International Journal of *Geo-Information*

MDPI

Article

Robust Visual SLAM with Multi-Level Adaptive Image Enhancement

Qiaobin Dai [1], Zhe Yue [1,2,*], Wangyang Yu [1], Xuerong Zhang [1] and Zengzeng Lian [1]

[1] School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China; dqb@home.hpu.edu.cn (Q.D.); ywyang@home.hpu.edu.cn (W.Y.); zxuer@home.hpu.edu.cn (X.Z.); zengzenglian@163.com (Z.L.)
[2] Henan Province Spatial Big Data Acquisition Equipment Development and Application Engineering Technology Research Center, Henan Polytechnic University, Jiaozuo 454003, China
* Correspondence: yuezhe@hpu.edu.cn

Abstract

To address the limitation that existing Visual Simultaneous Localization and Mapping (VSLAM) methods fail under complex and variable illumination conditions due to the inability to extract sufficient feature points, this paper proposes a robust V SLAM localization method based on multi-level adaptive image enhancement. First, the method employs dynamic brightness compensation to preprocess the original image, initially improving the global brightness distribution. Second, through RGB-to-HSV color space conversion, the brightness V-channel is separated to eliminate the interference of color information in the enhancement process. On this basis, to overcome the limitation of the existing CLAHE algorithm that relies on a fixed clipping threshold and cannot adapt to the local brightness distribution and texture complexity of different image regions, we propose an improved adaptive-threshold CLAHE algorithm based on local statistical characteristics, providing a stable image foundation for feature extraction. Meanwhile, to handle the interference of moving objects in dynamic environments, we incorporate a YOLOv5 object detection thread into the ORB-SLAM3 framework to remove feature points on dynamic objects. This detection module works in synergy with the multi-level image enhancement module, further improving localization robustness in dynamic scenarios. Extensive experiments on the public EuRoC and TUM datasets demonstrate that our method reduces the root mean square error of absolute trajectory error by 29.60% compared to ORB-SLAM3, with a reduction of up to 97.85% on high-dynamic sequences. Our method achieves better localization accuracy and robustness under complex illumination conditions, offering a new solution for visual localization in challenging illumination scenarios.

Keywords: VSLAM; complex lighting; image enhancement; adaptive CLAHE; dynamic environments

Academic Editors: Eliseo Clementini, Zhiyong Zhou and Wolfgang Kainz

Received: 10 June 2026
Revised: 4 July 2026
Accepted: 10 July 2026
Published: 11 July 2026

1. Introduction

Visual Simultaneous Localization and Mapping (VSLAM) is a key technology applied in the fields of intelligent robotics, computer vision, and autonomous driving [1–4]. It mainly relies on visual sensors (e.g., various cameras) to acquire environmental information. Common VSLAM methods include LSD-SLAM [5], DSO [6], ORB-SLAM [7], and SVO [8]. Depending on the principle of odometry computation, the use of VSLAM technology can be divided into the direct method and the feature-point method. In feature-point method-based VSLAM systems, stable and sufficient feature point extraction is the foundation for achieving accurate pose estimation and map construction. The number, quality, and distribution of feature points directly affect the localization accuracy and robustness of the system. However, in real-world application scenarios, feature extraction in VSLAM systems faces severe challenges.

Specifically, existing VSLAM methods face the following illumination-related issues that require further improvement: under extremely low light conditions, the number of feature points is insufficient to support stable pose estimation; under sudden illumination changes, the exposure time cannot be adjusted in time, causing the image to become instantly too dark or overexposed, leading to tracking failure; under non-uniform illumination, feature points tend to be concentrated in certain regions, resulting in incomplete constraints for pose estimation; under overexposure, pixel values in bright regions reach saturation, causing a loss of texture details. Specifically, existing VSLAM methods face the following illumination-related issues that require further improvement: under extremely low light conditions, the number of feature points is insufficient to support stable pose estimation; under sudden illumination changes, the exposure time cannot be adjusted in time, causing the image to become instantly too dark or overexposed, leading to tracking failure; under non-uniform illumination, feature points tend to be concentrated in certain regions, resulting in incomplete constraints for pose estimation; under overexposure, pixel values in bright regions reach saturation, causing a loss of texture details. In complex and variable illumination environments, images captured by cameras are prone to issues such as low contrast, texture deficiency, and excessively low brightness, all of which can degrade the localization accuracy of VSLAM and adversely affect the subsequent mapping accuracy [9]. To address this problem, researchers have conducted extensive studies from two directions: deep learning and image processing. In terms of deep learning-based methods, research has primarily focused on improving feature extraction quality under low-light conditions by employing learnable features or enhancement networks. Zhu [10] was the first to introduce the R2D2 feature extraction network into RGB-D VSLAM, replacing traditional hand-crafted features, and achieved significant improvements in low-texture scenarios; however, the inference latency of deep networks limits its real-time performance. To address this issue, Jiang [11] and Liu [12] made improvements from different perspectives. Jiang combined Retinex enhancement with deep feature optimization to design a system specifically targeting feature matching in low-light environments, while Liu proposed LDFE-SLAM, which integrates three collaborative components to improve the success rate of low-light tracking. Nevertheless, the former suffers from insufficient detail preservation under extreme conditions, and the latter imposes high computational resource requirements due to the parallel operation of multiple networks. Zhao [13] further proposed Light-SLAM, which adopts deep local feature descriptors in place of complete networks, improving efficiency while maintaining illumination adaptability; however, matching quality still drops sharply under instantaneous extreme brightness transitions. Overall, deep learning-based methods still face a trade-off between generalization and real-time performance. In terms of image processing-based methods, Dai [14] proposed ELL-ORB-SLAM, which employs a hybrid dynamic illumination enhancement framework and automatically lowers the detection threshold in dark environments to extract more feature points; however, excessive enhancement may amplify noise in smooth regions. Chen [15] proposed EAMT-SLAM, which utilizes thermal imaging sensors to maintain operation in the absence of illumination but is susceptible to interference from heat sources. Yu [16], from a signal processing perspective, proposed AFE-ORB-SLAM combining truncated AGC with unsharp masking, which reduces the impact of varying illumination conditions on VSLAM, but the multi-layer processing incurs high computational

https://doi.org/10.3390/ijgi15070315

overhead. Canh [17] further introduced an adaptive feature extraction mechanism driven by image entropy and gradient, which dynamically adjusts detection sensitivity based on image content and filters unreliable feature points; however, it relies on empirical parameters and has limited cross-scenario generalization capability. Hou [18] approached the problem from the camera parameter side, employing Bayesian optimization to dynamically adjust exposure settings, effectively improving SLAM performance under changing illumination conditions, yet suffers from optimization lag under rapid illumination changes. Xu [19] proposed AirSLAM, which fuses point and line features to construct a multi-stage relocalization method, but degenerates to point-only features in scenarios with extremely sparse texture. In the main line of contrast enhancement, research has undergone a clear evolution from global equalization to local adaptive enhancement. Fang [20] proposed HE-SLAM, which was the first to introduce histogram equalization into the SLAM front end. It offers fast computation and effectively increases the number of extractable feature points; however, HE indiscriminately enhances global contrast, which tends to amplify noise in complex scenes. To mitigate this drawback, Yang [21] introduced a contrast-limiting mechanism and proposed CLAHE-SLAM, which avoids excessive gain in noisy regions through block-wise processing and clipping limits, thereby improving visual naturalness. Nevertheless, CLAHE with fixed parameters provides limited improvement under extreme illumination conditions and incurs higher computational complexity. On this basis, Aihua [22] adaptively adjusted the clipping limit using image standard deviation, eliminating the need for manual parameter tuning; however, in scenes with coexisting bright and dark regions, a luminance-only adaptive strategy still struggles to balance both ends simultaneously. In summary, existing methods either adopt a single enhancement strategy that cannot simultaneously address brightness recovery, contrast enhancement, and noise suppression, or rely on deep networks and empirical parameters, leading to a contradiction between real-time performance and generalization.

Therefore, this paper proposes a robust visual SLAM localization method based on multi-level adaptive image enhancement, which handles different degradation conditions through a hierarchical processing scheme to compensate for the deficiencies of existing single enhancement approaches. This paper takes ORB-SLAM3 [23] as the basic framework and adopts dynamic brightness compensation, color space conversion, improved CLAHE enhancement, and YOLOv5-based dynamic feature removal to improve the system's adaptability and localization accuracy.

The contributions of this paper are as follows:

1. To address image quality issues, based on the brightness distribution and characteristics of the image, brightness compensation is applied to images that do not meet quality standards. The brightness value of each pixel is dynamically adjusted to ensure high image quality under different lighting conditions.
2. Based on the existing CLAHE image enhancement algorithm, this paper proposes an adaptive CLAHE algorithm that adaptively adjusts the contrast limit by incorporating image gradient and texture complexity information, achieving superior image enhancement.
3. In dynamic scenes, this paper organically integrates the multi-level adaptive image enhancement method with an object detection network. The enhanced data enable more stable recognition and extraction of sufficient feature points, while also accurately identifying dynamic objects and eliminating feature points on dynamic objects, thereby achieving accurate subsequent camera pose estimation.
4. To verify the accuracy and robustness of the proposed method, experimental validation is conducted on different datasets. The experimental results demonstrate that the proposed method achieves good localization accuracy and robustness in challenging lighting scenes.

