# LI-SLAM: Fusing LiDAR and Infrared Camera for Simultaneous Localization and Mapping

Baoding Zhou [1,2,3], Doudou Xie [1], Shoubin Chen [3], Chunyu Li [1] and Haoquan MO [1]

[1] *College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China*
[2] *Institute of Urban Smart Transportation & Safety Maintenance, Shenzhen University, Shenzhen 518060, China*
[3] *Guangdong Key Laboratory of Urban Informatics, Shenzhen University, Shenzhen 518060, China*

**Abstract**

As an active sensor, light detection and ranging (LiDAR) has an important position in Simultaneous Localization and Mapping (SLAM). In particular, it is not affected by lighting conditions, so that it can work well at night. Infrared image is also less affected by illumination, so it can be well fused with LiDAR. Generally, the LiDAR SLAM system consists of a front-end odometer and a back-end optimization module. The back-end with loop closure detection plays an important role in systematically improving the positioning and mapping accuracy of LiDAR SLAM. However, LiDAR works at a single wavelength (such as 905nm), and few features are extracted, which limits the performance of loop closure detection and graphics optimization based on point cloud matching. In order to improve the performance of LiDAR SLAM, a SLAM back-end which is not affected by illumination conditions is proposed in this paper. This method combines infrared image and the geometric features of LiDAR as loop detection. Firstly, the bag of word (BOW) model, describing the visual similarities, was constructed to assist in the loop closure detection. Then, through the time interval of infrared image and the spatial distance of point cloud, the loop closure detection is verified, and accomplish graph optimization. We conducted experiments in different scenes at night to evaluate our method. The results show that the addition of infrared images effectively helps the loop closure detection and improves the performance of LiDAR SLAM.

**Keywords**

LiDAR, Infrared Camera, loop closure detection, graph optimization

## 1. Introduction

With the coming of the "New Technology Revolution", human beings have higher and higher requirements for automation, intelligence and even intelligence, and the demand for positioning services is also increasing day by day. In outdoor environments, the global navigation satellite system (GNSS) is a very mature localization technology that can provide accurate outdoor localization services [1]. However, it is greatly affected by satellite signals and is not suitable for working in weak signal areas, especially in indoor scenes with satellite e signal occlusion [2]. For indoor environment, SLAM has the characteristics of high positioning accuracy and no need to lay infrastructure in advance. With these characteristics, SLAM technology is widely used in the field of positioning. Traditionally, SLAM technology is generally divided into visual SLAM and LiDAR SLAM according to different sensors [3][4][5][6]. With the continuous research on SLAM technology by scholars, there are more and more SLAM schemes for the fusion of the two sensors.

Graeter proposed the LiDAR-monocular visual odometry [7], The depth information extracted from LiDAR point cloud is used for camera feature point tracking, which makes up for the defect of monocular vision scale. However, the core framework of this method is still based on visual SLAM,

IPIN 2022 WiP Proceedings, September 05–Septe mber 07, Beijing, China
EMAIL: bdzhou@szu.edu.cn (B.Zhou)
ORCID:0000-0003-1607-2626 (B.Zhou)


and the role of LiDAR is only a supplement, which does not give full play to the advantages of LiDAR. Similar schemes include binocular vision inertial navigation LiDAR SLAM proposed by Shao and Shin's direct vision slam using sparse depth [8][9]. Zhang proposed the slam method of visual LiDAR fusion [10]. The visual odometer provides the initial value for laser point cloud matching, its accuracy and robustness are further improved than LOAM [11]. But this method lacks back-end loop detection and global graph optimization. In order to solve this problem, Chen proposed a scheme to provide loop detection with LiDAR odometer as the front-end and image as the back-end, and achieved good results [12].

Most of the visual information used in the above scheme is RGB image. In the case of poor lighting conditions such as at night, these schemes often have poor effects or even useless. Infrared image is less affected by light and can be used normally at night. It can be well fused with LiDAR. At present, scholars have few SLAM schemes for the fusion of infrared image and LiDAR. Pierre proposed a SLAM method, which is able to solve various difficulties (lack of GPS signal, lighting conditions, smoke) [13]. In this method, the infrared camera and LiDAR sensor are fused when the illumination is insufficient. However, infrared images are different from RGB images. Directly using BOW to perform loop detection will lead to inaccurate loop candidates [14]. Shin propose a thermal-infrared SLAM system enhanced by sparse depth measurements from LiDAR [14]. Compared with the SLAM method of monocular camera and LiDAR fusion, this method has good robustness in day and night. But the core framework of this method is still based on visual SLAM, which does not give full play to the advantages of LiDAR. Kataoka proposed a SLAM method combining LiDAR intensity and near-infrared according to the near-infrared information of common puddles in the damaged nuclear power plant and the characteristics of LiDAR reflection intensity [15]. This method may be limited in other scenarios.

