Application of LiDAR and Visual SLAM Technology in Automatic Driving

Tailong Shi [1]

[1] School of International Education, Wuhan University of Technology, Wuhan, 430070, China
tailongs@whut.edu.cn

Abstract. In recent years, autonomous driving has attracted a great deal of attention in many fields, such as academia and industry. If autonomous vehicles want to operate safely and effectively in complex and dynamic real environments, they must solve their own precise positioning problems. SLAM technology makes a big difference in the technology of automatic driving and automatic parking. This paper discusses the principle and application of slam technology based on laser and vision in automatic driving. This paper also sets a specific scene in an underground parking lot, evaluates the positioning methods proposed by LiDAR and vision camera to realize automatic driving, and compares them with relevant existing research methods. It is found that SLAM technology based on LiDAR and vision can greatly optimize the function of automatic driving. Finally, this paper presents the existing problems and future development direction of SLAM technology in the field of automatic driving.

Keywords: SLAM, Automatic driving, LiDAR, Vision, Point cloud.

1 Introduction

With the rapid development of 5G and big data, autonomous driving technology is an important development direction of today's automotive industry. At the same time, automatic driving is inseparable from the collaborative work of a variety of sensors, for example, automatic driving technology uses cameras to obtain information such as pedestrian conditions on the road and traffic light changes; The radar sensor is used to obtain the road condition information such as vehicle speed and distance. Among them, how to generate high precision map quickly and accurately has become a hot research direction at present. With the continuous development and progress of high and new technologies such as computers and sensors, SLAM technology plays an indispensable role in solving important fields such as high-precision real-time positioning and map construction. Therefore, slam technology can be applied to the realization of automatic driving, and the high-precision map formed with sensor data can improve the accuracy and safety of automatic driving.

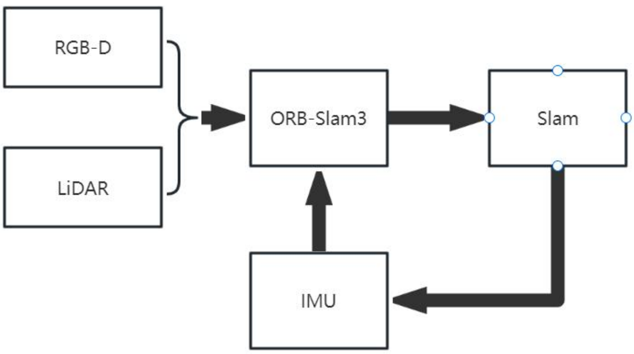
Slam technology needs to fuse multi-source sensors to observe a large amount of environmental information in order to optimize strategies more flexibly. For example, Jin Kaile, et al. proposed an optimization method for autonomous navigation of mobile robots under laser SLAM technology to accurately estimate the pose of mobile

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_19

robots and achieve better navigation effects [1]. Lin Lili proposed an automatic driving vehicle positioning method for underground parking lots based on point cloud registration. This method can provide better positioning accuracy in a few seconds and greatly save storage space. It can be seen that SLAM technology plays an important role in the development of autonomous driving. In addition, in some specific scenarios of autonomous driving, there are two implementation schemes of vision and laser. This paper discusses the feasibility of different schemes, analyzes their advantages and disadvantages, in order to look forward to providing help in the realization of full-scene autonomous driving in the future. Finally, the future development is prospected.

2 Process Strategies to Achieve Scenario-specific Autonomous Driving

Self-driving cars start by using a variety of sensors installed in different locations on their own, such as radar sensors to monitor and detect the position of nearby vehicles, LiDAR sensors to measure distance, identify lane markings, and detect road edges. Optical cameras to recognize road markings, detect traffic signal lights, find nearby pedestrians and track and detect surrounding cars as well as ultrasonic sensors on wheels to detect vehicles next to them and the edge of the road as they stop, to build and maintain a map of their surroundings. Secondly, the advanced positioning system accurately locates the current vehicle position based on the information obtained by the sensor. The traditional method or machine learning method is utilized to extraction of the LiDAR point cloud data, and the driving route and obstacle avoidance scheme are rationally planned by the on-board CPU. Huang Guangwei et al. [2] proposed a multi-sensor fusion SLAM strategy using "RGB-D+IMU+ LiDAR", using depth camera RGB-D and LiDAR as collectors, and applying the ORB-SLAM3 algorithm to automatically and gradually build a map consistent with the environment. Rapid autonomous positioning and mapping (SLAM) of moving vehicles in unknown environments can be realized, and high-precision inertial measurement sensor IMU is used to feed information such as vehicle speed and acceleration to the main controller for better attitude adjustment. The basic principles



of SLAM are shown in Fig.1.

