



AFE-ORB-SLAM: Robust Monocular VSLAM Based on Adaptive FAST Threshold and Image Enhancement for Complex Lighting Environments

Leijian Yu¹ · Erfu Yang¹ · Beiya Yang¹

Received: 8 November 2021 / Accepted: 23 April 2022 / Published online: 17 May 2022
© The Author(s) 2022

Abstract

Monocular Visual Simultaneous Localisation and Mapping (VSLAM) systems are widely utilised for intelligent mobile robots to work in unknown environments. However, complex and varying illuminations challenge the accuracy and robustness of VSLAM systems significantly. Existing feature-based VSLAM methods often fail due to the insufficient feature points that can be extracted in those challenging illumination environments. Therefore, this paper proposes an improved ORB-SLAM algorithm based on adaptive FAST threshold and image enhancement (AFE-ORB-SLAM), which works in the environments with complex lighting conditions. An improved truncated Adaptive Gamma Correction (AGC) is combined with unsharp masking to reduce the effect caused by different illuminations. What is more, an improved ORB feature extraction method with the adaptive FAST threshold is proposed and adopted to obtain more reliable feature points. To verify the performance of the AFE-ORB-SLAM, three public datasets (the extended Imperial College London and National University of Ireland Maynooth (ICL-NUIM) dataset with different lighting conditions, Onboard Illumination Visual-Inertial Odometry (OIVIO) dataset and the European Robotics Challenge (EuRoC) dataset) are utilised. The results are compared with other state-of-the-art monocular VSLAM methods. The experimental results demonstrate that the AFE-ORB-SLAM could achieve the highest average localisation accuracy with robust performance in the environments with complex lighting conditions while keeping similar performance in the normal lighting scenarios.

Keywords Monocular VSLAM · Adaptive FAST threshold · Image enhancement · Complex lighting environments

1 Introduction

With the rapid development and wide applications of autonomous robotic systems, Simultaneous Localization and Mapping (SLAM) has attracted a lot of attention in the robotics research field [1, 2]. SLAM, which estimates the position of the robot while reconstructing the surrounding

environment simultaneously in the unknown environment, has vital theoretical significance and application value. Moreover, it is the core technology of autonomous robots in the unknown environment [3].

Lots of sensor modalities such as the sonar, Light Detection and Ranging (LiDAR) and camera can be used for different SLAM systems [4–6]. Among them, Visual SLAM (VSLAM) researches are blooming due to their convenience and relatively low requirements for sensors. VSLAM only relies on the camera, which obtains plenty of texture information and has been widely deployed on robotic platforms [7]. The accuracy and robustness of VSLAM are vital for autonomous navigation, especially in the complex lighting environment. For this reason, the stereo camera, depth camera or Inertial Measurement Unit (IMU) have been adopted to VSLAM systems to improve their performances [8]. However, these methods will cause inconvenience and extra cost to the system. Monocular VSLAM systems only rely on the lightweight camera, and

✉ Erfu Yang
erfu.yang@strath.ac.uk

Leijian Yu
leijian.yu@strath.ac.uk

Beiya Yang
beiya.yang@strath.ac.uk

¹ Department of Design, Manufacturing and Engineering Management, University of Strathclyde, Glasgow, G1 1XJ, United Kingdom

simple calibration makes them particularly attractive for many robotic applications [9].

Depending on the image matching methods, VSLAM systems can be divided into the feature-based method and direct method [10]. The former extracts feature points from images and finds their corresponding based on geometric constraints, while the latter finds the corresponding of different frames based on their pixel intensities directly. The VSLAM has been investigated from different perspectives, and lots of cutting-edge VSLAM methods such as the ORB-SLAM3 [11] and Direct Sparse Mapping (DSM) [12] have been developed. However, most advanced VSLAM systems are evaluated in well-lighted environments without considering challenging lighting conditions, such as dark, over-bright or dynamic illumination conditions. Visual blur or feature changes because of different illumination conditions occur in these complex lighting environments. Thereby, the feature matching or frame-to-frame matching process is significantly affected by the changes of illumination conditions. As a result, VSLAM systems may fail in these environments. Thus, developing a robust monocular VSLAM system for the challenging scenario with complex light has significant research and application value.

Towards this end, this paper presents a robust monocular VSLAM system (AFE-ORB-SLAM) through adopting the proposed adaptive FAST threshold and image enhancement techniques. In the proposed AFE-ORB-SLAM, the ORB-SLAM3 is chosen as the framework due to its excellent performance in well-lit environments. In order to handle the poor performance of the ORB-SLAM3 in challenging lighting environments, the truncated Adaptive Gamma Correction (AGC) is enhanced and combined with the unsharp masking method in the AFE-ORB-SLAM. Meanwhile, the proposed system is improved by the proposed efficient and adaptive FAST threshold method. The main contributions of this work are surmised as follows:

1. A truncated AGC is improved and combined with the unsharp masking technique as the image enhancement method to reduce the effect caused by different illumination conditions for the ORB-SLAM3.
2. An improved ORB feature points extraction method based on the adaptive FAST threshold is proposed and adopted to the ORB-SLAM3, which obtains more reliable feature points.
3. Extensive experiments are conducted to validate the accuracy and robustness of the AFE-ORB-SLAM. Experimental results evidence that the AFE-ORB-SLAM outperforms other state-of-the-art algorithms in challenging lighting scenarios while keeping similar performance in the normal lighting scenarios.

The rest of the paper is structured as follows: Section 2 reviews related works. Section 3 provides the framework of the AFE-ORB-SLAM. Section 4 presents the details of improved image enhancement and feature extraction. Experimental results and analysis are given in Section 5. Section 6 concludes the whole work.

2 Literature Review

2.1 Feature-Based VSLAM Methods

For feature-based VSLAM methods, image enhancement has been widely utilised to handle the challenging lighting environment. The Histogram Equalisation (HE) was adopted to the HE-SLAM, which improved the contrast of captured low-contrast images [13]. Compared to the ORB-SLAM2 [14], the HE-SLAM was more robust in a harsh environment. However, the HE is affected by background noises significantly. To improve the robustness of the HE-SLAM, Yang et al. [15] adopted the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm to the ORB-SLAM2 framework. The trajectory generated by this method was closer to the ground truth than that calculated by the HE-SLAM and ORB-SLAM2. The Dim-light Robust Monocular SLAM (DRMS) was proposed in [16], which utilised the linear transformation and CLAHE algorithms as the image pre-processing to enhance the brightness and contrast of input images. After that, the performance of the proposed VSLAM in the dim-light conditions was improved. However, the CLAHE algorithm calculates the neighbourhood histogram for each pixel and performs histogram equalisation processing for sub-regions of the image, which is computationally extensive [17]. Meanwhile, the aforementioned methods mainly focus on either global or local enhancement for all kinds of images, and it may not perform well for different types of images [18]. Moreover, the decreased sharpness of images owing to image transformations should also be taken into consideration [19]. Recently, the deep neural network was adopted to increase representations of captured images for the VSLAM algorithm, which improved the robustness of the VSLAM system in High Dynamic Range (HDR) environments [20]. However, deep-learning based methods need even more computing resources than traditional methods. What is more, even with the enhanced images, sufficient feature points still could not be extracted in some challenging environments. Thus, investigating the VSLAM combined with multiple features has gained research interest. Pumarola et al. [21] proposed the PL-SLAM that relied on line features as well as the ORB features. To this end, a more robust performance

in environments with challenging illumination conditions was achieved. Huang et al. [22] processed ORB and Brisk feature points at the same time for a low-lighting environment to improve the robustness of the VSLAM system. However, extracting multiple features needs extra computing resources. Thereby, their applications for mobile robots are restricted due to robots' limited onboard computing capabilities.

2.2 Direct VSLAM Methods

In terms of direct VSLAM methods, several works have been done towards complex lighting environments. Extensive experiments have been conducted to verify direct VSLAM systems towards changing illumination environments in [23]. Experiments showed that most direct VSLAM systems failed due to abrupt illumination changes while the brightness constancy assumption was adopted. Sun et al. [24] combined the RGB channel linearly to compensate for the lighting changes. In [25], illumination changes were modelled for affine lighting correction. Thereby, the illumination invariance was handled for the direct VSLAM system. As these methods still rely on the brightness constancy assumption, direct VSLAM systems still cannot handle complex lighting environments.