This paper constructs a LiDAR / infrared SLAM based on loop closure detection and global graph optimization to improve the accuracy of localization trajectories and the consistency of point cloud images in the night environment. In LiDAR / infrared SLAM system, a loop closure detection method based on infrared image information and point cloud re-matching is realized. We conducted experiments in different scenes at night to evaluate our method. The results show that the addition of infrared image effectively helps the loop closure detection and improves the SLAM performance of LiDAR in night environment.

## 2. MODEL FOR LOCALISATION IN NIGHT

### 2.1. System Overview

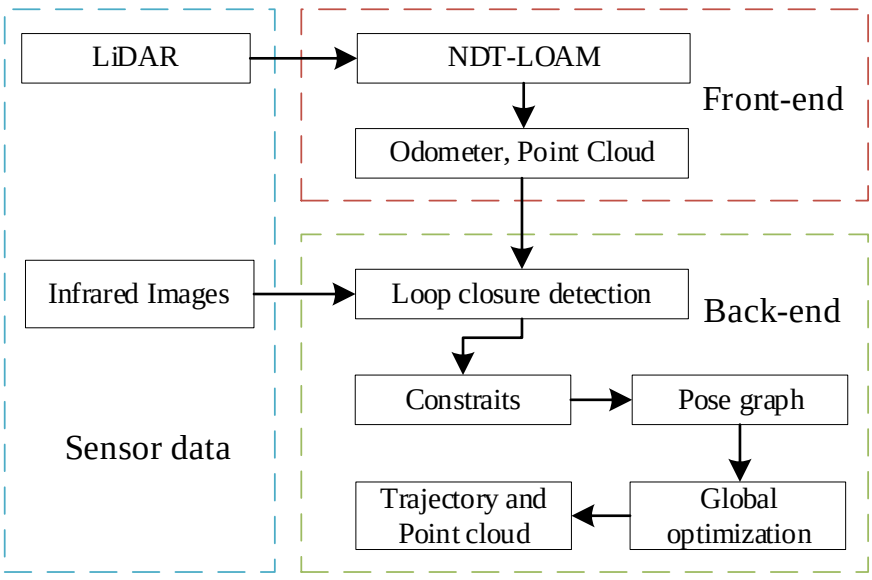



**Figure 1**: The whole system structure

In this section, we will introduce the system and structure of our method in detail. As shown in Figure 1, our method consists of sensor data, algorithm front-end and algorithm back-end.

The sensor data of our algorithm includes LiDAR point cloud data and infrared image data. LiDAR point cloud data is collected by 16-line LiDAR, and infrared image data is collected by Azure Kinect DK camera.

The front-end input of our algorithm is LiDAR sensor data. Through NDT-LOAM algorithm [16] to process LiDAR data, we finally get the front-end odometer. The back-end input of our algorithm is infrared image data and odometer data obtained from the front-end. Firstly, Bow is used to evaluate the similarity of infrared images. Then judge the similarity results from the time interval of the infrared image, and then judge the spatial distance from the data obtained from the front end of the algorithm to realize loop detection. Then, the edge constraints between key frames are calculated, and the attitude graph of global key frames is constructed. Finally, the graph optimization theory is used to reduce the global cumulative error, improve the global Trajectory Accuracy and map consistency, and obtain the final global motion trajectory and point cloud map.

## 2.2. Infrared Loop Closure Detection

The technical route of loop closure detection is mainly based on geometry and appearance. The loop closure detection method based on geometry refers to verifying whether there is a loop closure when the platform returns to a previous position according to the front-end odometer or other known information [17]. The loop closure detection method based on appearance is independent of the state estimation of the front-end and back-end. It uses the similarity detection of two images or point clouds to get rid of the influence of cumulative error [18][19][20].