2. Technical Route of the Proposed Method

Visual SLAM localization technology relies on visual sensors to capture and match multi-dimensional features of the environment to achieve real-time localization and map construction. Under ideal lighting conditions, this technology can provide high-precision localization services, demonstrating excellent localization accuracy and system robustness. However, under complex and variable lighting conditions, the difficulty of feature extraction increases significantly, leading to accumulated localization errors and, in severe cases, even complete loss of localization capability. Furthermore, feature points on dynamic objects in the environment may be misjudged as static landmark points, further interfering with feature matching and camera pose estimation, resulting in trajectory drift and localization bias. To address this challenge, this paper proposes a visual SLAM localization method based on multi-level adaptive image enhancement. The overall technical route is shown in Figure 1.

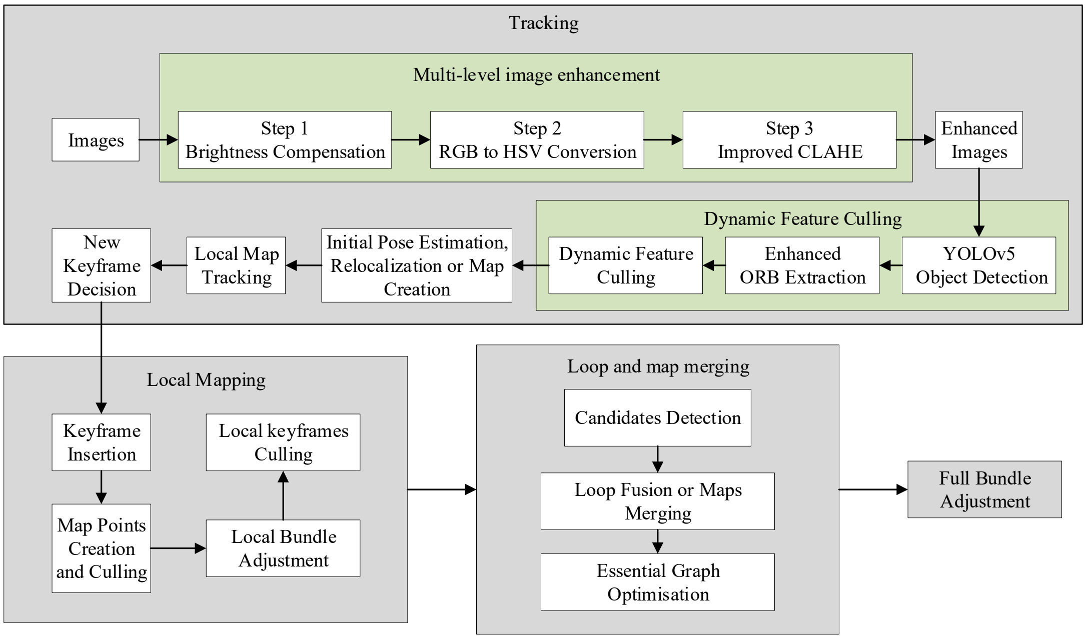


Figure 1. Overall algorithm structure (green indicates improved components).

As shown in Figure 1, the process first dynamically processes the image through an adaptive adjustment algorithm to effectively enhance image details and improve contrast. First, by analyzing the brightness distribution and characteristics of the image, the mean brightness and standard deviation of brightness are calculated to assess image quality. Adaptive brightness compensation is applied to images that do not meet quality standards, dynamically adjusting the brightness value of each pixel to ensure high image quality under different lighting conditions. Subsequently, the enhanced image is converted from the RGB color space to the HSV color space [24], and an improved CLAHE algorithm is applied to the brightness (V) channel to further enhance image brightness and contrast while maintaining image smoothness. This multi-level optimization strategy significantly improves image quality, which not only enhances the stability of feature extraction but also, by improving contrast and detail, makes object features clearer and boundaries more distinct, thereby enabling more accurate identification and selection of feature points under complex lighting conditions. On this basis, the algorithm introduces YOLO [25] object detection technology to accurately identify dynamic objects in complex lighting scenes, eliminating interference from dynamic feature points. By removing feature points on dynamic objects, the system reduces the impact of dynamic factors on pose estimation. This method effectively reduces trajectory deviation and drift, enabling the visual SLAM system to maintain high-precision localization capability even in dynamic environments. The method proposed in this paper significantly improves the robustness and accuracy of visual SLAM in scenarios with complex lighting conditions.

In summary, compared with existing CLAHE-based and illumination-robust SLAM methods, our method has the following key distinctions: unlike single-stage enhancement performed in the RGB space, our method employs a multi-level architecture comprising brightness compensation, HSV color space conversion, and improved CLAHE to progressively improve image quality; standard CLAHE uses a fixed clipping limit, whereas our method dynamically adjusts the limit according to local gradient and texture complexity of the image, suppressing noise amplification in texture-sparse regions while preserving details in texture-rich regions; existing illumination-robust SLAM methods mostly focus solely on enhancement itself, while our method further synergizes the enhancement module with YOLOv5-based dynamic feature removal, improving detection recall rate and enhancing localization accuracy.

3. Overview of Principles and Methods

This section presents the principles and techniques of the proposed method in detail, including dynamic brightness compensation, color space conversion, improved CLAHE algorithm, and dynamic feature detection and removal. The overall structure of the algorithm is shown in Figure 2.

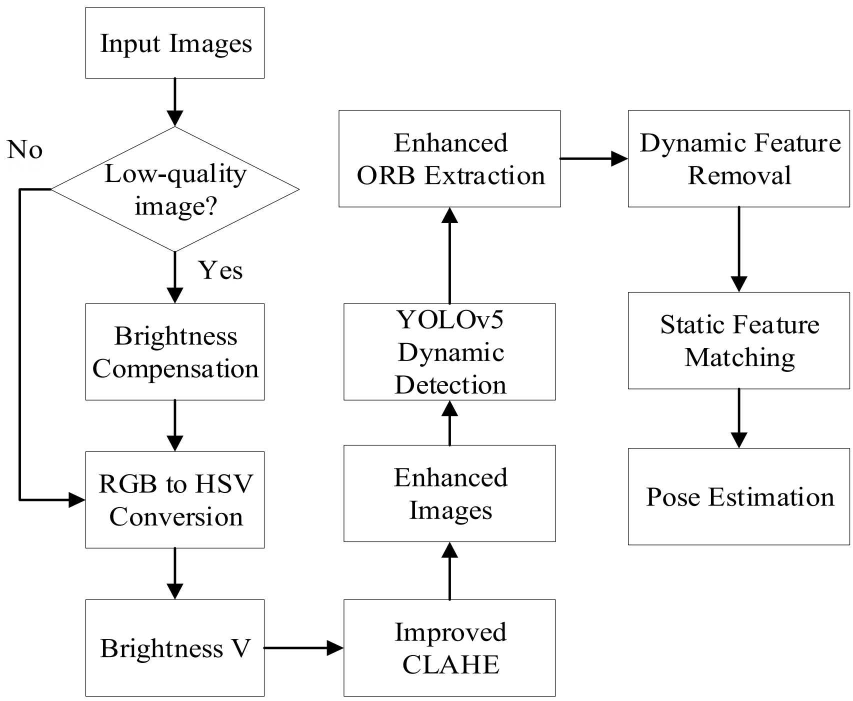


Figure 2. Structure of the proposed algorithm.

As shown in Figure 2, the input image first undergoes quality assessment, followed sequentially by dynamic brightness compensation, HSV color space conversion, and improved adaptive CLAHE enhancement, producing a high-quality enhanced image for subsequent feature extraction. Meanwhile, the enhanced image is fed into the YOLOv5 object detection thread to identify dynamic objects. After the ORB feature points within the detected dynamic regions are removed, only the static feature points are retained for pose estimation.

https://doi.org/10.3390/ijgi15070315

3.1. Image Enhancement

To address the problem of complex illumination affecting the localization accuracy of VSLAM, this paper adopts a multi-level image enhancement method, which mainly consists of three components: dynamic brightness compensation, color space conversion, and improved CLAHE enhancement.

3.1.1. Dynamic Brightness Compensation

Under complex lighting conditions, the captured images may suffer from excessively low brightness or uneven illumination, which severely affects the accuracy of feature point extraction and localization precision. In such cases, simply applying image enhancement techniques often fails to achieve the desired results. Therefore, targeted brightness compensation is required for such images to improve image quality and provide a better image foundation for subsequent feature point extraction and localization.

To address this issue, this paper proposes a dynamic brightness compensation method, which dynamically adjusts the brightness value of each pixel by analyzing the brightness distribution and characteristics of the image. Before performing brightness adjustment, the image quality is first evaluated. Brightness adjustment is only applied when the image quality falls below a certain threshold, thereby avoiding unnecessary processing on high-quality images. The specific steps are as follows:

First, the luminance mean L_{mean} of the image is calculated as shown in Equation (1):

$$L_{mean} = \frac{\sum_{x=1}^{n}\sum_{y=1}^{m} L(x,y)}{m \times n} \tag{1}$$

where n and m represent the width and height of the image, respectively. where $L(x,y)$ denotes the luminance value at point (x,y).

