Dean&Francis

ISSN 2959-6157

Research on Navigation Technology of Indoor Warehousing and Logistics Robots

Zhiqing Deng

School of Automation, Nanjing University of Science and Technology, Nanjing, Jiangsu Province, 210094, China
Corresponding author Email: dengzhiqing@lsu.edu.gn

Abstract:

With the surging demand for warehousing, indoor logistics robots have become the key to enhancing logistics efficiency. In the current highly dynamic warehouse operation environment, the traditional navigation methods based on fixed paths, such as magnetic strips and QR codes, require the pre-laying of guidance facilities. Problems such as path conflicts and location failures often occur. It is difficult to meet the requirements of smart warehousing. The existing Simultaneous Localization and Mapping robot navigation technology has made considerable progress, but still faces core challenges such as dynamic interference and computational efficiency. This paper focuses on the navigation technology of warehouse robots. By systematically analyzing the performance of visual Simultaneous Localization and Mapping (SLAM), laser SLAM, and multi-sensor fusion technology, it reveals the advantages and disadvantages of various algorithms in terms of accuracy, robustness, and adaptability. Research findings show that deep learning and multi-sensor fusion can significantly enhance the stability of the system in complex environments, but the contradiction between algorithm complexity and real-time performance still needs to be resolved. The research suggests that in the future, lightweight neural network models should be developed and combined with new types of sensors to achieve more efficient and reliable autonomous navigation systems, providing new ideas for the construction of smart logistics.

Keywords: intelligent navigation; Logistics robot; Indoor warehousing

1. Introduction

In recent years, e-commerce has witnessed explosive growth, which has greatly driven the demand for intelligent warehousing. Data shows that in 2024, the cumulative volume of express delivery services in China exceeded 175 billion pieces [1]. With the rapid development of e-commerce, the logistics industry is facing an urgent need for efficiency improvement and cost reduction. However, traditional manual sorting and fixed path navigation have been unable to adapt to complex dynamic scenarios. Warehousing and logistics robots achieve automatic handling and sorting of goods through technologies such as environmental perception and autonomous navigation, which can effectively improve logistics efficiency. Due to reasons such as a high rate of human-machine mixed movement, high similarity of shelves, and generally small aisle widths, the current warehousing environment has typical characteristics such as high dynamics and strict spatial constraints. This places extremely high demands on the environmental perception and navigation capabilities of robots. The traditional guidance methods based on magnetic strips or QR codes are gradually being replaced by autonomous navigation technologies based on Simultaneous Localization and Mapping (SLAM) due to their lack of flexibility and high maintenance costs [2].

Nowadays, significant progress has been made in the research of navigation technology for indoor warehousing and logistics robots. In terms of sensor fusion, research shows that combining lidar with visual data can effectively enhance environmental perception capabilities. For instance, the bird's-eye view mapping and data-level fusion method proposed by Ou et al. significantly improves the accuracy of obstacle detection by compressing 3D point clouds into 2D planes and matching them with lidar and camera data. Experiments show that the obstacle recognition range after fusion is more comprehensive than that of a single lidar. The navigation success rate in complex environments has increased to 97.5%, and the average number of collisions is only 0.063 [3]. In the field of SLAM technology, Kalman filtering and particle filtering algorithms are widely used in environmental modeling. For instance, Zhang et al. 's research is based on a robot mapping and navigation system that integrates multiple sensors. It uses laser SLAM technology to scan environmental geometric features in real time and combines point cloud registration and pose optimization algorithms to construct high-precision maps [3]. Furthermore, the application of loopback detection technology has further improved the map accuracy and positioning stability. Important breakthroughs have also been made in the application of deep reinforcement learning in navigation. Ou et al. further proposed an autonomous navigation system based on SAC, combined with heuristic global planning, to achieve efficient path planning in unknown environments. The experimental results show that its average navigation time is shortened by 15.6% compared with the traditional TD3 algorithm, and the path length is reduced by 8.4% [4]. Although progress has been made in navigation research, dynamic environmental adaptability, limitations of computing resources, and the generalization ability of algorithms remain the challenges of current research. For instance, the research on logistics inspection robots by Zhang et al. pointed out that the feature matching performance of the navigation system integrating multi-line lidar and vision would significantly decline when moving at high speed [4]. This has prompted the academic and industrial communities to continuously explore the application of emerging technologies such as deep learning, multi-sensor fusion, and edge computing in navigation systems.