3 Structure of the AFE-ORB-SLAM

This work proposes a novel monocular VSLAM system (AFE-ORB-SLAM) based on the ORB-SLAM3 framework for complex lighting environments. The overall schematic architecture is presented in Fig. 1. The two blocks with words in red are the main novel works proposed in this work. Three parallel threads (tracking, local mapping and loop and map merging) are utilised by the ORB-SLAM3. Besides, all the generated maps are managed by the Atlas, which is a novel multiple-map system described below.

3.1 Atlas

The Atlas is a multi-map technology that manages all the sub-maps generated by the ORB-SLAM3. The sub-map, utilised by the tracking thread to locate incoming frames, is called the active map. All other sub-maps are called non-active maps. Both active and non-active maps consist of map points, keyframes, covisibility graphs and spanning trees.

3.2 Tracking Thread

The tracking thread is responsible for computing the camera pose and deciding new keyframes. To improve the robustness for illumination variation, the incoming frame is pre-processed by the improved image enhancement method. Next, FAST keypoints are detected with the adaptive threshold and then described with rotated BRIEF descriptor. The details of improved image enhancement and keypoints extraction methods will be given in Section 4. Afterwards, feature points are matched to initialise, track or re-localise the current camera pose. Finally, the tracking thread decides whether the current frame is a keyframe.

3.3 Local Mapping Thread

After a new keyframe is deemed by the tracking thread, the local mapping thread is invoked. It eliminates bad map points and adds the keyframe and associated map points into the active map. Poses of keyframes and map points observed by the newly added keyframe will be optimised by a local bundle adjustment. To maintain a scalable and reliable map, the redundant keyframes are deleted.

3.4 Loop and Map Merging Thread

In the loop and map merging thread, the newly added keyframe is compared to all frames stored in the Atlas. If two similar frames exist in different maps, both maps are

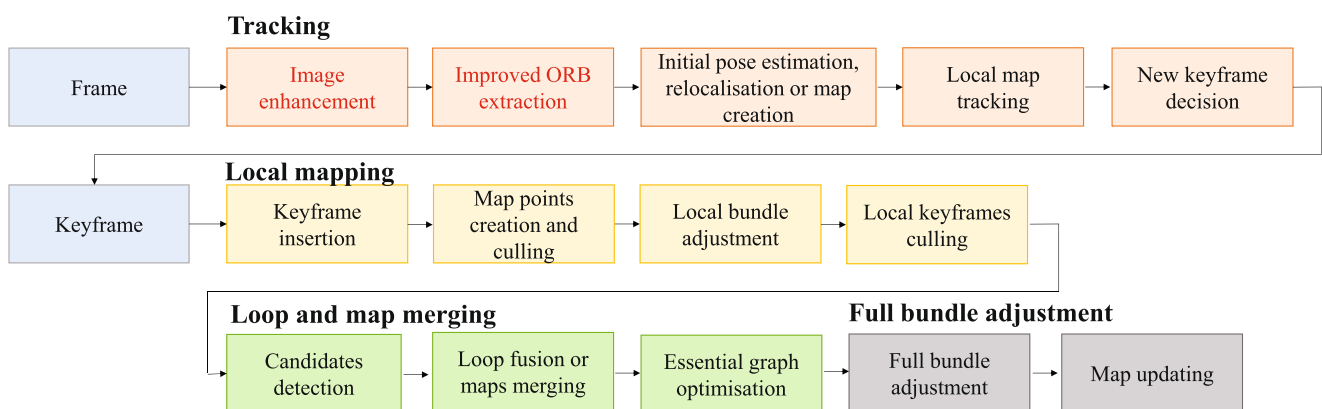


Fig. 1 Structure of the AFE-ORB-SLAM

merged as a new active map. On the contrary, a pose-graph optimisation is carried out to reduce the accumulated drift error within the active map. After that, the global bundle adjustment is launched in an independent thread to refine the map further.

Although the ORB-SLAM3 with the monocular sensor achieves impressive performance in well-lit environments, its accuracy and robustness still suffer a lot in complex lighting environments. In these environments, the performances of feature extraction and matching drop significantly. When there are not enough matched ORB feature points obtained from the surrounding environment, the pose estimation process cannot be implemented, even leading to initialisation and tracking failures [19]. For the reason mentioned above, it is crucial to incorporate the algorithm that can handle the variations of illumination in the ORB-SLAM3. In this paper, the image enhancement technology and ORB feature extraction are improved and deployed in the tracking thread to solve this problem. After the image enhancement, more distinguished texture information is revealed. Besides, more stable ORB feature points could be obtained with the adaptive threshold FAST feature extraction in complex lighting environments. Eventually, the accuracy and robustness of the ORB-SLAM3 are enhanced in complex lighting environments.

4 Image Enhancement and Feature Extraction

4.1 Image Enhancement

The texture information is decreased in the dimmed or over-bright image. Thus, the captured images suffer from poor contrast. Contrast enhancement algorithms improve the visibility of objects in the dimmed or bright area by directly modifying pixel values based on the proper regulation [26]. Gamma correction [27] has gained lots of interest due to easy adjustment and efficient implementation. The Adaptive Gamma Correction algorithm with Weighting Distribution (AGCWD) [28] behaves well to enhance the images captured in the low-lighting environment. As the AGCWD focuses on improving contrast of dimmed images, detail loss occurs in the bright area. Inspired by Cao et al’s work [29], truncated Cumulative Distribution Function (CDF) based AGCWD (IAGCWD) is improved and adopted to process both dimmed and over-bright images. Thereby, the local over-enhancement can be reduced. To compensate for decreased sharpness because of the image transmission and transformation, details and contours of the image are enhanced through unsharp masking technology. Eventually, with the combination of image contrast enhancement and image sharpening adjustment technologies, texture

information, especially for contours contained in the image, will be more prominent.

The overall structure of the proposed image enhancement method is shown in Fig. 2. It consists of the contrast adjustment module and sharpening adjustment module.

4.1.1 Image Contrast Enhancement

The standard deviation of the image intensity denotes the average contrast of the image [30], and it can be used to divide one image as the low contrast image and the high contrast image. The standard deviation of the image intensity is represented by λ .

$$I(x, y) = \begin{cases} I_{low}(x, y) & \lambda \leq 0.25 \\ I_{high}(x, y) & otherwise \end{cases} \quad (1)$$

The Probability Density Function (PDF) can be calculated by

$$P(l) = \frac{n_l}{N} \quad (2)$$

where l is the pixel intensity in the (x, y) position. n_l represents the number of pixels with intensity l , and N indicates the total pixels contained in the image. After the histogram distribution is smoothed by the weighting distribution function [31], the weighting distributed PDF can be formulated as

$$P_{wd}(l) = P_{max} \left(\frac{P(l) - P_{min}}{P_{max} - P_{min}} \right)^i \quad (3)$$

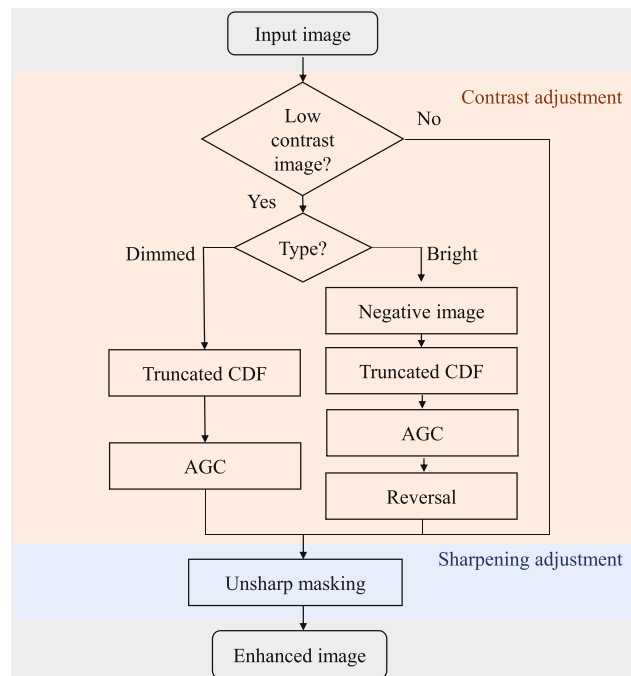


Fig. 2 Structure of the proposed image enhancement method

where i is used to adjust the smooth level. The sum of $P_{wd}(l)$ can be calculated through

$$\sum P_{wd} = \sum_{l=0}^{l_{max}} P_{wd}(l) \tag{4}$$

Thereby, the CDF can be formulated as

$$C_{wd}(l) = \frac{\sum_{l=0}^{l_{max}} P_{wd}(l)}{\sum P_{wd}} \tag{5}$$

Eventually, the parameter γ_{wd} can be obtained through

$$\gamma_{wd} = 1 - C_{wd}(l) \tag{6}$$

To improve the performance of the image enhancement, the dimmed and bright images should be processed differently. Thus, based on the average pixel intensity m_I , t is calculated to represent the overall brightness of the image

$$t = \frac{m_I - 128}{128} \tag{7}$$

Finally, the image is divided into the bright and dark sub-classes based on the value of t .

