An Adaptive Lighting Indoor vSLAM With Limited On-Device Resources

Zhuhua Hu, *Senior Member, IEEE*, Wenlu Qi, Kunkun Ding, Guangfeng Liu, and Yaochi Zhao

***Abstract*—Visual simultaneous localization and mapping (vSLAM) is a critical technology for enabling robots to achieve self-awareness in unknown environments. However, the robustness and precision of vSLAM are continuously challenged by issues, such as sudden changes in lighting, reflections from shadows, and reduced contrast within indoor settings under constrained hardware resources. This article introduces an improved ORB-SLAM3 algorithm based on image enhancement for constrained hardware resources adaptive-light enhanced visual simultaneous localization and mapping (ALE-vSLAM), aimed at addressing challenges in varying lighting conditions. First, we have released a proprietary data set specifically targeting issues related to lighting variations, combined with image enhancement techniques to categorize and process the textural and structural features of images under different lighting conditions. Second, to more accurately extract feature points across various scales, we propose an adaptive contrast-guided contrast limited adaptive histogram equalization (CLAHE) enhancement method, in conjunction with an image pyramid. Finally, we have developed a CLAHE-integrated features from the accelerated segment test corner detection strategy that reincorporates important feature points filtered out due to lighting changes, while mitigating noise impact as much as possible. The experiments utilized the European robotics challenge data set and our own data set (SLAMLightClass data set). Results indicate that under conditions of limited resources, ALE-vSLAM consistently outperforms the ORB-SLAM3 algorithm in both positioning accuracy and the robustness of the feature extraction.**

***Index Terms*—Image enhancement, limited resources, ORB-SLAM3, variable lighting, visual simultaneous localization and mapping (vSLAM).**

I. Introduction

SIMULTANEOUS localization and mapping (SLAM) has become a mature interdisciplinary technology, incorporating advanced hardware, sensor technologies, kinematic models, and artificial intelligence algorithms [1], [2], [3]. Subdomains, such as Visual SLAM, LiDAR SLAM, sonar SLAM, and multisensor SLAM have been gaining considerable attention in recent years [4]. Particularly, with improvements in GPU computational capabilities and low-cost visual sensors, Visual SLAM has gradually revealed its significant commercial potential [5].

Visual SLAM primarily relies on the visual sensors, such as monocular cameras, stereo cameras, and RGBD cameras to obtain rich texture information for efficient localization and mapping [6]. The progress of the field has been significantly propelled by various open source algorithms. However, for Visual SLAM highly dependent on lighting conditions, algorithms based on the traditional stereo cameras often lead to additional cost wastage [7]. In contrast, monocular Visual SLAM offers a more cost-effective solution, with the reduction of dependency on normal lighting conditions emerging as a key research focus [8].

Visual SLAM systems are mainly divided into direct methods and feature-based methods [9]. The former optimizes camera motion and map building by minimizing the pixel-level reprojection errors, while the latter depends on the extraction, tracking, and matching of feature points to estimate the camera motion and the 3-D structure of the environment [10]. The current state-of-the-art Visual SLAM technologies are primarily designed for environments with consistent lighting and are run on mid-to-high-end hardware equipped with efficient GPU and high-quality cameras. Although these can meet most of the users' needs, the high cost is a bottleneck for the widespread deployment of intelligent robots. Therefore, there is an urgent need to develop a Visual SLAM system suitable for constrained resources, which also means that its practicality in variable lighting environments faces even greater challenges [11].

1) *Lack of GPU Support:* The parallel processing capabilities of GPU greatly assist in the detection of image feature points. Without a GPU, there is an increase in the time delay for feature extraction and matching, particularly when dealing with a large number of feature points, which can severely limit matching efficiency. Additionally, the absence of GPU support can restrict the system's ability to handle the complex algorithms. This might lead to a loss of real-time performance when dealing with rapid changes in lighting, such as transitions between light and dark, consequently affecting the robot's real-time navigation and decision-making capabilities.
2) *Low-Resolution Monocular Cameras:* First, monocular cameras rely more on visible light compared to the

Manuscript received 18 March 2024; revised 28 April 2024; accepted 19 May 2024. Date of publication 19 June 2024; date of current version 23 August 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62161010 and Grant 62361024; in part by the Key Research and Development Project of Hainan Province under Grant ZDYF2022GXJS348 and Grant ZDYF2024GXJS021; in part by the National Key Research and Development Program of China under Grant 2022YFD2400504; and in part by the Hainan Province Natural Science Foundation under Grant 623RC446. *(Corresponding author: Yaochi Zhao.)*

Zhuhua Hu, Wenlu Qi, and Kunkun Ding are with the School of Information and Communication Engineering, Hainan University, Haikou 570288, China.

Guangfeng Liu and Yaochi Zhao are with the School of Cyberspace Security (School of Cryptology), Hainan University, Haikou 570288, China (e-mail: zhyc@hainanu.edu.cn).

Digital Object Identifier 10.1109/JIOT.2024.3406816

2327-4662 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

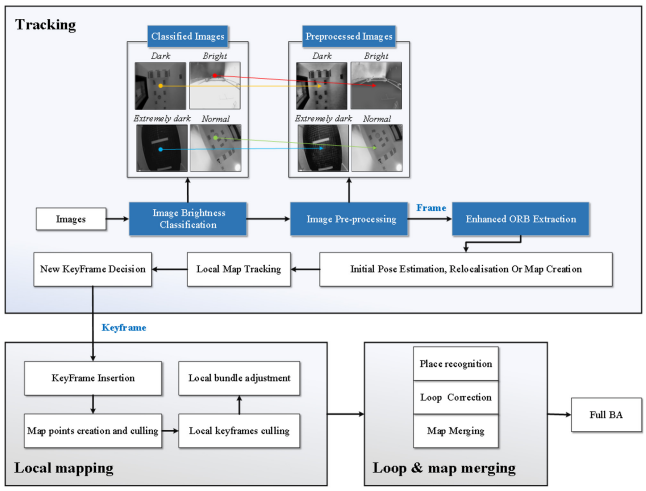


Fig. 1. Structure of the ALE-vSLAM algorithm (modules in blue indicate improvements or additions).

stereo, depth, or RGBD cameras. In low-light conditions, it is necessary to increase ISO or decrease shutter speed to capture more light, which can result in increased image noise. Furthermore, monocular cameras lack depth information and depend on motion estimation from the consecutive frames. Poor lighting can decrease the accuracy of depth estimation. Additionally, the low pixel count further restricts the image resolution, resulting in additional loss of details.

3) *Low Frame Rate:* A low frame rate means increased time intervals between the consecutive frames, leading to more scene changes and weakening the continuity and accuracy of feature matching. It may even introduce image blur due to rapid motion.
4) *Variable Indoor Lighting Conditions:* Unstable lighting conditions can significantly alter the appearance and perspective of objects. Under such lighting, changes in object scale can occur, negatively impacting the distance estimation of feature points. When illumination is uneven, meaningful feature points may be difficult to extract in certain areas of the image, resulting in an uneven distribution of feature points.

Therefore, for resource-constrained indoor robots, the development of a monocular Visual SLAM method that can adapt to variable lighting conditions is of significant research and practical importance.

This article introduces an improved ORB-SLAM3 algorithm for the monocular mode based on an image enhancement strategy, named adaptive-light enhanced visual SLAM (ALE-vSLAM). The specific algorithm structure is shown in Fig. 1. Although the ORB-SLAM3 system has exhibited excellent performance in localization and map construction under ideal lighting conditions, its performance still faces challenges in complex and variable lighting situations [12]. To enhance the adaptability and accuracy of the ORB-SLAM3 algorithm in variable lighting conditions, this study has employed methods, such as image brightness grading, multiscale contrast limited adaptive histogram equalization (CLAHE) enhancement, and improved features from the accelerated segment test (FAST) corner detection. The main contributions of this work are as follows.

1) We have developed and made available a proprietary data set that introduces an image brightness grading strategy. Catering to the performance characteristics of different brightness features on hardware with limited resources, we have integrated the advantages of techniques, such as CLAHE and gamma correction (GC) to implement a more targeted enhancement strategy. This improves the adaptability of ORB-SLAM3 to variable lighting conditions.
2) We have utilized an adaptive contrast-guided CLAHE enhancement in conjunction with a Gaussian pyramid. By conducting contrast analysis, we exploit the local adaptive features of CLAHE to address detail issues across various scales. This balances the amplification of noise with the enhancement of details, thereby reducing feature matching errors in monocular cameras that arise due to the lack of depth information in suboptimal lighting conditions.
3) We have integrated CLAHE with threshold adjustment for FAST corner detection. A new fixed threshold is introduced to retain the computational efficiency of FAST corner detection on hardware with limited resources, while filtering to preserve feature points that would normally be retained but are excluded due to changes in illumination. This aids in achieving more accurate localization and map construction under dynamic lighting conditions.
4) Experiments validate the precision and robustness of ALE-vSLAM, showing that our method surpasses both the original approach and techniques utilizing traditional image enhancement in positioning accuracy, while also sustaining a high level of precision in environments with standard lighting conditions.

