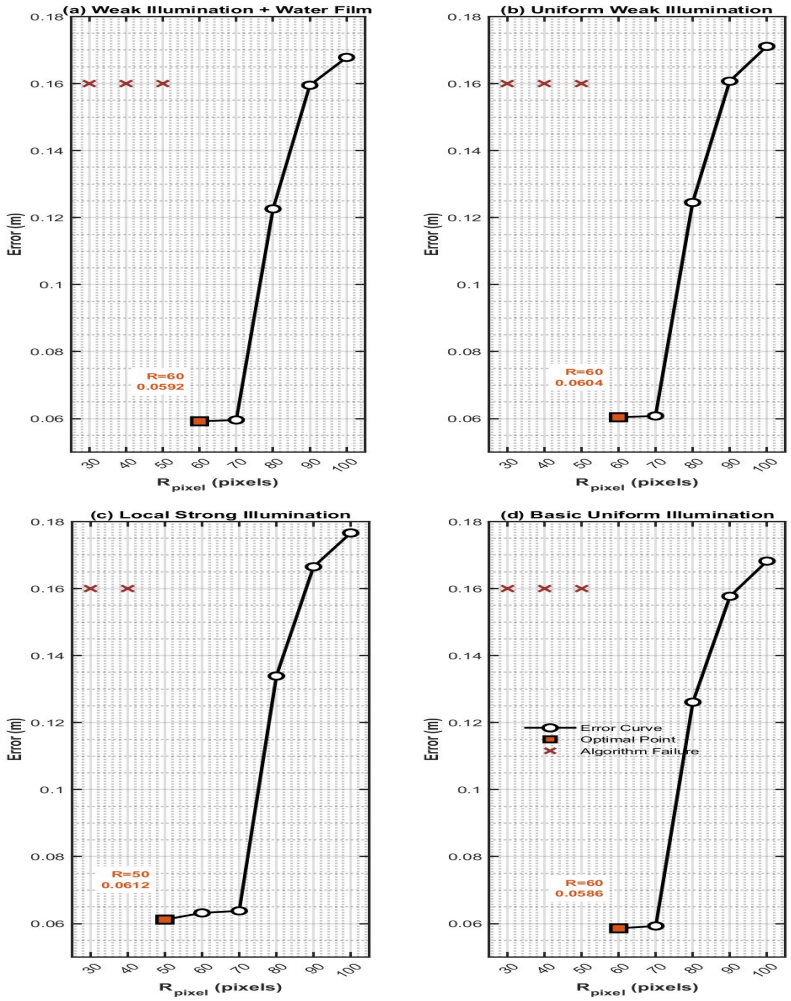

Figure 9. Ablation Results Under Four Illumination Scenarios.

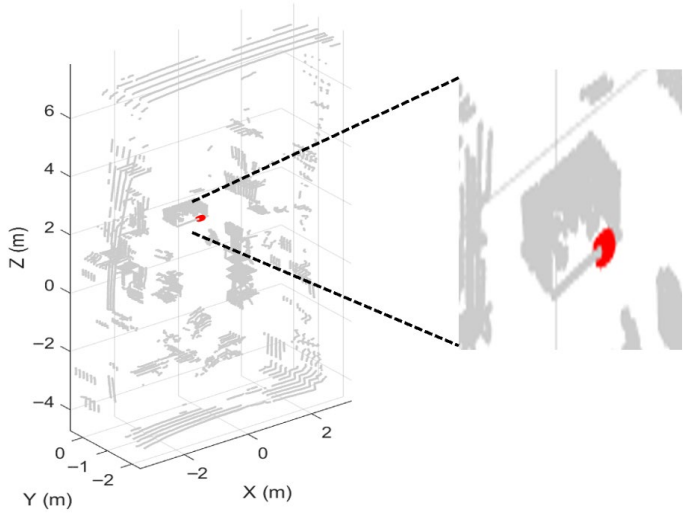

Figure 10. Flange ROI Extraction. The red region indicates the Region of Interest (ROI) of the flange.

3.5.3. Flange Plane Fitting

As a rigid planar structure, the flange undergoes plane fitting to obtain its spatial attitude parameters, which provides a reference for subsequent 3D circle center fitting of

the flange. In this study, the least squares method is adopted for flange plane fitting, and the fitted plane equation is given as follows:

$$aX + bY + cZ + d = 0 \tag{22}$$

where (a, b, c) is the plane normal vector, which reflects the spatial attitude of the plane after unitization; d is the plane intercept, which reflects the position of the plane relative to the coordinate origin. The spatial position of the flange can be fully represented by these four parameters.

After ROI extraction constrained by vision, the point cloud may still contain noise such as ground points. To eliminate residual noise, this study sets a distance threshold of 0.01 m based on flange machining accuracy and measurement errors and only retains valid points whose perpendicular distance to the plane is less than or equal to 0.01 m for fitting. This ensures that the plane parameters can accurately characterize the actual pose of the flange.

Justification of Plane-Fitting Distance Threshold (0.01 m):

The selection of the 0.01 m distance threshold is grounded in three key considerations tied to the experimental setup and system requirements:

LiDAR Measurement Precision: The RoboSense 16-line LiDAR used in this study has a specified ranging accuracy of ±2 cm (Section 4.2). The 0.01 m threshold is approximately half of this value, ensuring that only points with minimal measurement deviation are retained, mitigating the impact of sensor noise without overly restricting valid data.

Point Cloud Density Balance: Within the effective working distance (1.0–2.5 m), the 16-line LiDAR yields an average of 19.5 valid points per flange after ROI extraction [Section 3.5.2]. A threshold larger than 0.01 m would significantly increase residual noise (e.g., ground or pipe interference) as reflected by the 2.1-fold positioning error increase [Section 3.5.2], while a smaller threshold (<0.005 m) would exclude a considerable portion of valid points (reduced from 19.5 to ~16), leading to unstable fitting. The 0.01 m value strikes a balance, maintaining sufficient inliers for robust plane estimation.

Adaptation to flange machining tolerance: The flatness tolerance of the flange is ±0.5 mm [Section 3.5.4], which is much lower than the 0.01 m threshold, avoiding the impact of minor surface defects on the fitting result. Meanwhile, this threshold prevents excessive error propagation to subsequent 3D circle center fitting—critical for sparse point cloud scenarios.

This threshold is further validated by consistent performance across four illumination scenarios, confirming its suitability for maintaining fitting stability while preserving valid data.

3.5.4. Circle Fitting and 3D Circle Center Calculation

Given that the point cloud generated by the 16-channel LiDAR used in this experiment is relatively sparse, a dimensionality reduction method that converts sparse 3D point clouds into dense 2D features is adopted. Combined with constraints derived from the physical dimensions of the flange, the stability of fitting is improved, and 3D coordinate mapping is ultimately achieved.

First, based on Equation (22), the mapping point of the coordinate origin on the plane is selected as the reference point P_0. By setting $X = 0$ and $Y = 0$ and substituting them into the equation, $Z = -\frac{d}{C}$ is obtained; thus, the coordinates of the reference point are $P_0\left(0, 0, -\frac{d}{C}\right)$, which is used as the original datum of the local coordinate system of the plane.

Then, the plane normal vector $(a,\ b,\ c)$ is normalized to obtain the unit normal vector $\boldsymbol{n}$. The cross product of the obtained unit normal vector $\boldsymbol{n}$ and the X-axis unit vector i = (1,0,0) is calculated to yield the initial direction vector within the plane, which is subsequently

normalized to generate the first unit vector $\boldsymbol{u}$. A cross-product operation is then performed again to obtain the second unit vector $\boldsymbol{v}$ that is perpendicular to both the unit normal vector $\boldsymbol{n}$ and the first unit vector $\boldsymbol{u}$. Thus, a right-hand coordinate system where $\boldsymbol{n}$, $\boldsymbol{u}$, and $\boldsymbol{v}$ are orthogonal to each other pairwise is established, and the vectors $\boldsymbol{n}$ and $\boldsymbol{u}$ together form the local Cartesian coordinate system within the plane.

Furthermore, an arbitrary point $P(X, Y, Z)$ is selected from the point cloud after ROI extraction, and the position vector $\boldsymbol{r} = P - P_0$ relative to the reference point P_0 is calculated. The 2D coordinates (s, t) of the position vector $\boldsymbol{r}$ in the local coordinate system of the plane are obtained via vector dot product:

$$s = r \cdot u, \; t = r \cdot v \tag{23}$$

where s denotes the projection length of $\boldsymbol{r}$ along the direction of $\boldsymbol{u}$, and t denotes the projection length along the direction of $\boldsymbol{v}$. Through such transformation, the 3D point cloud contour is accurately mapped to a 2D circular feature within the plane, which reduces the fitting difficulty caused by sparse point clouds.