$$I(x, y) = \begin{cases} I_{bright}(x, y) & t \geq 0 \\ I_{dark}(x, y) & t < 0 \end{cases} \tag{8}$$

Following the image classification, the contrast of dimmed and bright images will be restored separately. The bright region in the dimmed image will be degraded due to an overly low gamma value. To this end, a truncated CDF is utilised.

$$\gamma'_{wd} = \max(\tau, \gamma_{wd}) \tag{9}$$

τ is the threshold used for CDF truncation. It makes sure that bright regions are not adjusted by a low gamma value. From plentiful experimental observations, it is set to 0.3 in this work. Thereby, the detailed contour information in the bright area could be reserved. With the adoption of CDF truncation, the dimmed pixels will be processed by a small gamma value, while the restricted adjustment is applied to bright pixels. Thereby, the pixel intensity could be transformed by the following equation:

$$I_{ce}(l) = 255 \left(\frac{l}{255} \right)^{\gamma'_{wd}} \tag{10}$$

Specifically, the process to enhance the contrast of the dimmed image is introduced in Algorithm 1.

Algorithm 1 Contrast enhancement for the dimmed image

- Step1: Calculate the $P(l)$ of the input image I .
 - Step2: Compute $P_{wd}(l)$ with the weighting distribution function.
 - Step3: Obtain $C_{wd}(l)$ according to (5)
 - Step4: Calculate and choose proper γ'_{wd} .
 - Step5: Output the contrast enhanced image I_{ce} .
-

Massive pixels in the dimmed or over-bright images have similar intensities. Over-bright images have high pixel intensities, and their negative images contain an enormous number of pixels with low-intensity values. Thus, the negative image of the over-bright image can be treated as a dimmed image, and it is formed by:

$$I' = 255 - I(x, y) \tag{11}$$

Then, Algorithm 1 can be utilised directly to enhance I' . After that, the final contrast enhanced image I_{ce} could be obtained through the reverse of the enhanced negative image.

Finally, the contrast enhancement mask $T_{mask}(x, y)$ can be obtained as.

$$T_{mask}(x, y) = I_{ce}(x, y) - I(x, y) \tag{12}$$

4.1.2 Sharpening Adjustment

Image sharpening enhancement highlights the contour and makes the textures of the image clear. Unsharp masking is a typical image sharpening technique. This technique utilises a low-pass filter to get a blurred image. Based on that, a mask is created and combined with the original image to make the texture of the image clear. Specifically, the process of unsharp masking can be realised through the following steps:

The input image is processed by one low-pass filter

$$f(x, y) = I(x, y) * h(m, n) \tag{13}$$

where $*$ denotes the convolution operator, and $h(m, n)$ is a low-pass filter.

Unsharp mask $g_{mask}(x, y)$ can be calculated through

$$g_{mask}(x, y) = I(x, y) - f(x, y) \tag{14}$$

The sharpened image can be obtained through

$$g_{sa}(x, y) = I(x, y) + k \cdot g_{mask}(x, y) \tag{15}$$

where k represents the sharpening level. For the unsharp masking technique, k is set to 1. In this work, a Gaussian low-pass filter is used, which could be represented by

$$G_0(x, y) = e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{16}$$

in which σ is the standard deviation of the normal distribution.

Finally, the enhanced image can be represented by

$$I_{enhanced} = \begin{cases} I(x, y) + \alpha \cdot g_{mask}(x, y) + \beta \cdot T_{mask}(x, y) & \lambda \leq 0.25 \\ I(x, y) + \alpha \cdot g_{mask}(x, y) & otherwise \end{cases} \tag{17}$$

α and β are two adjustable parameters, which control the level of image contrast enhancement and image sharpening adjustment, respectively. As the unsharp masking technique is utilised in this work, α should be set to 1.0. β is obtained

by carrying out extensive experiments under different scenarios to achieve relatively good results, and it is set to 0.3. Users can adjust them to acquire a more preferable result in a specific environment.

4.2 Adaptive FAST Threshold for ORB Extraction

The ORB feature consists of the FAST keypoint and BRIEF descriptor. If the pixel intensity significantly differs from that of surrounding pixels, this pixel will be treated as a keypoint. To detect whether a pixel p is a FAST keypoint, the pixel intensity I_p will be compared with that of 16 pixels on a circle with a radius of 3 pixels (as shown in Fig. 3). A threshold θ is set manually to distinguish the current and surrounding 16 pixels. If there are over 12 contiguous pixels brighter than $I_p + \theta$ or darker than $I_p - \theta$, the current pixel will be considered as a FAST keypoint. To improve the detection efficiency, the differences between the current pixel and pixels on the circle with numbers 1, 5, 9 and 13 will be detected first. Wherein at least 3 points meet the condition that the pixel intensity difference is larger than θ or smaller than $-\theta$, the remaining 12 pixels on the circle will be detected. Otherwise, the pixel p will be discarded. Then, the scale, orientation invariance and BRIEF descriptor will be calculated by following the approach in [32].

Through the analysis above, the threshold θ is vital for the feature extraction process. Thereby, the performance of the whole VSLAM system will be improved through a proper θ value. However, a fixed θ cannot be adjusted to different illumination conditions. Thus, the feature extraction is degraded in different environments. To overcome this problem, an adaptive FAST threshold calculation method is proposed and adopted to the AFE-ORB-SLAM. Considering the computing efficiency, the λ used for the image enhancement is utilised to control the value of θ . Following the feature extraction process utilised in the ORB-SLAM3, two adaptive threshold values are set. The values of θ are set to 20 and 7 by the ORB-SLAM3. Similarly, if enough feature points can be extracted from

the environment, a relatively large θ is used to obtain more reliable feature points.

$$\theta = \omega \cdot \lambda + 20 \quad (18)$$

If the number of extracted feature points is not enough in a quite low contrast image, a relatively small θ will be set.

$$\theta = \frac{\omega \cdot \lambda}{2} \quad (19)$$

ω is a parameter to control the threshold for ORB feature points extraction. In our work, as the texture information is enriched, we set ω to 128 which is the median value of the pixel intensity. To a specific scenario, users can adjust ω accordingly to obtain the best result.

5 Experiments and Analysis

5.1 Experimental Environment

To verify the performance of the proposed AFE-ORB-SLAM, a laptop with Ubuntu 16.04 is used. The processor is Intel(R) Core(TM) i7-8750H and the program uses C++ 17 compilation. Besides, the laptop is equipped with 12GB RAM. The Imperial College London and National University of Ireland Maynooth (ICL-NUIM) dataset with simulated lighting changes [23], Onboard Illumination Visual-Inertial Odometry (OIVIO) dataset [33] and the European Robotics Challenge (EuRoC) dataset [34] are utilised to verify the localisation accuracy and illumination robustness of the proposed AFE-ORB-SLAM.