The structure of this article is organized as follows. Section II reviews the related research. Section III provides a detailed introduction to the improved image enhancement methods and system algorithms. Section IV analyses the experimental results. Section V concludes the work of this article.

II. Related Work

In the field of feature-based Visual SLAM research, Davison et al. [13] first achieved real-time performance under monocular cameras with Mono-SLAM. In the same year, Klein and Murray [14] introduced PTAM, which implemented parallel computation and nonlinear optimization in two threads. Mur-Artal et al. [15], [16] further optimized the system performance with ORB-SLAM and ORB-SLAM2, utilizing a three-thread parallel architecture and oriented FAST and rotationally invariant BRIEF (ORB) features. In 2020, Campos et al. [17] introduced ORB-SLAM3, which building upon ORB-SLAM2, introduced support for visual-inertial integration and new methods for map merging. This allowed for the construction of a new map after tracking loss and its merger with the old map upon relocalization. It was the

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

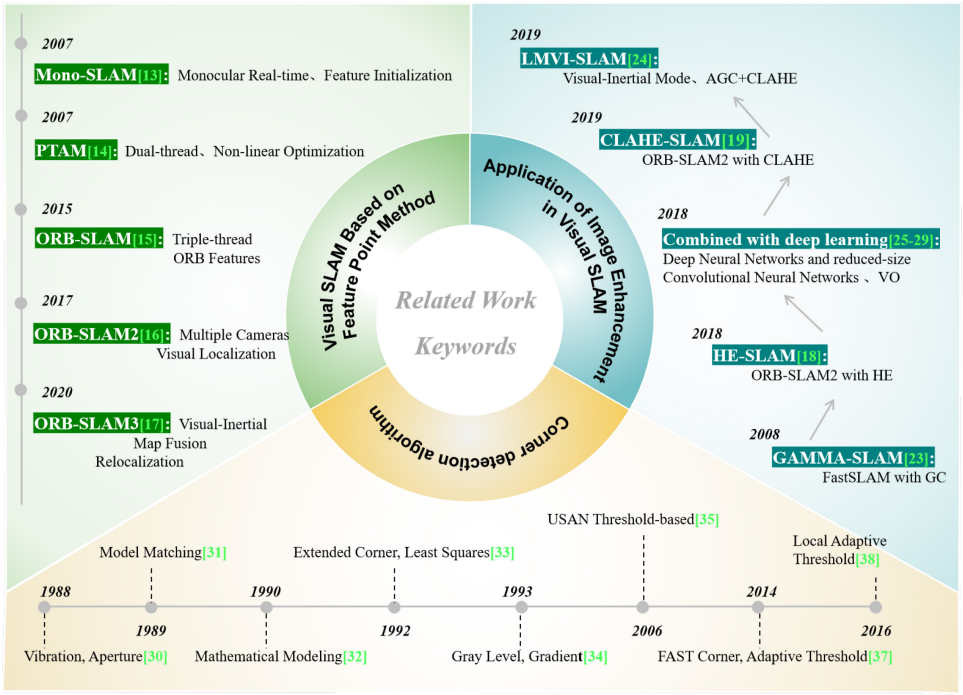


Fig. 2. Compilation of related work.

first Visual SLAM system that comprehensively utilized prior information, enhancing the system's robustness in variable lighting conditions after tracking loss and relocalization.

A. Image Enhancement Applications in Visual SLAM

To tackle challenging lighting conditions, HE-SLAM [18] incorporates histogram equalization (HE) into the ORB-SLAM2 framework. While it performed well in low-illumination and low-contrast scenes, the HE technique is susceptible to noise and limited to specific environments. To enhance the system robustness, Yang and Zhai [19] incorporated CLAHE into ORB-SLAM2. CLAHE, a local image enhancement algorithm, performs meticulous balancing for different areas, effectively harmonizing local details with global contrast. However, CLAHE requires additional computational resources [20], which can impact real-time performance and may overemphasize noise in low signal-to-noise ratio scenes, disturbing feature point matching. Therefore, CLAHE-SLAM may need to be combined with other image preprocessing methods to achieve optimal performance [21]. Based on the FastSLAM [22] framework, Marks et al. [23] applied GC to optimize SLAM performance in irregular outdoor environments, but this method relies on visual ranging from stereo cameras and is susceptible to lighting and weather conditions. Hao et al. [24] used adaptive GC in conjunction with CLAHE and visual-inertial tightly coupled optimization. Although this approach performs well in high-illumination environments, it lacks adaptability to changes in ambient lighting. Yu et al. [12] introduced Afe-ORB-SLAM, a methodology that classifies illumination conditions and utilizes adaptive thresholding to comprehensively address most variations in lighting encountered in practical settings. Nonetheless, its efficacy is compromised under extreme lighting situations, like severely dark or overexposed environments, attributed to the coarse nature of its lighting categorization strategy and the inadequate accommodation for the specific needs of these challenging conditions. In recent years, image enhancement methods based on deep learning have gradually gained scholars' attention [25], [26], [27], [28], [29]. While they have increased the precision of the SLAM algorithms in dynamic environments and under variable lighting conditions, their application on the mobile robotic platforms is restricted due to computational resource limitations.

B. FAST Corner Detection Algorithm

Guiducci [30] was among the first to employ characteristics, such as vibrancy and aperture to detect wedge corners, and subsequent scholars [31], [32] enhanced the accuracy of corner localization through template matching and mathematical modeling. Rohr [33] expanded the types of corners and employed the least-squares method for corner positioning. Deriche and Blaszka [34] introduced grayscale and gradient parameters to optimize the accuracy of corner detection. In the SUSAN algorithm [35], Smith used a threshold-based USAN region, which showed good noise resistance, but still performed poorly against complex image backgrounds. Rosten and Drummond [36] proposed a corner detection algorithm that incorporated machine learning FAST. Following that, Florentz and Aldea [37] added an adaptive threshold to FAST corner detection to reduce computational load, but a single global threshold struggled to adapt to local variations within images. In contrast, Xing et al. [38] proposed a corner detection algorithm based on local adaptive thresholds, which

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

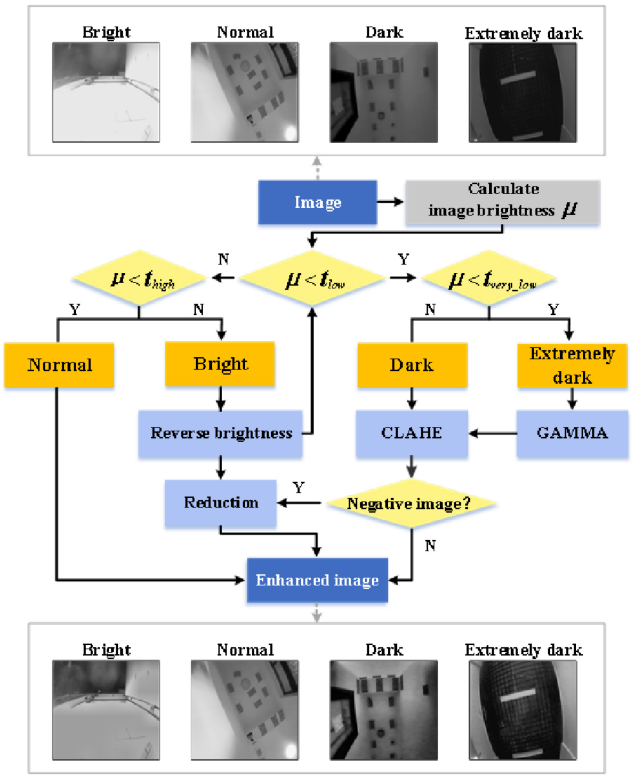


Fig. 3. Image enhancement algorithm based on brightness leveling.

dynamically adjusted thresholds according to local image characteristics. While this improved adaptability to diverse image regions, it increased computational complexity and performed poorly in scenarios with a lot of noise or uneven distribution of feature points.