Given that the physical radius of the actual flange is known and its machining accuracy is within ± 0.5 mm, which is much smaller than the measurement error of the LiDAR, constraints are imposed on the flange radius during the fitting process. Through such constraints, the fitting algorithm preferentially matches circles that conform to the physical dimensions, thus avoiding circle center offset or elliptical distortion, and ensuring the consistency between the fitting results and the flange contour. The standard circle equation is adopted for fitting, which is given as follows:

$$(s - s_0)^2 + (t - t_0)^2 = r^2 \tag{24}$$

where (s_0, t_0) denotes the 2D circle center coordinates in the local coordinate system of the plane, and r denotes the radius of the fitted circle. The fitting results are shown in Figure 11.

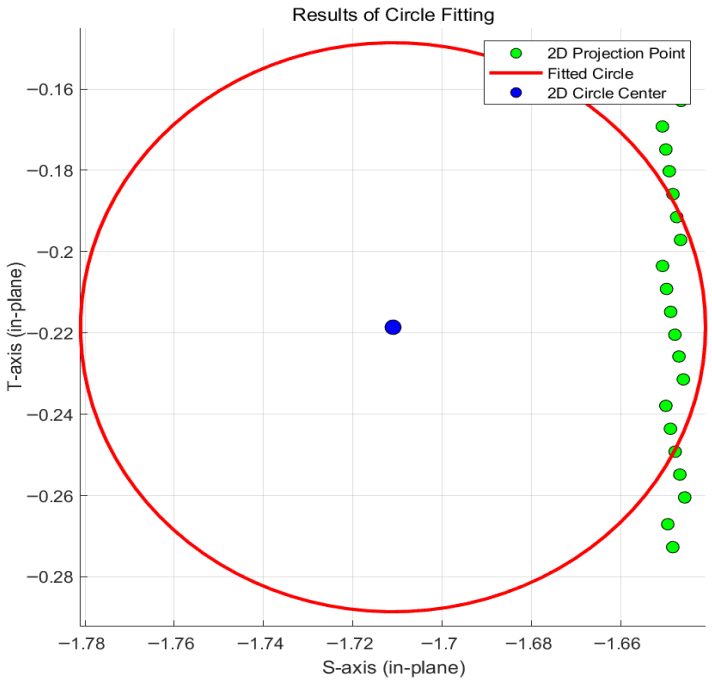

Figure 11. Circle Fitting Result.

Finally, according to the inverse projection equation:

$$(X_0, Y_0, Z_0) = P_0 + s_0 \cdot u + t_0 \cdot v \tag{25}$$

The 2D circle center (s_0, t_0) in the local coordinate system of the plane is inversely mapped to the 3D space, and the 3D circle center coordinates (X_0, Y_0, Z_0) of the flange are ultimately obtained.

3.6. Overall Algorithm Flowchart and Pseudocode

Figure 12 illustrates the complete workflow of the proposed monocular vision-LiDAR fusion localization algorithm. Based on an illumination-adaptive multi-modal data fusion strategy, this algorithm aims to achieve robust 3D localization of LNG flange interfaces under extreme lighting conditions. As shown in the figure, the system first synchronously acquires image and point cloud data, dynamically adjusts detection parameters through illumination evaluation, and then sequentially performs 2D circle center detection, LiDAR ROI extraction, plane fitting, and 3D circle center solving, ultimately outputting high-precision localization results.

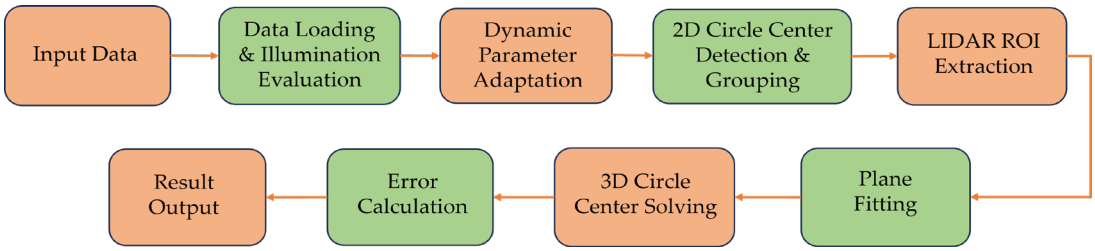

Figure 12. Workflow of the fusion localization algorithm.

For a clear and formal description of the proposed method, detailed pseudocode is provided in Algorithm 1. The inputs to the algorithm include the scene image, the synchronized LiDAR point cloud, the camera intrinsic matrix, the physical dimensions of the flange, and the preset calibration coefficient. The outputs are the 2D pixel coordinates of the centers for both the left and right flanges, the precise 3D coordinates of the right flange center, and the corresponding localization error.

Algorithm 1. Robust 3D Localization of Flange Center via Monocular Vision and LiDAR Fusion

Input:
I: RGB image from the monocular camera.
P_L: Raw LiDAR point cloud ($3 \times$ N matrix) in the LiDAR coordinate system.
K: Camera intrinsic matrix (3×3).
T_C←L: Extrinsic transformation matrix from LiDAR to camera coordinates.
D_outer, D_inner: Known physical diameters of the flange's outer edge and central hole (in meters).
P_true: Ground-truth 3D coordinate of the target (right) flange center for validation.
preset_mm2pixel: Baseline millimeter-to-pixel conversion coefficient.

Output:
C_pixel_L, C_pixel_R: 2D pixel coordinates of the left and right flange centers.
C_3d_est: Estimated 3D coordinate of the target (right) flange center.
error_3d: Localization error (C_3d_est − P_true).

Algorithm 1. *Cont.*

```
// --- Step 1: Data Loading & Illumination Assessment ---
I_gray ← RGB2GRAY(I)
L ← mean(I_gray) / 255                          // Evaluate illumination level L ∈ [0, 1]
P_C ← TRANSFORM_POINT_CLOUD(P_L, T_C←L) // Transform LiDAR points to camera coord.

// --- Step 2: Illumination-Adaptive Parameter Calibration (Innovation 1) ---
params_edge, params_circle ← ADAPT_PARAMETERS(L) // Dynamically set Canny thresholds, filter size, etc.
mm2pixel ← preset_mm2pixel // Use the pre-calibrated baseline coefficient (stable under fixed setup)
R_pixel_outer ← (D_outer / 2) * mm2pixel
R_pixel_inner ← (D_inner / 2) * mm2pixel
radius_range ← [R_pixel_inner * 0.8, R_pixel_outer * 1.2]

// --- Step 3: Vision-Based 2D Flange Detection & Grouping (Innovation 2) ---
I_enh ← ENHANCE_IMAGE(I_gray, params_edge)       // Median filtering & contrast stretching
edges ← CANNY_DETECT(I_enh, params_edge)
candidate_circles ← GRADIENT_VOTE_CIRCLE_DETECT(edges, radius_range, params_circle)
candidate_circles ← FILTER_BY_RADIUS(candidate_circles, R_pixel_outer, R_pixel_inner, tolerance = 0.2)
(circles_left, circles_right) ← KMEANS_CLUSTER_2D(candidate_circles.center) // Separate left/right flanges
C_pixel_R ← WEIGHTED_CENTER_FIT(circles_right, w_outer = 1.0, w_inner = 0.6) // Weighted fitting
C_pixel_L ← WEIGHTED_CENTER_FIT(circles_left, w_outer = 1.0, w_inner = 0.6)

// --- Step 4: LiDAR ROI Extraction & Plane Fitting ---
roi_indices ← []
for each point p in P_C do
    p_pixel ← PROJECT_TO_IMAGE(p, K)             // Project 3D point to 2D pixel
    if DIST(p_pixel, C_pixel_R) ≤ 60 and p.z ∈ [1.0, 2.5] then
        roi_indices.append(index(p))
    end if
end for
P_roi ← P_C[roi_indices]                         // Extract ROI point cloud
[a, b, c, d], inlier_mask ← RANSAC_PLANE_FIT(P_roi, dist_thresh = 0.01)
P_inliers ← P_roi[inlier_mask]                   // Points belonging to the flange plane

// --- Step 5: 3D Circle Fitting on Plane (Innovation 3) ---
n ← [a, b, c] / norm([a, b, c])                  // Unit normal vector of the plane
P0 ← [0, 0, −d/c]                                // A point on the plane (assuming c ≠ 0)
u ← CROSS(n, [1, 0, 0]); u ← u / norm(u)         // Define in-plane x-axis
v ← CROSS(n, u)                                  // Define in-plane y-axis
proj_2d ← []
for each point q in P_inliers do
    vec ← q − P0
    s ← DOT(vec, u)
    t ← DOT(vec, v)
    proj_2d.append([s, t])
end for
[s0, t0, r_fit] ← CIRCLE_FIT_WITH_CONSTRAINT(proj_2d, fixed_radius = D_outer / 2)
C_3d_est ← P0 + s0 * u + t0 * v                  // Back-project to 3D camera coordinates
```