The ICL-NUIM dataset with simulated lighting changes is a synthetic dataset, and the camera position is available as the ground truth. It contains image sequences under different illumination conditions. Thus, it is suitable for testing the performance of VSLAM systems under different lighting conditions. We use office room sequences with static, local variation, global variation and local and global

Fig. 3 FAST keypoint extraction

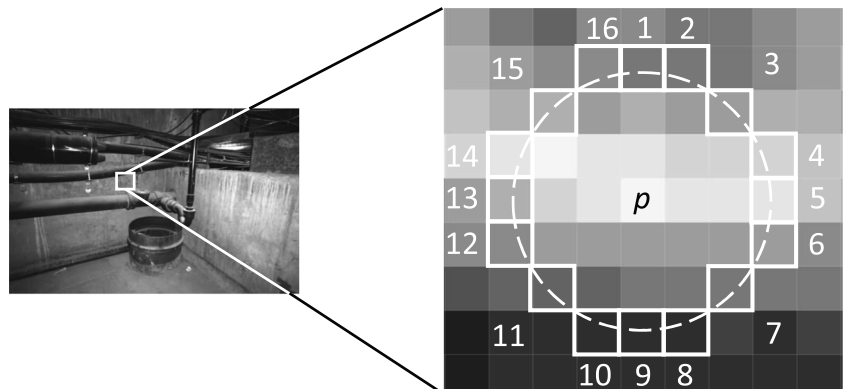


Fig. 4 Example images in the ICL-MUIM dataset with simulated lighting changes



variation lighting conditions in this work. Some sample images are shown in Fig. 4.

The OIVIO dataset contains 9 image sequences captured by the Clearpath Husky UGV in weakly lighted environments, such as mines, tunnels and other dark environments. There are 3 scenarios named “MINE GROUND-VEHICLE 1”, “MINE GROUND-VEHICLE 2” and “TUNNEL GROUND-VEHICLE 1” have the ground truth generated by the Leica TCRP1203 R300, and these sequences are utilised in our work to verify the performance of VSLAM systems. What is more, an onboard light of approximately 1350, 4500, or 9000 lumens is utilised to illuminate each scene.

The EuRoC dataset contains 11 sequences collected by the AscTec “Firefly” hex-rotor helicopter. Among them, 5 sequences are recorded in a large machine hall with ground truth provided by a Leica Multistation. The other 6 sequences are recorded in a small Vicon room with ground truth provided by the motion capture system. To complete the V103 sequence, the ORB-SLAM3 relies on the multi-map system significantly, and the ORB-SLAM3 cannot complete the V203 sequence. Tracking losses will lead to unpredictable threats to robot platforms. Thus, in this work, the other 9 sequences are chosen to validate VSLAM methods to simulate their performances on a robot platform.

If the trajectory has a loop, the motion trajectory generated by ORB-SLAM3 and other ORB-SLAM based VSLAM algorithms will be optimised by g2o [35].

5.2 Verification of Image Enhancement

To compare performances of different image contrast enhancement methods, the HE and CLAHE that are utilised in VSLAM systems, and the original IAGCWD are chosen. Figure 5 demonstrates the results of different image enhancement algorithms. Figure 5(a) indicates the original images selected from different scenarios. As shown in Fig. 5(b) and (c), some high contrast images are achieved. However, if there are some noises contained in the images, the noises will also be increased significantly. Figure 5(d) shows the results achieved by the IAGCWD, it incurs over-enhancement in some bright regions. Figure 5(e) proves that the contrast and visibility of the texture information contained in images are enhanced by the proposed method.

5.3 Evaluation on the ICL-NUIM Dataset with Simulated Lighting Changes

To verify a VSLAM system, the Absolute Trajectory Error (ATE) [36] is a common practice. The ATE represents the difference between the ground truth and the path estimated by the VSLAM system. Figure 6 shows the visible trajectory of the ORB-SLAM3 and AFE-ORB-SLAM in the office scenario with local and global variation lighting conditions. The results prove that the trajectory generated by the AFE-ORB-SLAM is closer to the ground truth compared to the ORB-SLAM3. A large offset occurs on the initialising stage

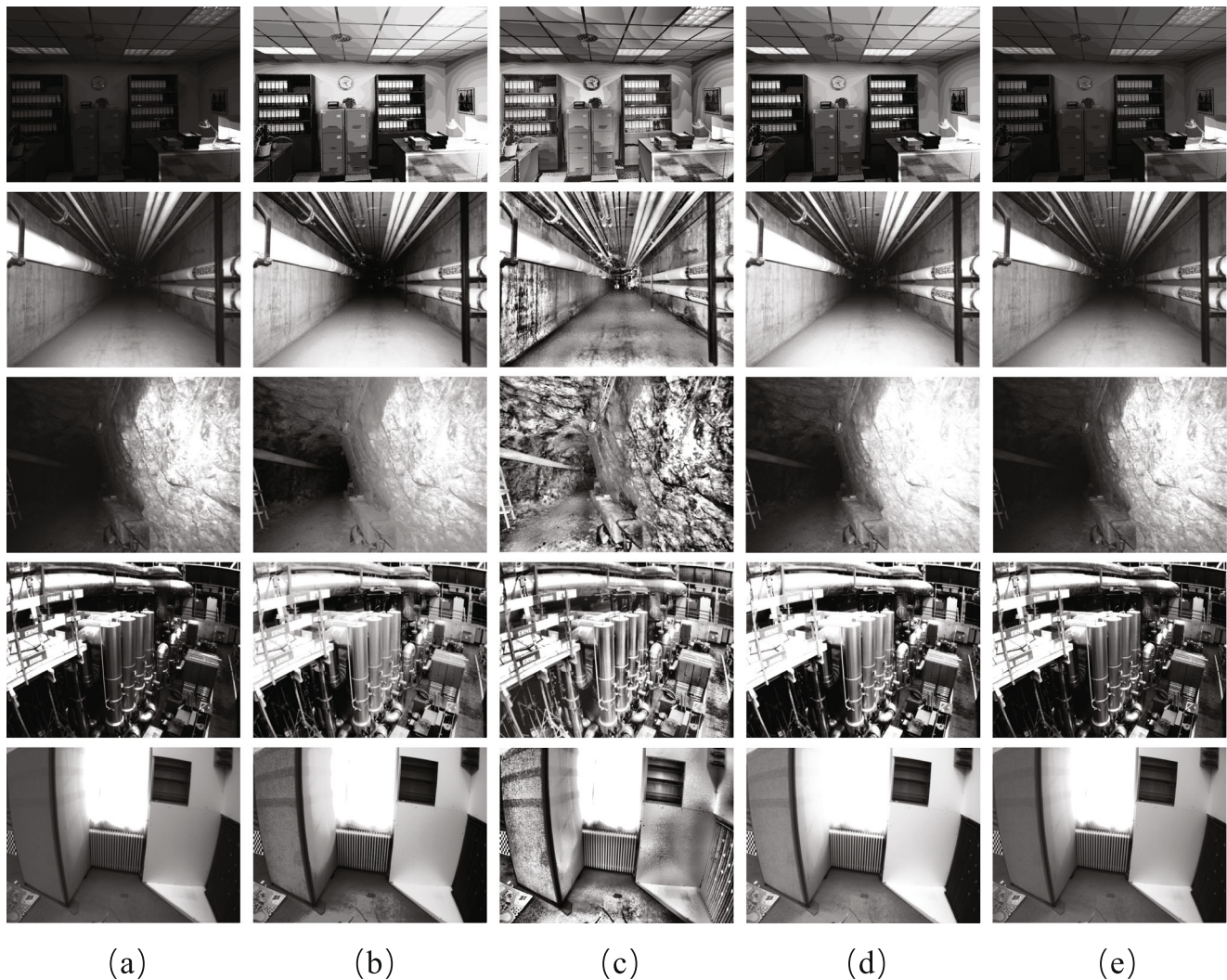


Fig. 5 Results of image enhancement. (a) Original images; Enhanced images by HE (b), CLAHE (c), IAGCWD (d) and the proposed method (e)

of the ORB-SLAM3 while the AFE-ORB-SLAM has a relatively smaller error compared to the ground truth, and our method outperforms the ORB-SLAM3 in all coordinate directions.

To further verify the pose estimation performance of the AFE-ORB-SLAM, the PL-SLAM, DSM and ORB-SLAM3 with default parameters are selected for comparison. The median value of the localisation results for each method from 10 times running is presented.

Table 1 indicates the mean ATE and Root Mean Square (RMS) ATE of keyframe trajectories. If the VSLAM system cannot complete all the sequences, the results will be marked by *. The AFE-ORB-SLAM outperforms the original ORB-SLAM3 in all video sequences. For the DSM, even it could achieve the best localisation accuracy in several sequences contained by the Syn1 scenario, it is still vulnerable to different illumination conditions. Moreover, our method shows the smallest error considering

the average performance of the same sequence under different illumination conditions. The overall results prove that the AFE-ORB-SLAM achieves accurate localisation with robustness to illumination variations.