Infrared image and RGB image have certain similarity, they can reflect the structural information of the scene, so the principle of loop closure detection using the two devices is the same. However, compared with RGB images, infrared images have less information and poor quality. The accuracy and recall of loop matching of infrared image are worse than that of RGB image. Therefore, improving the accuracy and recall of infrared loop matching is an important problem that need to be solved urgently before we optimize the back-end. In order to improve the accuracy and recall of infrared loop matching, the loop closure detection method adopted in this paper combines geometric detection method and appearance detection method. The algorithm flow chart is shown in Figure 2.

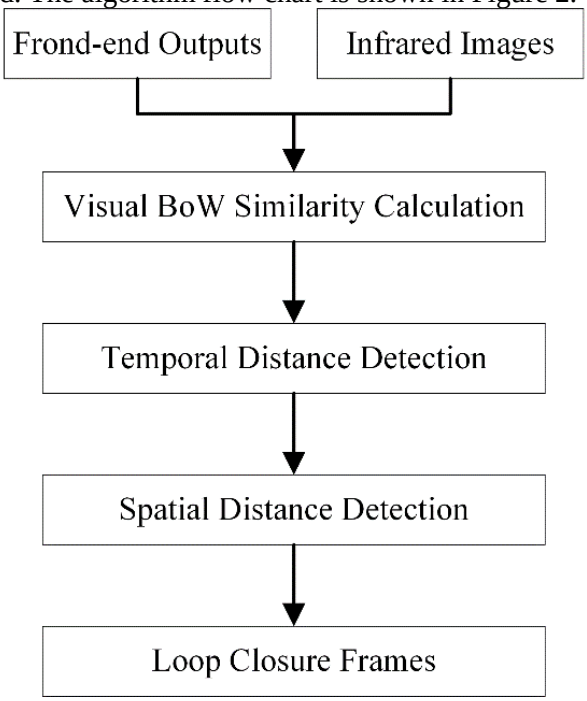



**Figure 2**: Flow chart of the loop closure detection method

Firstly, we use Bow model to test the similarity of infrared images and get the similarity score results of each image. The program we use here is the open source library DBOW3 (https://github.com/rmsalinas/DBow3, accessed on 7 June 2021) to assist the implementation. However, the characteristics of infrared image and RGB image are different. Direct loop closure detection will lead to inaccurate loop closure candidates[14]. In order to improve the computational efficiency and loopback accuracy of the algorithm, we eliminate the results of infrared image similarity less than 0.05. Then the results are verified in temporal distance and spatial distance. The first is to screen the loop closure results from the temporal distance. Details are shown in points 1 to 3. The spatial filtering is based on the odometer and LIDAR point cloud output from the front-end of the algorithm. Details are shown in points 4 to 7.

1) Because there is usually a standing time at the beginning and end of data acquisition, we eliminate the start and end $T_1$ second data;
2) The time length threshold $T_2$ of the loop, and the elapsed time of the loop is greater than this threshold;
3) The interval of each loop is $T_3$. In each time interval $T_3$, select the point with the highest similarity as the suspected loop closure frames.
4) The threshold L1 of the search area, and only the loop closure frames less than the threshold can be the potential loop closure frames;
5) The space interval threshold L2 of the loop, and the interval between two loops should be greater than this threshold;
6) The loop length threshold is L3, and the loop distance should be greater than this threshold;
7) After the above 6 conditions are met, the point cloud re-matching verification is carried out. Match the current frame with the suspected loop closure frame one by one in order. If the mean square of the distance from the source point cloud to the target point cloud is lower than the threshold L4, the frame is the final loop closure frame of the current frame.

When detecting the loop closure, there are usually four conditions: true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs). An FP indicates that the truth is not a loop but the algorithm determines that it is a loop. A TN indicates that the truth and algorithm detection are not loop closures. An FN indicates that the truth is a loop, but the algorithm determines that it is not a loop.

In our results, we want TP and TN to appear as much as possible, and FP and FN to appear as little as possible, or not at all. According to the frequency of the four possible results, we can calculate the accuracy rate and recall rate as two evaluation indexes:

$$Accuracy = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

## 2.3. Global Graph Optimization

With the accumulation of time, the trajectory of the robot will be longer and longer, and the scale of the map will continue to grow. Since the scanning space of LiDAR is always limited, the scale of point cloud or road sign can not grow indefinitely with the map, and the constraint relationship between the current frame and earlier historical data may no longer exist. In addition, there are errors in direct matching and feature point optimization, the cumulative error obtained at the front end of the algorithm will be larger and larger, and the inconsistency of the global map will be more and more obvious.