III. Proposed Method

Unconventional and rapidly changing lighting can result in the loss of texture information in images, leading to a significant reduction in the number of feature points extracted by the camera, thereby disrupting frame-to-frame matching and causing tracking failures. Moreover, porting the system to hardware with limited performance often necessitates more complex image processing strategies to overcome challenges in image processing capabilities due to hardware constraints. [39], [40] To address these issues, we have introduced an image enhancement technology to enhance the system's noise resilience and improve the stability of feature detection.

A. Brightness Level-Based Image Enhancement Algorithm

To avoid the over-enhancement of images under normal lighting conditions by the image enhancement algorithms, while not compromising the real-time performance of the system, this article proposes an image enhancement algorithm based on the brightness level assessment. The specific process is illustrated in Fig. 3, where $t_{\text{very\_low}}$ is the brightness threshold for extremely dark images, t_{low} for the dark images, and t_{high} for the bright images.

GC is utilized to rectify images that are overly bright or dark, obtaining a more balanced grayscale distribution to enhance image details. However, the grayscale values of natural images tend to cluster in a narrow range, leading to detail loss. To achieve a more uniform distribution of image grayscales and effectively process areas of both high and low brightness presented in an image, we introduce the CLAHE [41] technique. Compared to the traditional histogram equalization, CLAHE combines local contrast statistical information, achieving a more detailed balance of the histogram while also better handling large grayscale variations.

The brightness of an image is the mean value of all the pixel values, providing an important reference standard for the categorization of image brightness. Given an image I with M rows and N columns, making the size of the image $M \times N$. First, we accumulate all the pixel values and divide through the total number of pixels in the image to determine its average brightness μ

$$\mu = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} I(i,j) \tag{1}$$

where $I(i,j)$ is the pixel value at the position (i,j) in the image.

Subsequently, based on the calculated brightness level μ and predetermined brightness thresholds, the input image is classified into four brightness categories: 1) extremely dark; 2) dark; 3) normal; and 4) bright

$$\begin{cases} \mu < t_{very\_low}, & \text{Extremely dark image} \\ t_{very\_low} \le \mu < t_{\text{low}}, & \text{Dark image} \\ t_{\text{low}} \le \mu \le t_{\text{high}}, & \text{Normal image} \\ \mu > t_{\text{high}}, & \text{Bright image} \end{cases} \tag{2}$$

where t_{very_low} is the brightness threshold for extremely dark images, t_{low} for the dark images, and t_{high} for the bright images.

Based on this classification, images with normal brightness require no processing, while the other three categories of images undergo targeted enhancement to ensure that each type of image achieves an optimized visual effect after enhancement.

1) Processing of Extremely Dark and Dark Images: For extremely dark images, we utilize a combination of GC and CLAHE to enhance the details and contrast of the image, specifically augmenting the contours and detail information of extremely dark areas. Initially, we compute the cumulative distribution function (CDF)

$$c(i) = \sum_{j=0}^{i} \frac{h(j)}{\sum_{k=0}^{255} h(k)} \tag{3}$$

where $c(i)$ is the cumulative distribution value up to the pixel intensity i. $h(j)$ represents the frequency of the pixel intensity j, which is the number of pixels with intensity i in the histogram. $\sum_{k=0}^{255} h(k)$ represents the total number of pixels in an 8-bit grayscale image, and $\frac{h(j)}{\sum_{k=0}^{255} h(k)}$ is used to calculate the weighted distribution for a given pixel intensity.

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

Next, we determine the parameter γ based on the specific location of the CDF

$$\gamma = \frac{\alpha}{c(t)} \tag{4}$$

where α is a predefined constant. When $\alpha > 1$, the image becomes brighter; at this time, a larger α value makes the image brighter. When $0 < \alpha < 1$, the image becomes darker; a smaller α value will darken the image. When $\alpha = 1$, the brightness of the image is considered normal and remains unchanged. Additionally, t is also a variable that represents the selected pixel intensity threshold, and $c(t)$ is the CDF value when the pixel intensity is t.

Finally, GC is applied to the image using γ

$$I'_{\gamma}(x, y) = I(x, y)^{\gamma} \tag{5}$$

where $I(x, y)$ is the pixel value of the original image at (x, y). $I'_{\gamma}(x, y)$ is the pixel value after GC.

GC is a nonlinear intensity stretching method that is particularly effective for very dark images. However, when applied globally, it may overlook local details of the image, thereby affecting the extraction of feature points and the stability of tracking. To further enhance picture details, we apply the CLAHE technique to make up for the shortcomings of GC as a global operation, while avoiding over enhancement or noise amplification.

First, we use the CDF to equalize the pixel values of each subregion

$$I'_{\mathrm{CLAHE}}(x, y) = \frac{L \times c\big[I'(x, y)\big]}{M \times N} \tag{6}$$

where $M' \times N'$ represents the size of the subregion and L is the level of pixel intensity, which for an 8-bit image, $L = 255$.

Then, based on the set contrast limit threshold, pixels exceeding this threshold are uniformly distributed across the entire histogram. To eliminate the boundary effects between the blocks, the interpolation methods are used to ensure the smoothness and continuity of the image. Taking bilinear interpolation as an example, assume that I_{00}, I_{01}, I_{10}, and I_{11} are the equalized values of four adjacent subregions

$$\begin{cases} I'_{00} = I_{00}(x_1 - x)(y_1 - y) \\ I'_{10} = I_{10}(x - x_0)(y_1 - y) \\ I'_{01} = I_{01}(x_1 - x)(y - y_0) \\ I'_{11} = I_{11}(x - x_0)(y - y_0) \\ I'_b(x, y) = \dfrac{I'_{00} + I'_{10} + I'_{01} + I'_{11}}{(x_1 - x_0)(y_1 - y_0)} \end{cases} \tag{7}$$

where I'_{00}, I'_{10}, I'_{01}, and I'_{11} are the values obtained by linear interpolation at the upper-left, upper-right, lower-left, and lower-right corners, respectively. The entire interpolation process is performed on four adjacent subregions, each with a corresponding interpolation term. The final pixel value after interpolation is obtained by weighted averaging, denoted as $I'_b(x, y)$.

For dark images, which do not have as severe brightness issues as extremely dark images, GC is not required. Thus, by applying the CLAHE technology directly, it is possible to correct the loss of local details caused by insufficient lighting to achieve a visually optimized effect. This also avoids the risks of potential noise amplification and over enhancement and is more computationally efficient.

2) Processing of Bright Images: Overexposed images often lead to the loss of texture information, which can significantly impact the accuracy of location positioning and environmental mapping in applications, such as intelligent mobile robots.

This phenomenon is not solely caused by the camera's sudden exposure to bright light sources but is more commonly observed when capturing surfaces, such as white walls, smooth objects, or reflective areas. In these instances, image frames frequently exhibit extensive brightness or overexposure.

Because the brightness values in high-light areas approach or reach the maximum (255), the scope for optimizing image enhancement algorithms is considerably limited. To achieve more effective image enhancement suitable for camera lenses, frames with high brightness and indistinct details can be transformed into their darker counterparts before applying image enhancement techniques to restore details and contrast.

Therefore, for an image I with a given brightness, we first create a negative image I_{neg} by subtracting each pixel in the original frame from the maximum pixel value (255)

$$I_{\mathrm{neg}} = 255 - I. \tag{8}$$

This step darkens the bright areas of the image and lightens the dark areas, laying the groundwork for subsequent adjustments in brightness and contrast.

Then, based on the average brightness of the negative image, it is classified as either a dark image or a very dark image. For instance, if the average brightness of the negative image is below the very dark threshold($t_{\mathrm{very\_low}}$), it indicates that the original image is overly bright and requires further processing through GC and CLAHE.

Finally, the adjusted image is reconstructed by reversing the enhanced negative image with the maximum pixel value (255). This tiered processing approach not only ensures effective enhancement across various levels of brightness but also avoids over processing and reduces computational costs.

B. Multiscale Enhancement Algorithm

In variable lighting environments, such as direct sunlight or shaded areas, local regions of images can exhibit significant variations in lighting conditions. To ensure that the image pyramids capture features effectively across scales in these complex scenes, we have applied parameterized CLAHE at each level, based on the contrast properties of the images. This method not only prevents the amplification of noise and loss of detail that may result from excessive enhancement but also boosts the system's ability to adapt to lighting changes.

Although the computational cost is slightly higher compared to the simple histogram equalization, integrating this approach with the multiscale processing of the image pyramid allows for effective resource allocation without sacrificing performance. This ensures more consistent feature point representation under a variety of lighting conditions, reduces mismatches, and consequently improves the match rate between the feature points.