5.4 Evaluation on the OIVIO Dataset

To further evaluate the performance of the AFE-ORB-SLAM, apart from the VSLAM methods utilised in Section 5.3, VSLAM systems improved by the image contrast enhancement methods are also utilised for comparison. The HE-SLAM and CLAHE-SLAM represent the monocular version of [13] and [15], respectively. The IAGC-SLAM indicates the ORB-SLAM3 with the IAGCWD as the pre-processing technique. Meanwhile, the effect of the proposed image contrast enhancement method and the adaptive FAST threshold for ORB feature extraction are analysed separately. If only the proposed image contrast enhancement

Fig. 6 Comparison of the ORB-SLAM3 and AFE-ORB-SLAM. (a) and (b) are the overview of the whole trajectory. (c) and (d) are the detailed camera position in x, y and z directions

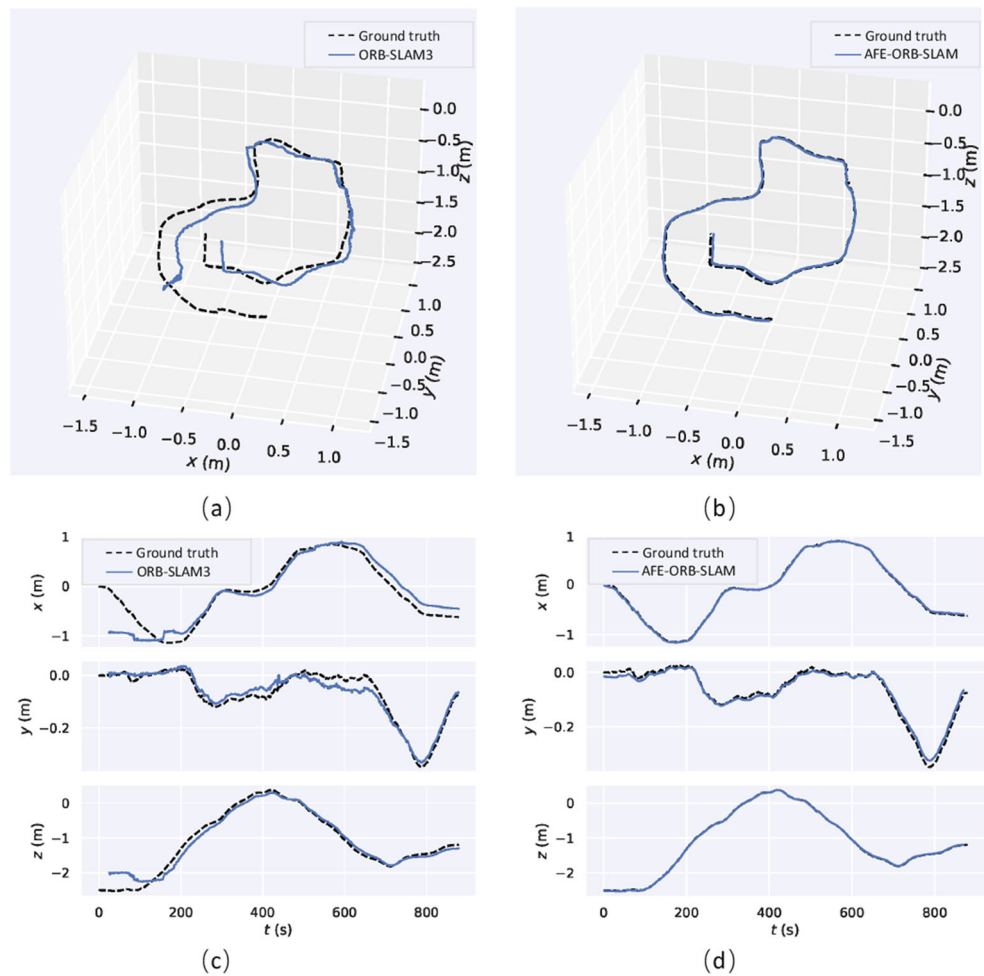


Table 1 Performance comparison on the ICL-NUIM dataset with simulated lighting changes for the mean ATE (m) and RMS ATE (m). The best results are highlighted in a bold font

ICL-NUIM benchmark	DSM		PL-SLAM		ORB-SLAM3		AFE-ORB-SLAM	
	Mean ATE	RMS ATE	Mean ATE	RMS ATE	Mean ATE	RMS ATE	Mean ATE	RMS ATE
Syn1	0.0028	0.0030	0.0271	0.0311	0.0684	0.0732	0.0340	0.0375
Syn1-local	0.8524	0.8937	0.0385	0.0429	0.2214	0.2952	0.0376	0.0407
Syn1-global	0.0031	0.0035	-	-	0.1543	0.2065	0.0329	0.0365
Syn1-local-global	0.0112	0.0119	0.2158	0.3132	0.1309	0.1684	0.0299	0.0333
Syn1-average	0.2174	0.2280	0.0938*	0.1291*	0.1438	0.1858	0.0336	0.0370
Syn2	0.5426	0.5997	0.0964	0.1093	0.0848	0.1273	0.0717	0.0817
Syn2-local	0.5438	0.5920	0.0841	0.1001	0.0758	0.0951	0.0889	0.1160
Syn2-global	0.5266	0.5722	0.0935	0.1046	0.1103	0.1458	0.0775	0.0839
Syn2-local-global	0.5198	0.5664	0.0858	0.0992	0.0858	0.0982	0.0670	0.0764
Syn2-average	0.5332	0.5826	0.0899	0.1033	0.0892	0.1166	0.0763	0.0895

method is adopted to the ORB-SLAM3, the VSLAM system is named the IE-SLAM. The TH-SLAM represents the ORB-SLAM3 improved by the adaptive FAST threshold for ORB feature extraction. Finally, the DSM, PL-SLAM, ORB-SLAM3, HE-SLAM, CLAHE-SLAM, IAGC-SLAM, IE-SLAM and TH-SLAM are selected to compare with the proposed AFE-ORB-SLAM.

To simulate the performance of different VSLAM systems on robot platforms, the full trajectories generated by VSLAM systems are used to calculate the RMS ATE. As the DSM and PL-SLAM only output the keyframe trajectory, the keyframe trajectory is still utilised in this section. The median RMS ATE of 10 executions is provided in Table 2. If the VSLAM system cannot complete all sequences, the results are marked by *. Compared with the ICL-NUIM dataset with simulated lighting changes, sequences in the OIVIO dataset have long trajectories, and there is no loop closure during the whole process. What is more, the texture information is not as rich as that of the ICL-NUIM dataset with simulated lighting changes, especially for the TUNNEL scenario. The RMSE ATE obtained in this dataset are larger than the ICL-NUIM dataset with simulated lighting changes.

The performance of DSM is influenced by the illumination conditions significantly, and it achieves the worst performance in almost all sequences. The PL-SLAM fails in the TN_015_GV_01 sequence due to the weak visual connectivity. The HE-SLAM, CLAHE-SLAM and IAGC-SLAM achieve higher localisation accuracy than the ORB-SLAM3 in several sequences. However, the noises contained in the images are also enhanced, and the over-enhancement exists in some regions. The average accuracy of the ORB-SLAM3 outperforms the HE-SLAM and IAGC-SLAM, while the CLAHE-SLAM and ORB-SLAM3 have similar average accuracy. The IE-SLAM and TH-SLAM obtain better results than the ORB-SLAM3 in most sequences. However, due to the less texture information contained

in the TUNNEL scenario, the TH-SLAM performs worse than the ORB-SLAM3. Apparently, the AFE-ORB-SLAM obtains the best localisation performance in all sequences.

The visualised localisation results are exhibited in Fig. 7 for the TUNNEL scenario. It shows that the low visibility of the environment has a great impact on the DSM. The PL-SLAM and ORB-SLAM3 rely on feature matching against neighbouring frames. When not enough reliable matched feature pairs are obtained, significant performance degradation can be observed. Considering the HE and CLAHE algorithms cannot handle different images properly, their performances are also influenced by different illumination conditions. The proposed AFE-ORB-SLAM could localise the robot accurately under different illumination conditions.