Fig. 1. Basic principles of SLAM [2].

3 SLAM Technology Based on Vision and LiDAR

3.1 LiDAR SLAM

At present, SLAM technology has two schemes based on vision localization and LiDAR localization. This chapter will introduce the localization principle and application of the two schemes in automatic driving. The existing 3D LiDAR positioning methods for autonomous vehicles can be divided into three categories: method based on 3D registration, method based on 3D features and method based on deep learning.

Traditional positioning methods generally refer to method based on 3D registration and method based on 3D features: they usually involve the following steps, namely, feature representation method, matching algorithm, outlier elimination (optional), matching cost function, spatial search or optimization framework and time optimization or filtering framework, as shown in Fig.2.

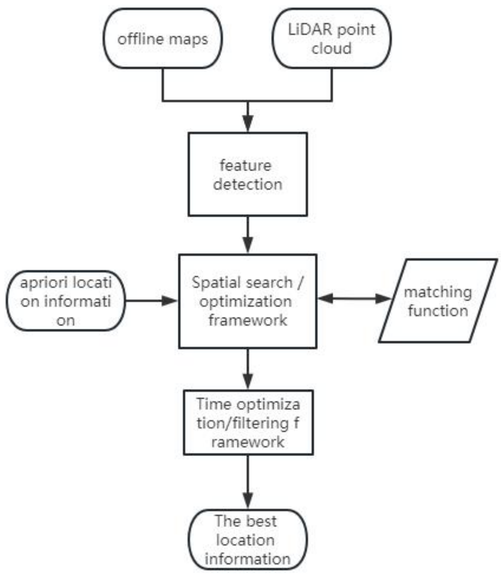


Fig. 2. Process of traditional positioning methods.

Method based on 3D registration is often combined with maps built offline, taking advantage of advances made in 3D point cloud registration. These methods, although very accurate, often fail to meet the requirements of real-time processing when using only LiDAR point cloud data. These methods can be considered intensive methods because they utilize all the points in the LiDAR point cloud data. In autonomous vehicle location research, registration is typically used in one of two ways: to register continuous scan data, or to register incoming scan data frames and pre-stored point cloud maps. The researchers came up with ways to reduce the cumulative errors brought on by continuous registration by incorporating widely used 3D registration algorithms, like ICP or NDT, into the positioning process. For instance, at work, ICP is coupled with a closed-loop system and a pose map construction process. Kovalenko Dmitri et al. odometer pipeline incorporates the physical understanding of LiDAR sensors and enhances the ICP algorithm through the use of the normal covariance filter-based down-sampling technique and the geometric antagonist-based outlier removal method [3]. To begin the optimization process and prevent local minimums, methods like ICP or NDT require a beginning position calculate. This frequently necessitates the inclusion of additional sensors, such IMUs, to create odometers which calculate the initial location. In a more classical SLAM method, IMLS-SLAM, a three-step technique described by Jean-Emmanuel Deschaud [4], under-samples the scanned data using a sampling strategy based on each point's observability after removing dynamic objects. Ultimately, the scanning-model matching technique defined by the implicit moving least square method is used to optimize the transformation's matching phases. In addition, some pretreatment methods, such as calculating the facet representation of point cloud, converting to line clouds and dimensionality reduction processing, are tried to improve the registration accuracy.

