# HLE-SLAM: SLAM for overexposed construction environment

**Xinyu Chen[1] and Yantao Yu[1]**

[1]Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology

xchengl@connect.ust.hk, ceyantao@ust.hk

**Abstract -**
**One of the main challenges in the visual simultaneous localization and mapping (V-SLAM) of construction robots is robustness to overexposure conditions. The main difficulties arise from sensor exposure limitations that cause images to lose information. In addition, construction robots can hardly track enough points in overexposure conditions due to the assumption of constant brightness in SLAM. We propose a High and Low Exposure SLAM (HLE-SLAM) system to recover missing information in overexposed frames. Our method uses frame exposure fusion to generate globally well-exposed frames. It uses exposure, contrast and information entropy as indicators to select the best part of brightness and information in high and low exposed frames. We adopt the Shi-Tomasi and Kanade-Lucas-Tomasi (KLT) sparse optical flow algorithms to improve the ability to detect and track feature points in the overexposed environment. Experimental results on data sets and real environments show that HLE-SLAM can effectively solve the overexposure problem.**

**Keywords -**

**Overexposure; Frames fusion; Simultaneous Localization and Mapping; Construction robots**

## 1 Introduction

The construction industry in industrialized countries has faced severe labor shortages in recent years. Improving the automation of construction projects can solve these challenges[1]. As a result, the development of autonomous robotics and vehicles is increasing for a variety of construction applications. Nowadays, SLAM systems are mainly divided into LiDAR SLAM and V-SLAM systems. The V-SLAM system is more suitable for low-cost construction robot platforms due to its low price, low computational complexity, and rich visual information. Although these V-SLAM algorithms have achieved impressive results in a controlled laboratory environment, the robustness of V-SLAM in real-world construction scenarios is a major challenge[2]. In an indoor scene with direct sunlight or when moving from a dimly lit scene to a highly lit scene, overexposure problems occur due to the dynamic limitations of the visual camera.

The dynamic range of a camera is usually determined by fixed parameters or proprietary algorithm. The built-in control algorithms are generally suitable for situations where lighting conditions are constant or change only slowly. For scenes with changing lighting conditions, the automatic exposure camera will result in poorer exposure images. The algorithms of V-SLAM are set for certain ideal environments, such as constant light intensity and rich textures. In other words, construction robots lose their localization and recognition capabilities in overexposed scenes. Complex lighting scenario on construction site is a major challenge for robot path planning. Nowadays, there are three solutions to reduce the impact of overexposed images on the V-SLAM. The first method relies on the post-processing of overexposed images to reduce variations in light intensity. The second approach uses the invariability of luminance changes in object detection for feature detection and matching. These methods can improve the performance of visual navigation to some extent. However, they cannot compensate for the loss of information caused by overexposed images, and these methods involve additional computational overhead. The third method is to adjust the image exposure parameters using an automatic exposure algorithm[3]. However, adjusting the camera response parameters is still a complex problem given different scenes. Improper exposure parameter settings can also result in low-light images.

In this paper, we propose a High and Low Exposure SLAM (HLE-SLAM) system to recover the overexposed images, which helps construction robots in localization and mapping. We use two cameras with different exposure settings to record videos simultaneously and repair the overexposed by exposure fusion. We calculate the exposure, contrast and information entropy values of different exposure images to generate the fusion weight. Then, we use fusion weight to retain rich frame information in different exposure pictures. After image fusion, we use the Shi-Tomasi[4] method for corner detection and uniformly assign feature points to track in ORB-SLAM2[5]. Finally, we test the HLE-SLAM system in EuRoc[6] dataset and construction environment. Experiments show that our proposed method has better location performance in both simulated overexposed environment datasets and real indoor overexposed construction environments.

## 2 Methodology

The flow of the proposed method is shown in Figure1. The HLE-SLAM system mainly consists of three parts: dual camera frame acquisition and alignment, frame ex-

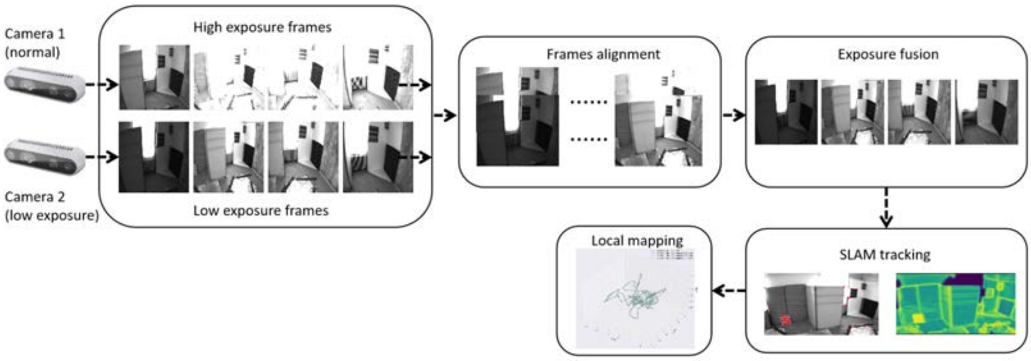


Figure 1. Architecture of the proposed HLE-SLAM.

posure fusion, feature extraction, tracking and local mapping.