In order to improve the pose accuracy of odometer and ensure the quality of global point cloud map, we save the trajectory obtained from the front-end of the algorithm and construct a back-end global map optimization with only trajectory to reduce the cumulative error. The global pose map takes the pose of odometer as the node. The relative motion estimator between the two pose nodes obtained by

point cloud matching is used as the constraint edge. Finally, the nonlinear least squares adjustment method is used to obtain the results with higher accuracy and stronger consistency.

## 3. EXPERIMENTS AND RESULTS

To evaluate our method and compare with other approaches, we designed a series of experiments to compare the performance of NDT-LOAM and LI-SLAM in positioning and mapping.

### 3.1. Dataset Description

We use mobile robots for data acquisition. The mobile machine is loaded with a variety of sensors, including Azure Kinect DK camera and RS-LiDAR-16, as shown in Figure 3. Through these sensors, we can obtain data such as LiDAR and infrared images.


**Figure 3**: Mobile robot equipment

In the experiments, we controlled the robot walking via a robot remote control handle. When the robot ran as planned, we recorded all sensor data using ROS commands. We compare the location and mapping results obtained by our method (LI-SLAM) with those obtained by other methods. In order to make the experiment more universal in the night, we conducted experiments in the underground parking lot of the Zhi Teng Building at Shenzhen University, indoor and some outdoor scenes on the first floor of Zhi Teng building and the first floor of Zhi Yuan building of Shenzhen University.

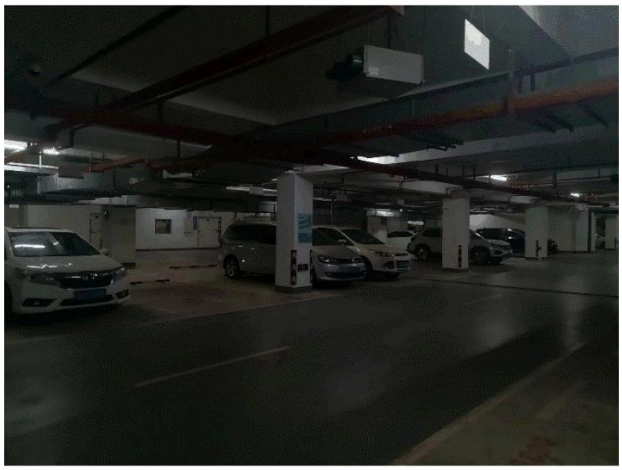


**Figure 4**: Experimental scene 1

As shown in Figure 4. Experimental scene 1 was an underground parking lot of the Zhi Teng Building at Shenzhen University, covering a total distance of 203.245 m. The feature points of the scene are obvious, and there are no dramatic changes in the scene. Generally speaking, LiDAR has high accuracy in this scene.

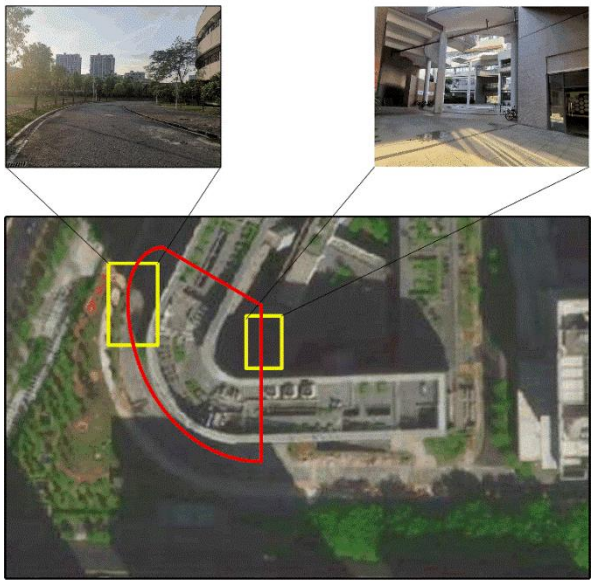


**Figure 5**: Experimental scene 2. For the convenience of display, here we put the real picture taken during the day.

As shown in Figure 5. Experimental scene 2 was an indoor and some outdoor scenes on the first floor of Zhi Teng building, covering a total distance of 422.863 m. The red line indicates the approximate movement track of the robot, and the intersection of the red line is the starting point and end point of the robot movement. The upper left and right corners of the picture are the outdoor real map and indoor real map of the scene respectively. It can be seen that the indoor geometric characteristics are obvious and there are many characteristics. Outdoors, the scene is relatively single, there are few feature points, and the scene changes greatly when switching between indoor and outdoor. Therefore, generally speaking, the accuracy of LiDAR positioning and mapping is not particularly high in this scene.

