

Binocular Vision Localization Method for Roadheader Based on Point-Line Features in Coal Mines

Abstract

Accurate and robust localization of roadheaders is essential for intelligent control in underground coal mining, where visual perception systems face significant challenges such as low illumination, abrupt lighting changes, dust interference, and complex backgrounds. This paper proposes a stereo vision-based localization method for roadheaders that integrates point and line features to enhance pose estimation accuracy and robustness under complex lighting conditions. To address image degradation caused by the coexistence of low-light and overexposed regions, an adaptive multi-region image enhancement approach is developed. The method dynamically segments the image into high-, medium-, and low-illumination regions based on average brightness and standard deviation, applying region-specific enhancement strategies such as gamma correction, CLAHE, and contrast adjustment. A gradient-based weighted fusion mechanism and bilateral filtering are introduced to improve image quality and feature extraction stability. Line features are extracted using EDLines and represented via Plücker coordinates. A joint point-line optimization model with a dynamic weighting scheme is constructed to improve localization robustness. Experiments on public datasets including EuRoc MAV and KITTI demonstrate that the proposed method achieves an APE of 0.10 m and RMSE of 0.13 m, outperforming existing methods such as PL-SLAM and PL-EDLines-SLAM. Further validation in a dust- and glare-simulated underground environment yields an APE of 0.15 m and RMSE of 0.17 m, confirming the method's environmental adaptability and practical applicability. This study provides a reliable visual localization solution for underground roadheaders, supporting downstream tasks such as path planning, tracking control, and autonomous operation.

Keywords: Stereo vision localization; Point-line feature integration; Underground coal mine; Image enhancement under complex lighting; Roadheader localization

1. Introduction

In coal mine roadway tunneling, high-precision and robust positioning technology for tunnel equipment is critical to ensuring operational safety and efficiency. Compared to other tunneling equipment, boom-type roadheader pose substantially greater localization challenges due to three key factors: frequent dynamic pose variations, geometrically complex roadway configurations, and operationally constrained visual environments featuring low illumination, high particulate density, and recurrent occlusions. Improving the positioning accuracy of the roadheader and simultaneously overcoming the adaptability to the complex working conditions in underground coal mines have become the key to solving the positioning problem of the roadheader at present.

In recent years, vision-based localization has gradually emerged as an effective approach to address the aforementioned challenges. Traditional positioning methods, such as Ultra-Wideband (UWB)-based systems[1], have improved localization accuracy to some extent. However, their stability remains limited due to susceptibility to multipath effects and noise interference. Although the fusion

of inertial navigation and LiDAR perception[2] can compensate for inertial drift, its performance degrades significantly under harsh conditions such as heavy dust or strong light reflections. Time Difference of Arrival (TDOA)-based UWB positioning systems[3] can achieve centimeter-level accuracy at short ranges; however, their localization error increases markedly in long-distance tunnel environments. Although multi-sensor fusion[4] have improved overall system performance, they impose high demands on sensor synchronization and environmental adaptability, thereby increasing the complexity and difficulty of system deployment. Compared to conventional methods, visual positioning systems exhibit enhanced environmental robustness, with distinct advantages in GPS-deprived underground mining operations. A LiDAR-based GPS-denied navigation system[5] tailored for room and pillar coal mines can achieve robust autonomous localization using only relative environmental positioning, without relying on global coordinates. Yang et al. [6] enhanced positioning accuracy in low-illumination and dusty environments through laser beam projection and a non-uniform blur modeling approach. Du et al. [7] developed a cross-laser-based attitude detection system that achieved high measurement accuracy, yet exhibited

insufficient stability in complex operating environments. Wu et al.'s CNN-based monocular end-to-end pose estimation method [9] significantly improves generalization capability; however, it still faces challenges in feature-impoverted environments or under severe occlusion conditions. The binocular vision system [10] shows certain advantages in complex backgrounds, but the measurement accuracy is still limited at a long distance. The monocular visual localization method [11] proposed by Yu et al., which combines feature matching with RANSAC, has achieved relatively optimal 3D reconstruction results. However, the matching accuracy and stability of depth estimation are still limited in environments with sparse features or poor image quality. Cui and Wang [12,13] proposed a vision-inertial system that enhances stability in low-texture environments through filtering optimization. In addition, the introduction of deep learning has brought further improvements to visual localization systems. The multi-scale variational autoencoder[14] proposed by Wu et al. has enhanced the accuracy of pose estimation. Zhang et al. [15] constructed high-precision maps by integrating RGB-D SLAM. However, in dynamic mine environments, visual information is susceptible to interference, leading to insufficient robustness of the system. Although existing methods have demonstrated satisfactory performance in laboratory settings or certain real-world scenarios, they still face multiple challenges in the complex underground coal mine environment. These challenges include image quality degradation caused by low illumination, high dust levels, and strong reflections, the complexity of deploying systems that rely on specific target markers, and decreased localization accuracy in long-distance, winding tunnels or dynamic scenes. Moreover, the need for substantial human involvement in processes such as feature extraction and sensor calibration affects the system's level of automation and practicality.

With the in-depth application of visual SLAM in the field of robotics, point-line fusion SLAM has gradually emerged as an effective approach to address the limitations of traditional point-feature-based SLAM. In low-texture, dynamically changing, and unstable lighting conditions, pure point-feature methods often perform poorly due to sparse features and unstable matching. In recent years, numerous studies have focused on the joint mapping and optimization of point and line features. The visual SLAM technology based on point-line feature fusion effectively mitigates occlusion failures and error accumulation issues, demonstrating both high sensitivity and real-time processing capabilities that are particularly suitable for dynamic complex environments in tunnels[16,17]. The RPL-SLAM which combines RGB-D and DBM matching[18,19], still exhibits instability in scenes with interference. Wang et al. [20] optimized the line segment angle observation model to cope with dynamic environments, but the robustness is still insufficient under varying lighting conditions. The point-line fusion SLAM [21] demonstrates

superior performance in static complex environments, but still faces challenges in large-scale dynamic scenes. The PLFF-SLAM [22] designed by Huang et al., which integrates GCNV2 and an enhanced LSD algorithm, improves accuracy under dynamic lighting conditions. The system proposed by Liu et al., based on enhanced VINS-fusion[23], exhibits strong real-time performance. However, the sparse features limit its performance in complex dynamic environments. The PLF-SLAM [24] method proposed by Fang and Zhan enhances localization robustness in low-texture scenes through point-line fusion, but it is still limited by the accuracy of feature extraction and tracking in dynamic environments. Jin et al. optimized the line feature processing in EPL-VINS [25], which improved performance in low-texture environments. However, in highly dynamic settings, it still faces issues of sparse features and mismatches.

In summary, point-line fusion visual SLAM has demonstrated stronger robustness and accuracy compared to traditional methods in complex environments. However, it is still limited by the stability of feature extraction, the accuracy of matching, and the overall robustness of the system. In response to the unique working conditions of underground coal mines, especially under conditions of strong light protrusion and overall insufficient illumination, there is an urgent need for a visual localization technology that can balance accuracy, stability, and adaptability.

To address the issue of insufficient localization stability in underground coal mines caused by strong light protrusion and overall insufficient illumination, this paper proposes a binocular visual localization method for roadheaders that integrates point and line environmental features. While point-line feature extraction algorithms have been widely investigated to reduce the computational complexity of traditional visual navigation methods, our approach advances the state of the art by incorporating a tailored low-illumination image enhancement process, gradient-weighted feature fusion, and a dynamic weighting strategy for joint point-line optimization under harsh underground conditions. The main contributions of this work including the following:

(1) A novel binocular visual localization method that integrates point and line features is proposed. This method overcomes the limitations of traditional cooperative-target-based visual localization methods, improving pose estimation accuracy and system robustness under complex lighting conditions.

(2) An adaptive image enhancement technique is developed, combining brightness dynamic partitioning, region-specific enhancement strategies, and gradient-guided edge-preserving fusion. This technique improves image quality and detail preservation in coal mine environments with low illumination and strong light contrast.

(3) A dynamic weighting-based joint optimization of point and line features is proposed. Leveraging EDLines and

Plücker representation, the method adaptively adjusts point and line feature weights according to visual observations, significantly improving localization accuracy and robustness in complex underground scenarios.

The rest of this paper is organized as follows. Section 2 provides a detailed description of the proposed binocular vision localization framework, including adaptive image enhancement, EDLines-based line feature extraction, and the fusion strategy of point and line features. Section 3 presents the experimental validation, including the evaluation of image enhancement, comparison with public datasets, and tests under complex underground conditions. Section 4 discusses the strengths and limitations of the proposed method, the impact of environmental factors on its performance, and directions for future improvements. Finally, Section 5 concludes this work.

2. Methodology

2.1 System Overview

To address the issue of difficult localization of roadheader in complex working conditions in underground coal mines, this paper proposes a binocular visual localization scheme integrating point and line features, as shown in Figure 1. The scheme mainly covers core components such as image preprocessing, extraction and matching of point and line features, feature tracking, and pose estimation with point-line fusion, aiming to improve the accuracy and robustness of visual localization in complex underground coal mine environments.