Algorithm 1. *Cont.*

44: C_3d_est.z ← CLAMP(C_3d_est.z, min(P_roi.z), max(P_roi.z)) // Stability enhancement

45: // --- Step 6: Error Calculation & Output ---
46: error_3d ← C_3d_est − P_true
47: return C_pixel_L, C_pixel_R, C_3d_est, error_3d

The core of the algorithm achieves robust localization through six steps: First, data loading and illumination assessment are performed (Step 1), followed by dynamic adjustment of detection parameters based on illumination intensity (Step 2). Subsequently, 2D circle center detection and left-right flange grouping are accomplished using monocular vision (Step 3). Next, the LiDAR ROI point cloud is extracted using the 2D center as a constraint, and plane fitting is applied (Step 4). Based on this, the 3D circle center coordinates are solved via plane projection and radius-constrained circle fitting, and the localization error is calculated (Step 5). Finally, the localization results are output and visualized (Step 6). This pipeline fully integrates the high-resolution advantage of vision with the robustness of LiDAR in depth measurement, effectively overcoming the limitations of a single sensor under extreme illumination.

4. Experiments and Results

4.1. Application Assumptions and Boundary Conditions

Prior to detailing the experimental hardware and metrics, we first explicitly define the key application assumptions and boundary conditions of the proposed fusion localization system. This clarification is essential for understanding the operational scope and inherent limitations of our method.

1. Working Distance Range: The system is designed for short- to medium-range operation. The effective working distance is between 1.0 and 2.5 m from the sensor suite to the target flange. This range is chosen because it (a) covers the typical standoff distance for LNG tanker loading/unloading interfaces, and (b) ensures sufficient point cloud density from the 16-line LiDAR for stable plane and circle fitting. Performance may degrade significantly outside this range.
2. Viewing Angle Constraint: The sensors are assumed to have a direct frontal or near-frontal view of the flange plane. The system's performance, particularly the visual circle detection module, is optimized for this perspective and may degrade under extreme oblique angles (e.g., >45°) due to significant perspective distortion and ellipsoidal projection of the circular features.
3. Target Object Specification: The method is tailored for standard circular flanges with known, fixed physical dimensions. In this study, a flange with an outer diameter of 140 mm and an inner hole diameter of 45 mm is used. The algorithm's geometric constraint filtering and scale determination critically depend on this prior dimensional knowledge.
4. Expected Performance Bounds: The system is engineered to achieve centimeter-level positioning accuracy suitable for automated guidance. The quantitative evaluation of accuracy under the tested illumination and distance conditions is presented later in this section.
5. Limitations on Target Geometry Variation: The current implementation has a strong dependency on the prior knowledge of the target's circular geometry and fixed dimensions. Significant deviations—such as encountering a flange of a different size, a non-circular interface, or a heavily occluded target that breaks the circular

contour—are beyond the scope of the current algorithm and would require recalibration or structural modifications to the detection pipeline.

Following these clarifications, the specifics of the experimental setup are described below.

4.2. Experimental Setup and Implementation Details

The experimental environment consisted of flanges, connecting pipes, and a chassis bracket; two sets of flanges were configured to simulate the height and quantity of the rear flanges of an LNG tank truck. Additional pipes were installed behind the two flange sets to replicate the structural complexity of the rear section of an LNG tank truck. The data acquisition system was composed of a Logitech C270 USB monocular camera with a resolution of 720 $\times$ 1280 and a frame rate of 30 fps, and a RoboSense 16-Line LiDAR adopting the TOF ranging method with a frame rate of 10 fps, a field of view of $\pm 15^{\circ}$, and a ranging accuracy of ± 2 cm. The configuration of the computing platform used in the experiment was as follows: the processor was a 12th Gen Intel(R) Core (TM) i5-12500H (2.50 GHz), the memory capacity was 16 G, and the operating systems were Windows 11 and Ubuntu 20.04 with ROS Noetic.

The true 3D coordinates of the target flange center were measured in the Camera Coordinate System (CCS). The measurement was performed using a Deli DL331040C lithium-ion laser (Deli Group Co., Ltd., Ningbo, Zhejiang, China) rangefinder with a nominal accuracy of ± 1.5 mm/m. The CCS is defined as follows: the origin is located at the camera's optical center, the Z-axis points along the optical axis toward the measured flange, the X-axis is horizontally to the right (consistent with the image coordinate system), and the Y-axis is vertically downward (consistent with the image coordinate system).

To ensure the precise alignment between the rangefinder measurements and the CCS, the following calibration and measurement procedures were implemented:

Coordinate System Alignment: A fixed reference point was placed directly in front of the camera, at the same height as the optical axis. First, the initial coordinates of this reference point in the CCS were accurately measured using the rangefinder near the camera installation position, serving as the spatial benchmark for all subsequent relative measurements.

Flange Center Coordinate Measurement: Based on the aforementioned spatial benchmark, the indirect 3D coordinate measurement of the target flange center was conducted by measuring the offset distances of the flange center relative to the spatial benchmark in the X (horizontal), Y (vertical), and Z (depth) directions, respectively:

- Horizontal offset (X-direction): Measure the linear distance between the flange center and the reference point in the left-right direction.
- Vertical offset (Y-direction): Measure the linear distance between the flange center and the reference point in the up-down direction.
- Depth offset (Z-direction): Along the direction of the camera lens pointing to the flange, directly measure the linear distance from the measurement position to the plane where the flange center is located.

Data Acquisition and Processing: Each directional offset was independently measured three times under stable support conditions. After eliminating gross errors, the average value was taken as the optimal offset estimation for that direction. Finally, these average offsets were combined with the coordinates of the spatial benchmark to calculate the final true 3D coordinates (X_{true}, Y_{true}, Z_{true}) of the flange center in the CCS.

The combined uncertainty of this measurement method mainly stems from the rangefinder's nominal accuracy (± 1.5 mm/m) and minor deviations from manual alignment. At the typical working distance of approximately 1.5 m, the estimated combined standard uncertainty of each axis is less than ± 3 mm. This level of uncertainty is an order of magnitude smaller than the centimeter-level positioning errors evaluated in our experi-

ments. Therefore, the obtained ground-truth coordinates can be reliably used as a precise reference for validating the accuracy of the proposed fusion-based localization method.

Subsequently, a control group and an experimental group were set up. The experiment of the experimental group was conducted indoors; the curtains and lights in the room were turned off to eliminate the interference of natural light and indoor lighting, and artificial light sources were used to illuminate the entire experimental area. The specific grouping of the experiment and the controlled variables are shown in Table 1.

Table 1. Experimental Grouping Details.

Grouping Type	Experimental Scenarios	Illumination Settings
Experimental Group 1	Uniform Illumination	One 40 W fill light was placed at a distance of 1.5 m from the left and right sides at 45°, respectively.
Experimental Group 2	Local Strong Illumination	A 40 W fill light was placed at 45° to the left at a distance of 1.5 m, and another 40 W fill light was placed at 45° to the right at a distance of 0.5 m.
Experimental Group 3	Uniform Low Illumination	One 10 W fill light was placed at a distance of 1.5 m at 45° to the left and right, respectively.
Experimental Group 4	Uniform Low Illumination with Water Film	One 10 W fill light was placed at a distance of 1.5 m at 45° to the left and right, respectively.

Monocular camera positioning was adopted as the control group in the experiments. This positioning scheme is constructed based on the pinhole camera imaging model, and its core lies in capturing images of the flange scene via a monocular camera and achieving 3D positioning by combining the geometric constraint of the known actual radius of the flange. The specific process is as follows: first, the Zhang Zhengyou camera calibration method is employed to complete camera calibration and obtain the intrinsic parameter matrix; then, edge detection and ellipse fitting are performed on the preprocessed images to extract characteristic parameters such as the central pixel coordinates of the projected ellipse of the flange; subsequently, relying solely on the geometric constraint of the known flange radius R, the depth information is calculated by combining the mapping relationship between the spatial circle and the projected ellipse, and then the 3D coordinates of the circle center in the world coordinate system are derived through inverse projection model calculation and coordinate transformation [27]. The monocular camera positioning scheme features advantages including simple structure, low hardware cost, and no requirement for multi-sensor synchronization and calibration, thus being widely applied in low-cost industrial scenarios. However, it is limited by the scale ambiguity problem inherent to monocular imaging. Even with the introduction of the flange radius constraint, its positioning accuracy remains susceptible to the errors of image feature extraction, camera calibration accuracy, and changes in shooting perspective, and cumulative errors are prone to occur in depth information calculation [28]. This is precisely the core comparative significance of setting up the LiDAR-camera fusion positioning experimental group.