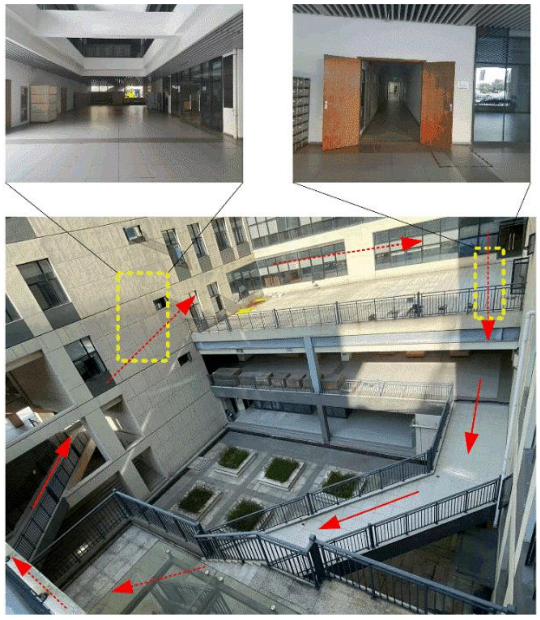


**Figure 6**: Experimental scene3. For the convenience of display, here we put the real picture taken during the day

As shown in Figure 6. Experimental scene 3 was the first floor of Zhi Yuan building of Shenzhen University. We manipulated the robot to walk twice on the first floor of Zhi Yuan building, with a total

distance of 323.552m. In Figure 6, the red arrow indicates the approximate walking route of the robot. In the process of walking, the scenes that the robot passes through include railing scene, hall scene and narrow corridor scene. Generally speaking, there are few geometric features in the corridor, and the location accuracy of the laser radar is very poor.

We tested the data of these three experiments. Some important threshold parameters used in the experiment are set as follows:

Parameter $T_1, T_2, T_3$ are 10s, 25s and 3s; Parameters L1, L2, L3 and L4 are 5m, 7m, 20m and 0.2m.

## 3.2. Results and Analysis

At night, we conducted experiment 1, experiment 2 and experiment 3 in the above experimental scene 1, experimental scene 2 and experimental scene 3 respectively, and all experiments have loops. We compare the results with loops (LI-SLAM) and without loops (NDT-LOAM). The positioning accuracy results are shown in Table 1.

**Table 1**

Comparison of error results

| Group | Distance(m) | Amount of Detected Loop Closure | Accuracy (%) | NDT-LOAM | LI-SLAM |
|---|---|---|---|---|---|
| #01 | 203.245 | 8 | 100% | 0.107 | 0.024 |
| #02 | 422.863 | 6 | 100% | 1.735 | 0.245 |
| #03 | 323.552 | 10 | 100% | 4.575 | 1.176 |

Experiment 1 is an underground parking lot scene, including many columns, beams and other conventional buildings, with good environmental perception structure. It can be seen from Table I. and Figure 7 that in this experiment, the positioning accuracy of NDT-LOAM has been relatively high, and the positioning error is 0.1072m. In this experiment, LI-SLAM detects 8 loop closures, with an accuracy of 100% and a positioning error of 0.0244m. Since the positioning accuracy of NDT-LOAM is relatively high, point cloud comparison is not done here. The experiment shows that the positioning accuracy of LI-SLAM is much higher than that of NDT-LOAM in the environment with simple scene and good LiDAR positioning and mapping effect.