The time usage of the DSM, PL-SLAM, ORB-SLAM3, HE-SLAM, CLAHE-SLAM and AFE-ORB-SLAM for the scenario under different illumination conditions are averaged and shown in Fig. 8. The results further confirm that the AFE-ORB-SLAM is able to deal with the challenging scenes that provide less visual information effectively and efficiently. The average accuracy is improved by 34.65% with the comparison to the ORB-SLAM3. In contrast, the average processing time is only increased by 3.38%.

5.5 Evaluation on the EuRoC Dataset

The AFE-ORB-SLAM is further validated on the EuRoC dataset. Comparisons of the AFE-ORB-SLAM against the DSO [37], SVO [38], DSM, ORB-SLAM3, HE-SLAM and CLAHE-SLAM are presented in this section. The results published in [37] for the DSO, in [38] for the SVO and in [12] for the DSM are utilised. For other VSLAM systems, the median RMS ATE of 10 executions for the full trajectories is obtained. The results are shown in Table 3. As the most images in the EuRoC dataset contains rich texture information with regular lighting

Table 2 Performance comparison on the OIVIO dataset for the RMS ATE (m). The best results are highlighted in a bold font

OIVIO benchmark	DSM	PL-SLAM	ORB-SLAM3	HE-SLAM	CLAHE-SLAM	IAGC-SLAM	IE-SLAM	TH-SLAM	AFE-ORB-SLAM
MN_015_GV_01	5.5781	0.1858	0.1780	0.2848	0.2033	0.1746	0.1528	0.1461	0.1265
MN_050_GV_01	0.9177	0.2707	0.2218	0.2128	0.1965	0.2465	0.1810	0.1873	0.1431
MN_100_GV_01	0.6433	0.2494	0.1714	0.1765	0.1467	0.1799	0.1779	0.1578	0.1254
MN_015_GV_02	3.3304	0.2150	0.1186	0.1285	0.1014	0.1527	0.0937	0.1041	0.0855
MN_050_GV_02	0.8907	0.1977	0.1040	0.1203	0.1256	0.1587	0.0921	0.0969	0.0891
MN_100_GV_02	0.4740	0.1428	0.0964	0.1298	0.0886	0.1315	0.0936	0.0928	0.0854
TN_015_GV_01	0.8458	-	0.3231	1.1108	0.4548	0.4131	0.2751	0.2382	0.1728
TN_050_GV_01	1.0153	0.5171	0.2693	0.5378	0.2591	0.4320	0.2271	0.3393	0.1569
TN_100_GV_01	0.4948	0.3394	0.2551	0.2425	0.1608	0.2039	0.1649	0.3241	0.1511
Average	1.5767	0.2647*	0.1931	0.3281	0.1930	0.2325	0.1620	0.1874	0.1262

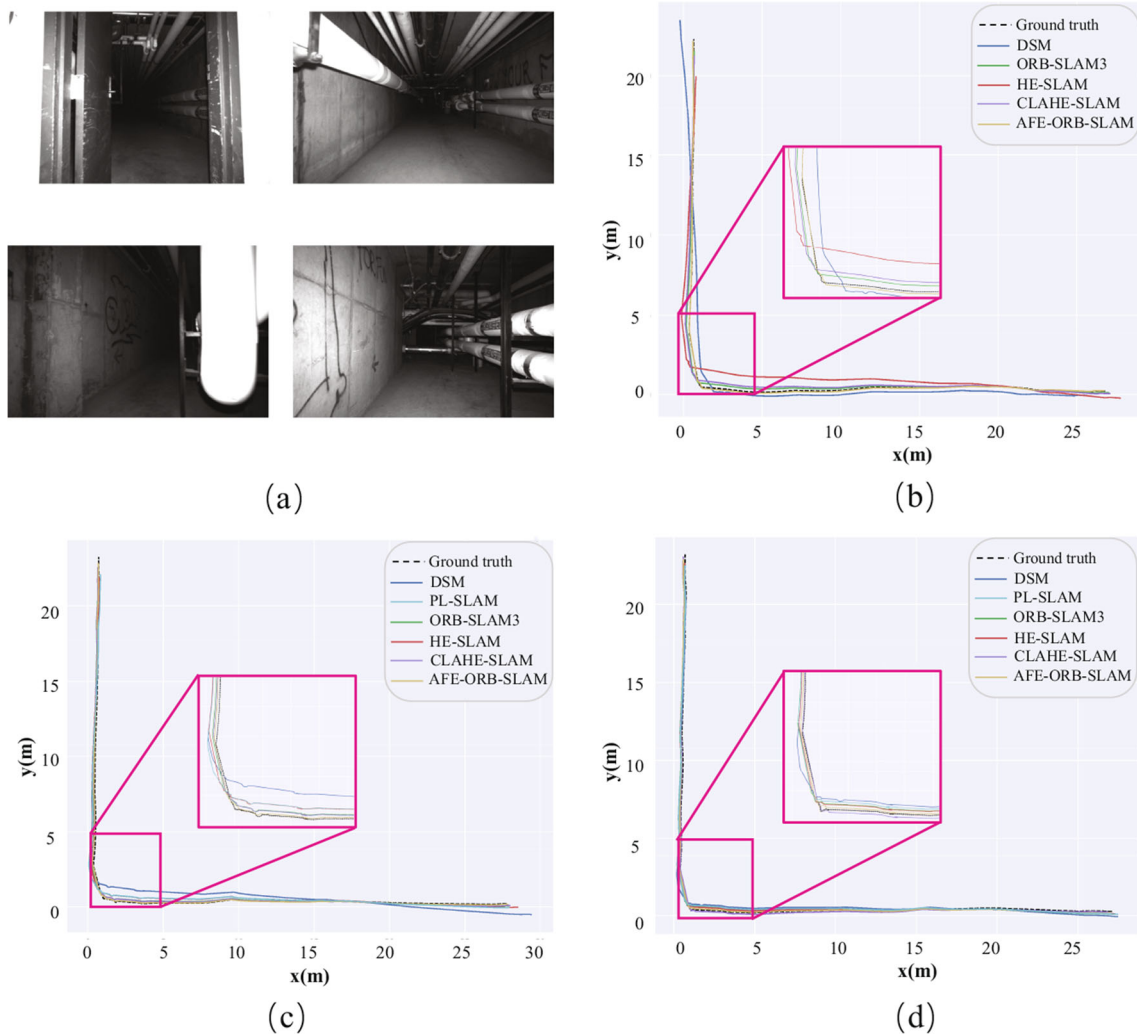


Fig. 7 Trajectory comparison on the TUNNEL sequences. (a) represents some sample images. (b), (c) and (d) indicate the localisation performance with the light of 1350, 4500, or 9000 lumens, respectively

Fig. 8 Comparison of time usage for different VSLAM systems

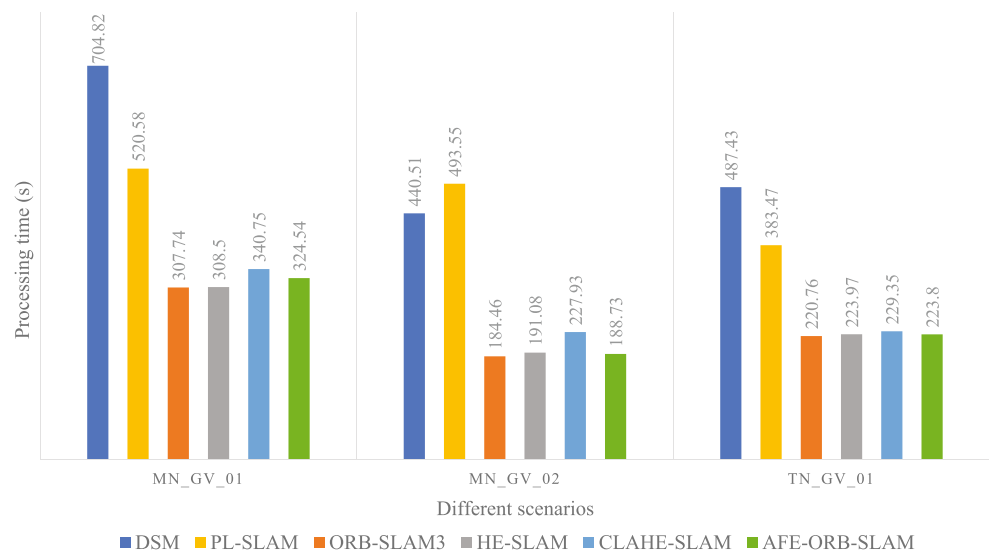


Table 3 Performance comparison on the EuRoC dataset for the RMS ATE (m). The best results are highlighted in a bold font