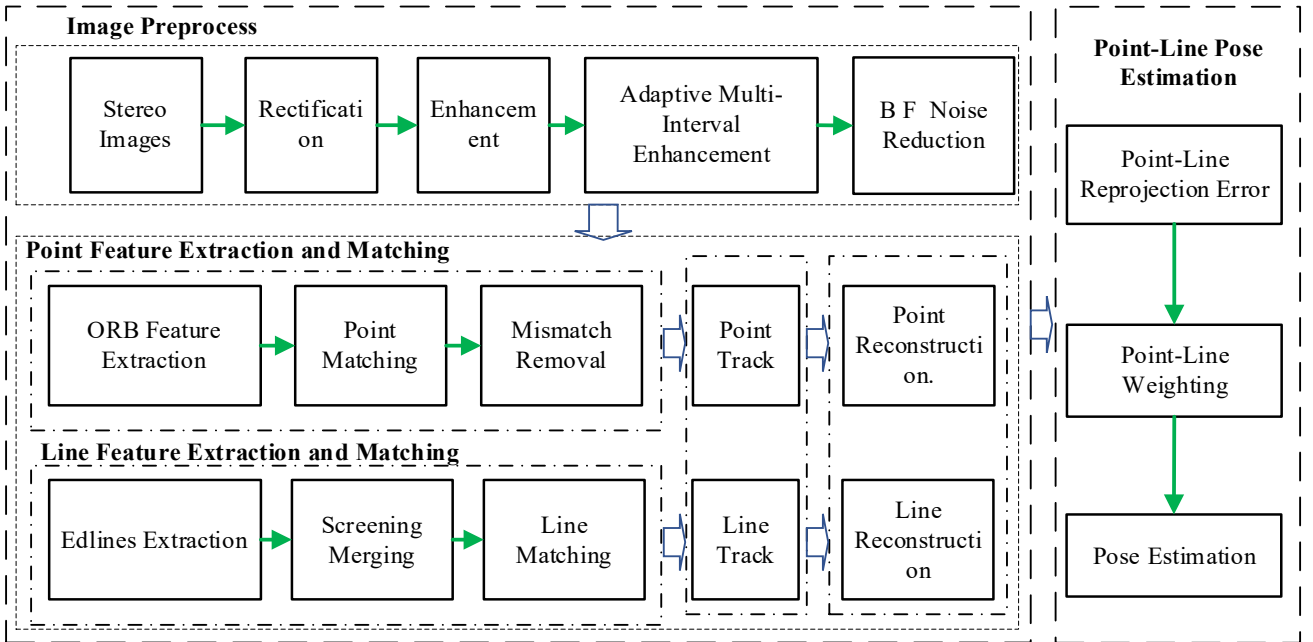


Figure 1 Binocular Visual Localization Scheme with Point-Line Feature Fusion under Complex Working Conditions

In the image preprocessing stage, to address the issues of low illumination and strong light interference in underground coal mines, methods such as stereo image rectification, adaptive multi-interval image enhancement, and bilateral filtering for noise reduction are employed. These methods are used to improve image quality and provide more stable visual input for subsequent feature extraction. In the stage of point and line feature extraction and matching, point features are extracted and matched using the ORB algorithm, and the matching quality is improved by a mismatch removal algorithm. For line features, they are extracted based on the EDLines algorithm [26] and optimized by a line segment selection and merging strategy to enhance the stability of line features. During the point-line feature tracking process, point features and line features are tracked separately to maintain the continuity of visual information and achieve the association between point and line features, thereby further completing the 3D reconstruction. Based on this feature

information, in the point-line fusion pose estimation stage, the Jacobian matrix of point-line projection errors is constructed. A dynamic weighted fusion method for pose estimation using point and line features is proposed. This method fully utilizes the local stability of point features and the geometric constraints of line features to improve the accuracy of pose estimation. Finally, optimization is performed through the analysis of point-line reprojection errors to enhance the accuracy and robustness of localization.

2.2 Adaptive Image Enhancement

When conducting cooperative target-free visual localization in underground coal mines, the issue of poor image quality due to low illumination with strong light protrusion is encountered. To address this issue, an adaptive multi-interval image enhancement method is proposed to tackle the problem of image enhancement in underground coal

mines with uneven illumination and strong light protrusion. This method dynamically thresholds the image by calculating its mean brightness and standard deviation, segmenting it into over-illuminated regions, under-illuminated regions, and normally illuminated regions. Different image enhancement strategies or parameters are applied to regions with varying brightness levels to address overexposure in over-illuminated areas and achieve significant enhancement in under-illuminated areas, thereby laying a solid image foundation for subsequent feature extraction.

Calculate the basic statistical features of the image to determine its brightness and contrast. Specifically, the image I is converted to a grayscale image I_{gray} , which contains N pixels. The average brightness μ and standard deviation σ of the image are calculated through the following steps:

$$\mu = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N I_{\text{gray}}(x, y) \quad (1)$$

$$\sigma = \sqrt{\frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N (I_{\text{gray}}(x, y) - \mu)^2} \quad (2)$$

Based on the average brightness and standard deviation of the image, the high and low dynamic thresholds for image brightness segmentation are calculated. The high threshold is used to segment the brighter regions of the image, while the low threshold is used to segment the darker regions. The low threshold is calculated as follows:

$$T_{\text{low}} = \max(\mu - \alpha \cdot \sigma, \min n) \quad (3)$$

where $\min n = \min(I_{\text{gray}})$ is the minimum grayscale value of the image, and α is a constant that controls the sensitivity of the low threshold, typically set to 0.5 but can be adjusted according to the specific situation.

The calculation method for the high threshold is:

$$T_{\text{high}} = \min(\mu + \beta \cdot \sigma, \max x) \quad (4)$$

where $\max x = \max(I_{\text{gray}})$ is the maximum grayscale value of the image.

By calculating the masks for three brightness regions using adaptive thresholds, the image can be segmented into over-illuminated, normally illuminated, and under-illuminated regions. Different image enhancement strategies are applied to these three brightness regions. Specifically, gamma correction is performed on the over-illuminated regions to reduce the brightness differences in high-luminance areas.

$$I_{\text{high}} = a \cdot I^\gamma \quad (5)$$

Where a is a constant, and γ is the gamma value. By adjusting γ the brightness differences in the brightness region are compressed.

The normally illuminated region, with brightness close to the normal visual range and rich in details, should be enhanced moderately to maintain good contrast and texture details, avoiding over-enhancement that may cause detail loss or image harshness.

To this end, Contrast Limited Adaptive Histogram Equalization (CLAHE) can be employed to adjust local contrast, effectively enhancing details without causing drastic brightness changes. The under-illuminated region, often

suffering from detail loss, increased noise, or underexposure due to low brightness, requires an enhancement strategy focusing on boosting brightness and contrast to restore details while preventing noise amplification and overexposure. Since the enhancement strategies for different regions may cause uneven brightness during image fusion, this paper calculates the gradient weights for each region and employs weighted fusion techniques to achieve smooth transitions, thereby ensuring the overall image is rich in details and has natural boundaries. After image partitioning, the transitions between over-illuminated, normally illuminated, and under-illuminated regions are regulated based on gradient information. Gradients are used to identify detailed areas, ensuring detail retention and brightness smoothness during the enhancement process. Weights are assigned to each region based on their light sensitivity characteristics: low for over-illuminated areas, high for under-illuminated areas, and moderate for normally illuminated areas. The enhanced results are then weighted and fused to achieve natural transitions and clear details. The enhanced result for each pixel is obtained by weighted averaging:

$$I_{\text{enhanced}}(x, y) = \sum_{k \in \{\text{low}, \text{mid}, \text{high}\}} \omega_k(x, y) \cdot I_k(x, y) \quad (6)$$

Where, $\omega_k(x, y)$ is the weight value for the region, which is typically adjusted based on gradient information. $I_k(x, y)$ are the enhanced results for the over-illuminated, normally illuminated, and under-illuminated regions, respectively.

To better smooth the brightness transitions between the three different brightness regions, adaptive fusion using image gradients as weights is employed. The weight $\omega(x, y)$ can be obtained from the gradient of the image.

$$\omega(x, y) = \frac{G(x, y)}{\sum G} \quad (7)$$

Where $G(x, y)$ represents the gradient information, and $\sum G$ denotes the total gradient of the entire image.

$$G(x, y) = \sqrt{G_x^2 + G_y^2}, \quad G_x = \frac{\partial I(x, y)}{\partial x}, \quad G_y = \frac{\partial I(x, y)}{\partial y}.$$

Bilateral filtering is applied to the synthesized image for smoothing[27]. Compared with traditional linear filters, bilateral filtering can better preserve the edge information of the image, avoiding blurring and distortion.

2.3 EDLines-based Feature Extraction

EDLines is an efficient line segment extraction algorithm based on edge detection. It identifies the position and orientation of straight lines in images through edge tracking and the Hough transform. Compared with the LSD algorithm, EDLines achieves a detection speed improvement of approximately 10 times while maintaining similar detection accuracy, thanks to the introduction of its anchor point selection mechanism[28]. Compared with the redundant operations of LSD on all image pixels, EDLines significantly reduces computational overhead by connecting anchor points instead of using bounding rectangles and calculating the Number of False Alarms (NFA). In the complex environment of coal mines, line segment detection is prone to generating a

large number of short line segments, which increases the risk of redundant computation and mismatching.

To enhance the quality of the features, further optimization processes such as short line removal, nearby line merging, and broken line stitching are applied to the line segment results. During the inter-frame line feature matching process, long line segments exhibit higher stability, while short line segments are more susceptible to interference and have a higher rate of repeated matching. To address this, a length threshold method is employed to remove short line segments. The specific criteria for short line segment removal are as follows:

$$L_i = \{\forall l_i \in L, i \in [1, k] : l_i \geq \alpha \min(W_i, H_i)\} \quad (8)$$

Where L_i represents the i -th line segment extracted from the image initially, and the scale factor $\alpha \in (0, 1)$ is used to adjust the proportion of the screening. H_i is the height of the image, and W_i is the width.

To enhance the coherence and completeness of line features, line segment stitching is implemented using Breadth-First Search (BFS). Using the endpoints of the line segments as seed points, adjacent line segments are stitched together if their directions are approximately aligned.

$$L = \{\forall L_i \in L : (\theta_i - \theta_0) < \beta * \theta_{\min}\} \quad (9)$$

Where θ_0 is the starting endpoint of the seed node, and β is a proportional coefficient.

After distinguishing the states of line segments, the same screening criteria are applied for merging or stitching. both the horizontal distance d_1 and the vertical distance d_2 must be less than the set distance screening threshold. Taking the starting endpoint of the seed node as (x_0, y_0) , and the start and end points of the two line segments as (x_i, y_i) and (x_j, y_j) respectively, the optimized line segments are stored in L' :

$$L' = \{\forall L_i, L_j \in L : (|x_i - x_0| + |x_j - x_0| < d_1) \quad (10)$$

$$\&(|y_i - y_0| + |y_j - y_0| < d_2)\}$$

In this study, the horizontal and vertical distance thresholds d_1 and d_2 for line segment merging were empirically set to 30 pixels, based on the average inter-feature distance observed in multiple underground image sequences. These values ensure that only line segments with sufficient spatial proximity and continuity are merged, while avoiding erroneous connections across unrelated structures. Experimental observations confirmed that this setting balances the trade-off between over-segmentation and false merging, and achieves stable performance across different lighting and structural conditions.