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

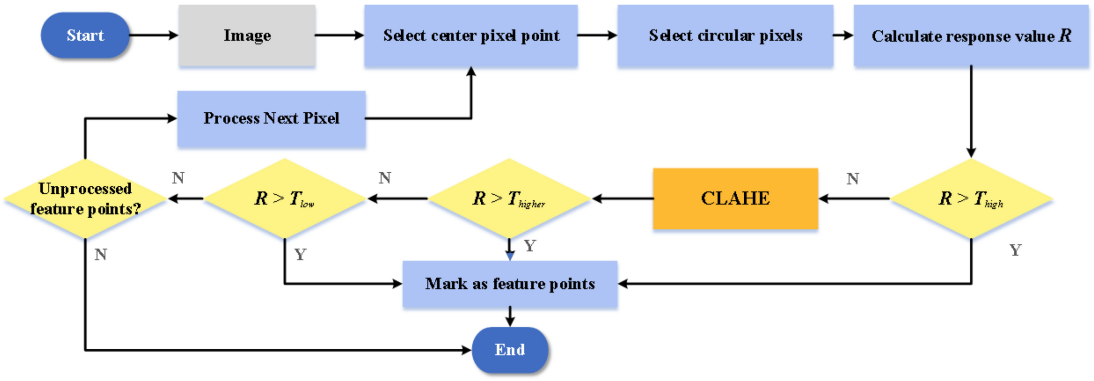


Fig. 4. Enhanced FAST corner detection algorithm.

Image contrast can be quantified by its standard deviation. The standard deviation is a statistical measure that describes the range of data distribution. For images, a larger standard deviation indicates higher contrast. Therefore, we first calculate the image's standard deviation σ

$$\begin{cases} \mu = \frac{1}{N}\sum_{i=1}^{N} I_i \\ \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_i - \mu)^2} \end{cases} \tag{9}$$

where μ is the mean value of the image, reflecting the overall brightness of the image. N is the total number of pixels in the image, which is equal to the width of the image multiplied by the height of the image. The intensity value of pixel number i, represented as I_i in the image.

Then, specific CLAHE image enhancements are applied for different scales and contrasts of the image to adapt to its characteristics

$$T = \text{int}(T_{\text{base}} + T_{\text{scale}} \times s) \tag{10}$$

where T is the size of the subarea. s is the scaling factor of the image and is related to the image pyramid scale, which typically ranges between $[0, 1]$. T_{base} is the base adjustment coefficient and T_{scale} indicates how the variation in s affects the change amplitude represented by T. $\text{int}(\beta)$ is an integer operation

$$C = C_{\text{base}} + C_{\text{contrast}} \times \frac{\sigma}{M} \tag{11}$$

where C is the maximum permissible degree of contrast enhancement. C_{base} is the base adjustment coefficient, C_{contrast} indicates the variation amplitude of C with the change of contrast, and M is the median value of the image intensity.

C controls contrast to enhance intensity, increasing C will highlight detail but may result in excessive sharpness, decreasing C will enhance smoothness and maintain a natural feel. T controls the size of the subregion, increasing T is suitable for processing a large range of features to enhance overall contrast, but may lose detail. Reduced T is suitable for processing local details, retaining more information but with less overall contrast enhancement.

The dynamic adjustment of CLAHE parameters based on the image contrast ensures enhanced effectiveness across various image scales, optimizing computational efficiency and reducing redundant enhancement operations. Furthermore, in uncontrollable and variable environments, images may undergo frequent changes in lighting. For instance, parts of an image may be directly exposed to sunlight while other areas are in shadow. The strategy of multiscale enhancement with dynamic parameter adjustment ensures that all regions are appropriately processed.

By combining the image brightness stratification strategy, even in the face of scene transitions or single images with lighting issues, the stability of the ORB-SLAM3 system can be ensured. This means that our approach ensures the stability of the system, whether it is a data set facing diverse lighting conditions or an undesirable brightness area in a single image.

C. Improvements to FAST Corner Detection

The FAST corner detection algorithm determines if a pixel should be considered a corner by evaluating the difference in intensity between that pixel and a majority of pixels in its neighborhood [36]. Although the implementation method of FAST corner detection is concise and the calculation speed is relatively fast, there are still shortcomings.

In the Visual SLAM system, the number and quality of feature points are central factors in ensuring the stability and accuracy of the system. If the number of feature points is insufficient or their quality is poor, tracking errors or inaccuracies in map construction can still occur. This is particularly the case with ORB-SLAM3, where the original FAST method may underperform under certain complex lighting conditions, especially in scenes with high light variation or low contrast. Although issues can be somewhat mitigated through luminance grading and multiscale enhancement, it is still not guaranteed that all high-value image blocks will extract high-quality feature points.

To retrieve high-quality feature points that have been filtered out due to lighting, we have introduced CLAHE image enhancement for those image blocks that are difficult to extract feature points under high thresholds. To reduce the risk of noise introduction, we have set a stricter corner detection threshold for the enhanced image blocks.

The structural diagram of the algorithm is shown in Fig. 4. Where T_{high} is used as a high threshold for initial feature point detection and T_{low} is the corresponding low threshold used to

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON OF RMSE ATE (M) PERFORMANCE ON THE EUROC DATA SET (BEST RESULTS ARE SHOWN IN BOLD)

Dataset	ORB-SLAM3 [17]	HE-SLAM [18]	CLAHE-SLAM [19]	Gamma-SLAM [23]	ALE-vSLAM(Ours)	Enhancement rate (ALE-vSLAM compared to ORB-SLAM3)
MH01	0.046894	0.046251	0.046513	0.046754	**0.039600**	15.55%
MH02	0.035921	0.034308	0.037618	0.035308	**0.031765**	11.57%
MH03	0.045658	0.046252	**0.041640**	0.044455	0.042196	7.58%
MH04	0.114419	0.112453	0.073918	0.082545	**0.060078**	47.49%
MH05	0.075141	0.321414	0.057720	0.062865	**0.055435**	26.23%
V101	0.095272	**0.092781**	0.095224	0.094575	0.093336	2.03%
V102	0.072214	0.073658	0.090130	0.079404	**0.065685**	9.04%
V201	0.085742	0.086681	**0.058756**	0.061092	0.060736	29.16%
V202	0.084844	0.101563	**0.063896**	0.074036	0.069165	18.48%
Average	0.072901	0.101818	0.062824	0.064559	**0.057555**	21.05%

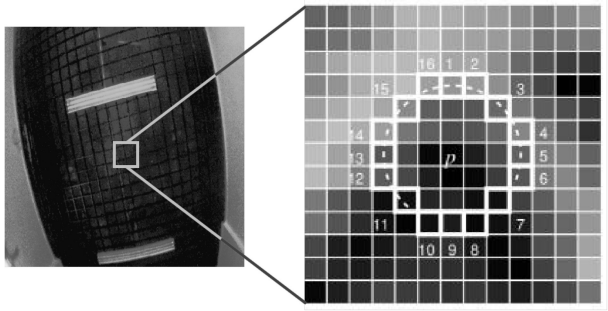


Fig. 5. FAST corner detection.

filter more candidate feature points. T_{higher} represents an even higher filtering criterion to offset the additional noise that may be introduced by the image enhancement algorithm, ensuring accuracy and reliability during the feature point detection process.

Thus, whether in bright, dim, or scenes with significant lighting fluctuations, the system can extract feature points more stably, thereby ensuring the stable operation of the SLAM system.

IV. EXPERIMENTAL RESULT AND ANALYSIS

To validate the performance of ALE-vSLAM, we designed four sets of experiments. The first set of experiments used the EVO evaluation tool to verify the improvements in ALE-vSLAM's localization accuracy and trajectory. The second set of experiments involved feature point extraction from single images under different lighting conditions to confirm the effectiveness and robustness of the ALE-vSLAM's feature extraction. The third set of experiments calculated the average success rate of feature point matching to verify whether the ALE-vSLAM algorithm improved the detection and matching performance of feature points. The fourth set of experiments assessed the optimization effect of the ALE-vSLAM algorithm on the hardware-limited visual sweeping robots.

The experimental environment for this study was a Lenovo Y9000P laptop with an Intel Core i7-12700H processor at 2.30 GHz and 16 G of RAM. The platform was Ubuntu 18.04 and the compilation used C ++ 7.5.

The data set used in this study and individual images all come from the European robotics challenge (EuRoC) data set [42], specifically, including the MH01-MHO5, V101-V102, and V201-V202 sequences. We chose the EuRoC data set for three reasons as follows.

1) It covers a variety of scenes, such as offices and industrial environments, which have different motion patterns and environmental characteristics, providing a benchmark for assessing the application scenarios of our algorithm.
2) It provides high-precision ground truth data, which is convenient for the comparative experiments.
3) The EuRoC data set also reflects a variety of real-world lighting changes, providing a crucial empirical basis for evaluating the performance of our algorithm in variable lighting conditions.