A control group was set up for each experimental scenario to serve as a comparison with the experimental group adopting the LiDAR-camera fusion positioning method. Given the difference between the monocular positioning error and the fusion positioning error, the error reduction rate was introduced, defined as:

$$Error\ Reduction\ Rate = \frac{Monocular\ Positioning\ Error\ -\ Fusion\ Positioning\ Error}{Monocular\ Positioning\ Error} \quad (26)$$

Root Mean Square Error (RMSE), Mean Absolute Error (MAE) [29], and error distribution characteristics were selected as evaluation metrics for a comprehensive assessment.

The root means square error (RMSE) and mean absolute error (MAE) were adopted to quantify the positioning errors of each experimental group in the three individual dimensions of X, Y, and Z, with the specific formulas given as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{27}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{28}$$

where $\hat{y}_i$ denotes the measured value, y_i denotes the true value, and n denotes the number of data samples (for a single group of experiments, n = 20; for the comprehensive experiments of four groups, n = 80).

The 3D means absolute error (3D MAE) and 3D root mean square error (3D RMSE) were employed to quantify the overall 3D positioning errors of each individual experimental group as well as the combined overall 3D positioning errors across all groups, with the specific formulas given as follows:

$$3D\ RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{3D})^2} \tag{29}$$

$$3D\ MAE = \frac{1}{n}\sum_{i=1}^{n}|d_{3D}| \tag{30}$$

where d_{3D} denotes the 3D distance error of each sample datum, and n denotes the number of samples (for a single group of experiments, n = 20; for the comprehensive experiments of four groups, n = 80).

$$d_{3D} = \sqrt{(X_{meas} - X_{true})^2 + (Y_{meas} - Y_{true})^2 + (Z_{meas} - Z_{true})^2} \tag{31}$$

where X_{meas}, Y_{meas}, Z_{meas} denote the measured values of the X, Y, and Z axes, and X_{true}, Y_{true}, Z_{true} denote the true values of the X, Y, and Z axes. Furthermore, box plots were used to visualize the median, interquartile range, and outliers of the error distribution, which comprehensively reflect the positioning performance of the two methods in each dimension and the overall 3D space. For each group of experiments, data were collected 20 times to obtain statistically stable sample data, and the final results were visualized using Matlab 2024a.

4.3. Experimental Results and Analysis

4.3.1. Localization Accuracy Analysis Under Variable Illumination

According to the measurement results, the statistical results of localization errors under different lighting scenarios are presented in Table 2. The error reduction rates (RMSE/MAE) for each scenario are detailed in Table 3, and the overall optimization effects combining the four scenarios are shown in Table 4.

As can be seen from Table 2, the monocular vision positioning method achieves acceptable accuracy in the X and Y axes under uniform illumination, but the Z-axis error reaches 0.2192 m, which stems from its inherent limitation of inferring 3D depth from 2D images. Under conditions of local strong illumination and low illumination, the errors of the X and Y axes increase to the range of 0.0450–0.0555 m due to edge detection offset caused by illumination interference, indicating significant scenario dependence. In contrast, the fusion method maintains stable mean errors in the X, Y, and Z axes within the ranges of 0.0343–0.0385 m, 0.0334–0.0345 m and 0.0266–0.0327 m, respectively, across the four scenarios, with all standard deviations less than 0.0005 m. Combined with the overall optimization effects presented in Table 4, the MAE values of the $X/Y/Z$ axes are reduced

by 33.08%, 30.57% and 75.91%, respectively, the overall 3D MAE is reduced by 61.69%, and the *Z*-axis RMSE is reduced by 79.88%. This method precisely compensates for the shortcomings of monocular vision in depth measurement and achieves dual improvements in both accuracy and stability.

Table 2. Statistical Results of Localization Errors Under Different Lighting Scenarios (n = 20, Unit: m).

Experimental Scenarios	Localization Method	ΔX (Mean ± 95% CI)	ΔY (Mean ± 95%CI)	ΔZ (Mean ± 95%CI)
Uniform Illumination	Monocular Vision	0.0371 ± (0.0354,0.0388)	0.0679 ± (0.0664,0.0695)	0.2192 ± (0.2179,0.2206)
	Fusion Method	0.0243 ± (0.0230,0.0256)	0.0343 ± (0.0330,0.0356)	0.0278 ± (0.0264,0.0293)
Uniform Low Illumination	Monocular Vision	0.0555 ± (0.0527,0.0582)	0.0451 ± (0.0429,0.0473)	0.0390 ± (0.0340,0.0439)
	Fusion Method	0.0355 ± (0.0340,0.0370)	0.0334 ± (0.0322,0.0346)	0.0305 ± (0.0291,0.0319)
Local Strong Illumination	Monocular Vision	0.0450 ± (0.0422,0.0478)	0.0525 ± (0.0501,0.0548)	0.0674 ± (0.0520,0.0827)
	Fusion Method	0.0360 ± (0.0348,0.0372)	0.0336 ± (0.0322,0.0350)	0.0327 ± (0.0313,0.0340)
Uniform Low Illumination with Water Film	Monocular Vision	0.0548 ± (0.0523,0.0573)	0.0468 ± (0.0449,0.0487)	0.0296 ± (0.0188,0.0403)
	Fusion Method	0.0385 ± (0.0365,0.0404)	0.0345 ± (0.0330,0.0361)	0.0266 ± (0.0249,0.0282)

Note: (1) Each experiment was repeated 20 times (n = 20), resulting in a total of 80 valid data sets across 4 scenarios, with raw data available in the Supplementary Materials; (2) Data are presented in the format of "mean ± 95% confidence interval" (calculated based on t-distribution, df = 19), reflecting statistical reliability.

Table 3. Error Reduction Rate of Each Scenario.

Experimental Scenarios	Dimension	MAE Error Reduction Rate (%)	RMSE Error Reduction Rate (%)
Uniform Illumination	X	35.12	17.48
	Y	41.83	43.05
	Z	88.69	88.70
Local Strong Illumination	X	65.79	65.79
	Y	11.03	11.03
	Z	79.45	79.45
Uniform Low Illumination	X	38.67	38.81
	Y	28.89	29.05
	Z	48.92	49.08
Uniform Low Illumination with Water Film	X	19.17	19.17
	Y	37.22	37.35
	Z	35.33	36.32

Table 4. Overall Optimization Effect of the Four Scenarios.

Dimension	MAE Error Reduction Rate (%)	RMSE Error Reduction Rate (%)
X	33.08	33.65
Y	30.57	32.71
Z	75.91	79.88
3D Overall	61.69	64.79

Under the uniform illumination scenario, the core limitation of monocular vision is concentrated in the depth dimension. As can be seen from the box plot in Figure 13, its *Z*-axis error shows no fluctuation, with the median as high as 0.21942 m, which is a typical systematic deviation directly related to the inherent limitation of monocular vision in inferring 3D depth from 2D images. In contrast, the box plot of the *Z*-axis error of the fusion method is highly compact, with the median reduced to 0.02798 m. According to Table 4, the MAE is reduced by 88.69%, which completely addresses the problem of uncontrolled depth measurement.

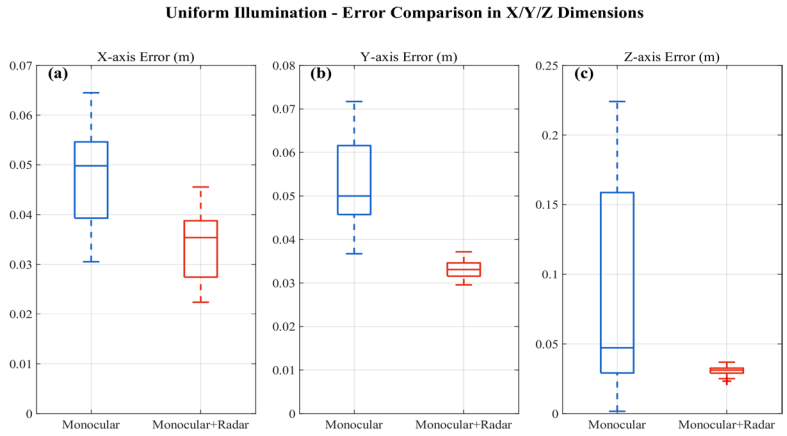

Figure 13. Basic Uniform Illumination. (**a**) X-axis Error; (**b**) Y-axis Error; (**c**) Z-axis Error.