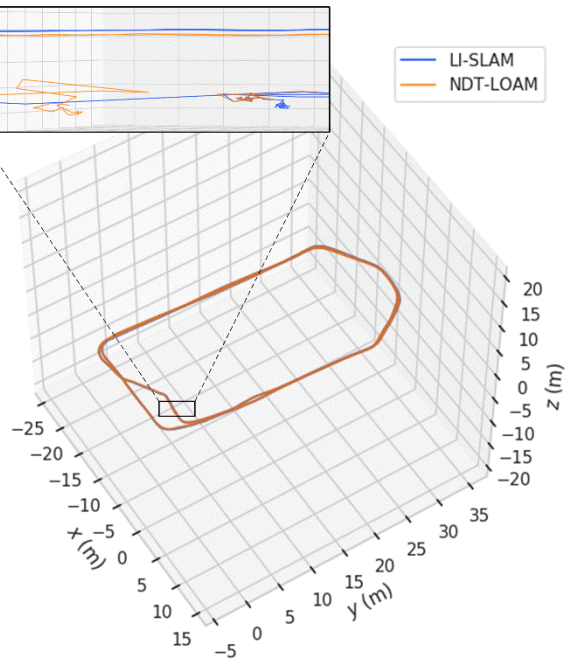



**Figure 7**: Experimental scene1 trajectory comparison

Experiment 2 includes indoor and outdoor scenes. The interior includes many columns, beams and other conventional buildings, with good environmental perception structure. However, the outdoor

space is relatively wide, the feature points are relatively few, and there are several speed bumps. Figure 8 shows the positioning trajectory obtained by NDT-LOAM and LI-SLAM, in which the red circle represents the starting point of the trajectory. In Table 1 and Figure 7, it can be seen that in this experiment the positioning error of NDT-LOAM is 1.735m. The LI-SLAM algorithm proposed by us has a positioning error of 0.245m. Compared with the algorithm without loop closure optimization, the positioning accuracy of LI-SLAM has been greatly improved. Figure 9 (a) and Figure 9 (b) are the point cloud images obtained by NDT-LOAM and LI-SLAM respectively. It can be clearly seen that the mapping accuracy of LI-SLAM is higher than that of NDT-LOAM.

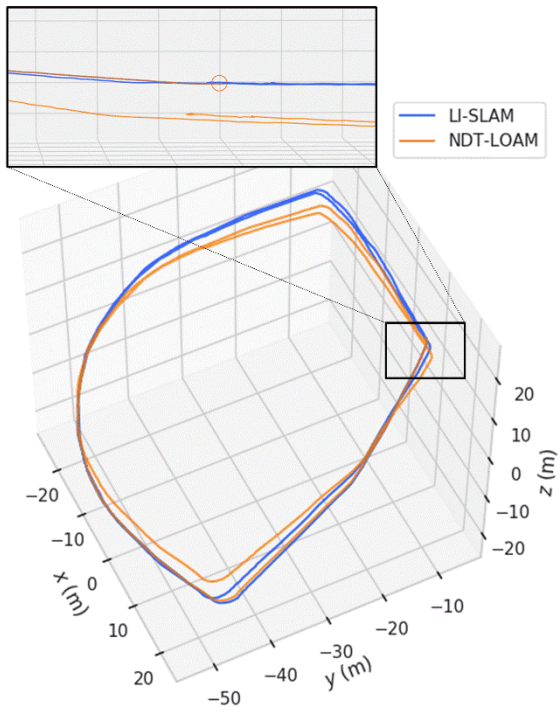
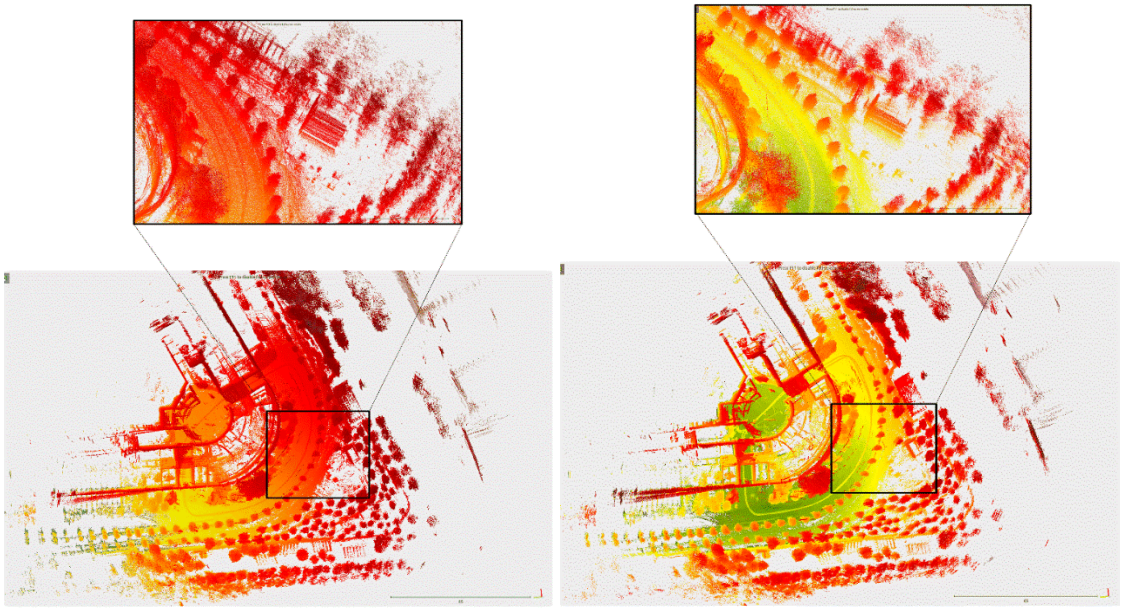