Secondly, the luminance standard deviation L_{std} of the image is calculated as shown in Equation (2):

$$L_{std} = \sqrt{\frac{\sum_{x=1}^{n}\sum_{y=1}^{m}(L(x,y) - L_{mean})^2}{m \times n}} \tag{2}$$

Subsequently, a decision function Q is defined to determine whether luminance adjustment is required, shown in Equation (3):

$$\mathrm{Q} = \begin{cases} 0, \text{if } |L_{mean} - 0.5| < \varepsilon \text{ and } L_{std} > \theta \\ 1, \text{otherwise} \end{cases} \tag{3}$$

where ε is a small positive constant used to assess the proximity of L_{mean} to 0.5, and θ is a positive constant for determining whether L_{std} is sufficiently large. When the mean brightness is far from 0.5 and the standard deviation of brightness is large, output $\mathrm{Q}=1$, perform brightness adjustment on the image; output $\mathrm{Q}=0$ indicates that the image quality is good and no adjustment is required. We selected thousands of images from the EuRoC dataset used in this paper. After trigger validation using multiple sets of parameters, we found that the trigger accuracy is relatively high when ε is between 0.10 and 0.15 and θ is between 0.20 and 0.25. Specifically, we selected $\varepsilon = 0.12$ and $\theta = 0.22$ as the thresholds for triggering dynamic compensation.

The adjustment formula used is shown in Equation (4):

$$L(x,y) = L_{mean} \times L(x,y)^{\alpha} + (1 - L_{mean}) \times L(x,y)^{\beta} \quad (4)$$

Equation (4) refers to reference [26], where α is the enhancement correction parameter for low-brightness regions, and β is the suppression correction parameter for high-brightness regions. In this paper, α and β are set to 0.20 and 1.20, respectively.

3.1.2. Improved CLAHE Algorithm Based on HSV

After performing dynamic brightness compensation, this section proceeds with image enhancement to improve image contrast and detail. For RGB image data, this paper introduces color space conversion technology to transform the RGB color space into the HSV color space. The HSV color space describes color information in an image through three dimensions: Hue (H), Saturation (S), and Value (V). Among these, Hue defines the type of color, Saturation measures the purity of the color, and Value describes the brightness of the color. Enhancing the V channel is equivalent to directly enhancing the grayscale image. ORB feature extraction relies on image grayscale information to compute FAST corners and BRIEF descriptors. The V channel of HSV directly reflects pixel brightness, so enhancing the V channel directly improves the identifiability of the grayscale structural information on which feature extraction depends. If enhancement is performed separately on the three RGB channels before conversion to grayscale, the varying enhancement amplitudes across different channels may cause the composite grayscale value to deviate from the true brightness distribution, leading to degraded grayscale consistency of the same scene under different viewpoints, which in turn compromises the stability of feature matching. Moreover, HSV conversion involves only simple arithmetic operations and does not significantly increase the processing latency of the SLAM front end, ensuring efficient integration between the enhancement module and feature extraction.

The conversion from the RGB color space to the HSV color space can be achieved using the following mathematical formulas:

Step 1: Calculate the maximum and minimum values of RGB.

$$\begin{aligned} M &= \max(R,G,B) \\ m &= \min(R,G,B) \\ C &= M - m \end{aligned} \quad (5)$$

where M denotes the maximum value among the three RGB color components, m represents the minimum value among the three RGB color components, C indicates the chroma difference, R corresponds to the red color component, G to the green color component, and B to the blue color component.

Step 2: Calculate H

$$H = \begin{cases} 0^{\circ} & \text{if } C = 0 \\ 60^{\circ} \times \left(\dfrac{G-B}{C} \bmod 6 \right) & \text{if } M = R \\ 60^{\circ} \times \left(\dfrac{B-R}{C} + 2 \right) & \text{if } M = G \\ 60^{\circ} \times \left(\dfrac{R-G}{C} + 4 \right) & \text{if } M = B \end{cases} \quad (6)$$

Step 3 Calculate S

$$S = \begin{cases} 0, \text{if } M = 0 \\ \dfrac{C}{M}, \text{otherwise} \end{cases} \tag{7}$$

Step 4 Calculate V

$$V = M \tag{8}$$

where H denotes hue, S represents saturation, and V indicates brightness.

Through the above steps, the conversion from the RGB color space to the HSV color space can be achieved. The improved CLAHE image enhancement technique is applied to the Value (V) channel of the converted HSV color space, which can more effectively improve contrast in low-light or illumination-varying scenes, thereby making feature point extraction more stable and increasing the number of feature points. CLAHE is an improved strategy over the HE algorithm, overcoming the problem of excessive noise enhancement that HE may cause when processing images. However, because the fixed-parameter CLAHE algorithm uses a fixed contrast limit without considering the specific content and characteristics of the image, it may fail to effectively preserve details in high-texture regions and may excessively enhance noise in low-texture regions. To address this issue, we further propose an adaptive CLAHE image enhancement algorithm.

The specific procedure is as follows:

First, local histogram equalization is applied. For each local region, its gray-level distribution is statistically analyzed to compute the grayscale histogram, which is achieved by the following Equation (9):

$$h_m(g) = \sum_{(x,y)\in R} \delta(I(x,y) - g), \quad g = 0,1,\ldots,255 \tag{9}$$

where $I(x,y)$ denotes the gray-level value of the input image at coordinate (x,y), δ is the Kronecker delta function defined as: it equals $\delta = 1$ when the input is 0, and $\delta = 0$ otherwise, and $h_m(g)$ represents the frequency of occurrence of gray level g within the local region.

Secondly, contrast limitation is applied. A parameter limit is introduced to prevent local over-enhancement caused by excessively high frequencies of certain gray levels. The CLAHE parameter limit is set adaptively according to the complexity of the image content: if the image contains rich textures or abundant details, limit is assigned a smaller value to avoid excessive enhancement of fine structures; conversely, for relatively uniform images, limit is increased. When the frequency of a histogram—that is, the pixel count of a given gray level exceeds the threshold limit, clipping is performed on that gray-level frequency, and the excess portion is redistributed smoothly to suppress noise over-amplification. The clipping of histogram $h_m(g)$ is implemented as shown in Equation (10).

$$h_{\text{clip}}(g) = \min\left(h_m(g), \text{limit}\right) \tag{10}$$

where limit denotes the histogram clipping upper limit, and $h_{\text{clip}}(g)$ represents the histogram after clipping.

To address the limitations of a fixed limit, this paper proposes an adaptive limit parameter setting method. By dynamically adjusting the contrast limit based on image gradient and texture complexity, the proposed approach can adaptively respond to the characteristics of different images, offering improved adaptability and flexibility. Furthermore, the resulting image enhancement is more uniform and natural, optimized according to the specific features of each image. This makes the method suitable for various complex scenarios, particularly those requiring fine-tuned contrast limitation to preserve details while suppressing noise. The specific implementation is given by Equation (11).

$$\text{limit} = \left(1 - \frac{\sigma}{\sigma_{\max}}\right) + \left(1 - \frac{\mathrm{d}}{\mathrm{d}_{\max}}\right) + \mathrm{c} \tag{11}$$

where σ is the gradient standard deviation of the current image, d is the texture complexity of the current image, $\sigma_{\max}$ is the maximum gradient standard deviation in the dataset, and $\mathrm{d}_{\max}$ is the maximum texture complexity in the dataset. Both of these values can be obtained through statistical analysis of the dataset. where c is the base contrast limit, which ensures the fundamental clipping enhancement effect. When both the gradient and texture complexity in Equation (11) reach their maximum values, the limit result is approximately c itself. Referring to the classical CLAHE [27] algorithm, c is set to 1.0. For other cases of gradient and texture complexity, adaptive clipping limits can be achieved through Equation (11). The calculation formulas for σ and d are given in Equations (12) and (13), respectively.

$$\sigma = \sqrt{\frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (I(i,j) - \bar{I})^2} \tag{12}$$

where $I(i,j)$ is the pixel value at position (i,j) in the image $\bar{I}$ is the average pixel value of the image, and m and n represent the width and height of the image, respectively.

$$d = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (L(i,j) - \bar{L})^2 \tag{13}$$

where $L(i,j)$ denotes the pixel value at position (i,j) after processing with the Laplacian operator, and $\bar{L}$ represents the mean pixel value of the processed image, defined as the arithmetic average of all Laplacian-derived pixel values. Equation (12) dynamically adjusts the clip limit limit by integrating both gradient magnitude and texture complexity information from the image, thereby achieving adaptive enhancement performance.

After clipping, the total number of pixels a in the excess portion is calculated as shown in Equation (14):

$$E = \sum_{g=0}^{255} \max\left(0, h_m(g) - \text{limit}\right) \tag{14}$$

The value E is uniformly redistributed across all gray levels, resulting in the adjusted histogram $h_{\text{redist}}(g)$, as shown in Equation (15):

$$h_{\text{redist}}(g) = h_{\text{clip}}(g) + \frac{E}{256} \tag{15}$$

where the redistributed histogram ensures that the total pixel count remains unchanged and the distribution becomes more uniform.