Method based on 3D feature approaches designs and extracts relevant features in 3D space, and then calculates the displacement between continuous scan data based on these features, also leveraging the advances in 3D point cloud feature extraction and registration. These methods have achieved satisfactory accuracy and real-time performance, but the effect is poor when dealing with sudden motion or high-speed motion. These techniques are categorized as sparse methods due to their reliance on only a chosen subset of points from the LiDAR point cloud data. In the article by Keisuke Yoneda [5], the researchers explore the types of data and features necessary for precise autonomous vehicle positioning. The team contends that feature construction and extraction should depend on the distribution of point clusters. Yet, their findings indicate significant variations in the distribution across different sites, rendering the method quite unstable.

Deep learning-based approaches: Locating is often viewed as a regression problem and attempts to use an end-to-end way to solve, either through feeding an original point cloud into a single network and outputting the predicted vehicle location, or by replacing certain modules in traditional methods with a learning network. The input point cloud frame is first projected into two-dimensional space to produce a panoramic depth image, and this image is then inputted into a simple two-branch convolutional network, designed to calculate the vehicle's displacement and directional shifts between the two provided frames. Lin Lili proposed a location method for autonomous vehicles in underground parking lots based on plane primitive assembly registration [6].

Most of the existing indoor location technology is used for personal location. These technologies typically include three broad categories: wireless based, vision based and measurement sensors based. However, the location method based on wireless has multi-path effect, and the location result is not stable enough. Vision-based positioning methods are susceptible to light and obstacles: positioning methods based on measurement sensors have large drift errors and cumulative errors. In view of the above problems and the lack of locating signals in underground parking lots, Huang Gang et al. proposed a multi-view and multi-scale positioning approach for intelligent vehicles positioning in underground parking lots [7].

3.2 Vision SLAM

For vision schemes, there are 3D target detection based on monocular camera images, 3D target detection based on binocular camera images, 3D target detection based on RGB images, and 3D target detection based on instance camera images [8].

Each visual camera has its own merits and demerits, which are gathered up in Table 1.

Table 1. Merits and demerits of each visual camera.

Monocular camera	SLAM using only one camera is simple in structure and low in cost.	It is impossible to provide accurate scene depth information, and the camera needs to be moved to estimate the distance and size of the object in the scene.
Binocular camera	It can provide accurate scene depth information.	Configuration and calibration are more complex; the amount of calculation is large, and the parallax calculation consumes a lot of computing resources.
RGB Camera	Captures colors and colors in the real world, restores image authenticity, provides better detail and accuracy, and is useful for high-precision measurement and inspection tasks	It can 't record depth information and can 't obtain three-dimensional information. For shooting in low light environment, RGB camera may not perform as well as black and white camera.

Event-based camera	High speed capture, high dynamic range, low latency, high resolution	Data asynchronous is not easy to deal with, single event effective information is less, data sparse incomplete.

4 Extraction Methods of LiDAR Point Cloud Features

Point cloud feature extraction is to extract the feature information of the point cloud itself from the point cloud with redundant data and unobvious features.

Traditional feature extraction methods mainly include:

The first one is called based on image feature detection method. Such methods transform raw point cloud data into feature descriptors, such as scale-invariant feature transformations

Commutation operators, accelerated robust feature descriptors and three-dimensional corner detection descriptors, etc. Then, the feature information of the point cloud is extracted through these feature descriptors [9].

The second one is called based on point cloud edge segmentation method. By segmenting the point cloud, this method extracts the intersection points and features of the line features in the point cloud.

The third one is called based on discrete point cloud geometric feature method. Such methods extract geometric features such as spatial distance, Angle, normal vector and curvature of point cloud based on the topological relationship of point cloud [10].

The fourth one is called point cloud clustering discrimination method. This method firstly divides the point cloud into clusters, then calculates the geometric and topological attributes of each cluster, and identifies and extracts the required features according to the distribution law of these attributes.

The fifth one is called based on deep learning Point cloud feature extraction method. This is one of the cutting-edge types of research in the field of point cloud processing, and its main idea.