A. Trajectory Localization Accuracy Comparison

To comprehensively evaluate the performance of ALE-vSLAM on the EuRoC data set, we conducted an exhaustive comparison with ORB-SLAM3 [17]. In addition, we integrated the image enhancement strategies mentioned in [18], [19], and [23] into ORB-SLAM3 and named them HE-SLAM [18], CLAHE-SLAM [19], and GAMMA-SLAM [23], respectively. Our main evaluation criteria include absolute trajectory error (ATE) and relative pose error (RPE). We assessed the parameters for ATE and RPE using the root mean-squared error (RMSE), calculating the error of the estimated poses at all times. To ensure the reliability of the experimental results, each SLAM system was executed ten times, and the median of the absolute trajectory error (RMSE ATE) and relative pose error (RMSE RPE) was calculated based on the complete trajectory. The specific comparison results are presented in Tables I and II.

The EuRoC data set is divided into two scenarios, machine hall (MH) and Vicon Room (V1 and V2), each representing different environmental characteristics and motion patterns. The MH series (MH01–MH05) was collected inside a mechanical hall, featuring complex flight trajectories, wide-ranging movement, and noticeable changes in lighting. The Vicon Room sequences (V101, V102, V201, and V202) were recorded inside a laboratory equipped with a Vicon motion capture system, featuring stable operational trajectories and uniform lighting conditions.

According to the data from Tables I and II, it is evident that the ORB-SLAM3 experiences a significant performance

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

TABLE II
COMPARISON OF RMSE RPE (M) PERFORMANCE ON THE EUROC DATA SET (BEST RESULTS ARE SHOWN IN BOLD)

Dataset	ORB-SLAM3 [17]	HE-SLAM [18]	CLAHE-SLAM [19]	Gamma-SLAM [23]	ALE-vSLAM(Ours)	Enhancement rate (ALE-vSLAM compared to ORB-SLAM3)
MH01	0.414724	0.373817	0.388208	0.412444	**0.327717**	20.98%
MH02	0.388187	0.345242	0.363366	0.351229	**0.332032**	14.47%
MH03	1.084901	0.984938	1.080318	1.024289	**0.825794**	23.88%
MH04	0.601529	0.570458	0.595573	0.639563	**0.517003**	14.05%
MH05	0.592632	0.520636	0.595775	0.540437	**0.505824**	14.65%
V101	0.446050	0.415213	0.451668	0.396962	**0.380433**	14.71%
V102	0.507373	0.483955	0.554589	**0.374719**	0.383290	24.46%
V201	0.185935	0.179712	0.199095	0.164651	**0.152107**	18.19%
V202	0.493197	0.380374	0.483323	0.374019	**0.325383**	34.03%
Average	0.523836	0.472705	0.523546	0.475368	**0.416620**	20.47%

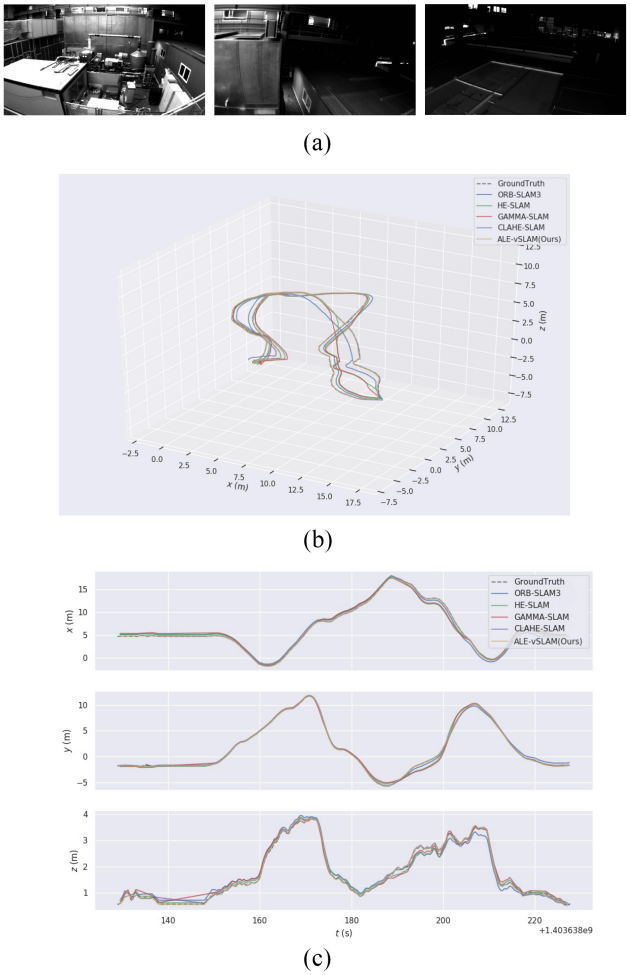


Fig. 6. MH04 trajectory comparison. (a) Typical images. (b) Trajectory comparison. (c) Comparative trajectories in the X, Y, and Z axes.

drop in extremely dark environments, such as the one shown in Fig. 6(a), with an absolute trajectory error (RMSE ATE) of 11.44 cm and a relative pose error (RMSE RPE) of 60.15 cm. The main reason for this performance decline is the impact of lighting on image contrast, which affects the quality of the detected keypoints and descriptors, leading to less accurate feature matching.

In the same data set, ALE-vSLAM reduced the absolute trajectory error (RMSE ATE) to 6.01 cm and the relative pose error (RMSE RPE) to 51.70 cm, achieving improvement rates of 47.79% and 14.05%, respectively. This confirms the suitability of our preprocessing method for low-light environments. Even on the V101 data set, ALE-vSLAM achieved an absolute trajectory error (RMSE ATE) of 9.33 cm, which still represented an improvement rate of 2.03%, further proving the enhanced algorithm's capability to capture visual information across different lighting scenarios.

From Table I, it is noted that the HE-SLAM performed best on the V101 data set, surpassing all the comparative methods with an improvement rate of 2.61%, but its performance degraded on the low-light data sets like MH05, with an average improvement rate dropping to −39.67%. CLAHE-SLAM had a particularly notable performance on the MH04 data set, with an improvement rate of 35.40%. GAMMA-SLAM achieved a 27.86% improvement on MH04, however, similar to CLAHE, it also faced negative optimization on the dynamic texture-rich V102 data set. Meanwhile, ALE-vSLAM maintained an improvement rate of 9.04% on V102, which also demonstrates that the improved algorithm is better suited to adapt to different lighting changes and texture distributions caused by movement.

The combined experimental data indicate that the ALE-vSLAM achieves consistent improvements under various lighting conditions. Compared to the ORB-SLAM3, it shows an enhancement of 21.05% in absolute trajectory error (ATE) and 20.47% in RPE.

Fig. 6 is a comparative trajectory graph of ALE-vSLAM and other algorithms on the MH04 data set, where the dashed line represents the true trajectory provided by the EuRoC data set, and the other solid lines represent the operational trajectories of different algorithms on the same data set. The closer a solid line is to the dashed line, the smaller the system error. Fig. 6(a) shows that the MH04 data set contains a rich variety of lighting changes.

Fig. 6 reveals that the trajectory generated by ORB-SLAM3 (the blue solid line) significantly differs from the ground truth trajectory provided by the data set, indicating compromised system performance in low-light conditions. This is mainly due to the ORB-SLAM3's reliance on feature matching between the consecutive frames for positioning, which becomes challenging under the extremely dark conditions shown in Fig. 6(a), resulting in a noticeable decline in positioning accuracy.

Fig. 6(c) shows that the operational trajectories of HE-SLAM (the green solid line), GAMMA-SLAM (the red

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

solid line), and CLAHE-SLAM (the purple solid line) also deviate to varying extents from the actual trajectory, likely due to the introduction of extra noise from over-enhanced images in normal lighting conditions, adversely impacting the system's trajectory precision. In contrast, the trajectory of the ALE-vSLAM algorithm proposed in this article (the yellow solid line) aligns more closely with the real trajectory, particularly evident in the lower right corner of Fig. 6(c), where the yellow solid line nearly coincides with the dashed line. This suggests a smaller localization error for our algorithm, confirming the accuracy and robustness of the ALE-vSLAM algorithm's trajectory under different lighting conditions.

B. Feature Point Extraction Experiment

To validate the effectiveness of the ALE-vSLAM algorithm in improving the number of feature extractions and its adaptability to a variety of complex lighting conditions, we conducted a feature extraction comparison experiment using single images against the ORB-SLAM3 algorithm. To ensure the reproducibility and comparability of the experiment, the reference images used were all taken from the EuRoC data set.