After detecting line features, the LBD (Line Binary Descriptor) is used for feature description. The K-Nearest Neighbors (KNN) algorithm is employed to complete inter-frame matching. To improve matching accuracy and eliminate outliers, two constraints are set: the Hamming distance must be greater than a certain threshold to ensure matching reliability, and the angular difference must be less than a certain threshold to ensure geometric consistency. These

strategies effectively suppress outlier line features, enhancing the stability and robustness of matching.

2.4 Pose Estimation with Point-Line Feature Fusion

2.4.1 Line Feature Parameterization and Observation Model.

In visual localization systems, line features can compensate for the deficiencies of point features in low-texture or repetitive structure scenes. However, the efficient parameterization and observation model design of line features are key challenges. A spatial line has four degrees of freedom, and an appropriate representation method must be chosen to balance computational complexity and optimization accuracy. A spatial point is represented by coordinates in $\mathbb{R}^3$, denoted as $\tilde{\mathbf{X}} = [x, y, z]^T$. In homogeneous coordinates, it can be represented as $\mathbf{X} = [x, y, z, w]^T$. A line segment is represented by its two endpoints, and its homogeneous coordinates are used as the row vectors of the matrix $\mathbf{W}$.

$$\mathbf{W} = \begin{bmatrix} \mathbf{X}_1^T & \mathbf{X}_2^T \end{bmatrix}^T \quad (11)$$

The above representation includes eight parameters, while a spatial line has only four degrees of freedom, resulting in parameter redundancy. In contrast, Plücker coordinates[29] can conveniently represent the projection relationship of line features. The Plücker matrix is defined as:

$$\mathbf{L} = \mathbf{X}_2 \mathbf{X}_1^T - \mathbf{X}_1 \mathbf{X}_2^T \in \mathbb{R}^{4 \times 4} \quad (12)$$

Where the diagonal elements of the Plücker matrix are all zero, and it is an anti-symmetric matrix, so it has six non-zero elements. These are the Plücker coordinates:

$$\ell = \begin{bmatrix} \tilde{\mathbf{X}}_1 \times \tilde{\mathbf{X}}_2 \\ w_1 \tilde{\mathbf{X}}_2 - w_2 \tilde{\mathbf{X}}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{n} \\ \mathbf{d} \end{bmatrix} \in \mathbb{R}^6 \quad (13)$$

Where, $\mathbf{n}$ represents the normal vector of the plane formed by the line feature and the coordinate origin; $\mathbf{d}$ represents the direction vector of the line feature.

The normal vector of the plane formed by the line feature and the coordinate origin, and the direction vector of the line feature, must satisfy the constraint $\mathbf{n}^T \mathbf{d} = 0$. The scalar $|\mathbf{n} \times \mathbf{d}| / \|\mathbf{n}\| \|\mathbf{d}\|$ represents the distance from the origin to the line. Figure 2 illustrates the Plücker coordinate representation of the line feature.

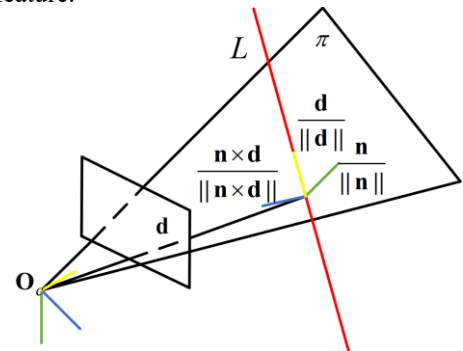


Figure 2 Schematic Diagram of Plücker Representation of Line

Projecting a spatial line onto the current camera image yields the equation of its projection line:

$$\mathbf{l}_c = \mathbf{K}\mathbf{n}_c \quad (14)$$

Where, $\mathbf{K}$ represents the camera's intrinsic parameter matrix.

Assuming the Plücker coordinates are $\ell = [\mathbf{n}^T, \mathbf{d}^T]^T$, their orthogonal representation is $(\mathbf{U}, \mathbf{W}) \in SO(3) \times SO(2)$. The QR decomposition of matrix $[\mathbf{n} | \mathbf{d}]$ can be computed. Due to the orthogonality constraint $\mathbf{n}^T \mathbf{d} = 0$, the decomposition process can be represented as:

$$QR([\mathbf{n} | \mathbf{d}]) = \mathbf{U} \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \\ 0 & 0 \end{bmatrix}, \mathbf{W} = \begin{bmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{bmatrix} \quad (15)$$

Where $w_1, w_2 > 0$. In fact, the conversion from Plücker coordinates to the orthogonal representation does not require QR decomposition and can be directly expressed as:

$$[\mathbf{n} | \mathbf{d}] = \begin{bmatrix} \mathbf{n} & \mathbf{d} & \mathbf{n} \times \mathbf{d} \\ \mathbf{n} \times \mathbf{n} & \mathbf{d} \times \mathbf{d} & \mathbf{n} \times \mathbf{d} \end{bmatrix} \begin{bmatrix} \mathbf{n} \times \mathbf{n} & 0 \\ 0 & \mathbf{d} \times \mathbf{d} \\ 0 & 0 \end{bmatrix} \quad (16)$$

In addition to the orthogonal matrix, there are other methods. The observation error of line features is usually defined as the 3D-2D projection error, which is the geometric distance between the projected line and the matched line segment in the image. Let the matched line segment in the image be represented by endpoints $\mathbf{X}_s$ and $\mathbf{X}_e$, as shown in Figure 3.

The error between the projected line and the matched line is defined as:

$$\mathbf{e} = [e_1 \quad e_2]^T = \left[\frac{\mathbf{X}_s^T \mathbf{l}'}{\sqrt{l_1^2 + l_2^2}}, \frac{\mathbf{X}_e^T \mathbf{l}'}{\sqrt{l_1^2 + l_2^2}} \right]^T \quad (17)$$

Where $\mathbf{l}' = [l_1, l_2, l_3]^T$ is the projected line. $\mathbf{X}_s^T = [u_s, v_s, 1]^T$ and $\mathbf{X}_e^T = [u_e, v_e, 1]^T$ are the homogeneous coordinates of the two endpoints of the matched line segment.

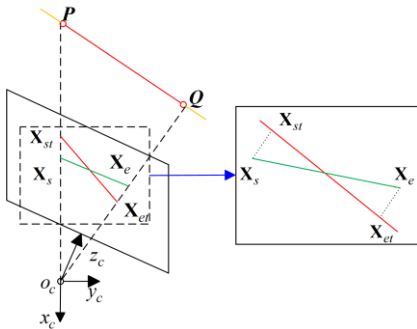


Figure 3 Reprojection Error of Line Features

2.4.2 Jacobian Matrix of Point and Line Feature Projection Error.

(1) Jacobian Matrix of Point Feature Projection Error

Assuming the camera pose is $\mathbf{T}_{cw}$, which represents the pose of the world coordinate system relative to the camera

coordinate system, it can be expressed by the rotation matrix $\mathbf{R}_{cw}$ and the translation vector $\mathbf{t}_{cw}$. Let the 3D point $\mathbf{X}_w$ be transformed into the camera coordinate system and represented as $\mathbf{X}_c$:

$$\mathbf{X}_c = \mathbf{R}_{cw}\mathbf{X}_w + \mathbf{t}_{cw} \quad (18)$$

Then, the pixel coordinates $\mathbf{x}'$ of $\mathbf{X}_c$ on the image are calculated based on the camera model:

$$\mathbf{x}' = \mathbf{n}(\mathbf{K}\mathbf{X}_c) \quad (19)$$

Where $\mathbf{n}(\cdot)$ represents the conversion from homogeneous coordinates to non-homogeneous coordinates. The error $\mathbf{e}_p$ between the projected point $\mathbf{x}'$ and the matched feature point $\mathbf{x}$ is calculated.

$$\mathbf{e}_p = \mathbf{x} - \mathbf{x}' \quad (20)$$

In the process of solving Bundle Adjustment (BA), methods such as the Gauss-Newton method or the Levenberg-Marquardt algorithm are typically used, which require linearization of the cost function near the current state estimate. Therefore, the Jacobian matrix $\mathbf{J} = \frac{\partial \mathbf{e}}{\partial \mathbf{x}}|_{\mathbf{x}}$ of the error function with respect to the state variables must be computed.

The Jacobian matrix of the re-projection error of a 3D point with respect to the pose increment δ_ξ and the 3D point coordinates $\mathbf{X}_w$ can be solved using the finite difference method. For the coordinate transformation $\mathbf{X}_c = \mathbf{R}_{cw}\mathbf{X}_w + \mathbf{t}_{cw}$, a small perturbation $\exp(\delta_\xi^\wedge)$ is added to compute:

$$\frac{\partial \mathbf{X}_c}{\partial \delta_\xi} = \begin{bmatrix} [\mathbf{X}_c]_\times & \mathbf{I} \end{bmatrix}_{3 \times 6} \quad (21)$$

$$\frac{\partial \mathbf{X}_c}{\partial \mathbf{X}_w} = \lim_{\delta \delta_\xi \rightarrow 0} \frac{\partial \exp(\delta_\xi^\wedge) \mathbf{T}_{cw} \mathbf{X}_w}{\partial \mathbf{X}_w} = \mathbf{R}_{cw} \quad (22)$$

For the camera projection point $\mathbf{x}' = \mathbf{n}(\mathbf{K}\mathbf{X}_c)$:

$$\frac{\partial \mathbf{x}'}{\partial \mathbf{X}_c} = \begin{bmatrix} f_u Z_c & 0 & -f_u X_c / Z_c^2 \\ 0 & f_v Z_c & -f_v Y_c / Z_c^2 \end{bmatrix}_{2 \times 3} \quad (23)$$

Where $\mathbf{X}_c = [X_c, Y_c, Z_c]^T$ is the projection point on the image.