**Figure 8**: Experimental scene2 trajectory comparison



(a) (b)

**Figure 9**: Point cloud maps of experimental scene2

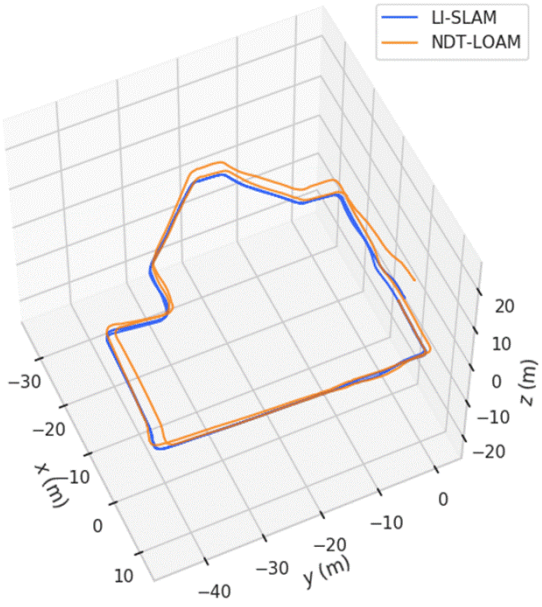
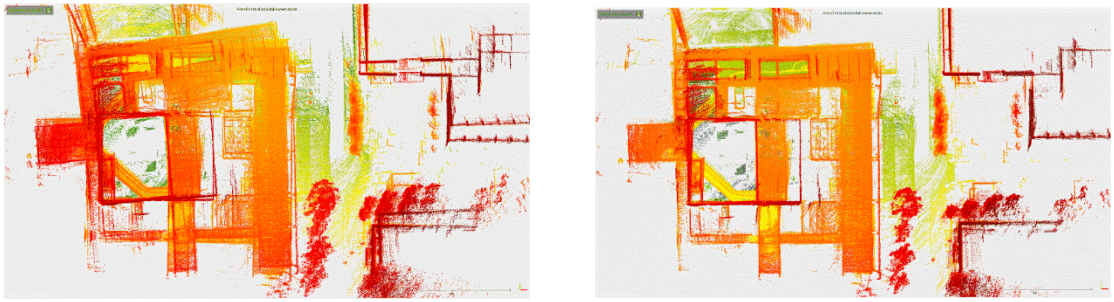



**Figure 10**: Experimental scene3 trajectory comparison



(a) (b)

**Figure 11**: Point cloud maps of experimental scene3.

The scene of experiment 3 is complex. The scene includes narrow corridors, halls and railings. Especially in narrow corridors, the structure of environmental perception is poor. in Table 1 and Figure 8, It can be seen that the positioning accuracy of LI-SLAM is significantly higher than that of NDT-LOAM. The positioning accuracy of NDT-LOAM is relatively poor, and the positioning error is 4.575m. The LI-SLAM proposed in this paper detects 8 loop closures, with an accuracy of 100% and a positioning error of 1.1758m. The experiment shows that the positioning accuracy of LI-SLAM is much higher than that of NDT-LOAM in the environment of complex scene and poor effect of LiDAR positioning and mapping.

Figure 11 is the point cloud diagram of experiment 3, in which figure 11 (a) is the point cloud diagram obtained by NDT-LOAM and Figure 11 (b) is the point cloud diagram obtained by LI-SLAM. From the point cloud map, it can be clearly seen that the point cloud map obtained by NDT-LOAM has wall ghosting and bending, which does not appear in LI-SLAM. This shows that, the mapping accuracy of LI-SLAM is significantly higher than that of NDT-LOAM in this environment.