EuRoC benchmark	DSO [37]	SVO [38]	DSM [12]	ORB-SLAM3	HE- SLAM	CLAHE- SLAM	AFE-ORB- SLAM
MH01	0.046	0.100	0.039	0.017	0.022	0.030	0.018
MH02	0.046	0.120	0.036	0.032	0.047	0.047	0.032
MH03	0.172	0.410	0.055	0.028	0.036	0.037	0.028
MH04	3.810	0.430	0.057	0.088	0.125	0.139	0.087
MH05	0.110	0.300	0.067	0.103	0.045	0.061	0.041
V101	0.089	0.070	0.095	0.033	0.033	0.033	0.033
V102	0.107	0.210	0.059	0.018	0.016	0.016	0.016
V201	0.044	0.110	0.056	0.022	0.023	0.022	0.023
V202	0.132	0.110	0.057	0.037	0.027	0.040	0.017
Average	0.506	0.207	0.058	0.042	0.042	0.047	0.033

conditions, the ORB-SLAM3 achieves similar localisation accuracy to the HE-SLAM and CLAHE-SLAM. Owing to the MH05 sequence containing images collected at night, the localisation performance improvement by applying the image contrast enhancement could be observed. The AFE-ORB-SLAM achieves a higher localisation accuracy in the MH05, V102 and V202 sequences than the ORB-SLAM3.

Moreover, the proposed AFE-ORB-SLAM still achieves the best average accuracy on selected scenarios.

To verify the robustness of different VSLAM systems, the results with 10 times execution are presented in Fig. 9. Different colour squares represent the RMS ATE obtained in each of the 10 executions. The results demonstrate that the precision of the ORB-SLAM3 could be improved by

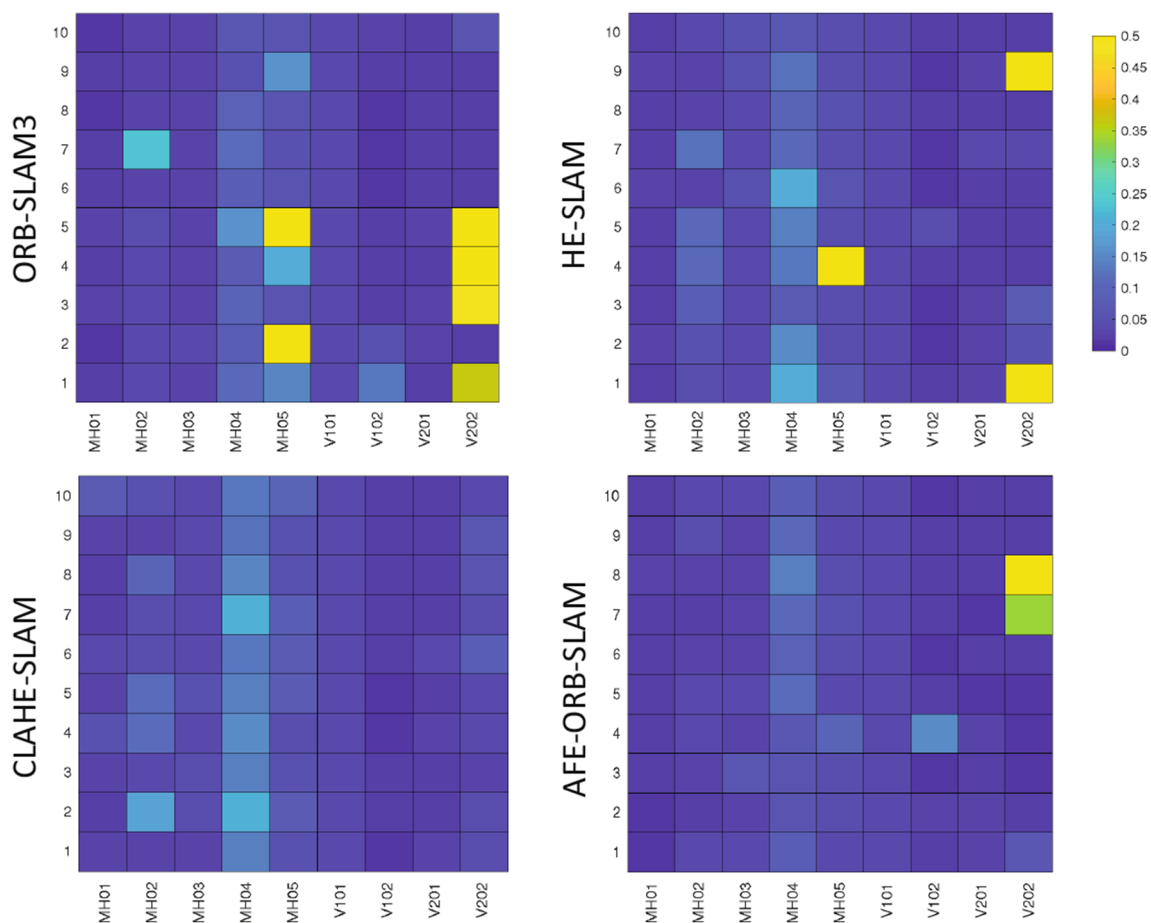


Fig. 9 Precision comparison of different VSLAM methods

adopting the proper image contrast enhancement method, and the CLAHE-SLAM achieves the best results. The proposed AFE-ORB-SLAM outperforms the ORB-SLAM3 in terms of not only the accuracy but also the precision.

6 Conclusion

In this paper, the AFE-ORB-SLAM based on adaptive FAST threshold and image enhancement for challenging lighting environments has been proposed. In contrast to other VSLAM methods, we aimed to extract more reliable feature points from the images captured in challenging lighting conditions. To realise this goal, the image was enhanced from both the contrast adjustment and sharpening adjustment. Moreover, the ORB feature extraction was improved by adopting the adaptive FAST threshold. Experiments on publicly available datasets demonstrated that the AFE-ORB-SLAM was capable of achieving accurate and robust localisation performance in the environments where the images were captured under different illumination conditions, even with less visual information. In addition, the AFE-ORB-SLAM outperformed other state-of-the-art monocular VSLAM methods in most sequences in terms of localisation accuracy.

Acknowledgements Thanks are given to the supports from Strathclyde NZTC robotics research team, especially Dr Gordon Dobie, Dr Charles MacLeod, Mr Mark Robertson and Prof Xiutian Yan.

Author Contributions Idea conception, L.Y.; project supervision, E.Y.; formal analysis, L.Y., E.Y. and B.Y.; original draft writing, L.Y.; review and editing, L.Y., E.Y. and B.Y.

Funding This work is supported in part by the UK Net Zero Technology Centre (NZTC) under the LOCUST research project (2019–2021, Grant No.: AI-P-028). It is also supported partially by the Royal Society under the MOEA/D-PPR research project (2022–2024, Grant No.: IECNSFC211434). Mr. Leijian Yu is funded by the China Scholarship Council and the International Fees Only Studentship from the University of Strathclyde (2018–2021).