Under the scenarios of uniform low illumination and low illumination coupled with water film, the low illumination condition weakens edge contrast. As shown in the box plot in Figure 14, the median of the *X*-axis error of monocular vision increases to 0.0555 m with a significant increase in dispersion. The error box plot of the fusion method is concentrated around 0.0355 m, with the *X*-axis MAE reduced by 38.67%. When low illumination is combined with water film, the box plot in Figure 15 indicates that the fluctuation of the *Z*-axis error of monocular vision intensifies. However, the fusion method filters valid points through plane fitting, so its error box plot remains compact, with the *X*-axis and *Z*-axis MAE reduced by 19.17% and 35.33%, respectively, thus adapting to the compound interference scenarios.

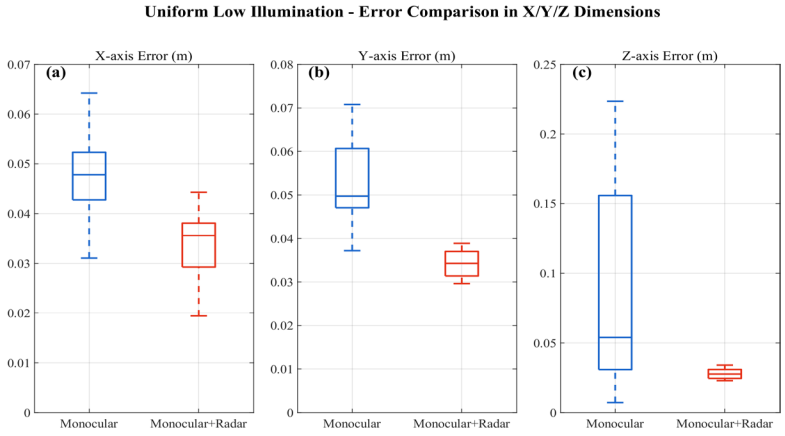

Figure 14. Uniform Low Illumination. (**a**) X-axis Error; (**b**) Y-axis Error; (**c**) Z-axis Error.

Under the local strong illumination scenario, the flange reflection caused by strong illumination will introduce false edge interference, which is clearly reflected in the box plot in Figure 16: the interquartile range of the Z-axis error of monocular vision reaches 0.0828 m with obvious outliers, indicating that the illumination interference leads to severe error fluctuation. By virtue of adaptive parameters for strong illumination and point cloud ROI constraints, the fusion method shows no discrete points in its error box plot, with the interquartile range only 0.0016 m. Table 4 shows that the overall 3D MAE is reduced by 61.69%, demonstrating significant robustness advantages.

The comprehensive error comparison across the four scenarios can be intuitively observed from the box plots in Figures 17–19. The error box plots of the fusion method exhibit the characteristics of low median, narrow interquartile range and no outliers in all three dimensions: the *X*-axis error is concentrated in the range of 0.0343–0.0385 m, which can avoid

lateral collision; the upper limit of the Z-axis error is reduced from 0.2192 m to 0.0327 m, meeting the requirements of centimeter-level docking. Combined with Table 4, the overall 3D RMSE is reduced by 64.79%, which verifies its consistency advantage in batch operations and fully meets the engineering requirements for the automated loading and unloading of LNG tank trucks.

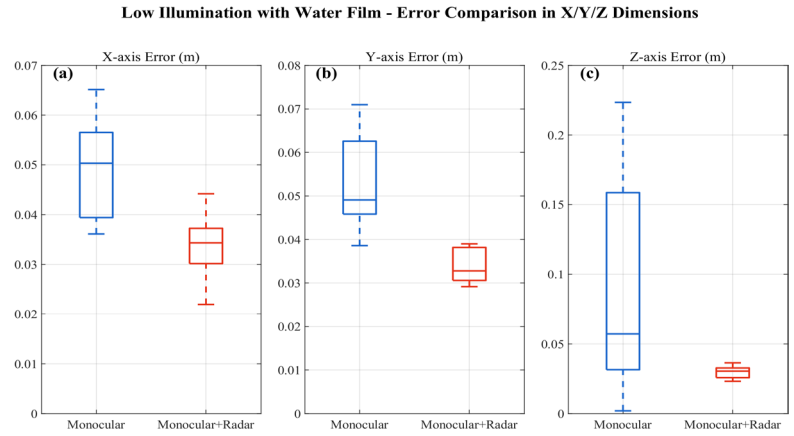

Figure 15. Uniform Low Illumination with Water Film. (**a**) X-axis Error; (**b**) Y-axis Error; (**c**) Z-axis Error.

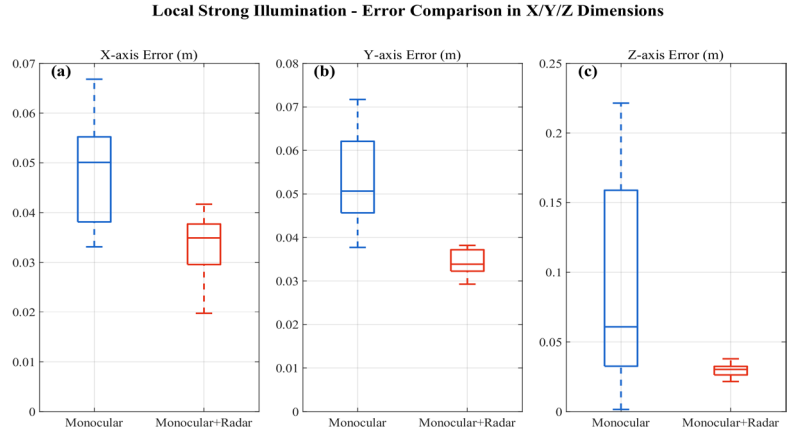

Figure 16. Local Strong Illumination. (**a**) X-axis Error; (**b**) Y-axis Error; (**c**) Z-axis Error.

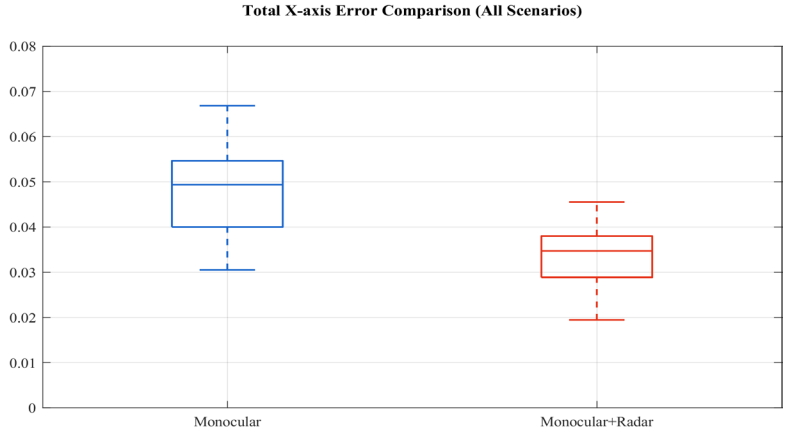

Figure 17. X-axis total error comparison (monocular vs. fusion).

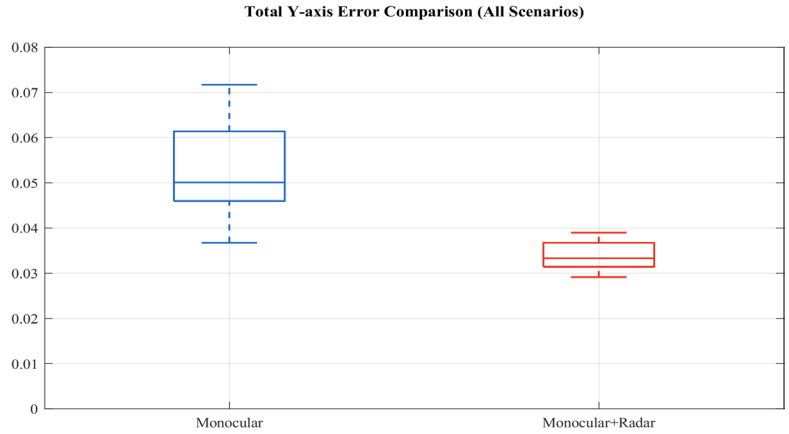

Figure 18. Y-axis total error comparison (monocular vs. fusion).