**A. Dual camera frame acquisition and alignment**

We use two Intel RealSense Depth D435 cameras for synchronous data acquisition, and the acquisition frequency is both 20Hz. One uses built-in program parameters, and the other controls the overall brightness of the image by limiting the maximum exposure time of the camera. Since the two cameras are arranged vertically, there is a difference of several pixel values in the vertical direction. We achieve the overall alignment of the image by removing the non-overlapping pixel values in the top and bottom frames.

**B. Frame exposure fusion**

Frame exposure fusion produces globally well-exposed images by preserving the best-displayed parts of the high and low-exposed sequences. The frame contains colourless areas due to underexposure and overexposure. These areas should be weighted less, while areas containing good light intensity and picture information should be retained. We select the best brightness part of the two pictures using three index parameters: exposure, contrast, and information entropy. The final image is determined using an average weighting method.

**Exposure:** By calculating the intensity of the gray image, the exposure of the pixel value can be judged. We want to keep frame intensities that are not near zero (underexposed) or one (overexposed). We use the following Gaussian curve:

$$\exp\left(-\frac{(i-0.5)^2}{2\sigma^2}\right) \tag{1}$$

to set the weight of each pixel according to how close its intensity i is to 0.5. The $\sigma$ equals 0.2 in our implementation. In this way, we can get the E representing the exposure value.

**Contrast:** We apply the Laplace filter to each grayscale image and take the absolute value of the filter response. The Laplace filter is defined as:

$$\nabla f(x,y) = \frac{d^2 f}{dx^2} + \frac{d^2 f}{dy^2}, \tag{2}$$

where f(x,y) is a two-dimensional image. This results in a simple contrast indicator C. It tends to assign high weights to important elements like edges and textures.

**Information entropy value:** The one-dimensional entropy of the image represents the information contained in the aggregated features of gray distribution in an image. $P_i$ represents the proportion of pixels whose gray value is i in the image, and then the unary gray entropy is defined as:

$$H = \sum_{i=0}^{255} P_i \cdot \log P_i, \tag{3}$$

Where H is the unary gray information entropy, and $P_i$ is the proportion of grey value i. When the image is overexposed/dark (white or black), there is only one grey value. Then the entropy is the minimum, H = 0. The rich information of images is conducive to feature extraction of V-SLAM. We divide the whole image into 16x16 blocks and calculate the information entropy of each block. Then, we assign the calculated information entropy value of each block to all pixels within that block, representing their information entropy value. On this basis, we generate the corresponding weight value and save the larger information entropy part. By using multiplication, we combine the information from the different index parameters into a scalar weight map. The linear combination weighting method is used to control the influence of each measure part, and the weight of each picture is generated by the following formula:

$$W_{ij,k} = \omega_c(C_{ij,k}) \cdot \omega_e(E_{ij,k}) \cdot \omega_h(H_{ij,k}), \tag{4}$$

where C, E and H are contrast, exposure and information entropy. The $\omega_c$, $\omega_e$, $\omega_h$ represent the corresponding weight. The subscript i and j, refers to pixel (i, j) in k(high/low) exposure images. We use equally weighted quality measures ($\omega_c$=$\omega_e$=$\omega_h$=1). In extreme lighting situations, such as a large area facing direct sunlight, we adjust the parameters based on field tests.

The fusion image R can then be obtained by a weighted blending of the input high/low exposure images:

$$R_{ij} = \sum_{high-exposure}^{low-exposure} W_{ij,k} \cdot I_{ij,k} , \tag{5}$$

where W means weight map, I means input images.

**C. Feature extraction, tracking and local mapping**

In this paper, ORB -SLAM2 is selected as a navigation algorithm for construction robots. The ORB -SLAM2 system is based on feature point extraction and matching, but in an overexposed environment, feature points are hardly detected and matched. Therefore, position and pose tracking will fail in a construction site scene with low texture and overexposed scenes. We adopt the Shi-Tomasi and Kanade-Lucas-Tomasi (KLT) sparse optical flow algorithms to solve these problems. The average distribution of feature points in an overexposed environment reduces feature point matching requirements. New points are added if there are not enough points left after the uniform procedure to reach the required 200. The next frame is considered a new keyframe if the difference between the parallaxes of the two frames is significant.


Figure 2. The results of image fusion in MAV datasets.

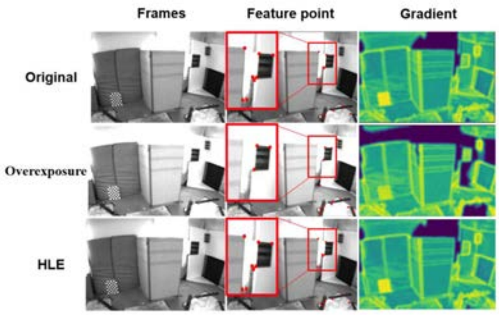


Figure 3. The gradient and feature point results of original, overexposure and HLE method.