This article aims to systematically review the current research status of navigation technology for indoor warehousing and logistics robots, deeply analyze key technological breakthroughs, and make forward-looking prospects for future development trends, providing theoretical references and technical route guidance for promoting the intelligent transformation of warehousing and logistics.

2. Navigation Technology for Warehousing and Logistics Robots

2.1 Visual SLAM Navigati on Technology

2.1.1 Feature points and direct SLAM

The visual SLAM navigation technology of warehousing and logistics robots has made remarkable progress in recent years. The visual SLAM technology mainly captures environmental images in real time through monocular, binocular, or RGB-Depth cameras. Extract Feature points such as Oriented FAST and Rotated BRIEF, Scale Invariant Feature Transform, or directly use a deep learning-based optical flow method to estimate camera motion. Meanwhile, sensor fusion is carried out in combination with IMU data to improve the stability of pose estimation [5]. Compared with laser SLAM, the vision solution has advantages in cost and scene understanding, and is particularly suitable for intelligent warehousing scenarios that require object recognition or human-machine collaboration. Its lightweight and semantic characteristics are promoting the development of Automated Guided Vehicles and Automatic Mobile robots towards higher-level autonomous decision-making.

The ORB-SLAM series algorithms based on the feature point method have achieved real-time positioning and mapping in the warehouse environment through efficient ORB feature extraction and matching. SLAM based on feature points has a relatively high computational efficiency through ORB feature extraction and optimization of the backend. Because it relies on stable feature point matching, it has high positioning accuracy in textured environments. However, in low-texture or repetitive texture areas,

ISSN 2959-6157

feature points are prone to being lost, and the performance will drop sharply, from 0.9 to 0.3 Root Mean Square Error, resulting in positioning failure [4]. Moreover, dynamic obstacles will interfere with the matching of feature points and require additional processing. While direct methods such as LSD-SLAM utilize the grayscale information of the image and show better robustness in the weak texture area, with the RMSE maintained at 0.8-0.9, but relying on high frame rate cameras [6]. The direct method is based on minimizing photometric error and does not rely on feature points, making it suitable for environments with weak textures or uniform lighting. The formula for minimizing photometric error is:

$$E_{photo} = \sum_{p \in \Omega} \left\| I_t(p) - I_{t-1}(w(p,\xi)) \right\|_{\gamma} \quad (1)$$

E_{photo} is the total luminosity error, p is the pixel coordinate, $I_t(p)$ is the intensity value of the image at the pixel at time t, ω is the deformation function, and ξ is the motion parameter of the camera from time $t-1$ to time t .Large-Scale Direct Monocular SLAM directly processes the grayscale information of pixels and can output semi-dense or dense maps, eliminating the feature extraction step, so it has a relatively high computational efficiency. However, this method is relatively sensitive to illumination, and the photometric error is easily affected by the variation of illumination. Moreover, when its camera moves rapidly, the image may be blurred, resulting in matching failure.

2.1.2 SLAM in deep learning

Deep learning-based SLAM, such as DeepSLAM, is adaptable to the environment. It can learn complex environmental features through training data and can also integrate multimodal data such as Light Detection and Ranging, cameras, and Inertial measurement units. The drawback of this method is that it requires a large amount of labeled data for training, and the gap between simulation and reality may affect performance. Moreover, it has a high demand for computing resources. Although the path planning efficiency of the AC algorithm in a dynamic environment is superior to that of traditional methods, reducing the average time by 30%, the training requires 2000 iterations [3].

Fig. 1 compares the positioning accuracy of three SLAM algorithms: ORB-SLAM2, LSD-SLAM and DeepSLAM in different texture environments. The vertical axis in Fig. 1 represents the positioning accuracy, and the horizontal axis represents the complexity of the environmental texture.