The Jacobian matrices of the re-projection error with respect to the pose and the 3D point are obtained using the chain rule:

$$\mathbf{J}p_\xi = \frac{\partial \mathbf{e}_p}{\partial \delta_\xi} = -\frac{\partial \mathbf{x}'}{\partial \mathbf{X}_c} \frac{\partial \mathbf{X}_c}{\partial \delta_\xi} \quad (24)$$

$$\mathbf{J}p_x = \frac{\partial \mathbf{e}_p}{\partial \mathbf{X}_w} = -\frac{\partial \mathbf{x}'}{\partial \mathbf{X}_c} \frac{\partial \mathbf{X}_c}{\partial \mathbf{X}_w} \quad (25)$$

(2) Jacobian Matrix of Line Feature Projection Error

Let there be a line in space ℓ_w , and convert its Plücker coordinates to the camera coordinate system to get ℓ_c .

$$\ell_c = \mathbf{H}_{cw} \ell_w \quad (26)$$

Then, the projection line $\mathbf{l}'$ of ℓ_c on the image is calculated based on the camera model.

$$\mathbf{l}' = \mathbf{K}\mathbf{n}_c \quad (27)$$

The distances from the endpoints $\mathbf{X}_s$ and $\mathbf{X}_e$ of the matched line segment $\mathbf{z}$ to the projected line $\mathbf{l}'$ are

calculated. The re-projection error of the line can be expressed as:

$$\mathbf{e}_l = \mathbf{d}(\mathbf{z}, \mathbf{l}') = \left[\frac{\mathbf{X}_s^T \mathbf{l}'}{\sqrt{l_1^2 + l_2^2}}, \frac{\mathbf{X}_e^T \mathbf{l}'}{\sqrt{l_1^2 + l_2^2}} \right]^T \quad (28)$$

Similarly, we continue to solve the Jacobian matrix of the re-projection error of the spatial line. For $\mathbf{e}_l = \mathbf{d}(\mathbf{z}, \mathbf{l}')$, let the coordinates of the line segment endpoints be $\mathbf{X}_s^T = [u_s, v_s, 1]^T$ and $\mathbf{X}_e^T = [u_e, v_e, 1]^T$. Therefore, the Jacobian matrix of the error with respect to the projected line $\mathbf{l}'$ is:

$$\frac{\partial \mathbf{e}_l}{\partial \mathbf{l}'} = \frac{1}{l_n} \begin{bmatrix} u_s - \frac{l_1 e_1}{l_n^2} & v_s - \frac{l_1 e_2}{l_n^2} & 1 \\ u_e - \frac{l_1 e_2}{l_n^2} & v_e - \frac{l_2 e_2}{l_n^2} & 1 \end{bmatrix}_{2 \times 3} \quad (29)$$

Where $e_1 = \mathbf{X}_s^T \mathbf{l}'$, $e_2 = \mathbf{X}_e^T \mathbf{l}'$, $l_n = \sqrt{l_1^2 + l_2^2}$.

For the camera projection $\mathbf{l}' = \mathbf{K} \mathbf{n}_c$, the Jacobian matrix of the projected line $\mathbf{l}'$ with respect to the Plücker coordinates ℓ_c is:

$$\frac{\partial \mathbf{l}'}{\partial \ell_c} = \frac{\partial \mathbf{K} \mathbf{n}_c}{\partial \ell_c} = [\mathbf{K} \quad \mathbf{0}]_{3 \times 6} \quad (30)$$

For the coordinate transformation $\ell_c = \mathbf{H}_{cw} \ell_w$, the Jacobian matrices with respect to the pose increment δ_ξ and the increment of the orthogonal representation of the spatial line δ_θ are computed. Let the orthogonal representation of ℓ_w

be $\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3]$, $\mathbf{W} = \begin{bmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{bmatrix}$.

$$\frac{\partial \ell_c}{\partial \ell_w} = \frac{\partial \mathbf{H}_{cw} \ell_w}{\partial \ell_w} = \mathbf{H}_{cw} \quad (31)$$

$$\frac{\partial \ell_w}{\partial \delta_\theta} = \frac{\partial \mathbf{H}_{cw} \ell_w}{\partial \delta_\theta} = \begin{bmatrix} -[w_1 \mathbf{u}_1]_x & -w_2 \mathbf{u}_1 \\ -[w_2 \mathbf{u}_2]_x & -w_1 \mathbf{u}_2 \end{bmatrix}_{6 \times 4} \quad (32)$$

The solution process for $\frac{\partial \ell_c}{\partial \delta_\xi}$ is decomposed into two parts, solving the Jacobian matrices for the first three elements δ_ϕ and the last three elements δ_p of δ_ξ . For δ_p , let the rotation part be $\delta_\phi = 0$, then $\delta_\xi = [\mathbf{0} \quad \delta_p]^T$. Then, calculate the updated spatial line ℓ_c^* .

$$\ell_c^* = \begin{bmatrix} \mathbf{R}_{cw} \mathbf{n}_w + [\delta_p + \mathbf{t}_{cw}]_x \mathbf{R}_{cw} \mathbf{d}_w \\ \mathbf{R}_{cw} \mathbf{d}_w \end{bmatrix} \quad (33)$$

Taking the partial derivative of δ_p

$$\frac{\partial \ell_c^*}{\partial \delta_p} = \begin{bmatrix} -[\mathbf{R}_{cw} \mathbf{d}_w]_x \\ \mathbf{0} \end{bmatrix}_{6 \times 3} \quad (34)$$

For the rotation δ_ϕ , let the translation part be $\delta_p = 0$, then.

$\delta_\xi = \begin{bmatrix} \delta_\phi \\ \mathbf{0} \end{bmatrix}$. Calculate the updated spatial line ℓ_c^* .

$$\ell_c^* = \begin{bmatrix} (\mathbf{I} + [\delta_\phi]_x) \mathbf{R}_{cw} \mathbf{n}_w + (\mathbf{I} + [\delta_\phi]_x) [\mathbf{t}_{cw}]_x \mathbf{R}_{cw} \mathbf{d}_w \\ (\mathbf{I} + [\delta_\phi]_x) \mathbf{R}_{cw} \mathbf{d}_w \end{bmatrix} \quad (35)$$

Taking the partial derivative of δ_ϕ :

$$\frac{\partial \ell_c^*}{\partial \delta_\phi} = \begin{bmatrix} -[\mathbf{R}_{cw} \mathbf{n}_w]_x & -[\mathbf{t}_{cw}]_x \mathbf{R}_{cw} \mathbf{d}_w \\ -[\mathbf{R}_{cw} \mathbf{d}_w]_x \end{bmatrix}_{6 \times 3} \quad (36)$$

Therefore, taking the partial derivative of δ_ξ , we get:

$$\frac{\partial \ell_c}{\partial \delta_\xi} = \frac{\partial \ell_c^*}{\partial \delta_\xi} = \begin{bmatrix} -[\mathbf{R}_{cw} \mathbf{n}_w]_x & -[\mathbf{t}_{cw}]_x \mathbf{R}_{cw} \mathbf{d}_w & -[\mathbf{R}_{cw} \mathbf{d}_w]_x \\ -[\mathbf{R}_{cw} \mathbf{d}_w]_x & \mathbf{0} \end{bmatrix} \quad (37)$$

Using the chain rule, the Jacobian matrices of the re-projection error $\mathbf{e}_l = \mathbf{d}(\mathbf{z}, \mathbf{l}')$ with respect to the pose and the 3D point can be obtained:

$$\mathbf{J}_{\ell_\xi} = \frac{\partial \mathbf{e}_l}{\partial \delta_\xi} = \frac{\partial \mathbf{e}_l}{\partial \mathbf{l}'} \frac{\partial \mathbf{l}'}{\partial \ell_c} \frac{\partial \ell_c}{\partial \delta_\xi} \quad (38)$$

$$\mathbf{J}_{\ell_\theta} = \frac{\partial \mathbf{e}_l}{\partial \delta_\theta} = \frac{\partial \mathbf{e}_l}{\partial \mathbf{l}'} \frac{\partial \mathbf{l}'}{\partial \ell_c} \frac{\partial \ell_c}{\partial \ell_w} \frac{\partial \ell_w}{\partial \delta_\theta} \quad (39)$$

2.4.3 Dynamic Weighted Fusion of Point-Line Features for Pose Estimation.

Considering that the system proposed in this paper is applied to low-texture environments such as underground coal mine tunnels, where point feature extraction is challenging, weighting of point and line features is performed to balance the significant disparity in their numbers that could otherwise degrade localization accuracy. This weighting approach ensures that even when point features are not extracted for several consecutive frames, the system can still rely on line features to complete map initialization, thereby maintaining its continuity and stability. To avoid complex weight allocation calculations and ensure real-time performance, this paper dynamically adjusts the weight ratio of point and line features based on the number of line features detected.

Let w_p and w_l represent the weights of point and line features, respectively, after their extraction. The distribution of the weights for point and line features is as follows:

$$w_p = \begin{cases} 0.2, N_l \geq N \\ \frac{N - N_p - 1}{N - N_p}, N_l < N \end{cases} \quad (40)$$

$$w_l = \begin{cases} 0.8, N_l \geq N \\ \frac{1}{N - N_l}, N_l < N \end{cases} \quad (41)$$

Where N_l represents the number of successful line matches, and N_p represents the number of successful point matches.

Using a uniform motion model, an initial camera pose is set for the current frame, and matching is performed between the current frame and the previous frame to obtain the matching relationship between 3D features and 2D features. Based on these matching relationships, the camera pose is solved using the PNP algorithm. After introducing line features, camera pose estimation needs to be performed jointly

using point and line features by minimizing the following cost function:

$$\{R_w^*, t_w^*\} = \arg_{R, t} \min \sum_{i=P} w_p L_p(\|e_{pi}\|^2 \sum_p^{-1}) + \sum_{j=L} w_l L_l(\|e_{lj}\|^2 \sum_l^{-1}) \quad (42)$$

Where, $\sum_p$ and $\sum_l$ represents the current frame's camera pose, which is the optimal solution. P and L are the sets of matched points and matched line segments in the current frame, respectively. e_{pi} and e_{lj} are the projection errors for points and lines, respectively. $\sum_p$ and $\sum_l$ are the covariance matrices for points and lines, respectively. w_p and w_l are the weights for point and line features introduced in this paper, and L_p and L_l are the Huber robust loss functions. Since the cost function may contain outliers during minimization, the Huber function is introduced to address this issue. The definition of the Huber function is shown:

$$L_\delta(a) = \begin{cases} \frac{1}{2} a^2, & |a| \leq \delta \\ \delta(|a| - \frac{1}{2}\delta), & |a| > \delta \end{cases} \quad (43)$$

Where, a is the residual, which represents the difference between the observed value and the estimated value. The goal function is minimized using the Gauss-Newton method to iteratively estimate the current frame's camera pose $\{R_w^*, t_w^*\}$.