Finally, bilinear interpolation is employed to achieve smooth transitions. For any pixel (x,y) in the image, the four nearest local region centers (x_i, y_j), (x_{i+1}, y_j), (x_i, y_{j+1}) and (x_{i+1}, y_{j+1}) are first determined. Normalized coordinates are then computed as shown in Equations (16) and (17):

$$u = \frac{x - x_i}{x_{i+1} - x_i} \tag{16}$$

$$v = \frac{y - y_j}{y_{j+1} - y_j} \tag{17}$$

where $u, v \in [0,1]$, u denote the normalized distances in the horizontal direction, representing the positional ratio of the pixel relative to the left region center to the right region center; v denotes the normalized distance in the vertical direction, representing the positional ratio of the pixel relative to the upper region center to the lower region center. The mapped results of the current pixel gray value $I(x, y)$ under the four adjacent local regions are respectively calculated as shown in Equation (18):

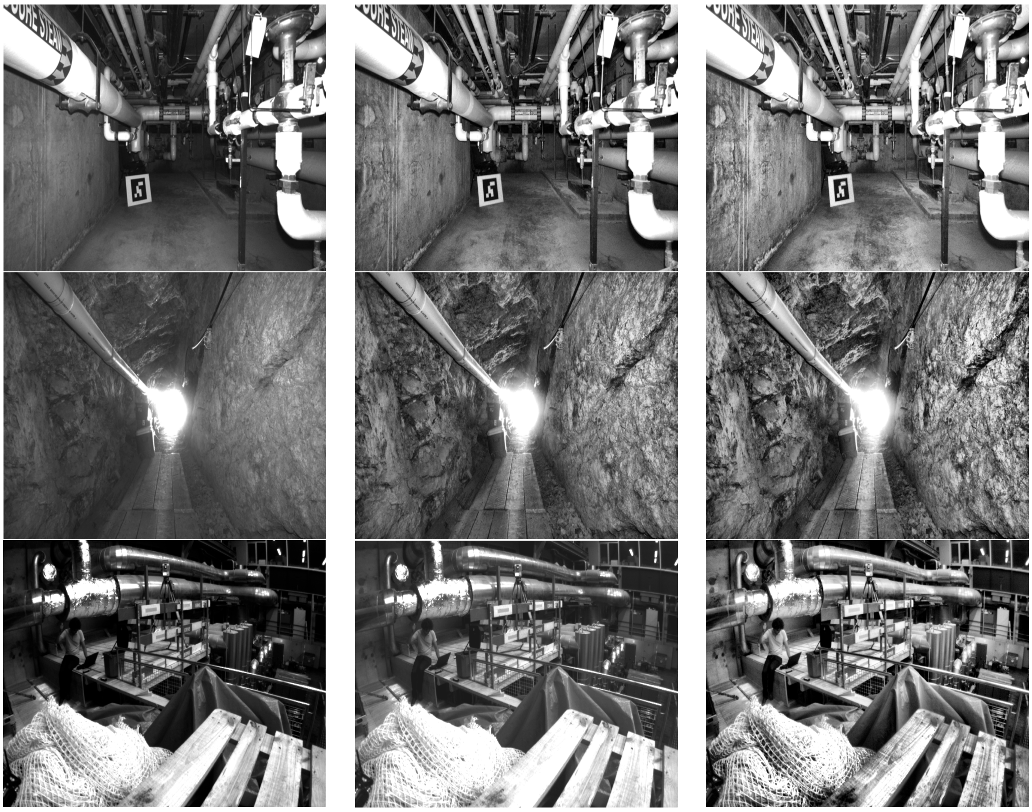
$$\begin{aligned} f_{00} &= T_{i,j}(I(x,y)) \\ f_{10} &= T_{i+1,j}(I(x,y)) \\ f_{01} &= T_{i,j+1}(I(x,y)) \\ f_{11} &= T_{i+1,j+1}(I(x,y)) \end{aligned} \tag{18}$$

where f_{00} represents the mapping result of the pixel in the upper-left region (i, j), f_{10} denotes the mapping result in the upper-right region $(i+1, j)$, f_{01} indicates the mapping result in the lower-left region $(i, j+1)$, and f_{11} corresponds to the mapping result in the lower-right region $(i+1, j+1)$. The bilinear interpolation formula is then applied to integrate these four mapped values, yielding the final enhanced gray value $h_{\text{out}}(x, y)$, as shown in Equation (19):

$$h_{\text{out}}(x,y) = (1-u)(1-v)f_{00} + u(1-v)f_{10} + (1-u)vf_{01} + uvf_{11} \tag{19}$$

This formula achieves smooth transitions through a weighted average of the mapping results from four adjacent local regions. The improved CLAHE achieves adaptive adjustment of the clipping amplitude, requiring only the updating of dataset statistics to automatically adapt to new image distributions. When an image contains rich texture or fine structures, its histogram is already naturally spread across a wide grayscale range without significant sharp peaks. The adaptive clipping limit is automatically raised to a lower level, and the clipping amplitude is small, having little impact on the histogram shape. This prevents the CDF mapping function from over-stretching the already well-spread histogram, thereby protecting fine structures from being excessively amplified. For image regions with gradual grayscale variations, pixels are highly concentrated within a limited grayscale range, and the histogram exhibits a pronounced peak distribution. The adaptive clipping limit is automatically raised to a higher level, allowing peaks to be moderately preserved rather than excessively flattened. This means that the CDF maintains a sufficiently steep slope at the peaks, retaining a stronger stretching effect of the mapping function, thereby mapping the originally concentrated grayscale levels more significantly and effectively enhancing the contrast of flat regions. By improving the CLAHE strategy, the stability of the ORB-SLAM3 system can be ensured even in scenes with illumination variations or when a single image suffers from lighting issues.

In summary of Sections 3.1, to address the degraded localization accuracy of visual SLAM in complex illumination environments, the proposed multi-level adaptive image enhancement method effectively improves image quality, as shown in Figure 3.



(**a**) Original (**b**) CLAHE (**c**) Our

Figure 3. Image enhancement results.

In Figure 3, Figure 3a displays the original image, Figure 3b presents the image enhanced by the classical CLAHE algorithm, and Figure 3c shows the result of our proposed adaptive enhancement method. It is evident that under complex illumination conditions, the quality of the original image in Figure 3a is significantly compromised, manifesting as deficiencies in both contrast and luminance. In comparison, Figure 3b demonstrates improved contrast and brightness through CLAHE enhancement. Our method employs a multi-level adaptive image enhancement approach, building upon dynamic luminance compensation and proposing an adaptive CLAHE algorithm. Unlike conventional CLAHE, our algorithm adopts an adaptive clip limit strategy, proving more effective in complex lighting conditions. As clearly observed in Figure 3, Figure 3c exhibits superior enhancement effectiveness in both contrast and luminance compared to Figure 3b, thereby establishing a high-quality image foundation for subsequent feature point identification.

3.2. Dynamic Feature Point Removal Based on Object Detection

In dynamic scenes, the localization robustness of traditional VSLAM systems faces fundamental challenges. When moving objects occupy a large proportion of the field of view or exhibit high-speed motion characteristics, visual odometry based on the static environment assumption can easily misclassify dynamic feature points as static environmental elements. Such mismatching undermines the accuracy foundation of the entire localization system.

To address the impact of dynamic environments on the localization results of ORB-SLAM3. In this paper, we incorporate a YOLOv5 object detection thread. The detection model is YOLOv5s, initialized with the COCO 2017 pre-trained weights. Our experiments are conducted in indoor environments, with humans as the primary dynamic objects to be detected. The specific integration strategy is as follows: after the tracking thread acquires a new frame, it pushes the frame into the detection queue. The YOLOv5 thread reads the frame from the queue and performs inference, outputting 2D bounding boxes of dynamic objects. After the tracking thread extracts ORB features, it removes the feature points that fall within the dynamic bounding boxes, retaining only the static feature points for subsequent matching and pose estimation. Meanwhile, the image enhancement module described in Section 3.1 and the dynamic feature removal module in this section do not operate independently, but rather form a closed-loop collaborative optimization. The synergy is manifested in two directions: under low illumination or non-uniform lighting conditions, the edge and texture information of dynamic objects in the original image is overwhelmed by noise, resulting in a significant drop in the detection recall rate of YOLOv5. The image enhancement in Section 3.1 improves the image quality at the source, making the contours of dynamic objects clearer and directly enhancing the detection recall rate and bounding box accuracy of YOLOv5 under degraded illumination. If the ORB feature points on dynamic objects are not removed, they will participate in pose estimation as pseudo-static features, introducing moving outlier errors that mask the feature quality improvements brought by enhancement. Dynamic removal eliminates these outliers, allowing the gains from enhancement to be fully translated into localization accuracy improvements. The two modules complement each other.as shown in Figure 4.

Figure 4. Comparison of recognition performance before and after enhancement.

To verify the object recognition capability of the proposed method under complex illumination, Gamma correction is applied to images from the dataset to artificially construct complex lighting scenes. As shown in Figure 4, under the left-side lighting condition, the recognition performance of YOLOv5 for various objects is subject to interference. Due to low image brightness and contrast, issues such as incomplete object recognition, misrecognition, and even failure to recognize objects occur. In such cases, it becomes difficult to accurately remove dynamic feature points, and subsequent localization will also be affected. In contrast, with the proposed method, image quality is improved, enabling accurate recognition of various objects, thereby providing strong support for subsequent dynamic removal. Based on the accurate recognition of dynamic objects in the environment, the proposed method achieves the removal of dynamic feature points, as shown in Figure 5.

Figure 5. Comparison of feature point extraction performance between ORB-SLAM3 and the proposed method.

As shown in Figure 5, in the presence of dynamic objects such as people, ORB-SLAM3 extracts dynamic feature points on these moving objects, and the use of such feature points severely affects localization accuracy. The proposed method removes these dynamic feature points while retaining more static feature points in the background. Based on a more stable number of static feature points, the subsequent localization performance becomes more accurate.

The workflow of the proposed VSLAM localization algorithm based on multi-level adaptive image enhancement is presented in Algorithm 1.