## 3 Experiments and results

Our work focuses on the visual navigation of construction robots to work reliably in overexposed environments. Therefore, we experiment with the EuRoc MAV dataset and real construction scenes.

**A.Public Dataset**

In EuRoc MAV datasets, there are some overexposed scenes. These indoor lighting problems are similar to overexposed problems in construction scenes, so they are used to test algorithms in overexposed environments. We artificially increased the exposure value in the MAV to create more obvious overexposure frames. In the dataset experiment, we use the original dataset as the low-light video stream and enhance it to simulate the overexposed video stream.

Therefore, our HLE system use the original and overexposed datasets as input. As shown in Figure 2, we use the HLE method to recover the overexposed frames in MAV datasets. To verify the availability of our method in overexposure, we compare the HLE with original and overexposure frames in SLAM feature point extraction and gradient calculation, as shown in Figure 3. Our method enhances the gradient information of frames and enables the algorithm to detect more feature points.

Table 1. Performance comparison in the EuRoC MH datasets (Root mean square error in m).

| Sequence | Original | Overexpose | HLE |
|---|---|---|---|
| V101 | 0.055 | 0.143 | 0.058 |
| V102 | 0.064 | 0.155 | 0.061 |
| V103 | 0.096 | 0.151 | 0.097 |
| V201 | 0.046 | 0.121 | 0.074 |
| V202 | 0.057 | 0.118 | 0.059 |

Table 1 compares the performance of the HLE in monocular sensors with original and overexposed; all tests are based on the KLT ORB-SLAM2. As shown in the table, raising the exposure value makes the overexposed frames show a more significant estimated error. Our method can reduce the interference caused by overexposure and restore image information to reduce errors. In the V102, the HLE shows better results than the original. Because some low-light images exist in the V102, it is difficult to locate them in the dark environment. The overexposed frames are fused with low-light frames to enhance the low-light frames in the original video. The HLE method outputs frames with stable and continuous brightness and more information, which makes the robot position with smaller error results.


Figure 4. High and low exposure frames captured by the Intel Realsense 435 camera and processed frame in the HLE method.

**B. Real-World Experiment**

We use the Intel Realsense 435 camera to conduct real experiments in overexposed indoor construction scenes. In our real-world experiment, we capture overexposed images in a direct sunlight environment using an auto-exposed camera, while we use a camera with a shorter exposure time to capture the low-exposure images. Low-exposure frames are difficult to initialize and map. So we only compare the HLE method with the original overexposure frames in the real-world experiment. In Figure 4, we compare the overexposed image with the HLE image. The camera produces some overexposed frames because of the intense light. In frames, overexposure is shown in the balcony doors and windows, and much information is missing. The HLE fuses overexposure frames with low-exposure frames to recover missing information. In order to further verify the performance of our algorithm, we conduct location estimation and mapping experiments. As shown in Figure5, in the left part of the navigation, the original frames cannot provide enough feature points for tracking and matching, so we cannot locate them effectively in this part. In addition, there is a large deviation in the right corner part.

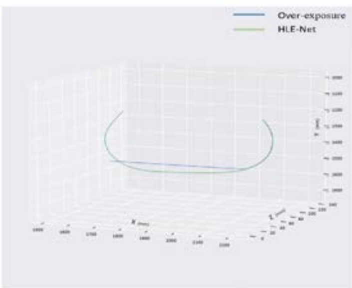


Figure 5. 3D pose graph. The green line is the HLE-SLAM. The blue line is the original overexposure frame.

# 4 Conclusions

In this paper, an exposure fusion-based construction robot SLAM system is proposed for overexposure problems. The exposure fusion method based on contrast, exposure and information entropy can recover missing information and help the construction robot track the feature points in the overexposed environment. We validate our HLE-SLAM in public datasets and real construction scenes. The results show that our method can effectively solve the problem of location errors caused by overexposed light. Our method can run in real-time on mobile computers and is suitable for various overexposed construction scenarios.

# 5 Acknowledgement

This work was financially supported by the HKUST-BDR Joint Research Institute Fund, No. OKT23EG01.

# References

[1] Sungjin Kim, Soowon Chang, and Daniel Castro-Lacouture. Dynamic modeling for analyzing impacts of skilled labor shortage on construction project management. *Journal of Management in Engineering*, 36 (1):04019035, 2020.

[2] Xinyu Chen and Yantao Yu. Image illumination enhancement for construction worker pose estimation in low-light conditions. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 147–162, 2022.

[3] Zichao Zhang, Christian Forster, and Davide Scaramuzza. Active exposure control for robust visual odometry in hdr environments. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 3894–3901. IEEE, 2017.

[4] Jianbo Shi et al. Tomasi. good features to track. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600. sn, 1994.

[5] Raul Mur-Artal and Juan D Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE transactions on robotics*, 33(5): 1255–1262, 2017.

[6] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10):1157–1163, 2016.