Initially, we selected an image from the V101 data set that represents typical indoor lighting conditions as the control group for normal illumination. Subsequently, we enhanced the brightness of this image by 65% to simulate potential overexposure scenarios that may occur during actual movement, in order to assess the performance of ALE-vSLAM under high lighting conditions. Finally, we chose an image with extremely low brightness from the MH04 data set to verify its robustness and effectiveness under low light conditions. The experimental results are illustrated in Fig. 7.

In monocular Visual SLAM systems, the quality and quantity of feature points play a crucial role in the system performance optimization. Especially in the monocular mode, the system primarily relies on feature points extracted by the camera to accomplish the tasks, such as mapping, loop closure detection, and relocalization. High-quality feature points help to reduce the risk of tracking loss, while a sufficient number of feature points provide a richer detail of the environmental information.

As observed from Fig. 7(b), ORB-SLAM3 maintains an uniform and good feature extraction effect under standard lighting conditions. However, under high or low lighting, the details of specific areas, such as the ground texture disappear, resulting in a drastic reduction in the number of feature points, which are mainly concentrated in areas rich in texture.

To address the problem of reduced feature point count under unconventional lighting, we integrated the CLAHE technique into the FAST corner detection, as shown in Fig. 7(c), which increased the number of feature points. However, in normal and low-light environments, many noises were misidentified as feature points, increasing the computational burden and possibly reducing the accuracy and stability of feature points. To further address the problem of excessive noise, COCO-SLAM introduced the concept of "enhanced thresholds." Comparing Fig. 7(d), the misidentification of noise as feature points in normal and low light conditions is significantly reduced compared to Fig. 7(c). At the same time, compared to Fig. 7(b), the number of feature points extracted under unconventional lighting has also been improved.

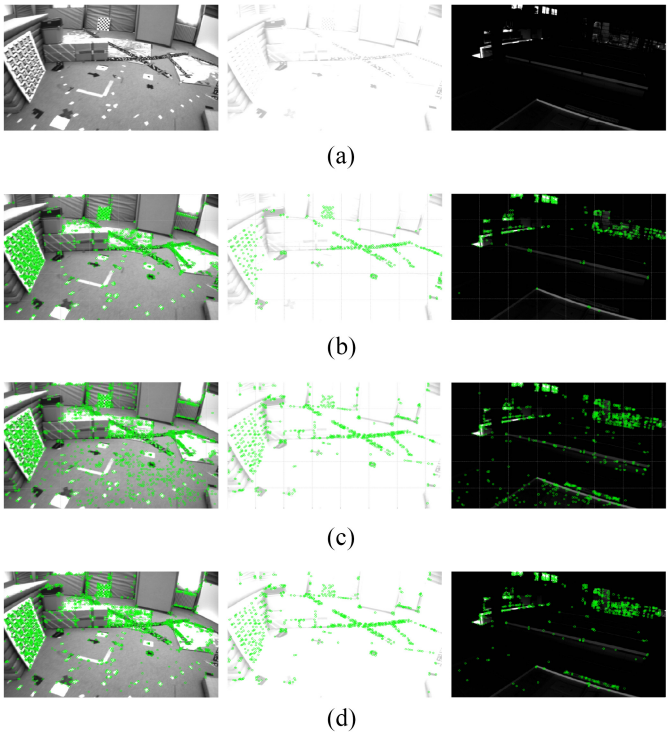


Fig. 7. Comparative analysis of feature point extraction efficacy. (a) Figure example. (b) Feature extraction in ORB-SLAM3. (c) Feature extraction integrating CLAHE. (d) Efficacy of feature extraction in ALE-vSLAM.

Overall, compared to the ORB-SLAM3, ALE-vSLAM is better suited to solve the balance between the number and quality of feature points under nonstandard lighting conditions.

C. Average Match Success Rate of Feature Points

Valid Inlier Average (VIA) Feature points that are consistent with the camera motion estimation model are termed inliers, and the quantity and quality of these inliers directly impact the accuracy and robustness of feature matching and pose estimation. To evaluate the improvements of the proposed algorithm, this study designed a quantitative experiment using the VIA metric to conduct a comparative analysis among the ALE-vSLAM (ours), ORB-SLAM3 [17], HE-SLAM [18], CLAHE-SLAM [19], and GAMMA-SLAM [23] algorithms. The experiment was conducted using the four subdata sets of the EuRoC data set–MH01, MH03, MH05, and V102–each characterized by different scene properties, difficulty levels, and lighting conditions.

To ensure consistency and reliability of the results, only the frames with inlier counts of ten or greater were considered. The experimental results are presented in Fig. 8.

As shown in Fig. 8, in the ORB-SLAM3 algorithm, the average VIA for the MH01 and MH05 data sets are 185.03 and 185.49, respectively, while the average for the MH03 data set is lower at 176.43. After incorporating image enhancement algorithms, the average values generally decreased. This

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

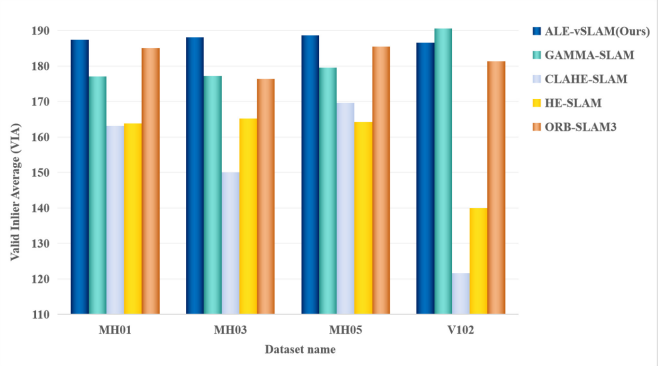


Fig. 8. Comparison of VIAs.

decrease is attributed to traditional image enhancement methods which, while enhancing image textures, may also alter the quality or distribution of feature points, thereby disrupting the RANSAC algorithm's ability to distinguish between inliers and outliers.

In contrast, ALE-vSLAM maintains higher and more stable average values between 186 and 189 across all the four data sets. Specifically, ALE-vSLAM performs better on the MH01, MH03, and MH05 data sets, with VIA scores of 187.38, 188.1, and 188.63, respectively, demonstrating the stability and effectiveness of our algorithm across different scenarios.

D. Empirical Experiments Based on Vision-Guided Sweeper

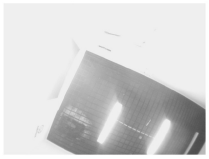
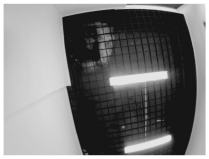
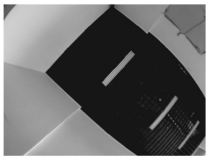
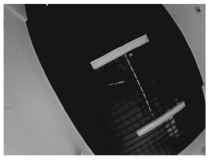
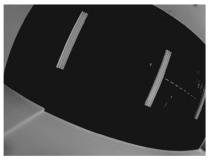
To verify the performance of the ALE-vSLAM algorithm on resource-constrained hardware, we ported both the ORB-SLAM3 and the ALE-vSLAM system to vision-guided sweepers for testing.

The robots utilized in this experiment are equipped with the Allwinner MR133 chip and the Tina Linux operating system, which is based on the Linux kernel and developed by the Allwinner Technology. The vision-guided sweeping robots use a 300 000-pixel monocular camera with a maximum framerate of 30.

To compare the performance differences between the ALE-vSLAM and ORB-SLAM3 systems within the same indoor setting, we conducted real-time testing along a predetermined closed-loop trajectory with the loop closure detection feature disabled. To emulate varying indoor visual environments, this study was divided into five separate tests, each conducted under conditions of progressively reduced indoor lighting. Fig. 10 displays the comparison of the motion trajectories generated by both the systems under five distinct lighting conditions.

Through the comparative analysis, it was observed that under bright lighting conditions [Fig. 10(a)], both ALE-vSLAM and ORB-SLAM3 maintained accurate trajectories during turns. However, deviations of varying degrees were noted at the endpoints along the closed-loop trajectories. In tests conducted under more ideal lighting conditions [Fig. 10(b)], both the systems successfully returned to the starting point after completing the circuit, exhibiting minimal trajectory errors.