More 3D features from the local map are matched with the current features, and more constraint terms are added to optimize the current frame's pose. The objective function is shown:

$$\{R_w^*, t_w^*\} = \arg_{R, t} \min \sum_{i=P_l} w_p L_p(\|e_{pi}\|^2 \sum_p^{-1}) + \sum_{j=L_l} w_l L_l(\|e_{lj}\|^2 \sum_l^{-1}) P_l \quad (44)$$

Where, P_l and L_l are the sets of matched points and matched line segments from the local map, respectively. 3. Experimental Validation

3.1 Image Enhancement and Feature Extraction

(1) Image enhancement results

To verify the effectiveness of the proposed enhancement method for low-light and high-contrast images, several sets of real scene images were selected for processing. To determine the optimal parameters for image enhancement, a sensitivity analysis was conducted using different image segmentation thresholds. By analyzing the grayscale histogram distributions of representative images, the threshold α coefficient in Eq. (3) was selected within the range of [30–60], and the parameter β in Eq. (4) was varied within [150–200].

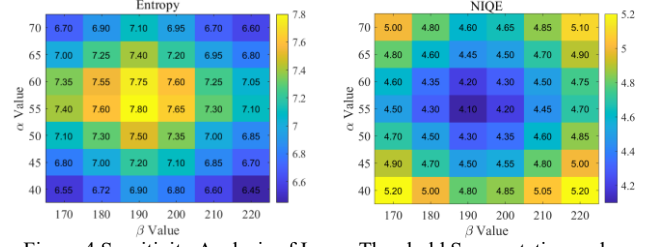


Figure 4 Sensitivity Analysis of Image Threshold Segmentation under Different Parameter Values

As shown in Figure 4, the Entropy heatmap indicates that as the parameters α and β in Eq. (3) and Eq. (4) increase, the overall image information content first increases and then levels off. The optimal parameter ranges are [45, 50] and [170, 180], within which the image exhibits the richest detail and the highest entropy values. Similarly, the NIQE heatmap demonstrates the same trend, suggesting that image quality peaks within the same parameter ranges. Based on this sensitivity analysis, the selected parameters were used for image enhancement, as shown in Figure 4. The proposed adaptive multi-region image enhancement method was applied to these images to generate high-quality outputs suitable for subsequent feature extraction.

To verify the effectiveness of the proposed image enhancement method for low-illumination images with strong light protrusion, several groups of real-scene images were selected for processing, and the results are shown in Figure 5. The adaptive multi-interval image enhancement method proposed in this paper was applied to these images to obtain high-quality images for feature extraction.

As shown in Figure 5, from the subjective image enhancement results, the proposed method not only maintains the image brightness but also significantly improves the detail clarity and contrast of the images. Compared with the CLAHE, HE, and SSR methods, the enhanced images are clearer and have better detail representation. To better compare the performance of the proposed adaptive multi-interval image enhancement method, metrics such as the Peak Signal-to-Noise Ratio (PSNR), Improved Signal-to-Noise Ratio (ISNR), and processing time were used for comparative analysis. The comparison results are shown in Figure 6.

As shown in Figure 6, in terms of the two key metrics for image quality, PSNR and ISNR, the proposed method achieved higher scores, indicating its significant advantages in image restoration and enhancement. In terms of processing time, although the method has been greatly improved, it is still relatively long. Given the significant improvement in image quality, this additional time cost is reasonable, especially in applications with high image quality requirements, such as underground coal mines, where the importance of image quality far outweighs processing time. Therefore, for applications that require high-quality image processing in low-illumination environments with strong light protrusion, the proposed method is an effective choice.

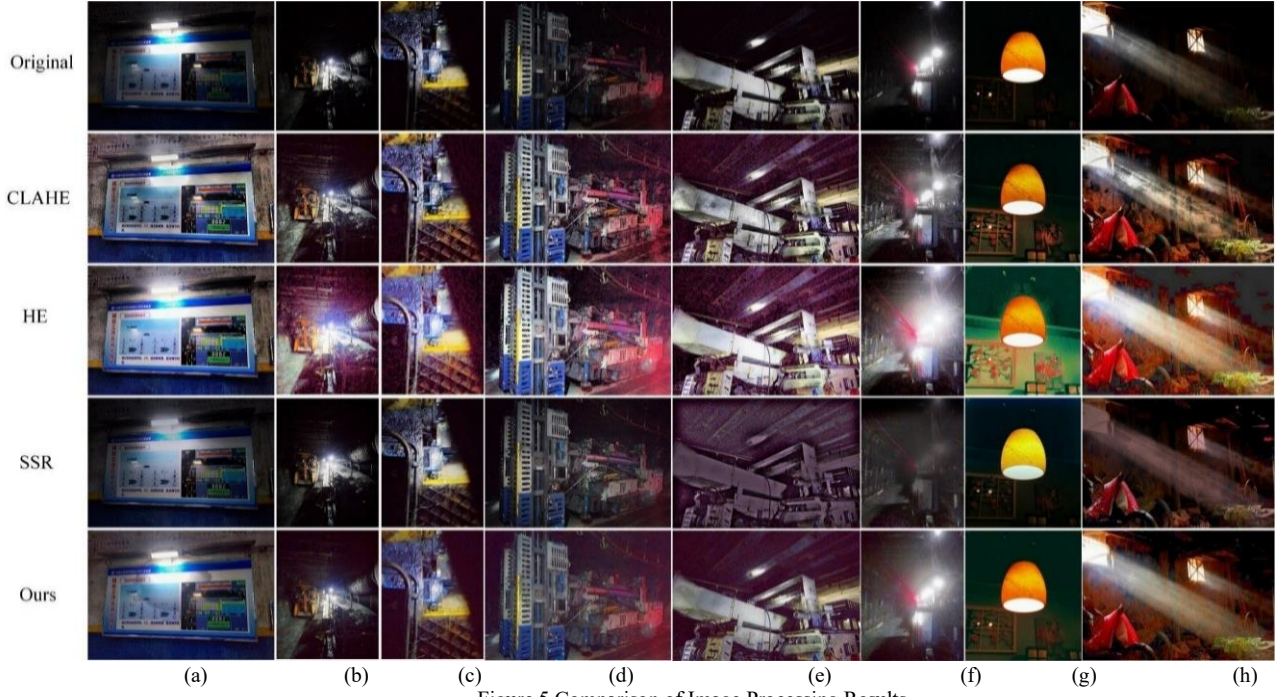


Figure 5 Comparison of Image Processing Results

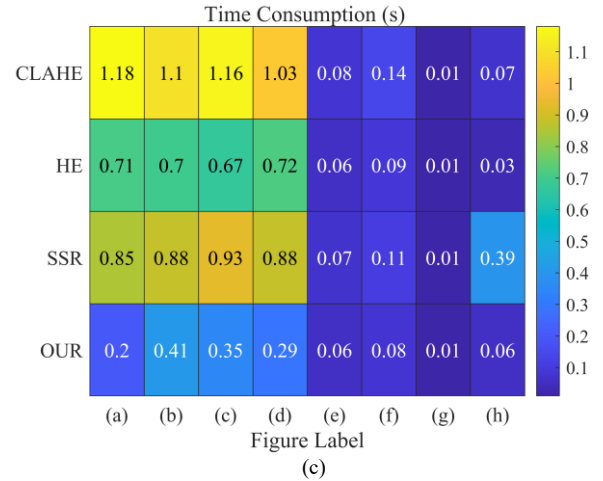
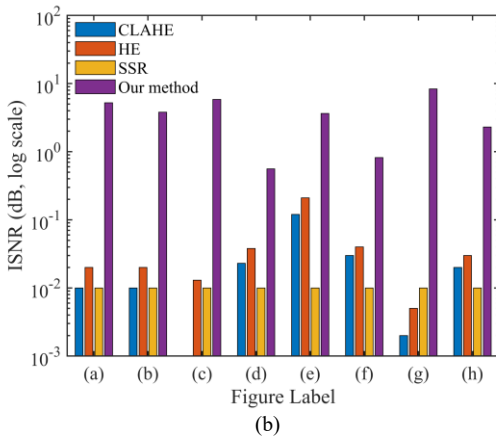
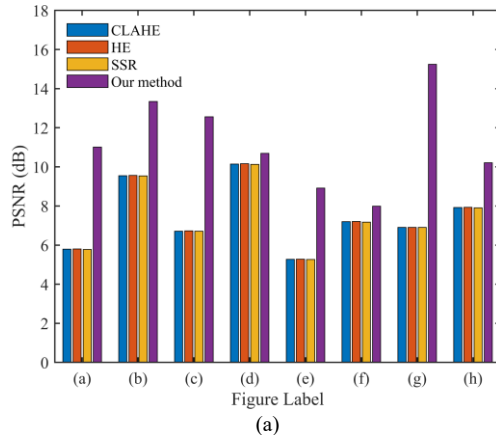


Figure 6 Comparison Results of Different Enhancement Methods. (a) PSNR results; (b) ISNR results; (c) Time results

Although designed for intra-frame enhancement, the method inherently provides partial inter-frame consistency due to its adaptive adjustment based on frame-wise brightness statistics. Nevertheless, for scenarios with extreme inter-frame contrast variation, additional temporal smoothing could be beneficial, which is a direction for future work.

(2) Results of line segment feature extraction

The feature extraction capabilities of LSD, FLD, and EDLines in complex environments were evaluated and compared, and the results are shown in Figure 7.

As shown in Figure 7, the number of line features extracted using EDLines is significantly higher than that obtained by LSD and FLD methods. The analysis of the number of features and processing time yielded the comparative results presented in Figure 8.