Algorithm 1: Robust Visual SLAM Localization Method Based on Multi-level Adaptive Image Enhancement

1. Dynamic Luminance Compensation
   1.1 Calculate the luminance mean L_{mean} of the image according to Equation (1):
   1.2 Calculate the luminance mean L_{std} of the image according to Equation (2):
   1.3 Define a decision function Q
      If $\left|L_{mean}-0.5\right|<\varepsilon and L_{std}>\theta$
      Obtain $\mathrm{Q}=0$ according to Equation (3).
      Elseif Obtain $\mathrm{Q}=1$ according to Equation (3).
   1.4 If $\mathrm{Q}=1$, perform luminance compensation according to Equation (4).
2. RGB Color Space Conversion
   2.1 Obtain the maximum and minimum values of RGB according to Equation (5).
   2.2 Compute HSV
      Obtain H according to Equation (6).
      Obtain S according to Equation (7).
      Obtain V according to Equation (8).
3. Adaptive CLAHE Image Enhancement
   3.1 Apply local histogram equalization to the image according to Equation (9).
   3.2 By substituting Equations (12) and (13) into Equation (11), the adaptive calculation of parameter limit is achieved.
   3.3 The clipping of the histogram is implemented according to Equation (10).
   3.4 The total number of pixels in the clipped portion is calculated according to Equations (14) and (15), yielding the corrected histogram $h_{\mathrm{redist}}(g)$.
   3.5 By substituting Equations (16)–(18) into Equation (19), the final enhanced grayscale value $h_{\mathrm{out}}(x,y)$ is obtained.
4. Enhanced Dynamic Feature Point Detection and Removal

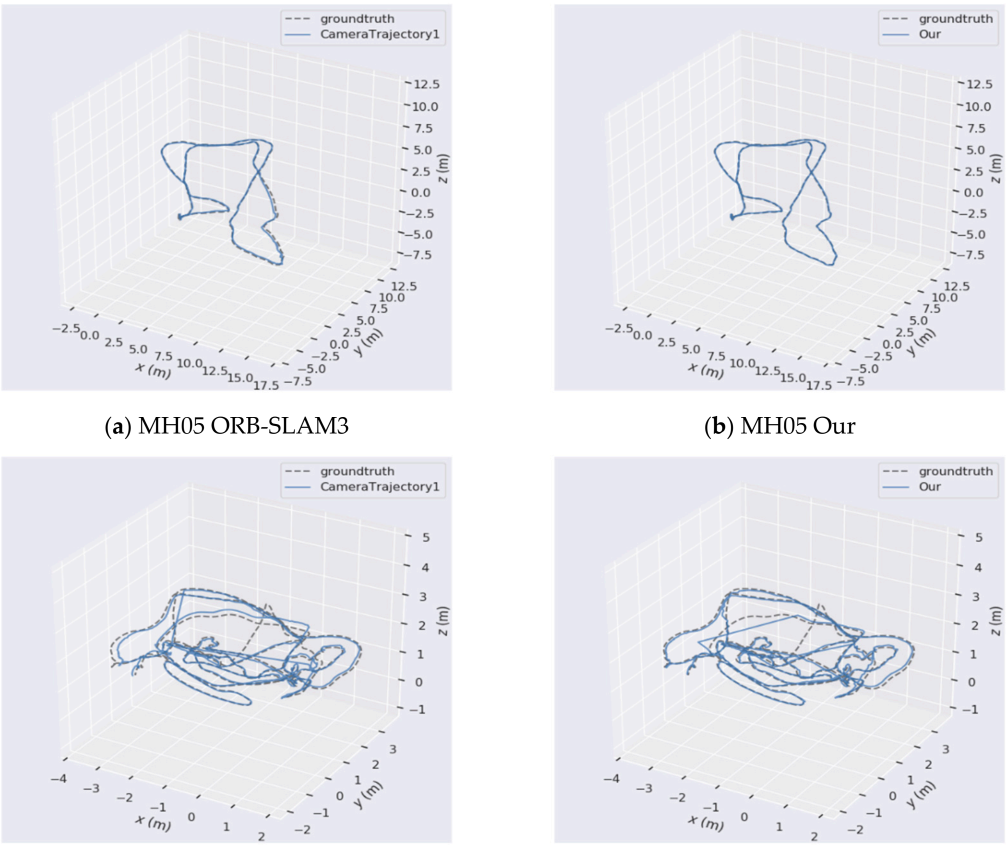
4.1 The image data undergo the aforementioned steps to achieve multi-level enhancement.
4.2 The enhanced data are processed through the YOLO object detection module to accurately identify dynamic objects within them.
4.3 Dynamic feature points located on the detected dynamic objects are removed.
4.4 Using the retained static feature points for feature matching to achieve high-precision pose estimation.

4. Experimental Results and Analysis

To address the problem that different lighting conditions negatively affect the localization accuracy of visual SLAM, this paper takes the ORB-SALM3 algorithm framework as the research object and achieves stable visual localization under different lighting conditions by integrating adaptive image enhancement techniques.

4.1. Experimental Validation Under Complex Lighting Conditions

Experiments were conducted on the MH05 and V203 sequences from the EuRoC dataset. These two sequences are classified as difficult level and contain challenging scenarios, such as varying illumination conditions, severe shaking, and dark environments. The experimental results were compared with those of ORB-SLAM3, and the overall trajectory comparisons are shown in Figure 6.

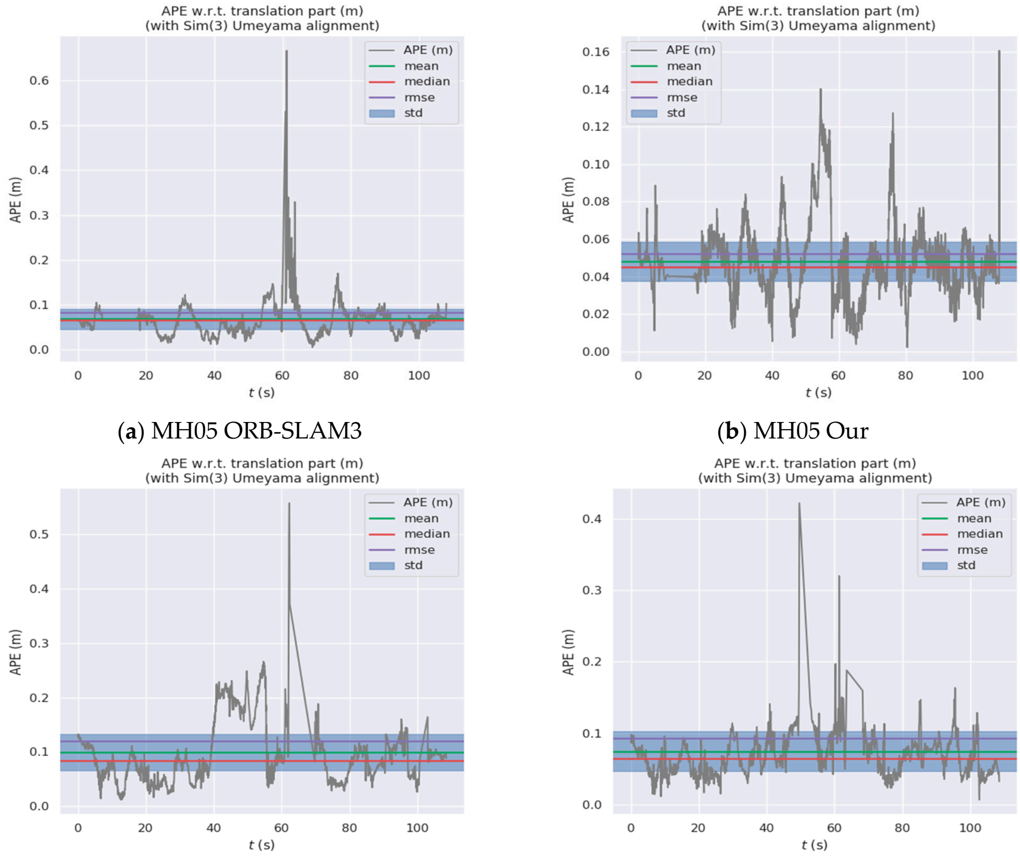


(**a**) MH05 ORB-SLAM3 (**b**) MH05 Our

(**c**) V203 ORB-SLAM3 (**d**) V203 Our

Figure 6. Comparison of overall trajectories between ORB-SLAM3 and the proposed method under complex lighting conditions.

From the overall trajectory comparison in Figure 6, it can be observed that the camera estimated trajectory of ORB-SLAM3 exhibits certain deviations from the ground truth trajectory (marked by the black dashed line) across the entire range, indicating that complex illumination conditions considerably affect the localization accuracy of the VSLAM system. In contrast, the camera estimated trajectory of our proposed method shows a high degree of overlap with the ground truth trajectory, demonstrating that our method achieves better robustness and higher localization accuracy under complex illumination conditions. The detailed absolute trajectory error comparison is shown in Figure 7.



(**a**) MH05 ORB-SLAM3 (**b**) MH05 Our

(**c**) V203 ORB-SLAM3 (**d**) V203 Our

Figure 7. Comparison of absolute trajectory errors.

In Figure 7, it can be observed that the absolute trajectory error of ORB-SLAM3 exhibits relatively large fluctuations overall, indicating that complex illumination conditions significantly affect its localization accuracy. The RMSE value is also relatively large, demonstrating that the localization accuracy of ORB-SLAM3 in this scenario is not satisfactory. In contrast, the absolute trajectory error of our proposed method shows smaller fluctuations overall, and the RMSE value is considerably lower than that of ORB-SLAM3, confirming that our method achieves higher localization accuracy and better robustness under complex illumination conditions. To verify the robustness of the proposed method, Figure 8 presents the results of ten executions on multiple sequences.