TABLE III
COMPARISON OF THE NUMBER OF FEATURE POINTS EXTRACTED USING THE HOMEMADE DATA SET

Dataset-corresponding images	Number of extracted feature points	
	ORB-SLAM3 [17]	ALE-vSLAM (Ours)
	636	1007
	864	1008
	389	1003
	488	1009
	226	1004

However, as the level of illumination decreased, both the systems exhibited varying degrees of scale error during turning actions. As seen in Fig. 10(d), ORB-SLAM3 experienced significant accumulation of scale error after multiple turns, preventing accurate return to the starting point. In contrast, ALE-vSLAM was able to return to the starting location under the same test conditions. With further reduction in lighting, as shown in Fig. 10(e), the trajectories of both the systems showed significant drift. Nonetheless, compared to the ORB-SLAM3, ALE-vSLAM exhibited smaller scale errors during turning actions. The comparison of real-time trajectories demonstrates that the ALE-vSLAM algorithm outperforms ORB-SLAM3 in handling relatively complex indoor trajectories and lighting conditions.

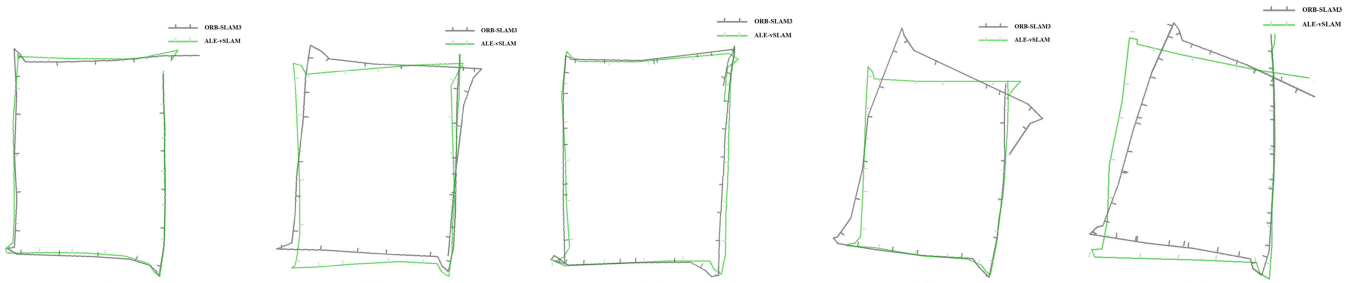
To further evaluate the performance of our algorithm under real-world varying lighting conditions, we conducted feature point extraction experiments using a homemade data set recorded by vision-guided sweeper robots. We applied both the ALE-vSLAM and ORB-SLAM3 algorithms to images captured at identical timestamps, to compare the number of feature points extracted by each algorithm under different lighting conditions.

Data analysis from Table III indicates that, in certain scenarios, such as overcast conditions or sudden illumination

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

(a) (b) (c)

Fig. 9. Hardware information and environment for empirical experiments. (a) Visual sweep robot. (b) MR133 chipset and parameters. (c) Top-down view of the experimental environment.



(a) (b) (c) (d) (e)

Fig. 10. Real-time closed-loop trajectory comparison under indoor light decline. (a) Bright light. (b) Normal bright light. (c) Normal light. (d) Normal dark light. (e) Dark light.

changes, the number of feature points extracted by the ORB-SLAM3 algorithm fluctuates between 200 and 800, suggesting that its feature point extraction performance may be affected under varying lighting conditions, which could adversely impact positioning accuracy and map building capabilities. In contrast, ALE-vSLAM was able to extract a relatively ample number of feature points in these scenarios, with the count of extracted feature points typically stabilizing around 1000. This further demonstrates ALE-vSLAM's enhanced robustness and reliability in handling scenes with changing lighting conditions.

E. Discussions

In comparative experiments, ALE-vSLAM outperforms ORB-SLAM3 and its versions combined with image enhancement technologies in terms of positioning accuracy. Especially in the MH04 and MH05 data sets, which have more complex lighting scenes, the improvement rates of ALE-vSLAM are 21.05% and 20.47%, respectively. This indicates that our code not only better fits the localization and mapping tasks under varying lighting conditions but also improves the accuracy of the algorithm in ordinary scenarios. This improvement is attributed to the following two points.

1) *Image Preclassification:* In response to the illumination invariance hypothesis of ORB-SLAM3, we preclassify images captured by the cameras in real-world scenes based on their brightness levels to ensure more suitable feature point extraction under various lighting conditions.
2) *Improved Feature Point Selection:* Beyond the BRIEF-based descriptors and pyramid structure of ORB-SLAM3, we further refine the feature point selection by incorporating higher thresholds, an enhanced FAST corner detection, and multiscale enhancement.

Combined with the results from Sections IV-B and IV-C, compared to the ORB-SLAM3, ALE-vSLAM ensures higher quality feature point extraction and matching in extreme lighting conditions, thereby improving the map accuracy and overall system efficiency in continuous frame estimation. Furthermore, experiments in Section IV-D have corroborated the practicality and reliability of ALE-vSLAM on resource-constrained hardware.

V. Conclusion

This article introduces an adaptive indoor lighting visual SLAM system, termed ALE-vSLAM, designed specifically for the resource-constrained hardware. Unlike the visual SLAM systems that incorporate traditional image enhancement methods, ALE-vSLAM employs a differentiated approach to processing image frames under varying lighting conditions, enhancing flexibility. Additionally, by optimizing the feature extraction process at the front end, ALE-vSLAM improves the efficiency and stability of feature point processing under fluctuating lighting conditions, enabling resource-limited hardware to adapt to a broader range of lighting environments.

The study found that the insufficient analysis of the lighting issues in input images and direct image processing could lead to a decrease in the number of matching inliers, thereby increasing the risk of losing low-texture areas, which adversely affects the system's accuracy assessment. For this reason, this article first proposes an image brightness grading strategy, to

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

select the appropriate image enhancement strategy based on the different brightness conditions. Second, for overexposed or underexposed areas in processed images, a dynamic parameter CLAHE multiscale enhancement method based on the contrast partitioning is introduced. Moreover, this article improves the FAST corner detection by combining image enhancement with higher thresholds, successfully filtering out features that would otherwise be eliminated due to lighting changes. This improvement preserves the robustness of the features while preventing unnecessary feature loss. Experiments conducted on the public data sets show that the ALE-vSLAM successfully reduces the impact of variable lighting in visually constrained situations, obtaining more accurate and high-quality feature points, thereby achieving superior positioning performance.

ACKNOWLEDGMENT

The authors would like to thank the referees for their constructive suggestions.

REFERENCES

[1] B. Huang, J. Zhao, and J. Liu, "A survey of simultaneous localization and mapping with an envision in 6G wireless networks," 2019, *arXiv:1909.05214*.

[2] P. Huang, L. Zeng, X. Chen, K. Luo, Z. Zhou, and S. Yu, "Edge robotics: Edge-computing-accelerated multirobot simultaneous localization and mapping," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 14087–14102, Aug. 2022.

[3] J. Wan, S. Tang, Q. Hua, D. Li, C. Liu, and J. Lloret, "Context-aware cloud robotics for material handling in cognitive Industrial Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2272–2281, Aug. 2018.

[4] D. Cai, R. Li, Z. Hu, J. Lu, S. Li, and Y. Zhao, "A comprehensive overview of core modules in visual SLAM framework," *Neurocomputing*, vol. 590, Jul. 2024, Art. no. 127760.

[5] R. Li, Y. Zhao, Z. Hu, W. Qi, and G. Liu, "TOHF: A feature extractor for resource-constrained indoor VSLAM," *J. Syst. Simul.*, 2024, doi: 10.16182/j.issn1004731x.Joss.23-1334.

[6] B. Fang and Z. Zhan, "A visual SLAM method based on point-line fusion in weak-matching scene," *Int. J. Adv. Robotic Syst.*, vol. 17, no. 2, 2020, Art. no. 1729881420904193.

[7] J. Nikolic, M. Burri, J. Rehder, S. Leutenegger, C. Huerzeler, and R. Siegwart, "A UAV system for inspection of industrial facilities," in *Proc. IEEE Aerosp. Conf.*, 2013, pp. 1–8.

[8] N. Brasch, A. Bozic, J. Lallemand, and F. Tombari, "Semantic monocular SLAM for highly dynamic environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2018, pp. 393–400.

[9] M. Servières, V. Renaudin, A. Dupuis, and N. Antigny, "Visual and visual-inertial slam: State of the art, classification, and experimental benchmarking," *J. Sens.*, vol. 2021, no. 1, 2021, Art. no. 2054828, doi: 10.1155/202/2054828.

[10] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—A modern synthesis," in *Proc. Int. Workshop Vis. Algorithms*, 2000, pp. 298–372.

[11] J. Zubizarreta, I. Aguinaga, and J. M. M. Montiel, "Direct sparse mapping," *IEEE Trans. Robot.*, vol. 36, no. 4, pp. 1363–1370, Aug. 2020.