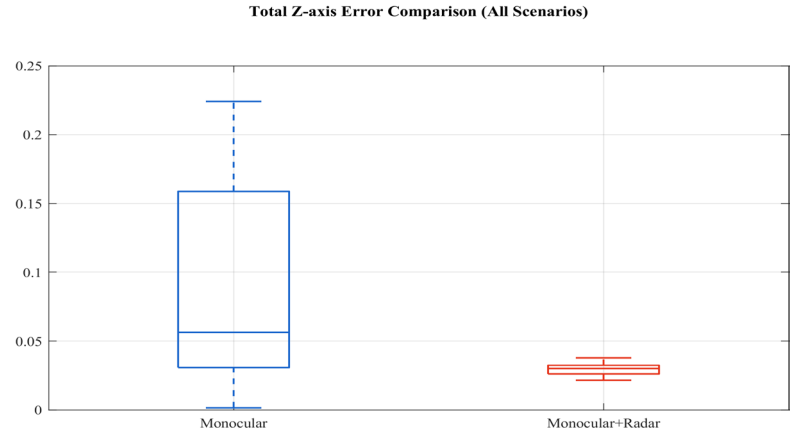

Figure 19. Z-axis total error comparison (monocular vs. fusion).

To evaluate the computational efficiency of the proposed fusion positioning system, a time-consuming test of the algorithm was conducted on the aforementioned experimental platform (12th Gen Intel i5-12500H, 16 GB RAM). A set of typical data from the scenario of low illumination coupled with water film was adopted, and the test was run continuously for 5 times with the average value calculated; the results are presented in Table 5.

Table 5. Time Consumption Analysis of the Fusion Positioning System.

Module	Time Consumption (ms)	Proportion (%)	Theoretical Frame Rate (fps)
Visual Circle Detection	278.2	95.96	3.6
Point Cloud Fusion	11.7	4.04	85.5
Full System	289.9	100	3.45

Although the frame rate of the current system has not yet met the industrial real-time requirements (typically $\geq$10 fps), the time consumption is mainly concentrated in the visual circle detection module (accounting for 95.96%), while the point cloud fusion module only accounts for 4.04%. This indicates that the fusion computation itself is not a bottleneck; the core performance limitation originates from the visual-perception front-end. This bottleneck is well-defined and readily optimizable. The delay of the vision module mainly stems from the sequential execution of a series of image-processing operations (median filtering, Canny edge detection, gradient calculation, and Hough-style voting) in the interpreted MATLAB environment. Importantly, these operations are standard operators in computer vision and possess mature parallelized implementations on underlying hardware.

Based on the above analysis, we propose the following concrete and feasible optimization roadmap, which provides a reliable engineering path for the system to achieve the target frame rate of $\geq$10 fps:

Algorithm porting and GPU acceleration: The visual detection algorithm will be migrated from MATLAB to an efficient C++14/OpenCV4.8.0 framework. Specifically, Canny edge detection and gradient calculation can be massively accelerated by exploiting GPU parallelism. The Hough circle detection can be implemented using optimized, hardware-aware voting algorithms.

Embedded-platform deployment: The final optimized system will be deployed on a high-performance embedded computing platform, such as an NVIDIA Jetson AGX Orin. This platform integrates powerful GPU and AI computing resources, is specifically designed for real-time robotic perception tasks, and can provide the hardware support required for the above optimizations.

The effectiveness of the aforementioned optimizations has been widely validated in the field of embedded vision. A real-time stereo-matching study conducted on the NVIDIA Jetson AGX platform [30] demonstrated that even a computational pipeline involving dense binocular matching can achieve a processing time of approximately 19.6 ms per frame (equivalent to 51 fps). In comparison, the core components of our visual module (namely Canny edge detection and Hough circle detection) are classical operators with lower computational density. Therefore, by adopting a comparable technical pathway (C++/GPU acceleration), it is reasonable to expect comparable or even superior performance gains. Conservatively estimated, the processing time of the optimized visual module could be reduced from 278 ms to 30–50 ms. Consequently, the overall system frame rate would stably exceed 10 fps, fully meeting the requirements for industrial real-time control.

In summary, the fusion method proposed in this study not only demonstrates significant advantages in accuracy and robustness, but its architecture also lays a solid foundation for achieving real-time operation through well-defined engineering optimizations, providing a technical path for automated LNG tanker loading/unloading that combines innovation with engineering feasibility.

4.3.2. Robustness Analysis

To address the requirement of quantifying robustness metrics (failure rate, sensitivity to reflections, and stability under noise or missing data), this subsection conducts a comprehensive analysis based on existing experimental data. To comprehensively evaluate the reliability of the proposed fusion localization system under extreme working conditions, we conduct a quantitative analysis from two core dimensions: operational reliability and output stability. The robustness assessment is based on the complete experimental process and data across the four illumination scenarios.

The core quantified robustness metrics of the proposed fusion method, in comparison with the monocular vision baseline, are summarized in Table 6. This comparison highlights the fusion method's superior performance in terms of operational reliability, stability under challenging lighting, and tolerance to sparse data.

Table 6. Comparison of robustness metrics between the proposed fusion method and the monocular vision baseline.

Robustness Metric	Proposed Fusion Method	Monocular Vision (Baseline)
Failure Rate (All scenarios, n = 80)	0% (100% task completion rate)	Detection failures observed, particularly under Local Strong Illumination
Reflection Sensitivity (Local Strong Illumination)	Z-axis error IQR: 0.0016 m; No outliers.	Z-axis error highly dispersed; IQR: 0.0828 m with obvious outliers.
Noise Stability (Low Illumination + Water Film)	All-axis errors remain at cm-level; Z-axis: 0.0266 m.	Error fluctuations intensified; performance degradation in some dimensions.
Data Sparsity Tolerance (Avg. 19.5 points per flange)	100% fitting success rate; Z-axis std. < 0.0005 m.	No explicit fitting success rate reported; stability inadequate under sparse data.

1. Operational Reliability

The primary aspect of a system's robustness lies in its ability to consistently provide valid outputs under various conditions. During the experimental process of this study, the monocular vision method occasionally experienced complete detection failure under extreme illumination (especially under the "local strong illumination" scenario), resulting in no localization output. To ensure a fair quantitative accuracy comparison (MAE/RMSE) in Section 4.3.1, that subsection only analyzed valid data where both methods succeeded in outputting results.

However, this very difference in "data availability" is itself a critical robustness metric. Under identical testing cycles, the proposed fusion method never experienced detection failure, successfully completing the entire pipeline from perception to localization in all attempts, achieving a 100% task completion rate. In contrast, the monocular method had to discard some data due to detection failures. This directly proves that the fusion method possesses a significant advantage in environmental adaptability to extreme illumination and task reliability for continuous operation.

2. Output Stability

Given that the system can produce outputs, the quality and consistency of these results constitute another key robustness indicator. We evaluate this from the following three aspects:

- Sensitivity to Reflections: The "local strong illumination" scenario is designed to simulate specular reflection interference from the metal flange surface. As shown in Figure 16, under this scenario, the Z-axis error box plot for the fusion method shows an interquartile range of only 0.0016 m with no outliers, whereas the error distribution for the monocular method is extremely dispersed (interquartile range = 0.0828 m) with obvious outliers. This indicates that the fusion method, through LiDAR point cloud ROI constraints and plane fitting, effectively suppresses false image edge interference introduced by reflections, exhibiting low sensitivity to reflections.
- Robustness to Image Noise: The "uniform low illumination" and "uniform low illumination with water film" scenarios introduce significant image noise (low contrast, blurred details). As shown in Table 2, under these two adverse conditions, the mean localization errors of the fusion method in the X, Y, and Z axes remain within the same order of magnitude (centimeter-level) as those under the ideal "uniform illumination" scenario, with no order-of-magnitude degradation. For instance, the Z-axis error, which is most sensitive to depth information, is 0.0278 m under "uniform illumination" and 0.0266 m under "uniform low illumination with water film", demonstrating stable performance. This proves the strong robustness of our method against image quality degradation caused by uneven illumination and water film.
- Stability with Sparse Point Clouds (Data Sparsity): The system employs a 16-line LiDAR, which inherently produces sparse point clouds (considered as structural data sparsity). As described in Section 3.5.2, the average number of valid point clouds per flange after ROI extraction is only 19.5. Under this sparse condition, the fusion method maintains a 100% success rate for plane and circle center fitting, and the standard deviation of the Z-axis error across all scenarios is less than 0.0005 m, indicating highly stable fitting accuracy. This confirms the robustness of the adopted "plane fitting—2D projection—radius-constrained fitting" pipeline to sparse point clouds.

In summary, the proposed fusion method demonstrates excellent performance in both operational reliability and output stability. These quantified metrics collectively confirm their high robustness and practical value when dealing with complex illumination variations, reflection interference, and limited data conditions in real-world industrial scenarios.