As shown in Figure 8, squares of different colors represent different ranges of RMSE values. The darker the color, the larger the RMSE value and the lower the relative localization accuracy; the lighter the color, the smaller the RMSE value and the higher the relative localization accuracy. In the left figure, it can be observed that in some executions of the MH04 and MH05 sequences, darker-colored squares appear, indicating that in these cases, ORB-SLAM3 has a larger RMSE and lower localization accuracy. Compared with the heatmap of ORB-SLAM3 on the left, the colors in the heatmap corresponding to the proposed method are generally lighter, and the number of dark squares is significantly reduced, indicating that the proposed method achieves a smaller RMSE, higher localization accuracy, and greater stability across multiple executions of multiple sequences.

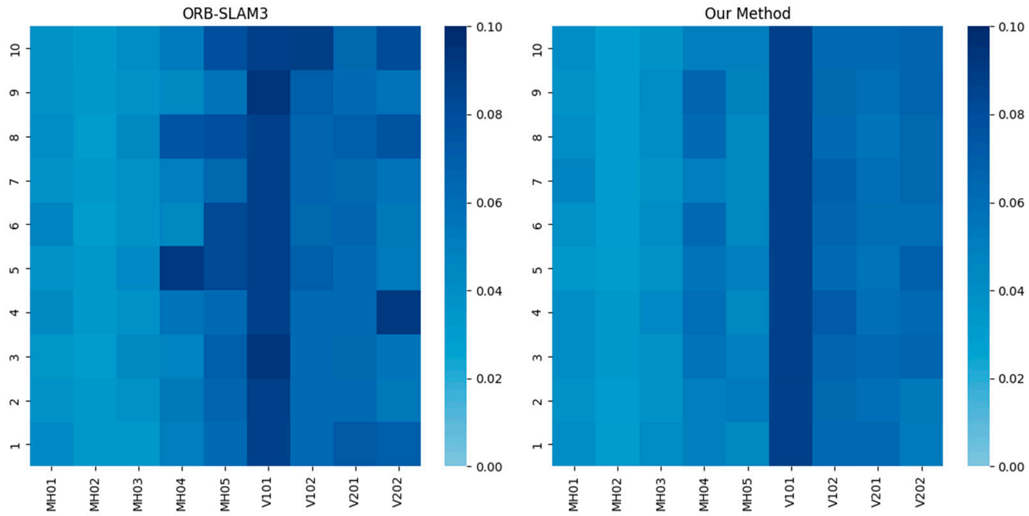


Figure 8. Comparison Between ORB-SLAM3 and the Proposed Method.

To enhance the reference value of our research and the persuasiveness of our proposed method, we have further extended the comparison methods, as shown in Table 1. To further validate the effectiveness of our proposed method under complex illumination scenarios, we compared it with multiple existing methods, including DSO, SVO, DSM [28], ORB-SLAM3, HE-SLAM, CLAHE-SLAM, Retinex-SLAM, and Light-SLAM. These methods encompass standard SLAM, image enhancement-based SLAM, and deep learning-based SLAM. The RMSE comparison results are presented in Table 1.

Table 1. Comparison of RMSE (m) performance on the EuRoC dataset.

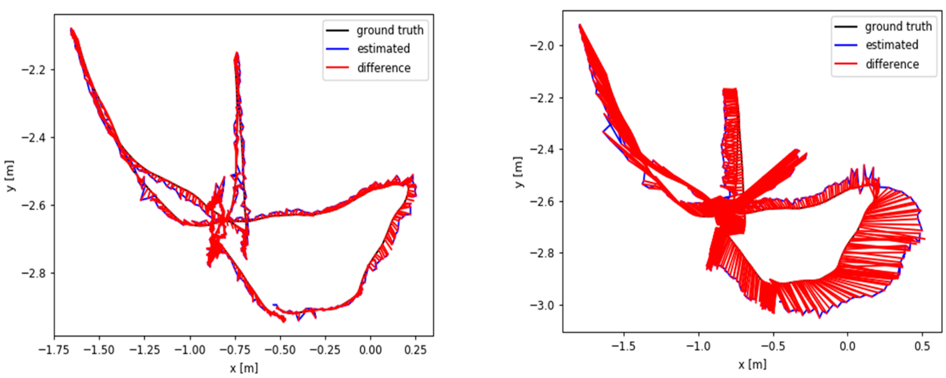
Sequence	DSO	SVO	DSM	ORB-SLAM3	HE-SLAM	CLAHE-SLAM	Retinex-SLAM	Light-SLAM	Our
MH01	0.046	0.100	0.039	0.038	0.041	0.038	0.035	0.035	0.035
MH02	0.046	0.120	0.036	0.032	0.040	0.038	0.032	0.037	0.031
MH03	0.172	0.410	0.055	0.037	0.050	0.040	0.039	0.033	0.037
MH04	3.810	0.430	0.057	0.062	0.105	0.063	0.071	0.112	0.057
V101	0.089	0.070	0.095	0.088	0.088	0.087	0.089	0.084	0.087
V102	0.107	0.210	0.059	0.076	0.064	0.065	0.063	0.058	0.066
V201	0.044	0.110	0.056	0.059	0.062	0.060	0.060	0.065	0.058
V202	0.132	0.110	0.057	0.061	0.063	0.060	0.057	0.052	0.056

Table 1 compares the performance of different VSLAM or visual odometry methods on multiple sequences of the EuRoC dataset. As can be seen from the table, our method achieves either the best or comparable accuracy to other state-of-the-art methods on the vast majority of sequences. Specifically, on the MH01, MH02, MH04, V201, and V202 sequences, our method yields the lowest RMSE values among all compared methods. Notably, MH04 is considered a challenging sequence in the EuRoC dataset with the most severe illumination conditions, and our method achieves the highest localization accuracy on this sequence. This benefits from the multi-level adaptive image enhancement algorithm adopted in our method, which dynamically adjusts the enhancement thresholds according to the structural characteristics of the image sequence, thereby providing more stable image feature points for localization and consequently improving localization accuracy.

Although the proposed multi-level adaptive image enhancement method can effectively improve image quality under complex illumination conditions, it should be noted that any image enhancement technique carries the potential risk of introducing artifacts. Excessive enhancement, such as overly high contrast stretching or improper CLAHE clipping threshold settings, may lead to halo artifacts at image edges, saturation distortion in textured regions, and even amplification of sensor noise in low-texture areas. These artificially introduced spurious structures may be misidentified as valid corners by feature detectors, resulting in a decreased inlier ratio in descriptor matching and consequently affecting pose estimation accuracy. To address this issue, our method dynamically adjusts the contrast limiting parameter of CLAHE by incorporating gradient standard deviation and texture complexity, achieving a balance between enhancement effectiveness and artifact suppression while avoiding the distortion of the original geometric and brightness features caused by excessive enhancement.

4.2. Experimental Validation in Dynamic Environments

To verify the effectiveness of the proposed method in dynamic scenes, the TUM dataset is selected for experimental validation. The TUM dataset, provided by the Technical University of Munich (TUM), is a series of datasets widely used in computer vision and robotics research, particularly for tasks such as visual odometry (VO) and SLAM. The TUM dataset contains image sequences from various scenes covering different parts of indoor environments. In this paper, the dynamic data sequences fr3_walking_half, fr3_walking_rpy, fr3_walking_static, and fr3_walking_xyz from the TUM dataset are selected for experimental validation, and the results are compared with those of ORB-SLAM3. The results are shown in Figures 9–12.



(**a**) Our (**b**) ORB-SLAM3

Figure 9. Comparison of trajectory errors on fr3_walking_half.

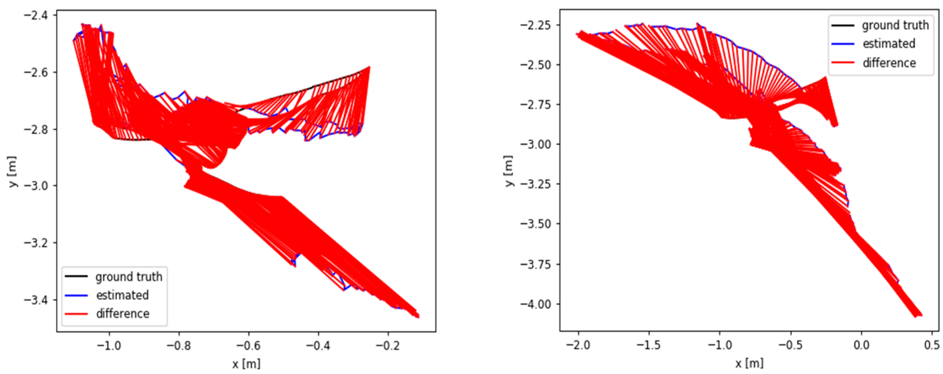
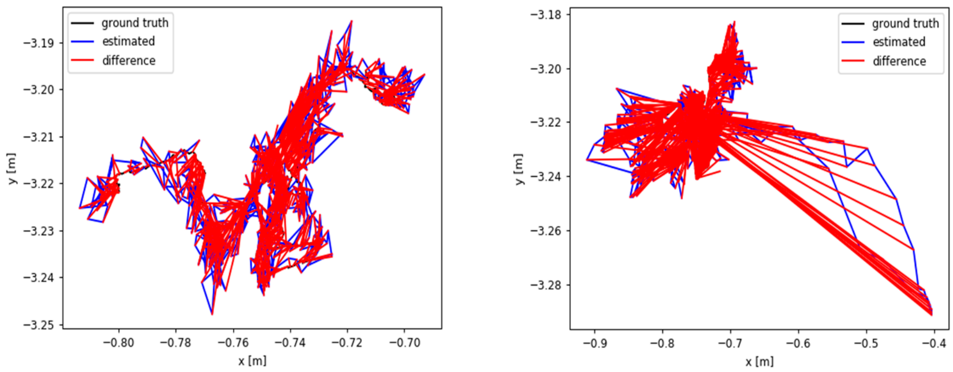
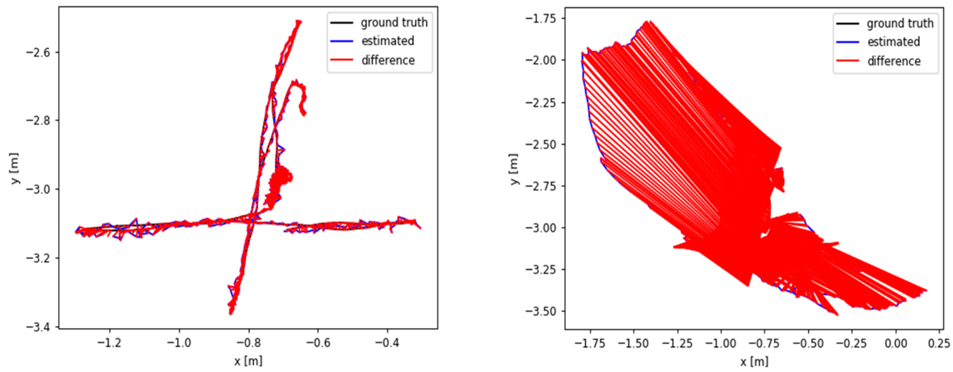