## 4. DISCUSSION

In this paper, we propose a LiDAR SLAM back-end used at night. This method combines LiDAR and infrared image. First, we use NDT-LOAM as the front-end to get the front-end odometer data. Then we use BOW to score the similarity of infrared images. Then we combine the infrared image, front-end odometer and LIDAR point cloud data to determine the loop closure. Finally, we optimize the front-end results and build the global map.

The main innovation of this paper is a new method of loop closure detection based on infrared image and LiDAR. This method is a good fusion of image and LiDAR information, and it can work normally in dark conditions. It judges the loop closure points in time and space and identifies the final loop closure. We selected several different scenes at night to verify our method. The results show that compared with the SLAM Based on LiDAR, our LiDAR/Infrared image SLAM with loop closure detection and global graphics optimization has achieved better performance, including better mapping and point location results. However, in some cases, the information extracted from infrared images is too little. Therefore, there is still room for improvement in the enhancement of infrared image information. In the future, we plan to use deep learning and other methods to enhance the infrared image information.

## References


[1] Zhou B, Li Q, Mao Q, et al. ALIMC: Activity landmark-based indoor mapping via crowdsourcing[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(5): 2774-2785.

[2] Zhou B, Li Q, Mao Q, et al. Activity sequence-based indoor pedestrian localization using smartphones[J]. IEEE Transactions on Human-Machine Systems, 2014, 45(5): 562-574.

[3] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]//2007 6th IEEE and ACM international symposium on mixed and augmented reality. IEEE, 2007: 225-234.

[4] Hess W, Kohler D, Rapp H, et al. Real-time loop closure in 2D LIDAR SLAM[C]//2016 IEEE international conference on robotics and automation (ICRA). IEEE, 2016: 1271-1278.

[5] Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters[J]. IEEE transactions on Robotics, 2007, 23(1): 34-46.

[6] Qin T, Li P, Shen S. Vins-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.

[7] Shin Y S, Park Y S, Kim A. Direct visual slam using sparse depth for camera-LiDAR system[C]//2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 5144-5151.

[8] Shao W, Vijayarangan S, Li C, et al. Stereo visual inertial LiDAR simultaneous localization and mapping[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019: 370-377.

[9] Zhang J, Singh S. Laser–visual–inertial odometry and mapping with high robustness and low drift[J]. Journal of Field Robotics, 2018, 35(8): 1242-1264.

[10] Hahnel D, Burgard W, Fox D, et al. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements[C]//Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453). IEEE, 2003, 1: 206-211.

[11] Zhang J, Singh S. LOAM: LiDAR Odometry and Mapping in Real-time[C]//Robotics: Science and Systems. 2014, 2(9): 1-9.

[12] Chen S, Zhou B, Jiang C, et al. A LiDAR/Visual SLAM Backend with Loop Closure Detection and Graph Optimization[J]. Remote Sensing, 2021, 13(14): 2720.

[13] Alliez P, Bonardi F, Bouchafa S, et al. Indoor localization and mapping: Towards tracking resilience through a multi-slam approach[C]//2020 28th Mediterranean Conference on Control and Automation (MED). IEEE, 2020: 465-470.

[14] Shin Y S, Kim A. Sparse depth enhanced direct thermal-infrared SLAM beyond the visible spectrum[J]. IEEE Robotics and Automation Letters, 2019, 4(3): 2918-2925.

[15] Kataoka R, Suzuki R, Ji Y, et al. ICP-based SLAM Using LiDAR Intensity and Near-infrared Data[C]//2021 IEEE/SICE International Symposium on System Integration (SII). IEEE, 2021: 100-104.

[16] Chen S, Ma H, Jiang C, et al. NDT-LOAM: A Real-time LiDAR odometry and mapping with weighted NDT and LFA[J]. IEEE Sensors Journal, 2021.

[17] Hahnel D, Burgard W, Fox D, et al. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements[C]//Proceedings 2003

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453). IEEE, 2003, 1: 206-211.
[18] Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE transactions on robotics, 2015, 31(5): 1147-1163.
[19] Latif Y, Cadena C, Neira J. Robust loop closing over time for pose graph SLAM[J]. The International Journal of Robotics Research, 2013, 32(14): 1611-1626.
[20] Ulrich I, Nourbakhsh I. Appearance-based place recognition for topological localization[C]//Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065). Ieee, 2000, 2: 1023-1029.