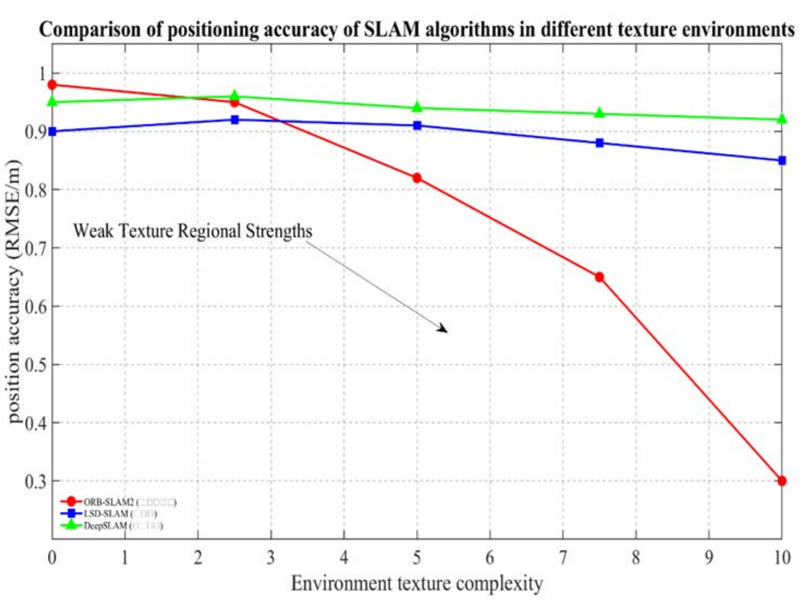

Fig. 1 Precision comparison curve of traditional SLAM and deep learning SLAM (Photo/Picture credit: Original).

The results show that when the texture complexity increases, the performance of ORB-SLAM2 decreases significantly, indicating its strong dependence on texture. LSD-SLAM performs stably, and DeepSLAM ultimately maintains the highest accuracy. It can be seen from the analysis that the selection of SLAM technology requires weighing the environmental characteristics, hardware resources and real-time requirements. Multi-sensor fusion and deep learning methods are a future trend, but traditional methods still have advantages in specific scenarios.

Visual SLAM relies on camera data. Its advantages include low cost, the ability to extract rich texture and semantic information, and suitability for indoor structured environments. The drawback is that it is sensitive to lighting and weak texture, and dynamic scenes are prone to failure. For instance, the changes in lighting conditions caused by the brightness differences in different areas of the warehouse can lead to unstable feature extraction. Repeatedly and neatly arranged shelves are likely to cause mismatching, and dynamic objects such as moving forklifts or workers can interfere with the accuracy of pose estimation. Furthermore, the problem of cumulative errors in large-scale environments still needs to be solved through a more efficient closed-loop detection mechanism.

2.2 Laser SLAM Navigation Technology

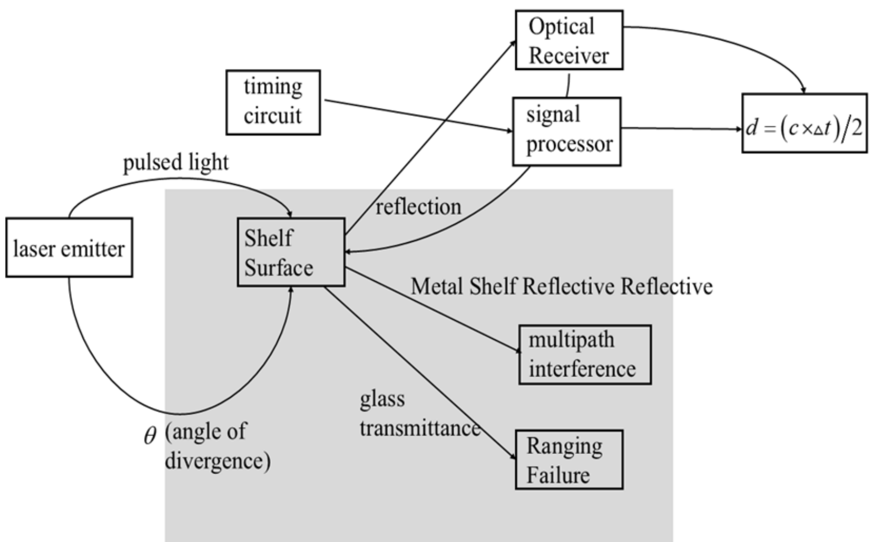

Fig. 2 Schematic diagram of Laser ranging principle in Storage (Photo/Picture credit: Original).

Fig. 2 shows the ranging principle of the warehouse laser. The laser SLAM navigation technology of warehousing and logistics robots has demonstrated remarkable technical maturity and practicality in warehousing scenarios in recent years. The core lies in using the high-precision ranging capability of lidar to construct the environmental geometric structure. A typical laser SLAM system acquires the environmental point cloud through the Time of Flight ranging principle to achieve the autonomous positioning and map construction of the robot.