5. Discussion

To contextualize our proposed fusion method within the broader research landscape, we explicitly compare it with two closely related studies that also address robustness in robotic or industrial localization under challenging conditions.

5.1. Comparison with Reference [31]

Tahiri et al. [31] proposed a Chaos-Artemisinin Optimizer (CAO) for PID tuning in a 3-DOF robotic manipulator, focusing on trajectory tracking under dynamic control optimization. Their work integrates chaotic maps into a metaheuristic algorithm to enhance exploration and avoid local minima, achieving superior tracking accuracy compared to PSO, GWO, and SMA.

- Methodology differences: While their approach is algorithm-centric (optimization-based control), ours is sensor-fusion-centric (LiDAR-vision fusion for geometric localization). They address control parameter tuning; we address perception and 3D localization under illumination variation.
- Robustness: Both studies emphasize robustness—theirs against controller convergence issues, ours against illumination extremes (glare, low light, water film).
- Evaluation: They use IAE (Integral Absolute Error) for tracking performance; we use MAE/RMSE in 3D Euclidean space for localization accuracy, complemented by scenario-specific robustness tests.

5.2. Comparison with Reference [32]

This study evaluates metaheuristic algorithms (HGS, SMA, ECO) for PID control of a differential mobile robot, focusing on trajectory tracking in simulated environments.

- Methodology differences: Their work is also control-oriented, using ITAE as the cost function for PID optimization. Our method, in contrast, is perception-oriented, fusing LiDAR and vision to achieve illumination-robust 3D localization without relying on control-loop tuning.
- Robustness: They test algorithm performance in tracking stability; we test sensor fusion performance under four extreme illumination scenarios, including water film interference—a common industrial challenge not addressed in their work.
- Evaluation protocols: They assess overshoot, settling time, and ITAE; we provide comprehensive 3D error statistics (MAE, RMSE) with confidence intervals and ablation studies on key parameters (e.g., R_{pixel}).

5.3. Synthesis and Distinct Contributions of Our Work

While the aforementioned studies advance optimization-based control in robotics, our work fills a gap in illumination-robust perception for industrial docking. Specifically:

- We propose an illumination-adaptive fusion framework that dynamically adjusts detection parameters based on grayscale evaluation, a feature absent in both compared works.
- We introduce multi-constraint flange detection (physical dimensions, K-means clustering, weighted fitting) tailored for fixed-size industrial targets, whereas the compared works focus on generic control optimization.
- Our evaluation includes real-world challenging scenarios (water film, glare) and a parameter sensitivity analysis, providing deeper insights into system robustness under operational variability.

In summary, while Tahiri et al. (both works) [31,32] contribute significantly to metaheuristic-enhanced control, our study advances multi-sensor fusion for robust in-

dustrial localization, offering a complementary perspective that addresses perception challenges in automated LNG tanker operations.

5.4. Comparison with Recent Baseline Methods

To further validate the advancement of the proposed fusion method, this section conducts a targeted comparison with three representative baseline methods from recent literature: a LiDAR-only approach (reference [33]), a monocular-only approach (reference [34]), and a hybrid multi-sensor approach (reference [35]). The comparative analysis is detailed as follows:

1. Comparison with the LiDAR-only baseline: The baseline method is a pure LiDAR incremental odometry approach, which demonstrates excellent performance in large-scale navigation and motion estimation, and is adaptable to different LiDAR types. However, it lacks the capability for fine-grained feature localization assisted by vision: first, it cannot achieve precise fitting for specific industrial targets such as flanges, resulting in relatively large localization errors; second, although LiDAR itself is unaffected by illumination, the method does not incorporate visual assistance for fine 2D positioning, making it difficult to meet the centimeter-level docking requirements in industrial scenarios. The method proposed in this paper combines high-resolution 2D feature detection from vision with depth compensation from LiDAR, achieving a balance between "large-scale stable positioning and small-scale precise docking", which better aligns with the needs of industrial loading and unloading operations.
2. Comparison with the Monocular-only baseline: The baseline method realizes 3D measurement in micro-baseline scenarios through single-camera rotation coupled with a neural network, achieving high accuracy (0.0864 mm absolute error) in static, small-range measurements. Nevertheless, it exhibits notable limitations: first, its performance depends on the training data distribution of the neural network, leading to weak generalization for untrained targets such as flanges; second, under low illumination or reflective conditions, the error in 2D feature point extraction increases significantly, causing degradation in 3D mapping accuracy. The proposed method in this study compensates for the depth ambiguity inherent in monocular vision by leveraging LiDAR's depth measurement advantage. Furthermore, through a multi-constraint fitting strategy, it enhances stability and maintains concentrated error distribution even under challenging conditions such as weak light coupled with water film.
3. Comparison with the Hybrid baseline: The baseline method achieves fused localization through the registration of LiDAR and a depth camera. While it meets the basic accuracy requirements for human–robot collaboration, it suffers from two major limitations: first, it does not account for the impact of illumination variations on depth camera data, rendering it unsuitable for extreme lighting conditions in LNG loading/unloading scenarios; second, it relies on high-density point clouds from a 64-line LiDAR, resulting in high hardware costs, and lacks a constraint mechanism designed for fixed-size industrial targets like flanges, leading to insufficient localization stability. In contrast, the proposed method employs a more cost-effective 16-line LiDAR fused with vision. By incorporating an illumination-adaptive framework and physical dimension constraints, it not only reduces hardware costs but also further lowers the 3D localization error to the centimeter level, with a 75.91% reduction in Z-axis depth error, better satisfying the precision requirements of industrial docking.

In summary, compared with the three aforementioned baseline methods, the core innovation of the proposed approach lies in its targeted optimization for the specific scenario of "industrial flange localization under variable illumination", achieving a triple enhancement in accuracy, robustness, and scenario adaptability. Specifically, it reduces

hardware costs and improves illumination adaptability compared to LiDAR-only methods; it resolves depth measurement ambiguity and illumination sensitivity issues compared to monocular-only methods; and it strengthens fine-grained localization capability for industrial targets compared to hybrid methods. Consequently, it provides a more targeted technical solution for the automated loading and unloading of LNG tankers.

6. Conclusions

Aiming at the positioning challenge of the loading/unloading port of LNG tank trucks under variable illumination conditions, this study proposes a robust positioning method that fuses monocular vision with LiDAR. Through multi-modal data synchronization, fused flange detection, and 3D circle center fitting, the system achieves stable and high-precision positioning under various complex illumination conditions, including uniform illumination, local strong illumination, uniform low illumination, and low illumination coupled with water film. Experimental results based on 20 samples per illumination scenario (80 valid data sets in total) demonstrate that: the proposed fusion method reduces the mean absolute error (MAE) and root mean square error (RMSE) in the Z-axis direction by 75.91% and 79.88%, respectively, which effectively overcomes the systematic deviation of monocular vision in depth estimation and significantly enhances the depth estimation capability; in the X, Y, and Z axes, the MAE is reduced by 33.08%, 30.57%, and 75.91%, respectively, the overall 3D error (3D RMSE) is reduced by 64.79%, and the 3D positioning accuracy is comprehensively improved; even under extreme scenarios such as local strong illumination and low illumination coupled with water film, the fusion method still maintains stable performance characterized by concentrated errors and no outliers, with the statistical reliability verified by the expanded sample size.

Although the proposed method exhibits excellent performance in terms of accuracy and robustness, it still has the following limitations: the current method requires prior knowledge of the physical dimensions of the flange. In practical applications, if the flange specifications change, re-calibration or an adaptive scale estimation mechanism must be introduced, indicating a dependency on the prior dimensions; under the condition of low-channel LiDAR, excessively sparse point clouds may affect the stability of plane fitting, thereby reducing the accuracy of circle center fitting; the current prototype system is implemented based on MATLAB with a theoretical frame rate of 3.45 fps, which has not yet met the industrial real-time requirements. The visual detection module is the main time-consuming component, and the real-time performance needs to be further improved. To address the above issues, future research work can be carried out in the following directions: To address more complex field interferences (e.g., severe oil stains, irregular structural occlusions), future work could explore introducing lightweight deep learning models (such as attention-enhanced CNNs) to assist in feature extraction or preliminary screening. However, their design objective should be to enhance the generalization capability of the existing geometric framework, not to replace the multi-constraint optimization process that relies on strong physical priors. To tackle the current real-time bottleneck of the system (where visual detection dominates the processing time), algorithm code optimization and porting the vision module to a GPU platform for parallel computing could be pursued to significantly improve the processing frame rate, meeting more stringent industrial real-time control requirements. Finally, embedded system deployment and long-term operational validation in real LNG loading/unloading sites are essential to promote the ultimate industrialization and practical application of this technology.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app16021128/s1.