Availability of data and materials The datasets used are publicly available in:

- OIVIO dataset: <https://arpg.github.io/oivio/>
- ICL-NUIM dataset with simulated lighting changes: <https://cvg.ethz.ch/research/illumination-change-robust-dslam/>
- EuRoC dataset: <https://projects.asl.ethz.ch/datasets/doku.php?id=knavvisualinertialdatasets>

Code Availability The code used to obtain the experimental results can be made available on reasonable request.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J.: Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* **32**(6), 1309–1332 (2016)
2. Yu, L., Yang, E., Ren, P., Luo, C., Dobie, G., Gu, D., Yan, X.: Inspection robots in oil and gas industry: a review of current solutions and future trends. In: 2019 25th International Conference on Automation and Computing (ICAC), pp. 1–6. IEEE (2019)
3. Fang, B., Mei, G., Yuan, X., Wang, L., Wang, Z., Wang, J.: Visual slam for robot navigation in healthcare facility. *Pattern Recogn.* **113**, 107822 (2021)
4. Fallon, M.F., Folkesson, J., McClelland, H., Leonard, J.J.: Relocating underwater features autonomously using sonar-based slam. *IEEE J. Ocean. Eng.* **38**(3), 500–513 (2013)
5. Hess, W., Kohler, D., Rapp, H., Andor, D.: Real-time loop closure in 2d lidar slam. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 1271–1278. IEEE (2016)
6. Forster, C., Pizzoli, M., Scaramuzza, D.: Svo: Fast semi-direct monocular visual odometry. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 15–22. IEEE (2014)
7. Fang, B., Zhan, Z.: A visual slam method based on point-line fusion in weak-matching scene. *International Journal of Advanced Robotic Systems* **17**(2), 1729881420904193 (2020)
8. Nikolic, J., Burri, M., Rehder, J., Leutenegger, S., Huerzeler, C., Siegwart, R.: A uav system for inspection of industrial facilities. In: 2013 IEEE Aerospace Conference, pp. 1–8. IEEE (2013)
9. Brasch, N., Bozic, A., Lallemand, J., Tombari, F.: Semantic monocular slam for highly dynamic environments. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 393–400. IEEE (2018)
10. Servières, M., Renaudin, V., Dupuis, A., Antigny, N.: Visual and visual-inertial slam: State of the art, classification, and experimental benchmarking. *Journal of Sensors*, **2021** (2021)
11. Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M., Tardós, J.D.: Or-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE Transactions on Robotics* (2021)
12. Zubizarreta, J., Aguinaga, I., Montiel, J.M.M.: Direct sparse mapping. *IEEE Trans. Robot.* **36** (4), 1363–1370 (2020)
13. Fang, Y., Shan, G., Wang, T., Li, X., Liu, W., Snoussi, H.: H-slam: A stereo slam system based on histogram equalization and orb features. In: 2018 Chinese Automation Congress (CAC), pp. 4272–4276. IEEE (2018)
14. Mur-Artal, R., Tardós, J.D.: Orb-slam2: an open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics* **33**(5), 1255–1262 (2017)

15. Yang, W., Zhai, X.: Contrast limited adaptive histogram equalization for an advanced stereo visual slam system. In: 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 131–134. IEEE (2019)
16. Gu, Q., Liu, P., Zhou, J., Peng, X., Zhang, Y.: Drms: Dim-light robust monocular simultaneous localization and mapping. In: 2021 International Conference on Computer, Control and Robotics (ICCCR), pp. 267–271. IEEE (2021)
17. Miao, Y., Song, D., Shi, W., Yang, H., Li, Y., Jiang, Z., He, W., Gu, W.: Application of the clahe algorithm based on optimized bilinear interpolation in near infrared vein image enhancement. In: Proceedings of the 2nd International Conference on Computer Science and Application Engineering, pp. 1–6 (2018)
18. Rahman, S., Rahman, M.M., Abdullah-Al-Wadud, M., Al-Quaderi, G.D., Shoyaib, M.: An adaptive gamma correction for image enhancement. EURASIP Journal on Image and Video Processing **2016**(1), 1–13 (2016)
19. Luo, K., Lin, M., Wang, P., Zhou, S., Yin, D., Zhang, H.: Improved orb-slam2 algorithm based on information entropy and image sharpening adjustment. Math. Probl. Eng. **2020** (2020)
20. Gomez-Ojeda, R., Zhang, Z., Gonzalez-Jimenez, J., Scaramuzza, D.: Learning-based image enhancement for visual odometry in challenging hdr environments. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 805–811. IEEE (2018)
21. Pumarola, A., Vakhitov, A., Agudo, A., Sanfeliu, A., Moreno-Noguer, F.: Pl-slam: Real-time monocular visual slam with points and lines. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 4503–4508. IEEE (2017)
22. Huang, J., Liu, S.: Robust simultaneous localization and mapping in low-light environment. Computer Animation and Virtual Worlds **30**(3-4), 1895 (2019)
23. Park, S., Schöps, T., Pollefeys, M.: Illumination change robustness in direct visual slam. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 4523–4530. IEEE (2017)
24. Sun, P., Lau, H.Y.: Rgb-channel-based illumination robust slam method. Journal of Automation and Control Engineering, **7**(2) (2019)
25. Engel, J., Stückler, J., Cremers, D.: Large-scale direct slam with stereo cameras. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1935–1942. IEEE (2015)
26. Maragatham, G., Roomi, S.M.M.: A review of image contrast enhancement methods and techniques. Res. J. Appl. Sci. Eng. Technol. **9**(5), 309–326 (2015)
27. Charles, P., et al: Digital Video and Hdtv Algorithms and Interfaces, vol. 260, p. 630. Morgan Kaufmann Publishers, San Francisco (2003)
28. Huang, S.-C., Cheng, F.-C., Chiu, Y.-S.: Efficient contrast enhancement using adaptive gamma correction with weighting distribution. IEEE Transactions on Image Processing **22**(3), 1032–1041 (2012)
29. Cao, G., Huang, L., Tian, H., Huang, X., Wang, Y., Zhi, R.: Contrast enhancement of brightness-distorted images by improved adaptive gamma correction. Computers & Electrical Engineering **66**, 569–582 (2018)
30. Bhandari, A., Kumar, A., Padhy, P.: Enhancement of low contrast satellite images using discrete cosine transform and singular value decomposition. World Academy of Science. Eng. Technol. **79**, 35–41 (2011)
31. Kim, M., Chung, M.G.: Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement. IEEE Trans. Consum. Electron. **54**(3), 1389–1397 (2008)
32. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: Orb-slam: a versatile and accurate monocular slam system. IEEE Transactions on Robotics **31**(5), 1147–1163 (2015)
33. Kasper, M., McGuire, S., Heckman, C.: A benchmark for visual-inertial odometry systems employing onboard illumination. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5256–5263. IEEE (2019)
34. Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The euroc micro aerial vehicle datasets. The International Journal of Robotics Research **35**(10), 1157–1163 (2016)
35. Grisetti, G., Kümmerle, R., Strasdat, H., Konolige, K.: G2o: a general framework for (hyper) graph optimization. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, pp. 9–13 (2011)
36. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of rgb-d slam systems. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 573–580. IEEE (2012)
37. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(3), 611–625 (2017)
38. Forster, C., Zhang, Z., Gassner, M., Werlberger, M., Scaramuzza, D.: Svo: Semidirect visual odometry for monocular and multicamera systems. IEEE Trans. Robot. **33**(2), 249–265 (2016)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Leijian Yu received the B.Eng. degree in electrical information engineering and the M.Sc degree in information and communication engineering from China University of Petroleum (East China), Qingdao, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree in the Department of Design, Manufacturing and Engineering Management (DMEM) at the University of Strathclyde, Glasgow, UK, since 2018. His current research interests include machine learning with applications on unmanned aerial vehicles (UAV), UAV vision-based autonomous navigation, and image contrast enhancement.

Erfu Yang received his Ph.D. degree in Robotics from the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK, in 2008. He is currently a Senior Lecturer in the Department of Design, Manufacturing and Engineering Management (DMEM), University of Strathclyde, Glasgow, UK. His main research interests include robotics, autonomous systems, mechatronics, manufacturing automation, signal and image processing, computer vision and applications of machine learning and artificial intelligence, etc. He has over 160 publications in these areas, including more than 80 journal papers and 10 book chapters. Dr. Yang has been awarded over 15 research grants as PI (principal investigator) or CI (co-investigator). He is the Fellow of the UK Higher Education Academy, member of IEEE and IEEE Society of Robotics and Automation, and the IET SCOTLAND Manufacturing Technical Network. He is also an associate editor for the Cognitive Computation journal published by Springer.

Beiya Yang received the B.Eng. degree in Electronic Information Engineering from Northwestern Polytechnical University, Xi'an, China, in 2013 and the M.Sc degree in Information and Communication Engineering from National University of Defense Technology, Changsha, China, in 2015. He is currently pursuing the Ph.D. degree in the Department of Design, Manufacturing and Engineering Management (DMEM) at University of Strathclyde, Glasgow, UK, since 2019. His current research interests include indoor localisation technology, unmanned aerial vehicles (UAV) localisation and wireless sensor networks.