[12] L. Yu, E. Yang, and B. Yang, "AFE-ORB-SLAM: Robust monocular VSLAM based on adaptive fast threshold and image enhancement for complex lighting environments," *J. Intell. Robotic Syst.*, vol. 105, no. 2, p. 26, 2022.

[13] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.

[14] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. 6th IEEE ACM Int. Symp. Mix. Augment. Real.*, 2007, pp. 225–234.

[15] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[16] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

[17] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.

[18] Y. Fang, G. Shan, T. Wang, X. Li, W. Liu, and H. Snoussi, "He-SLAM: A stereo slam system based on histogram equalization and orb features," in *Proc. Chin. Autom. Congr. (CAC)*, 2018, pp. 4272–4276.

[19] W. Yang and X. Zhai, "Contrast limited adaptive histogram equalization for an advanced stereo visual slam system," in *Proc. Int. Conf. Cyber-Enabl. Distrib. Comput. Knowl. Discovery (CyberC)*, 2019, pp. 131–134.

[20] Y. Miao et al., "Application of the CLAHE algorithm based on optimized bilinear interpolation in near infrared vein image enhancement," in *Proc. 2nd Int. Conf. Comput. Sci. Appl. Eng.*, 2018, pp. 1–6.

[21] S. Rahman, M. M. Rahman, M. Abdullah-Al-Wadud, G. D. Al-Quaderi, and M. Shoyaib, "An adaptive gamma correction for image enhancement," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, pp. 1–13, 2016, doi: 10.1186/s13640-016-0138-1.

[22] D. Hahnel, W. Burgard, D. Fox, and S. Thrun, "An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)(Cat. No. 03CH37453)*, 2003, pp. 206–211.

[23] T. K. Marks, A. Howard, M. Bajracharya, G. W. Cottrell, and L. Matthies, "Gamma-SLAM: Using stereo vision and variance grid maps for SLAM in unstructured environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2008, pp. 3717–3724.

[24] L. Hao, H. Li, Q. Zhang, X. Hu, and J. Cheng, "LMVI-SLAM: Robust low-light monocular visual-inertial simultaneous localization and mapping," in *Proc. IEEE Int. Conf. Robot. Biomimet. (ROBIO)*, 2019, pp. 272–277.

[25] R. Gomez-Ojeda, Z. Zhang, J. Gonzalez-Jimenez, and D. Scaramuzza, "Learning-based image enhancement for visual odometry in challenging HDR environments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2018, pp. 805–811.

[26] Y. Fu, B. Han, Z. Hu, X. Shen, and Y. Zhao, "CBAM-SLAM: A semantic SLAM based on attention module in dynamic environment," in *Proc. 6th Asian Conf. Artif. Intell. Technol. (ACAIT)*, 2022, pp. 1–6.

[27] X. Shen et al., "A closed-loop detection algorithm for online updating of bag-of-words model," in *Proc. 9th Int. Conf. Comput. Data Eng.*, 2023, pp. 34–40.

[28] H. Qi, Z. Hu, Y. Xiang, D. Cai, and Y. Zhao, "ATY-SLAM: A visual semantic SLAM for dynamic indoor environments," in *Proc. Int. Conf. Intell. Comput.*, 2023, pp. 3–14.

[29] D. Cai, Z. Hu, R. Li, H. Qi, Y. Xiang, and Y. Zhao, "AGAM-SLAM: An adaptive dynamic scene semantic SLAM method based on GAM," in *Proc. Int. Conf. Intell. Comput.*, 2023, pp. 27–39.

[30] A. Guiducci, "Corner characterization by differential geometry techniques," *Pattern Recognit. Lett.*, vol. 8, no. 5, pp. 311–318, 1988.

[31] K. Rangarajan, M. Shah, and D. Van Brackle, "Optimal corner detector," *Comput. Vis., Graph., Image Process.*, vol. 48, no. 2, pp. 230–245, 1989.

[32] A. Singh and M. Shneier, "Grey level corner detection: A generalization and a robust real time implementation," *Comput. Vis., Graph., Image Process.*, vol. 51, no. 1, pp. 54–69, 1990.

[33] K. Rohr, "Recognizing corners by fitting parametric models," *Int. J. Comput. Vis.*, vol. 9, no. 3, pp. 213–230, 1992.

[34] R. Deriche and T. Blaszka, "Recovering and characterizing image features using an efficient model based approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1993, pp. 530–535.

[35] S. M. Smith and J. M. Brady, "SUSAN—A new approach to low level image processing," *Int. J. Comput. Vis.*, vol. 23, no. 1, pp. 45–78, 1997.

[36] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. 9th Eur. Conf. Comput. Vis. (ECCV)*, 2006, pp. 430–443.

[37] G. Florentz and E. Aldea, "Superfast: Model-based adaptive corner detection for scalable robotic vision," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 1003–1010.

[38] Y. Xing, D. Zhang, J. Zhao, M. Sun, and W. Jia, "Robust fast corner detector based on filled circle and outer ring mask," *Inst. Eng. Technol. Image Process.*, vol. 10, no. 4, pp. 314–324, 2016.

[39] Y. Chen et al., "BEVSOC: Self-supervised contrastive learning for calibration-free BEV 3D object detection," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 22167–22182, Jun. 2024.

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.

[40] H. Qi, Y. Fu, Z. Hu, J. Wu, and Y. Zhao, "A lightweight semantic VSLAM approach based on adaptive thresholding and speed optimization," *J. Beijing Univ. Aeronaut. Astronaut.*, 2024, to be published.
[41] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems*, San Diego, CA, USA: Acad. Press, 1994, pp. 474–485.
[42] M. Burri et al., "The EuRoC micro aerial vehicle datasets," *Int. J. Robot. Res.*, vol. 35, no. 10, pp. 1157–1163, 2016.

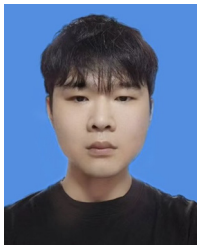
Zhuhua Hu (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from Jilin University, Changchun, China, in 2002 and 2005, respectively, and the Ph.D. degree from Hainan University, Haikou, China, in 2019.

He was a Software Engineer with Ningbo BIRD Research Institute of China, Ningbo, China, from 2005 to 2006. He was a Software Engineer with Nanjing Research Institute of ZTE, Nanjing, China, from 2006 to 2007. He was a Minister of Software Department with Shanghai Aoxun Information Technology Company, Ltd., Shanghai, China, from 2007 to 2009. He has been a Professor and a Doctorial Tutor with the School of Information and Communication Engineering, Hainan University since 2020. His current research interests include artificial intelligence and signal and information processing.

Prof. Hu is currently a high-level talent in Hainan Province. He acted as a reviewer for IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, IEEE/ACM TRANSACTIONS ON NETWORKING, *Engineering Applications of Artificial Intelligence*, *Knowledge-Based Systems*, *Frontiers in Marine Science*, *Remote Sensing*, *Applied Artificial Intelligence*, ICASSP2023, and ICME 2024. He is a Senior Member of CCF.

Wenlu Qi received the B.S. degree in electronic information from Taiyuan University of Science and Technology, Taiyuan, China, in 2021. She is currently pursuing the M.S. degree with the School of Information and Communication Engineering, Hainan University, Haikou, China.

Her research interest is visual SLAM.



Kunkun Ding received the B.S. degree in electronic information science and technology from Henan University Minsheng College, Kaifeng, China, in 2022. He is currently pursuing the M.S. degree with the School of Information and Communication Engineering, Hainan University, Haikou, China.

His research interest is visual SLAM.

Guangfeng Liu received the B.S. degree in information management and information systems from Shaoguan University, Shaoguan, China, in 2021. He is currently pursuing the M.S. degree with the School of Cyberspace Security (School of Cryptology), Hainan University, Haikou, China.

His research interest is visual SLAM.

Yaochi Zhao received the M.S. degree in pattern recognition and intelligent system from Central South University, Changsha, China, in 2005, and the Ph.D. degree from Tianjin University, Tianjin, China, in 2023.

She was with Ningbo BIRD Research Institute, Ningbo, China, and Shanghai Wingtech Communication Company, Ltd., Shanghai, China, for three years. She was engaged in teaching and research work with the College of Information Science and Technology, Hainan University, Haikou, China, where she is currently an Associate Professor with the School of Cyberspace Security (School of Cryptology). Her research interests include deep learning and intelligent information processing.

Authorized licensed use limited to: Hainan University. Downloaded on October 09,2025 at 08:19:32 UTC from IEEE Xplore. Restrictions apply.