Author Contributions: Conceptualization, M.L. and H.Z.; methodology, H.Z.; software, Y.Z.; validation, J.Z. and K.Z.; investigation, H.Z.; resources, H.Z.; data curation, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z.; project administration, M.L.; funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Lianyungang Municipal Science and Technology Bureau, grant number CG2417.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, M.Q.; Wang, J.C.; Zhang, H.; Zhang, Y.M.; Zhu, J.Q.; Zhu, K. The Evolution and Development Trends of LNG Loading and Unloading Arms. *Appl. Sci.* **2025**, *15*, 4316. [CrossRef]
2. Gao, Y.; Wang, B.; Hu, Y.D.; Gao, Y.J.; Hu, A.L. Development of China's Natural Gas: Review 2023 and Outlook 2024. *Nat. Gas Ind.* **2024**, *44*, 166–177.
3. Liu, Y.; Yang, L.; Jing, Y.X. Research and Analysis on the Technical Status of LNG Handling System. *Petrochem. Chem. Equip.* **2020**, *23*, 44–48.
4. Zu, Y.; Liu, M.Q.; Song, L.M.; Wang, J.C.; Li, X.B. Research on Flange Recognition and Positioning Algorithm of LNG Tanker Based on Binocular Vision Algorithm. *Mech. Eng.* **2025**, *5*, 54–58.
5. Dai, Z.X.; Yin, T.; Han, C.J.; Ma, K.D. Design and Sealing Performance Analysis of Low Temperature Fast Connection Device for Liquid Natural Gas Tanker. *Lubr. Eng.* **2025**, *50*, 172–178.
6. Liu, K.; Song, X.W.; Hu, X.J.; Zhang, X.D.; Li, X. A Three-claw Quick Connection Device Suitable for LNG Tanker. *Mech. Electr. Eng. Technol.* **2022**, *51*, 127–130.
7. Li, K.Q. Research on 3D Object Detection Method Based on Point Cloud and Image Data Fusion. Master's Thesis, Yantai University, Yantai, China, 2025.
8. Ahin, Y.S.; Park, Y.S.; Kim, A. Direct Visual SLAM Using Sparse Depth for Camera-LiDAR System. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018.
9. Xu, Y.; Ou, Y.; Xu, T. SLAM of Robot based on the Fusion of Vision and LIDAR. In Proceedings of the 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS), Shenzhen, China, 25–27 October 2018.
10. Xiao, L.; Chen, H.; Li, Y.; Liu, Y. Visual laser-SLAM in large-scale indoor environments. In Proceedings of the 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), Qingdao, China, 3–7 December 2016.
11. Seo, Y.; Chou, C.C. A Tight Coupling of Vision-Lidar Measurements for an Effective Odometry. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019.
12. Zhang, J.; Kaess, M.; Singh, S. Real-time depth enhanced Monocular Odometry. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014.
13. Lupton, T.; Sukkarieh, S. Visual-Inertial-Aided Navigation for High-Dynamic Motion in Built Environments Without Initial Conditions. *IEEE Trans. Robot.* **2012**, *28*, 61–76. [CrossRef]
14. Oh, T.; Lee, D.; Kim, H.; Myung, H. Graph Structure-Based Simultaneous Localization and Mapping Using a Hybrid Method of 2D Laser Scan and Monocular Camera Image in Environments with Laser Scan Ambiguity. *Sensors* **2015**, *15*, 15830–15852. [CrossRef]
15. Xiang, R.; Feng, W.W.; Song, S.L.; Zhang, H. Automated Docking System for LNG Loading Arm Based on Machine Vision and Multi-Sensor Fusion. *Appl. Sci.* **2025**, *15*, 2264. [CrossRef]
16. Qian, Z.D. Elliptical Visual Marker Detection and Its Application in Visual Localization. Master's Thesis, University of Chinese Academy of Sciences, Beijing, China, 2022.
17. Lv, T.Z.; Zhang, Y.Z.; Lu, C.; Zhu, J.J.; Wu, S. Targetless Intrinsics and Extrinsic Calibration of Multiple Lidars and Cameras with IMU using Continuous-Time Estimation. *arXiv* **2025**, arXiv:2501.02821. [CrossRef]
18. Yao, R.H.; Wang, H.G.; Guo, Y.A.; Xie, Z.Z. Robust real-time moving object detection on water surface: A LiDAR feature matching approach for maritime reliability enhancement. *Ocean Eng.* **2026**, *346*, 123860. [CrossRef]
19. Zhai, S.X.; Bai, Y.L.; Zhang, L.M.; Hu, Y.F.; Ren, S.N. Monocular Camera and LiDAR Point Cloud Fusion for Power Operating Tool Detection. *Mach. Build. Autom.* **2025**, *54*, 306–310+314.

20. Liu, Y.; Li, T.F. Research of the Important of Zhang's Camera Calibration Method. *Opt. Tech.* **2014**, *6*, 565–570.
21. Cao, L.; Gu, X.Y.; Zhu, H.Y.; Yuan, B.X.; Yang, H.Y. Automatic Calibration for Roadside Millimeter Wave Radar-Camera Fusion in Complex Traffic Scenarios. *Laser Optoelectron. Prog.* **2025**, *62*, 101–111.
22. Zhang, L. Research on Image Recognition and Filtering Algorithm of Salt and Pepper Noise. Master's Thesis, Shenyang University of Technology, Shenyang, China, 2023.
23. Li, Y.X. A new Edge Detection Method for Noisy Image Based on Discrete Fractional Wavelet Transform and Improved Canny Algorithm. *Expert Syst. Appl.* **2026**, *298*, 129668.
24. Miao, S.J.; Li, X.; Zhao, D.J.; Sun, Y. Synchronous Parallel Circle Detection Method Based on Hough Gradient. *Electron. Meas. Technol.* **2025**, *48*, 156–165.
25. Yun, O.; Deng, H.G.; Liu, Y.; Zhang, Z.Y.; Lan, X. An Anti-Noise Fast Circle Detection Method Using Five-Quadrant Segmentation. *Sensors* **2023**, *23*, 2732.
26. Fan, Y.Y.; Tian, D.P.; Xu, Q.H.; Sun, J.; Xu, Q.; Shi, Z.Z. Particle Swarm Optimization Based on K-means Clustering and Adaptive Dual-Groups Strategy. *Swarm Evol. Comput.* **2016**, *100*, 102226. [CrossRef]
27. Huang, S.W.; Guo, K.Y.; Song, X.Y.; Han, F.; Sun, S.J.; Song, H.S. Multi-target 3D Visual Grounding Method Based on Monocular Images. *J. Comput. Appl.* **2025**, 1–11. [CrossRef]
28. Zheng, T.X.; Jiang, M.Z.; Feng, M.C. Vision Based Target Recognition and Location for Picking Robot: A Review. *Chin. J. Sci. Instrum.* **2021**, *42*, 28–51.
29. Hodson, T.O. Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [CrossRef]
30. Chang, Q.; Xu, X.; Zha, A.; Meng, J.E.; Sun, T.Q.; Li, Y. TinyStereo: A Tiny Coarse-to-Fine Framework for Vision-Based Depth Estimation on Embedded GPUs. *IEEE Trans. Syst. Man Cybern. Syst.* **2024**, *54*, 5196–5208. [CrossRef]
31. Tahiri, M.; Kmich, M.; Bencherqui, A.; Sayyouri, M.; Khafaga, D.S.; Aldakheel, E.A. Modeling and PID tuning of a 3-DOF robotic manipulator using a novel Chaos-Artemisinin Optimizer. *Expert Syst. Appl.* **2026**, *297*, 129455. [CrossRef]
32. Tahiri, H.; Mchichou, I.; Ouabdou, M.; Sayyour, M. Advanced Control Strategies for Mobile Robots Using Artificial Intelligence. In Proceedings of the 2025 7th Global Power, Energy and Communication Conference (GPECOM), Bochum, Germany, 11–13 June 2025.
33. Fang, K.; Song, R.; Ho, W.H. Inc-DLOM: Incremental Direct LiDAR Odometry and Mapping. *IEEE Access* **2025**, *13*, 6527–6538. [CrossRef]
34. Chen, Q.Y.; Sui, G.R. Study on 3D Measurement based on Single Camera and Neural Network. *Opt. Tech.* **2022**, *48*, 214–222.
35. Wang, Z.K.; Li, P.C.; Zhang, Q.; Zhu, L.H.; Tian, W. A LiDAR-depth camera information fusion method for human robot collaboration environment. *Inf. Fusion* **2025**, *114*, 102717. [CrossRef]