Figure 10. Comparison of trajectory errors on fr3_walking_rpy.



(**a**) Our (**b**) ORB-SLAM3

Figure 11. Comparison of trajectory errors on fr3_walking_static.



(**a**) Our (**b**) ORB-SLAM3

Figure 12. Comparison of trajectory errors on fr3_walking_xyz.

As can be seen from the above figure, in dynamic scenes, the localization results of ORB-SLAM3 are severely disturbed by dynamic objects, and the camera estimated trajectory differs significantly from the ground truth trajectory. In highly dynamic environments, ORB-SLAM3 cannot achieve accurate localization, whereas the proposed method effectively solves this problem. Detailed evaluation results are shown in Tables 2 and 3.

Table 2. Comparison of ATE (m) performance.

Sequence	ORB-SLAM3			Our		
	MEAN	RMSE	STD	MEAN	RMSE	STD
fr3_walking_half	0.1622	0.2089	0.1315	0.0233	0.0273	0.0143
fr3_walking_rpy	0.5896	0.6841	0.3469	0.0242	0.0278	0.0136
fr3_walking_static	0.0431	0.0745	0.0607	0.0071	0.0080	0.0036
fr3_walking_xyz	0.6468	0.7481	0.3759	0.0140	0.0157	0.0079

Table 3. Comparison of RPE (m) performance.

Sequence	ORB-SLAM3			Our		
	MEAN	RMSE	STD	MEAN	RMSE	STD
fr3_walking_half	0.0157	0.0208	0.0137	0.0109	0.0134	0.0077
fr3_walking_rpy	0.0217	0.0280	0.0177	0.0159	0.0205	0.0129
fr3_walking_static	0.0108	0.0193	0.0159	0.0055	0.0065	0.0034
fr3_walking_xyz	0.0205	0.0247	0.0139	0.0095	0.0116	0.0066

From the analysis of the above tables, it can be seen that all accuracy evaluation metrics of the proposed method are superior to those of ORB-SLAM3. In the fr3_walking_half dataset, compared with ORB-SLAM3, the RMSE of ATE is reduced by 86.93%; in the fr3_walking_rpy dataset, the RMSE of APE is reduced by 95.94%; in the fr3_walking_static dataset, the RMSE of ATE is reduced by 89.26%; in the fr3_walking_xyz dataset, the RMSE of ATE is reduced by 97.90%. The experimental results demonstrate that, compared with ORB-SLAM3, the proposed method can effectively achieve stability and accuracy in visual localization under dynamic environments.

We conducted real-time performance tests on both the EuRoC dataset and the TUM dataset. On the EuRoC dataset, the per-frame processing time of ORB-SLAM3 is 16.3 ms, while that of our method is 25.7 ms, corresponding to a frame rate of 38.9 fps. On the TUM dataset, the per-frame processing time of ORB-SLAM3 is 21.3 ms, while that of our method is 40.7 ms, corresponding to a frame rate of 24.5 fps. Compared with the original ORB-SLAM3, our method incurs a certain increase in processing time. However, with the incorporation of the adaptive image enhancement module and the object detection module, the RMSE values are reduced by 26.9% and 92.5%, respectively, demonstrating a significant improvement in accuracy.

4.3. Ablation Experiment

To evaluate the individual contributions of each component in the proposed algorithm, we conducted an ablation study on the V103 sequence from the EuRoC dataset, as shown in Table 4.

Table 4. Ablation study RMSE (m) comparison of image enhancement modules.

	MEAN	RMSE
ORB-SLAM3	0.075	0.087
Brightness compensation	0.069	0.075
Improved CLAHE	0.062	0.067
V-channel + Improved CLAHE	0.060	0.064
Brightness compensation + Improved CLAHE	0.063	0.066
Full enhancement module	0.052	0.056

From the above table, it can be observed that, compared with ORB-SLAM3, each component of our proposed image enhancement algorithm contributes to improved accuracy. Specifically, brightness compensation provides balanced input, ensuring the effectiveness of the subsequent modules. Enhancing the V-channel directly improves the identifiability of the grayscale structural information on which feature extraction depends, ensuring the stability of feature matching, and thereby further enhancing the performance of Improved CLAHE. The synergy of all three components achieves the optimal localization results. This demonstrates that the accuracy gain of the image enhancement module is not dominated by a single component, but rather results from the synergistic effect of all three modules working together.

To further validate the collaborative effect between the image enhancement algorithm and the object detection algorithm, we conducted experiments on the fr3_walking_rpy sequence from the TUM dataset, with results presented in Table 5.

Table 5. Ablation RMSE (m) comparison of the overall image enhancement module and the tracking thread.

	MEAN	RMSE
ORB-SLAM3	0.5896	0.6841
YOLOv5-SLAM3	0.0314	0.0342
Our	0.0233	0.0273

From the above table, it can be seen that our proposed method achieves the highest accuracy, with a 20.2% improvement in accuracy over YOLOv5-SLAM3. This detection performance benefits from the front-end image enhancement, as the enhanced images enable YOLOv5 to recognize objects more completely and accurately, avoiding the retention of dynamic feature points caused by missed detections. This synergistic relationship indicates that image enhancement and object detection are not isolated modules but rather exhibit a positive mutual reinforcement effect.

Under complex illumination conditions, the image enhancement module effectively improves image quality, making feature extraction more stable and rendering the contours of dynamic objects clearer, which directly enhances the detection recall rate and bounding box accuracy of YOLOv5 in degraded illumination. When dynamic objects are present, YOLOv5 can detect them and remove the corresponding dynamic feature points, allowing the gains from image enhancement to be fully translated into localization accuracy improvements. The two modules complement each other. In summary, the accuracy gains of our proposed method arise from the synergistic effect of all components, with each module playing its role under its respective conditions.

5. Conclusions

To address the adverse effects of complex lighting conditions on the localization accuracy and robustness of VSLAM systems, this paper proposes a VSLAM localization method based on multi-level adaptive image enhancement, which is an improvement upon the ORB-SLAM3 framework. This method overcomes the limitations of conventional single-image enhancement techniques and provides a new solution for visual localization in complex lighting scenarios. First, dynamic brightness compensation is adopted to resolve issues related to image brightness. Then, to achieve separation of the brightness (V) channel, the image is converted from the RGB color space to the HSV color space. Subsequently, an adaptive-limit CLAHE algorithm is proposed to effectively enhance the image and highlight its texture information, providing a high-quality data foundation for accurate feature point extraction. Meanwhile, to address the challenges posed by dynamic environments, an object detection thread is added and organically integrated with the above multi-level adaptive image enhancement method. On the basis of achieving accurate identification of dynamic objects and elimination of dynamic feature points, this approach also preserves more static feature points in the background, thereby improving the robustness and accuracy of VSLAM localization in dynamic environments.

Although our method demonstrates favorable localization accuracy and robustness in complex illumination and dynamic environments, the following limitations remain to be addressed in future research. First, our experiments were conducted and validated only under monocular and RGB-D camera modes, and have not yet been extended to other camera types. Therefore, the applicability and generalization capability of our method across different camera types require systematic verification. Second, the datasets used in this study are primarily indoor environments, which may not fully represent the diversity of illumination challenges encountered in real-world applications. Consequently, the generalization ability of the method under more diverse and challenging environmental conditions warrants further investigation. Future work will focus on addressing the above limitations by incorporating more device types for extended evaluation. In addition, we plan to supplement our experiments with outdoor datasets to further assess and enhance the cross-scenario adaptability of our method, thereby improving its adaptability and reliability in a broader range of practical application scenarios.