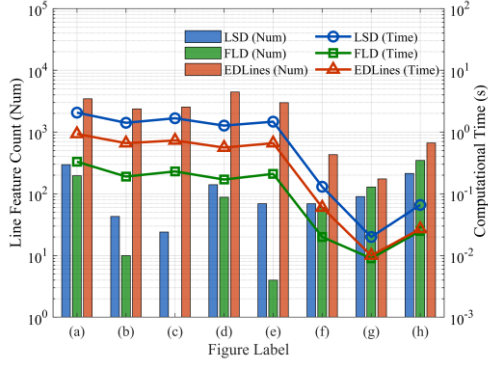


Figure 7 Comparison of Line Feature Extraction Quantity and Computational Time

As can be seen from the Figure 8, the number of lines extracted by EDLines is indeed much higher than that of the other methods, as shown in Figure 7, while also maintaining good timeliness. Therefore, the use of EDLines for feature



Figure 8 Comparison of Line Feature Extraction Methods in Complex Environments

As shown in Figure 9(a), compared with the ground truth trajectory, the global trajectory obtained using the proposed method is basically consistent. Compared with PL-SLAM and EDLines-SLAM, the trajectory obtained by the proposed method is closer to the ground truth. As shown in Figure 9(b), compared with the ground truth trajectory, the global trajectory obtained using the proposed method is basically consistent. Compared with PL-SLAM and EDLines-SLAM, the trajectory obtained by the proposed method is closer to the ground truth. To this end, the average APE (Absolute Pose Error) and average RMSE (Root Mean Square Error) compared with the ground truth trajectory were calculated to quantify the localization accuracy of the method. The results are shown in Figure 10.

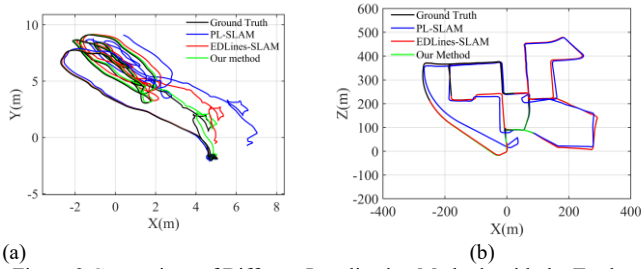


Figure 9 Comparison of Different Localization Methods with the Truth Trajectory. (a) the EuRoC Dataset; (b) the KITTI Dataset.

extraction in the complex underground coal mine environment is feasible and will serve as the foundation for the development of roadheader localization technology characterized by complex environments.

3.2 Experiments Based on Public Datasets

To verify the effectiveness of the proposed binocular visual localization technique integrating point and line features, global trajectory error comparison experiments were conducted using the EuRoC MAV dataset[31] and the KITTI dataset[32]. The accuracy of trajectory estimation was used to evaluate the performance of the algorithm. Furthermore, the proposed algorithm was compared with PL-SLAM[33] and other algorithms on the public datasets to test and verify the environmental adaptability of the algorithm. The global trajectory results obtained from processing the EuRoC MAV dataset are shown in Figure 9.

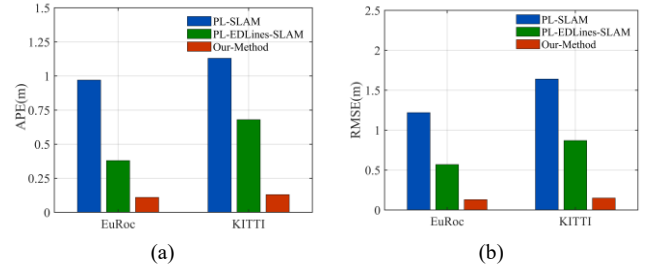


Figure 10 APE and RMSE of Different Algorithms on the EuRoC and KITTI Dataset

As shown in Figure 9(b), The proposed method in this chapter achieves the best performance in terms of trajectory estimation accuracy. Compared with PL-SLAM and PL-EDLines-SLAM, the APE of the proposed method is reduced to 0.13 m, and the RMSE is reduced to 0.15 m, which significantly improves the localization accuracy. This result indicates that the improvements in point-line feature extraction and optimization strategies in the proposed method effectively enhance the accuracy and stability of SLAM estimation.

As shown in Figure 10(a), there are significant differences in the performance of different algorithms on the EuRoC and KITTI dataset. PL-SLAM has an average absolute error (APE) of 0.97 m and a root mean square error (RMSE) of 1.22 m, which is relatively poor. PL-EDLines-SLAM,

which incorporates the EDLines feature extraction method, reduces the APE and RMSE to 0.38 m and 0.57 m, respectively, showing an improvement in accuracy. The proposed method in this chapter achieves the best performance with an APE of 0.11 m and an RMSE of 0.13 m, which are significantly better than the previous two methods, demonstrating a substantial improvement in accuracy. Therefore, the proposed method shows the best accuracy on the EuRoC dataset, verifying its effectiveness and reliability in practical applications.

3.3 Localization Experiments under Complex Working Conditions

To further verify the environmental adaptability and localization accuracy of the proposed binocular visual localization method integrating point and line features, an experimental verification platform in a low-illumination, narrow tunnel was designed, as shown in Figure 11. The platform consists of a mobile robot, a ZED stereo camera, a total station and prism, a light source, and a smoke machine. The stereo camera used in the experiment is the ZED2i, with a resolution of 1344×376, a frame rate of 15 Hz, a focal length of 2.1 mm, and a baseline length of 120 mm. All image processing was carried out on a laptop computer equipped with an Intel Core i7-12700H processor and 32 GB of RAM. The total station used is the Sokkia IX-1001, with a tracking accuracy of 1 mm.

To closely mimic the challenging underground coal mine conditions, the experimental tunnel was equipped with a controlled light source to generate both uniform and strong, uneven illumination, while a smoke machine was deployed to produce varying levels of airborne dust and haze. This setup replicates the harsh visual environment of real mining operations, characterized by low illumination, glare, and obscured visibility, ensuring that the proposed visual localization method is tested under realistic and adverse conditions.

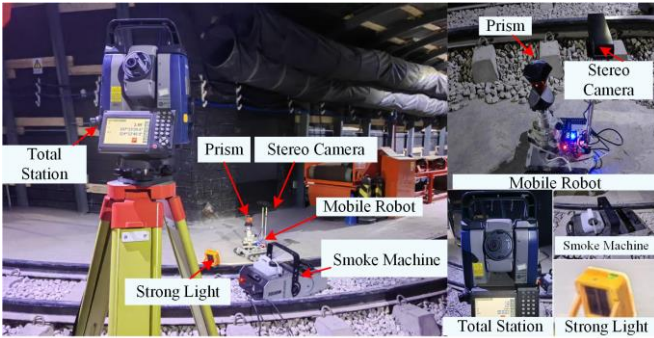


Figure 11 Localization Performance Experimental Verification Platform in Low-Illumination Narrow Tunnels

The stereo camera and prism were mounted on the mobile robot. During the movement of the robot, stereo images were captured in real time. Simultaneously, the total station was used to track the prism in real time and record the true trajectory of the movement.

Figure 12 to Figure 15 present the comparative results of positioning trajectories under four operational conditions. The black line represents the true trajectory, which serves as the benchmark for assessing the accuracy of other methods and is obtained by the total station tracking the prism fixed on the

mobile platform in real time. The blue line indicates the trajectory obtained using PL-SLAM. EDLines-SLAM represents the measurement results of PL-SLAM with EDLines line segment detection. The green line shows the trajectory obtained by the proposed point-line fusion localization method.

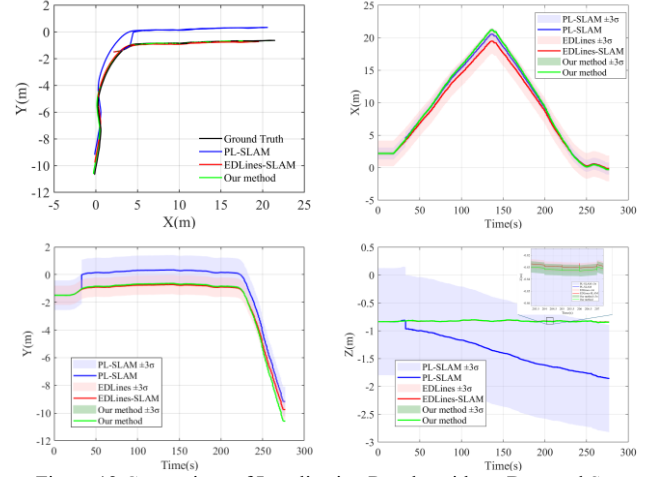


Figure 12 Comparison of Localization Results without Dust and Strong Light Interference

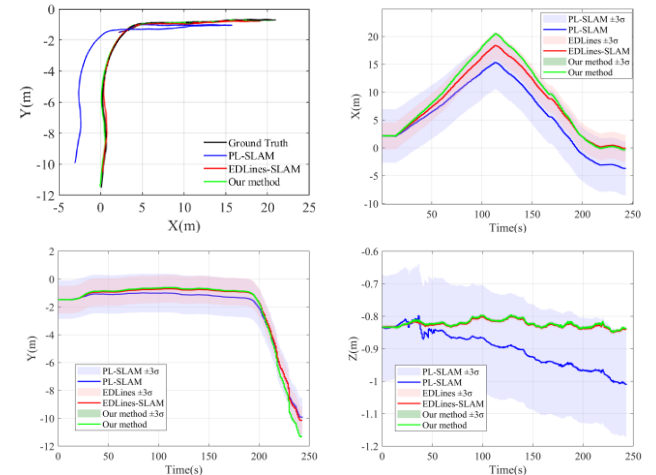


Figure 13 Comparison of Localization Results with Dust Interference

Figure 12 shows the localization results in a simulated low-illumination underground coal mine environment. As shown in the top-left subplot, compared with other methods, the global trajectory generated by the proposed localization method has the smallest deviation from the true trajectory, and the overall trajectory trend is highly consistent with the true path, demonstrating high localization accuracy. In contrast, both PL-SLAM and EDLines-SLAM exhibit significant trajectory offsets, indicating insufficient localization accuracy in complex environments.

In the top-right subplot, the proposed method consistently maintains the smallest X-direction localization error, while PL-SLAM and EDLines-SLAM suffer from higher deviations, especially near turning points. The bottom-left subplot shows that our method achieves the most stable Y-direction estimation, with tight alignment to ground truth and a narrow $\pm 3\sigma$ band, unlike the other methods. In the bottom-right subplot, the proposed method demonstrates excellent Z-

direction accuracy and robustness, avoiding the downward drift observed in PL-SLAM, thereby validating the system's overall 3D localization reliability in challenging underground conditions.