The Cartographer algorithm and Lightweight and Ground-Optimized Lidar Odometry and Mapping, as two representative algorithms in the field of laser SLAM, each have their distinct technical characteristics and applicable scenarios. The Cartographer algorithm has its powerful global optimization ability. This algorithm effectively solves the cumulative error problem during long-term operation through the graph optimization framework and loopback detection mechanism. It can maintain a high consistency in mapping in complex environments and is particularly suitable for the corridor structure in warehousing scenarios. Its objective function is [7]:

$$\arg\min_{\xi}\sum_{i,j}\left\|e_{ij}(\xi_i,\xi_j)\right\|^2_{\sum_{ij}} \quad (2)$$

Among them, e_{ij} is the error function, $\sum_{ij}$ is the covariance matrix, ξ is the pose set of the robot, and ξ_i, ξ_j are the i-th and j-th pose nodes. However, this advantage also brings about a relatively high computational complexity and requires reliance on high-performance computing platforms. These factors limit its application in resource-constrained scenarios.

In contrast, lightweight algorithms such as LeGO-LOAM, based on the core idea of feature extraction, enable the algorithms to run stably on embedded devices and are particularly suitable for application scenarios with high real-time requirements. For example, in a dynamic warehouse environment, LeGO-LOAM can effectively handle the interference of moving obstacles through point cloud segmentation and motion compensation technology. However, its lightweight design also brings the limitation of weak global consistency. In an environment with sparse features, this type of algorithm is prone to positioning drift. Experimental data show that the positioning error of pure visual SLAM is 0.2163 meters, while the multi-sen-

ISSN 2959-6157

sor solution integrating the wheeled odometer and IMU can reduce the error to 0.0594 meters. This indicates that the lightweight algorithm may not meet the high-precision requirements when operating independently and needs to rely on the assistance of other sensors [8].

The multi-lidar fusion scheme solves the blind zone problem of traditional 2D laser SLAM in high-level space by expanding the perception range. Modern laser SLAM systems combined with adaptive Monte Carlo positioning algorithms can achieve centimeter-level accuracy in dynamic warehouse environments. For example, in the rotating lidar solution, the point cloud density is increased by more than 35% [7]. However, this technology still faces problems such as ranging failure caused by the reflection of metal shelves and the transparency of glass partitions, and dense dynamic obstacles can interfere with point cloud registration.

Although laser SLAM performs well in terms of accuracy and real-time performance, it is limited by cost and the ability to handle dynamic environments. If combined with multiple sensors, its robustness can be significantly improved. Future research directions can focus on the development of low-cost lidars, such as solid-state lidars, and can also be dedicated to integrating lasers with vision, taking into account both geometric and semantic information. Meanwhile, optimize the lightweight algorithm to reduce the demand for computing resources and meet the navigation requirements of complex warehousing scenarios.

2.3 Multi-Sensor Fusion Navigation Technology

The multi-sensor fusion navigation technology of warehousing and logistics robots is becoming the mainstream solution in industrial applications. Its core lies in overcoming the limitations of a single sensor through the complementarity of multi-source information. Modern systems adopt tightly coupled Kalman filtering or factor graph optimization algorithms. Fig. 3 is a schematic diagram of multi-sensor fusion based on factor graphs. This technology deeply integrates the precise geometric measurements of lidar, the texture information of visual sensors, the high-frequency motion data of IMU, and the odometer information of wheel encoders to achieve centimeter-level positioning accuracy.

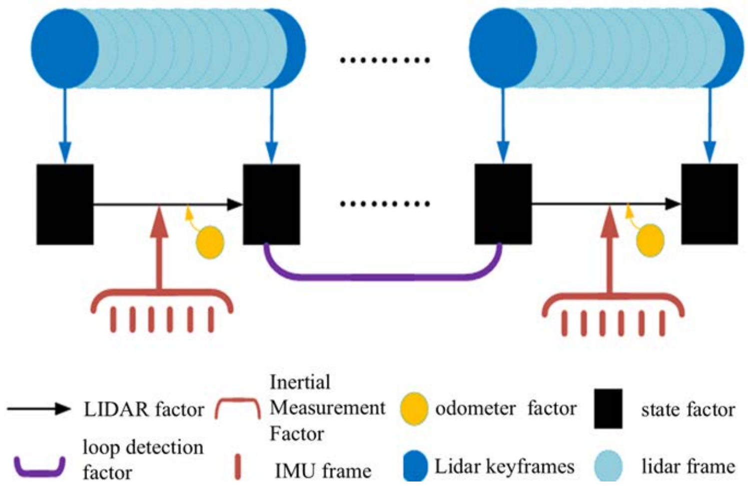

Fig. 3 Schematic diagram of multi-sensor fusion based on factor graph [9].