Author Contributions: Conceptualization, Qiaobin Dai and Zhe Yue; methodology, Qiaobin Dai and Zhe Yue; software, Qiaobin Dai; validation, Wangyang Yu; formal analysis, Qiaobin Dai;

investigation, Qiaobin Dai; resources, Qiaobin Dai; data curation, Qiaobin Dai, Zhe Yue and Zengzeng Lian; writing, original draft preparation, Qiaobin Dai; writing, review and editing, Zhe Yue, Wangyang Yu, Xuerong Zhang and Zengzeng Lian; visualization, Qiaobin Dai; supervision, Zhe Yue, Wangyang Yu, Xuerong Zhang and Zengzeng Lian; project administration, Zhe Yue; funding acquisition, Zhe Yue. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 42204040); the Henan Association for Science and Technology Young Talent Support Project (Grant No. 2025HYTP039); the Henan Polytechnic University Outstanding Youth Fund (Grant No. J2026-2); and the Henan Polytechnic University Young Backbone Teacher Training Program Project (Grant No. 2024XQG-15).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

1. Zheng, Z.; Su, K.; Lin, S.; Fu, Z.; Yang, C. Development of Vision–Based SLAM: From Traditional Methods to Multimodal Fusion. *Robot. Intell. Autom.* **2024**, *44*, 529–548. https://doi.org/10.1108/RIA-10-2023-0142.
2. Wang, K.; Yao, X.; Ma, N.; Chen, S.; Wang, L.; Hou, Z. PLMOT-SLAM: A Point-Line Features Fusion SLAM System with Moving Object Tracking. *Vis. Comput*. **2025**, *41*, 4547–4565. https://doi.org/10.1007/s00371-024-03677-9.
3. Wang, Y.; Liu, X.; Zhao, M.; Xu, X. VIS-SLAM: A Real-Time Dynamic SLAM Algorithm Based on the Fusion of Visual, Inertial, and Semantic Information. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 163. https://doi.org/10.3390/ijgi13050163.
4. Yin, Z.; Feng, D.; Fan, C.; Ju, C.; Zhang, F. SP-VSLAM: Monocular Visual-SLAM Algorithm Based on SuperPoint Network. In Proceedings of the 2023 15th International Conference on Communication Software and Networks (ICCSN), Shenyang, China, 21–23 July **2023**; pp. 456–459. https://doi.org/10.1109/ICCSN57992.2023.10297407.
5. Engel, J.; Stückler, J.; Cremers, D. Large-Scale Direct SLAM with Stereo Cameras. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October **2015**; pp. 1935–1942. https://doi.org/10.1109/IROS.2015.7353631.
6. Engel, J.; Koltun, V.; Cremers, D. Direct Sparse Odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 611–625. https://doi.org/10.1109/TPAMI.2017.2658577.
7. Mur-Artal, R.; Montiel, J.M.M.; Tardós, J.D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot*. **2015**, *31*, 1147–1163. https://doi.org/10.1109/TRO.2015.2463671.
8. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast Semi-Direct Monocular Visual Odometry. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–5 June **2014**; pp. 15–22. https://doi.org/10.1109/ICRA.2014.6906584.
9. Hu, Z.; Qi, W.; Ding, K.; Li, Y.; Liu, S.; Zhang, X. An Adaptive Lighting Indoor vSLAM with Limited On-Device Resources. *IEEE Internet Things J.* **2024**, *11*, 13. https://doi.org/10.1109/JIOT.2024.3406816.
10. Zhu, Q.; Li, X.; Zhao, Y.; Chen, W. VSLAM Based on Deep Learning in Low-Textured Scenes. *Meas. Sci. Technol.* **2024**, *35*, 126311. https://doi.org/10.1088/1361-6501/ad7be4.
11. Jiang, K.; Zhu, C.; Yang, J. Research on Low-Light Environment SLAM System via Retinex Enhancement and Deep Feature Optimization. *IEEE Sens. J.* **2025**, *25*, 29992–30004. https://doi.org/10.1109/JSEN.2025.3581532.
12. Liu, C.; Wang, Y.; Luo, W.; Zhang, H.; Li, M. LDFE-SLAM: Light-Aware Deep Front-End for Robust Visual SLAM under Challenging Illumination. *Machines* **2025**, *14*, 44. https://doi.org/10.3390/machines14010044.
13. Zhao, Z.; Liu, Y.; Wang, H.; Chen, L.; Sun, J.; Wu, Q. Light-SLAM: A Robust Deep-Learning Visual SLAM System Based on LightGlue under Challenging Lighting Conditions. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 9918–9931. https://doi.org/10.1109/TITS.2025.3562319.
14. Dai, J.; Yan, T.; Jiang, X.; Xiao, H.; Xin, H.; Zhang, C. ELL-ORB-SLAM: Improved Robust Monocular Visual SLAM Algorithm Based on Image Enhancement and Dynamic FAST Threshold for Low-Light Environment. In Proceedings of the 2025 5th Inter-

national Conference on Consumer Electronics and Computer Engineering (ICCECE), Dongguan, China, 28 February–2 March **2025**; pp. 699–704. https://doi.org/10.1109/ICCECE65250.2025.10985601.
15. Chen, J.; Hou, P.; Zhao, H.; Cao, Z.; Zhao, J. EAMT-SLAM: Edge-Aided Monocular Thermal Sensor SLAM in Low-Illumination Environments. *IEEE Sens. J.* **2024**, *24*, 33371–33386. https://doi.org/10.1109/JSEN.2024.3452557.
16. Yu, L.; Yang, E.; Yang, B. AFE-ORB-SLAM: Robust Monocular VSLAM Based on Adaptive FAST Threshold and Image Enhancement for Complex Lighting Environments. *J. Intell. Robot. Syst.* **2022**, *105*, 26. https://doi.org/10.1007/s10846-022-01645-w.
17. Canh, T.N.; Nguyen Quoc, B.; Zhang, H.; Veeraiah, B.R.; Hoang Van, X.; Chong, N.Y. IRAF-SLAM: An Illumination-Robust and Adaptive Feature-Culling Front-End for Visual SLAM in Challenging Environments. In Proceedings of the 2025 European Conference on Mobile Robots (ECMR), Padua, Italy, 2–5 September **2025**; pp. 1–7. https://doi.org/10.1109/ECMR65884.2025.11163050.
18. Hou, P.; Chen, X.; Chen, J.; Zhao, J. Efficient Camera Attribute Adjustment for Improved Visual SLAM in Extremely Changing Illumination Environment. *Meas. Sci. Technol.* **2025**, *36*, 066319. https://doi.org/10.1088/1361-6501/ade332.
19. Xu, K.; Hao, Y.; Yuan, S.; Wang, C.; Xie, L. AirSLAM: An Efficient and Illumination-Robust Point-Line Visual SLAM System. *IEEE Trans. Robot.* **2025**, *41*, 1673–1692. https://doi.org/10.1109/TRO.2025.3539171.
20. Fang, Y.; Shan, G.; Wang, T.; Li, Z.; Liu, H. He-SLAM: A Stereo SLAM System Based on Histogram Equalization and ORB Features. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December **2018**; pp. 4272–4276. https://doi.org/10.1109/CAC.2018.8623424.
21. Yang, W.; Zhai, X. Contrast Limited Adaptive Histogram Equalization for an Advanced Stereo Visual SLAM System. In Proceedings of the 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Guilin, China, 17–19 October **2019**; pp. 131–134. https://doi.org/10.1109/CyberC.2019.00030.
22. Zhang, A.; He, J. Light Adaptive Visual Inertial SLAM: A Dynamic Optimization Framework for Low Light Environments. In Proceedings of the 2025 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC), Zhengzhou, China, 17–19 May **2025**; pp. 2658–2663. https://doi.org/10.1109/YAC66630.2025.11149810.
23. Campos, C.; Elvira, R.; Rodríguez, J.J.G.; Montiel, J.M.M.; Tardós, J.D. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Trans. Robot.* **2021**, *37*, 1874–1890. https://doi.org/10.1109/TRO.2021.3075644.
24. Zhu, H.; Wang, F.; Li, X.; Chen, Y.; Zhang, L.; Liu, J. Computer Vision-Based Beef Grading for Marbling Beef Based on RGB-HSV Color Model and Python Software. In Proceedings of the 2023 3rd International Signal Processing, Communications and Engineering Management Conference (ISPCEM), Montreal, QC, Canada, 25–27 November **2023**; pp. 609–614. https://doi.org/10.1109/ispcem60569.2023.00115.
25. Wen, S.; Tao, S.; Liu, X.; Babiarz, A.; Yu, F.R. CD-SLAM: A Real-Time Stereo Visual–Inertial SLAM for Complex Dynamic Environments with Semantic and Geometric Information. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2517808. https://doi.org/10.1109/TIM.2024.3396858.
26. Shen, J.; Mei, J.; Shen, J. An Improved Image Enhancement Method Based on Histogram Equalization. *Comput. Digit. Eng.* **2024**, *52*, 3112–3115+3130.
27. Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. *Graph. Gems* **1994**, 474–485. https://doi.org/10.1016/B978-0-12-336156-1.50061-6. Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. In Graphics Gems IV; Heckbert, P.S., Ed.; Morgan Kaufmann: San Diego, CA, USA, **1994**; pp. 474–485. ISBN 0-12-336155-9. https://doi.org/10.1016/B978-0-12-336156-1.50061-6
28. Zubizarreta, J.; Aguinaga, I.; Montiel, J.M.M. Direct Sparse Mapping. *IEEE Trans. Robot.* **2020**, *36*, 1363–1370. https://doi.org/10.1109/TRO.2020.2991614.