Figure 13 shows the comparison of localization results in a low-illumination underground coal mine environment with dust interference. In the environment with dust interference, the localization results of the three different methods were compared. The results show that the proposed method exhibits better localization accuracy and stability in the X, Y, and Z directions. As shown in the top-left subplot, the trajectory of the proposed method in the X-Y plane is closely aligned with the true trajectory, outperforming both PL-SLAM and EDLines-SLAM.

In the top-right subplot, the proposed method maintains relatively low X-direction errors with minimal fluctuations. The bottom-left subplot shows that it also achieves more stable Y-direction performance, with a narrower uncertainty band compared to the other methods. In the bottom-right subplot, the proposed method exhibits limited vertical drift, whereas PL-SLAM shows a continuous downward trend. These results collectively suggest that the proposed method offers improved localization robustness and reduced error accumulation under visual degradation caused by dust interference.

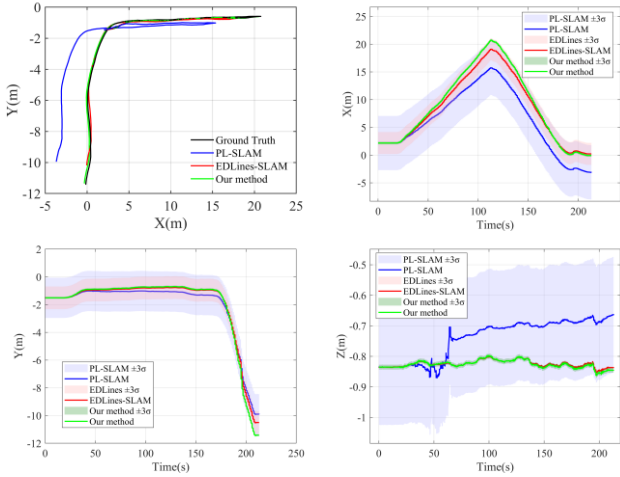


Figure 14 Comparison of Localization Results with Strong Light Source Interference.

In the environment with strong light interference, the localization results of the three methods were compared and analyzed, as shown in Figure 14. The top-left subplot shows the trajectory comparison in the X-Y plane. Compared with other methods, the proposed method is closest to the true trajectory. Although PL-SLAM and EDLines-SLAM can also track the general trajectory, they exhibit larger deviations in some areas, especially PL-SLAM.

In the top-right subplot, the proposed method achieves lower X-direction errors and maintains a relatively narrow uncertainty region, while PL-SLAM experiences increased deviation under brightness variation. The bottom-left subplot demonstrates that the proposed method maintains consistent Y-direction accuracy, whereas PL-SLAM shows greater fluctuations. The bottom-right subplot reveals that the

proposed method effectively limits Z-direction drift, in contrast to PL-SLAM, which shows vertical instability. Overall, these results indicate that the proposed method offers improved localization robustness and accuracy under strong light source interference.

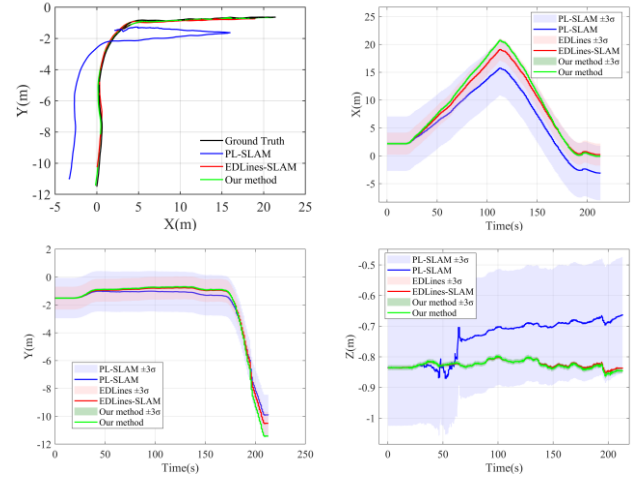


Figure 15 Comparison of Localization Results with Smoke and Strong Light Interference.

In the presence of simultaneous smoke and strong light interference, the localization performance of the three methods was evaluated, as illustrated in Figure 15. In the top-left subplot, the proposed method demonstrates the closest alignment with the ground truth trajectory in the X-Y plane, whereas PL-SLAM exhibits notable deviations and EDLines-SLAM shows moderate consistency, with observable errors in regions subject to severe visual degradation. The top-right subplot presents the X-direction localization accuracy. The proposed method maintains a low error profile and reduced uncertainty throughout, while PL-SLAM shows increasing divergence under light fluctuation and EDLines-SLAM suffers from intermittent instability. In the bottom-left subplot, the proposed method achieves stable Y-direction performance with minimal deviation. PL-SLAM yields larger fluctuations, particularly in high-interference intervals, and EDLines-SLAM performs comparably but with slightly reduced precision. The bottom-right subplot indicates that the proposed method effectively suppresses Z-direction drift, maintaining vertical consistency despite challenging conditions. PL-SLAM exhibits significant drift and widening uncertainty, while EDLines-SLAM maintains relatively stable performance but with reduced reliability. Overall, the proposed method consistently outperforms the other approaches in terms of localization accuracy and robustness across all spatial dimensions in complex underground scenarios involving smoke and intense lighting.

In summary, under the combined interference of dust and strong light, the proposed method demonstrates better robustness and accuracy in all directions, providing more reliable localization results. To comprehensively evaluate the effectiveness and robustness of the proposed binocular visual localization method integrating point and line features in practical applications, the localization performance was quantitatively compared and analyzed under different working

condition interferences. The localization accuracy and stability of the proposed method were compared with those of typical methods such as PL-SLAM and EDLines-SLAM in complex environments, as shown in Figure 16.

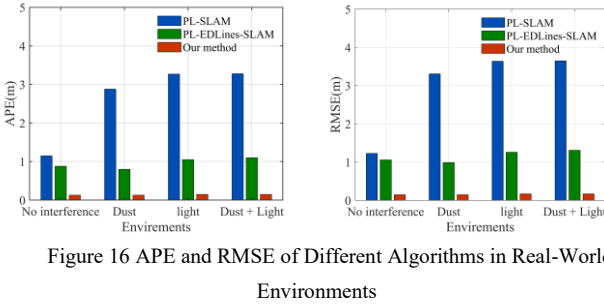


Figure 16 APE and RMSE of Different Algorithms in Real-World

As shown in Figure 16, the proposed binocular visual localization method integrating point and line features demonstrates superior localization accuracy and robustness under various environmental interferences. In the interference-free environment, the APE and RMSE of the proposed method are 0.13 m and 0.15 m, respectively, which are lower than those of PL-SLAM and PL-EDLines-SLAM, indicating that the proposed method can provide the most accurate localization results under ideal conditions. In the presence of dust interference, the APE and RMSE of PL-SLAM increase to 2.88 m and 3.31 m, respectively, reflecting its poor adaptability to environmental changes. In contrast, the errors of PL-EDLines-SLAM are reduced but still higher than those of the proposed method, indicating that the introduction of line segment features enhances the robustness of localization. The APE and RMSE of the proposed method remain at 0.13 m and 0.15 m in this environment, demonstrating its ability to maintain high-precision localization under dust interference. In the environment with strong light source interference, the localization accuracy of PL-SLAM further deteriorates, with APE and RMSE reaching 3.27 m and 3.64 m, respectively, showing a significant performance decline. While PL-EDLines-SLAM still outperforms PL-SLAM under these conditions, its APE and RMSE are still higher than those of the proposed method, at 1.05 m and 1.26 m, respectively. The proposed method has an APE of only 0.15 m and an RMSE of 0.17 m under strong light source interference, demonstrating its stability and robustness in environments with significant lighting changes. In the presence of both dust and strong light interference, the performance of PL-SLAM deteriorates the most significantly, with an APE of 3.28 m and an RMSE of 3.65 m, rendering it almost incapable of providing reliable localization results. PL-EDLines-SLAM performs slightly better than PL-SLAM but still shows much higher errors compared with the proposed method. The proposed method maintains an APE of 0.15 m and an RMSE of 0.17 m under these complex conditions, verifying its adaptability and accuracy in the face of multiple interferences.

To comprehensively evaluate the performance of the proposed method, not only localization accuracy and robustness but also computational efficiency were analyzed. We compared the average computational time per frame of the

proposed binocular visual localization method, PL-SLAM, and PL-EDLines-SLAM under different environmental conditions. All experiments were conducted on the same hardware platform, and the average time was measured for each condition.

Table 1 the average computational time per frame under different interference scenarios (ms).

Method	No	Dust	Strong Light	Dust+Strong Light
Proposed Method	35.2	36.1	36.8	37.5
PL-SLAM	20.4	23.0	23.5	24.2
PL-EDLines-SLAM	27.6	28.3	29.0	29.8

Table 1 shows the average computational time per frame of different methods under various interference conditions. The proposed method, while requiring higher computational time compared to PL-SLAM and PL-EDLines-SLAM due to its adaptive image enhancement and joint point-line optimization, maintains real-time feasibility and delivers superior localization accuracy and robustness.

In summary, the proposed method can stably provide high-precision localization results under various complex environmental interferences. It demonstrates particularly good robustness and lighting adaptability in the presence of dust and strong light interference. Therefore, the proposed binocular visual localization method integrating point and line features has significant advantages in complex environments such as underground coal mines, laying a reliable data foundation for further intelligent control of machine relocation.

4. Discussion

Accurate localization of roadheaders is critical for intelligent control in underground coal mine environments. Although there are no explicit standards for vision-based localization accuracy specific to roadheaders, reference can be made to roadway contour formation standards in coal mining, which typically require a maximum deviation of ± 0.1 m to ensure structural stability and safety. In this study, the proposed binocular visual localization method, based on point and line features, achieved a root-mean-square error (RMSE) of 0.17 m under challenging conditions involving strong lighting and airborne dust. Although this is slightly above the ± 0.1 m threshold, it is close to the practical engineering requirements, offering promising support for path planning, tracking control, and autonomous operation of roadheaders.