The current development status of multi-sensor fusion navigation technology shows a trend of evolving from traditional filtering methods to intelligent and adaptive directions. From the perspective of technical implementation, the current mainstream methods can be divided into two categories: the fusion framework based on optimization theory and the adaptive fusion based on machine learning. In the optimity-based method, depth constraints are provided for visual features through lidar point clouds to construct a nonlinear optimization objective function including polar line errors. The key lies in using the precise ranging ability of lidar to correct the scale uncertainty of monocular vision, and at the same time suppressing the cumulative error through keyframe binding adjustment [10]. This type of method performs well in structured environments, but it has a high dependence on the calibration accuracy of sensors and the stability of feature extraction.

In contrast, the application of deep reinforcement learning in machine learning in multi-sensor fusion navigation technology is gradually demonstrating unique advantages. It can solve complex decision-making problems that are difficult to handle by traditional fusion methods through autonomous learning and optimization strategies. Its adaptive weighting method can dynamically adjust the confidence level of each sensor according to environmental characteristics. For example, it can automatically reduce the visual weight under weak light conditions to enhance the dependence on laser data. Among them, Daytime Running Light can directly handle the input of multimodal sensors through end-to-end training, automatically learn the optimal feature extraction and fusion strategies, and avoid the limitations of traditional manual design of fusion rules. The Daytime Running Light-multiple model adaptive estimation framework represents a new direction of intelligent integration, combining multi-model adaptive estimation with deep reinforcement learning. The system monitors the residual difference constants of Point-to-line Iterative Closest Point and ORBSLAM2 in real time through an Unscented Kalman Filter. The weight distribution is dynamically adjusted by using the Proximal Policy Optimization algorithm. This method is particularly suitable for scenarios where the reliability of sensors changes dynamically. For example, when the lidar fails in a glass curtain wall environment, the system can automatically reduce the corresponding weight through residual difference constant detection. Experimental data show that in the warehouse environment with poor point cloud continuity, the positioning error of this method is stably reduced by 15%-20% compared with traditional methods [11].

It is worth noting that the two types of methods complement each other in actual deployment. The optimization method is more engineering-feasible on resource-constrained devices, while the learning method shows stronger adaptability in complex and unstructured environments. These advancements jointly drive multi-sensor fusion navigation from the laboratory to large-scale applications.

However, after analysis, it is found that the multi-sensor fusion technology also has many problems. At present, the performance of the fusion framework based on factor graph optimization highly depends on the accuracy of sensor calibration and the stability of environmental characteristics. Meanwhile, the error improvement shown by the experimental data is often achieved in a controlled environment. However, in a real warehouse, dynamic interferences such as the deformation of goods stacking and the occlusion of personnel flow can simultaneously disrupt the observation models of multiple sensors. At this time, the reinforcement learning strategy trained based on historical data may suffer from catastrophic forgetting due to the environmental state exceeding the distribution range.

3. Discussion on Future Directions

The existing technical challenges are mainly reflected in three dimensions: environmental adaptability, system robustness, and intelligence level. Although laser SLAM has the advantage of high precision, it is difficult to deal with reflective surfaces and transparent obstacles [4]. Moreover, visual SLAM has insufficient stability in complex lighting and texturally missing scenarios, while multi-sensor fusion is limited by calibration drift and computational complexity issues. Real-time obstacle avoidance decision-making in a dynamic environment has not yet broken through the bottleneck of millisecond-level response, especially the prediction ability for scenarios where irregularly shaped goods and personnel flow together is weak.