In real mining environments, lens contamination and vibration are two major factors that affect visual system performance. Contaminants such as coal dust, moisture, and oil can degrade image quality and reduce feature extraction accuracy. Vibrations may cause image jitter, further contributing to motion blur. Although artificial smoke was used in our experiments to simulate low-visibility dust conditions, the issue of lens contamination was not directly addressed. To mitigate this, strategies such as anti-fouling lens design and online calibration compensation are explored. Specifically, lenses with hydrophobic and oleophobic coatings can resist dust and oil adhesion, while protective covers with automatic cleaning mechanisms can alleviate lens fouling to

some extent. Additionally, when image quality degrades due to contamination or vibration, online calibration using visual-inertial fusion can be employed to compensate for camera intrinsic parameters and pose, enabling real-time adjustments without interrupting system operation.

To further validate the practicality and robustness of the proposed method, future work will focus on the collection of real-world underground data. Given the complex and harsh conditions of coal mines—dim lighting, high dust concentrations, and mechanical shocks—deploying the visual localization system in actual mines and analyzing its performance under real conditions will be crucial. These datasets will guide the optimization of anti-fouling optical systems and online compensation mechanisms, as well as support the development of more robust localization algorithms. Such efforts will contribute to the long-term reliability of roadheader localization systems and accelerate the advancement of intelligent tunneling technologies.

In summary, the proposed binocular visual localization method based on point and line features demonstrates accuracy approaching engineering-level requirements under challenging conditions. However, challenges such as lens contamination and vibration remain to be fully addressed. Through the integration of optical hardware design, online calibration strategies, and real-world validation, we aim to further enhance the performance and reliability of the localization system to meet the demands of future intelligent tunneling operations.

5. Conclusion and Future work

(1) The adaptive multi-interval image enhancement method proposed in this paper can effectively address the issue of low illumination with strong light protrusion in underground coal mines. Using real image data from underground coal mines for enhancement processing, the effectiveness of the image enhancement method was verified through both subjective and objective evaluations, providing stable image data for subsequent point and line feature extraction. The EDLines-based line segment feature extraction method proposed in this paper has better real-time performance and extraction results compared with LSD and FLD line segment feature extraction methods. Based on this, the method of short line merging and broken line splicing was used to eliminate short lines, laying the foundation for subsequent line segment matching.

(2) A binocular visual localization technique based on point-line feature fusion is proposed. Plücker coordinates are used for line segment feature representation, and the Jacobian matrix of point-line projection errors is constructed. A dynamic weighted fusion method for pose estimation using point and line features is proposed to achieve precise pose estimation. This overcomes the limitation of requiring external parameter calibration for cooperative targets in traditional vision-based schemes, ensuring the continuity of measurement.

(3) The localization method was tested using the EuRoc MAV and KITTI datasets. Compared with other algorithms,

the proposed localization method has an APE of 0.10 m and an RMSE of 0.13 m, demonstrating higher detection accuracy and robustness. Additionally, localization experiments were designed under complex working conditions. The results show that the proposed method has better adaptability in low-illumination environments with dust and strong light interference. Compared with other methods, the average APE and RMSE are the lowest, at 0.15 m and 0.17 m, respectively, indicating superior localization results. The proposed method achieves acceptable computational efficiency. Although the processing time is slightly higher than that of PL-SLAM and PL-EDLines-SLAM, it meets the requirements for real-time or near-real-time applications.

(4) This paper provides an approach to addressing the localization problem of roadheaders, but practical implementation still faces challenges such as vibrations and lens contamination. Therefore, future work should further enhance anti-vibration camera mounting structures, research anti-fouling lens technologies, and establish periodic calibration techniques.

References

- [1] Ma H, Zhang H, Yang K, Hu Y, Yang Z 2023 Research on the Positioning Accuracy of the Cutting Head of a Tunneling Machine Based on Ultra-Wideband Positioning Technology *Processes* **11** 2534.
- [2] Yin Z, Qi Y, Zhang K, Zheng R, Du M, Zhang Z, Tian S, Liu Y, Sun L and Wang Z 2025 Research on integrated positioning and monitoring system of roadheader based on inertial navigation/laser perception *Measurement*. 116718.
- [3] Fu S, Li Y, Zhang M, Zong K, Cheng L, and Wu M 2017 Ultra-wideband pose detection system for boom-type roadheader based on Caffery transform and Taylor series expansion *Meas. Sci. Technol.* **29** 015101.
- [4] Yan C, Zhao W, Lu X 2019 A multi-sensor based roadheader positioning model and arbitrary tunnel cross section automatic cutting *Sensors* **19** 4955.
- [5] Androulakis V, Sottile J, Schafrik S, and Agioutantis Z 2021 Navigation system for a semi-autonomous shuttle car in room and pillar coal mines based on 2D LiDAR scanners. *Tunn. Undergr. Space Technol.* **117**, 104149.
- [6] Yang W, Zhang X, Ma H and Zhang G 2020 Laser beams-based localization methods for boom-type roadheader using underground camera non-uniform blur model *IEEE Access* **8** 190327-190341.
- [7] Du Y, Tong M, Liu T and Dong H 2016 Visual measurement system for roadheaders pose detection in mines *Opt. Eng.* **55** 104107.
- [8] Yang W, Zhang X, Ma H, Zhang G 2019 Infrared LEDs-based pose estimation with underground camera model for boom-type roadheader in coal mining *IEEE Access* **7** 33698-33712.
- [9] Wu H, Liu S, Cheng C, Cao S, Cui Y, Zhang D, 2021 Multiscale variational autoencoder aided convolutional neural network for pose estimation of tunneling machine using a single monocular image *IEEE Trans. Ind. Inform.*

- 18** 5161–5170.
- [10] Cheng J, Wang D, Zheng W, Wang H, Shen Y, and Wu M 2023. Position measurement technology of boom-type roadheader based on binocular vision *Meas. Sci. Technol.* **35** 026301.
 - [11] Yu R, Fang X, Hu C, Yang X, Zhang X, Zhang C, Yang W, Mao Q and Wan J 2022 Research on positioning method of coal mine mining equipment based on monocular vision *Energies* **15** 8068.
 - [12] Cui Y, Pu J, Hu N, Guo Y, Zhou Y, and Liu S 2024 Autonomous Positioning for Mobile Vehicles Based on Visual-Inertial Fusion in Challenging Dark Roadway Scenes *J. Field Robot.* **41** 1–15
 - [13] Wang L, Hu C, Pan G, Zhang C, Liu B, Wan J, Yang H, Shi S, Dong Z, Du Y, Yang W and Zhang X 2022 Pose measurement technology of roadheader body based on fusion of visual and SINS *J. Phys.: Conf. Ser.* **2363** 012014.
 - [14] Wu H, Liu S, Cheng C, Cao S, Cui Y, Zhang D 2021 Multiscale variational autoencoder aided convolutional neural network for pose estimation of tunneling machine using a single monocular image *IEEE Trans. Ind. Inform.* **18** 5161-5170.
 - [15] Zhang W, Zhai G, Yue Z, Pan T and Cheng R 2021 Research on visual positioning of a roadheader and construction of an environment map *Appl. Sci.* **11** 4968.
 - [16] Cui Y, Liu S, Li H Gu C, Jiang H and Meng D 2023 Accurate integrated position measurement system for mobile applications in GPS-denied coal mine *ISA Trans.* **139** 621-634.
 - [17] Chen H, Yang W, Ma Y and Tian L 2021 Multi-sensor fusion method for roadheader pose detection *Mechatronics* **80** 102669.
 - [18] Li D, Liu S, Xiang W, Tan Q, Yuan K, Zhang Z and Hu Y 2021 A SLAM system based on RGBD image and point-line feature *IEEE Access* **9** 9012-9025.
 - [19] Li Q, Wang X, Wu T and Yang H 2022 Point-line feature fusion based field real-time RGB-D SLAM *Comput. Graph.* **107** 10-19.
 - [20] Wang R, Di K, Wan W and Wang Y 2018 Improved point-line feature based visual SLAM method for indoor scenes *Sensors* **18** 3559.
 - [21] Zhou F, Zhang L, Deng C and Fan X 2021 Improved point-line feature based visual SLAM method for complex environments *Sensors* **21** 4604.
 - [22] Huang S, Ren W, Li M 2025 PLFF-SLAM: A Point and Line Feature Fused Visual SLAM Algorithm for Dynamic Illumination Environments *IEEE Access* **13** 1–12.
 - [23] Liu X, Wen S, Zhang H 2023 A real-time stereo visual-inertial SLAM system based on point-and-line features *IEEE Trans. Veh. Technol.* **72** 5747-5758.
 - [24] Fang B and Zhan Z 2020 A visual SLAM method based on point-line fusion in weak-matching scene. *Int. J. Adv. Robot. Syst.* **17** 1729881420904193.
 - [25] Jin L, Zhang H, Meng X, Xie X, Wang G and Guo B 2025 Efficient point-line visual-inertial SLAM combined with EDLines and line flow optimization *Eng. Res. Express* **7** 015289.
 - [26] Ehret T and Morel J M 2024 Line segment detection: A review of the 2022 state of the art *Image Process. Online* **14** 41-63.
 - [27] Gavaskar R G and Chaudhury K N 2018 Fast adaptive bilateral filtering *IEEE Trans. Image Process.* **28** 779-790.
 - [28] Teplyakov L, Erlygin L and Shvets E 2022 Lsdnet: Trainable modification of lsd algorithm for real-time line segment detection *IEEE Access* **10** 45256-45265.
 - [29] Kumar U, Patel G A, Patel H P, Chudhari R P and Darji S S 2021 Impact of frontline demonstration programme on the yield of chickpea (*Cicer arietinum* L.) in Patan District of Gujarat, India. *Legume Res.* **44** 221-224.
 - [30] Dai J S and Sun J 2020 Geometrical revelation of correlated characteristics of the ray and axis order of the Plücker coordinates in line geometry *Mech. Mach. Theory* **153** 103983.
 - [31] Burri M, Nikolic J, Gohl P, Schneider T, Rehder J, Omari S and Siegwart R 2016 The EuRoC micro aerial vehicle datasets[J]. *Int. J. Robot. Res.* **35** 1157-1163.
 - [32] Geiger A, Lenz P, Stiller C and Urasun R 2013 Vision meets robotics: The kitti dataset. *Int. J. Robot. Res.* **32** 1231-1237.
 - [33] Gomez-Ojeda R, Moreno F A, Zuniga-Noël D, Scaramuzza D and Gonzalez-Jimenez J 2019 PL-SLAM: A stereo SLAM system through the combination of points and line segments *IEEE Trans. Robot.* **35** 734-746.