Based on the current research status and core challenges of navigation technology for warehousing and logistics robots, the navigation technology of future warehousing and logistics robots will develop in a more intelligent, adaptive and robust direction, deeply integrating multi-sensor data and advanced algorithms to break through the current bottlenecks of environmental adaptability and dynamic decision-making [3]. With the further combination of deep learning and SLAM technology, the online calibration and adaptive fusion methods based on neural networks will significantly improve the stability of the system in non-ideal environments such as complex lighting and dynamic interference. The introduction of new types of sensors, such as event cameras and millimeter-wave radars, will fill the perception blind spots of traditional vision and lidar in reflective, transparent objects and fast-moving scenarios [6]. The virtual-real mutual feedback training system will provide a high-fidelity simulation environment for deep reinforcement learning, solve the bottleneck of real data collection, and enable robots to have stronger real-time obstacle avoidance and path planning capabilities when facing complex scenarios such as the mixed flow of irregular goods and personnel. The coordinated development of these technologies will drive the evolution of warehousing and logistics robots from single navigation functions to intelligent agents with autonomous decision-making and collaborative operations, ultimately building a more efficient and flexible smart logistics system.

4. Conclusion

This paper reveals the key bottlenecks of the current warehouse robot navigation system in practical applications through technical comparisons. Although visual SLAM has the advantage of semantic understanding, it is prone to feature lock-loss in low-illumination ware-

ISSN 2959-6157

house areas. Although the measurement accuracy of laser SLAM can reach the centimeter level, it is limited by the multipath interference problem caused by metal shelves. Although the multi-sensor fusion scheme can enhance the robustness of the system, there are engineering implementation difficulties, such as high requirements for the time synchronization accuracy of each sensor and easy drift of calibration parameters. The research also particularly pointed out that when the existing algorithms deal with the unique compound interference of “high-dynamic human-machine mixed flow + dense shelf occlusion” in the warehousing scenario, their positioning errors will show a nonlinear growth trend, which becomes the main factor restricting the improvement of navigation performance. It can be seen that although the current technology has made remarkable progress in positioning accuracy and scene adaptability, it still faces key challenges such as environmental dynamics, sensor limitations, and algorithm robustness. Future technological breakthroughs will rely on the deep integration of multimodal perception fusion and cognitive intelligence, including the application of online calibration technology based on neural networks, new types of sensors such as event cameras, and new paradigms of reinforcement learning. These development directions will drive warehouse robots to evolve from single navigation execution to intelligent agents with environmental understanding and autonomous decision-making capabilities, ultimately achieving an industrial upgrade of the logistics system from “automation” to “intelligence” and providing important technical support for the construction of smart warehouses.

References

[1] CEIC data. China’s express delivery business volume. 2025.5.25. Available: https://ceidata.cei.cn

[2] Chen F. Research on AMR robot visual navigation technology for warehousing and logistics. North University of China, 2023.

[3] Ou Y., Cai Y., Sun Y., Qin T. Autonomous navigation by mobile robot with sensor fusion based on deep reinforcement learning. Sensors (Basel), 2024, 24(12): 3895.

[4] Zhang Y., Zhou Y., Li H., Hao H., Chen W., Zhan W. The navigation system of a logistics inspection robot based on multi-sensor fusion in a complex storage environment. Sensors (Basel), 2022, 22(20): 7794.

[5] Xu D. Research on key technologies of indoor autonomous mobile robot navigation for logistics operations. Zhejiang University, 2020.

[6] Arce D., Solano J., Beltrán C. A comparison study between traditional and deep-reinforcement-learning-based algorithms for indoor autonomous navigation in dynamic scenarios. Sensors (Basel), 2023, 23(24): 9672.

[7] Trybała P., Szrek J., Dębogórski B., Ziętek B., Blachowski J., Wodecki J., Zimroz R. Analysis of lidar actuator system influence on the quality of dense 3D point cloud obtained with SLAM. Sensors (Basel), 2023, 23: 721.

[8] Li F., Yu Z., Li H. Research on indoor positioning method of mobile robots based on multi-sensor fusion. Machinery Design & Manufacture, 2025, 1-4.

[9] Yang Y. Design of autonomous navigation system for warehouse robots based on multi-sensor information fusion. Harbin University of Science and Technology, 2024.

[10] Dai J., Li D., Li Y., Zhao J., Li W., Liu G. Mobile robot localization and mapping algorithm based on the fusion of image and laser point cloud. Sensors (Basel), 2022, 22: 4114.

[11] Wong C. C., Feng H. M., Kuo K. L. Multi-sensor fusion simultaneous localization mapping based on deep reinforcement learning and multi-model adaptive estimation. Sensors (Basel), 2023, 24(1): 48.