Deep learning network is utilized to extract features from point cloud data. There are three main methods for point cloud feature extraction based on deep learning: Firstly, by based on the multi-view method, the point cloud is converted into multi-view data, and feature learning is performed on the projected image [11]. Secondly, by based on the voxelization method, the point cloud is separated into regular voxels and input into the convolutional neural network for learning [12]. Lastly, by based on point cloud convolution method, that is, global processing of point cloud data directly [13].

The method to extract feature based on deep learning is devoted to optimizing the network model, thereby reducing the computational complexity and developing the accuracy of point cloud extracting feature in complex scenes.

This paper introduces in detail the upward growth method based on voxel used in vehicle point filtering and the plane structure extraction method based on region growth used in feature extraction. The algorithm framework is shown in Fig.3.

Firstly, the global point cloud data of the underground parking lot is obtained and preprocessed to remove the noise data. Then, the preprocessed global point cloud data is subjected to voxel-based upward growth to filter the three-dimensional points of the vehicle.

Secondly, the global point cloud data after filtering the vehicle is quickly segmented to obtain the initial plane set of the global point cloud, and the plane boundary of the initial plane set is refined to obtain the final plane set of the global point cloud.

Thirdly, the short-term point cloud data of the environment around the autonomous vehicle is obtained and preprocessed to remove the noise data. Then, the preprocessed short-term point cloud data is subjected to voxel-based upward growth to filter the three-dimensional points of the vehicle;

Finally, the short-term point cloud data after filtering the vehicle is quickly segmented to obtain the initial plane set of the short-term point cloud, and the plane boundary of the initial plane set is refined to obtain the final plane set of the short-term point cloud.

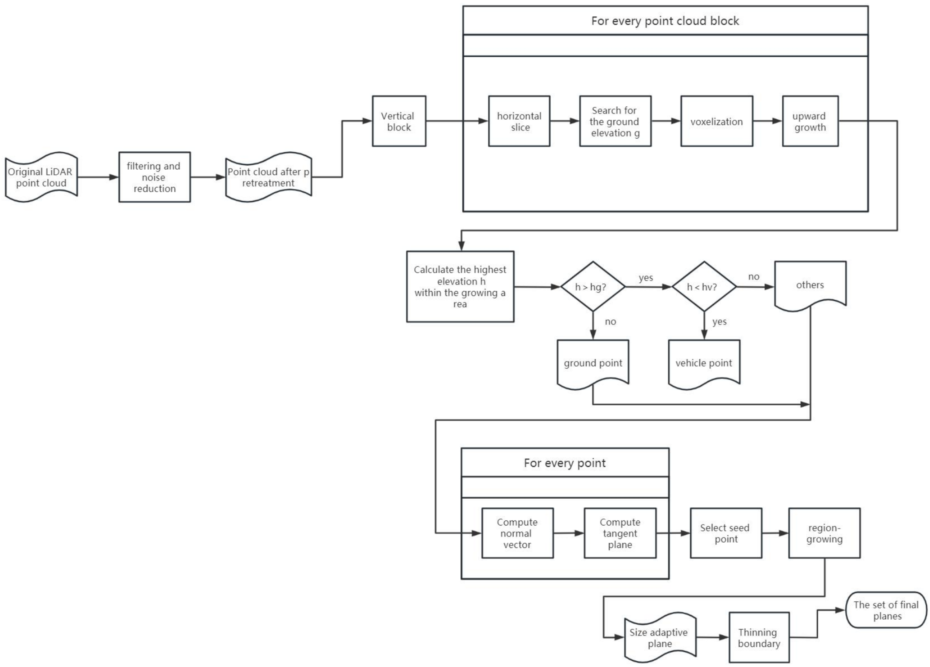


Fig. 3. The algorithm framework [6].

5 Feasible Scheme of Path Planning and Obstacle Avoidance

Huang Guangwei proposed a navigation framework, using Dijkstra algorithm as the global algorithm and DWA dynamic window algorithm as the local algorithm to realize the path planning function of autonomous mobile vehicles [2][14].

The global path planning is shown in Fig 4: Initialize the parameters after the start. After the driver sets the target point, the CPU plans the global driving path of the vehicle according to the Dijkstra algorithm, and plans the local driving path of the vehicle according to the DWA algorithm. After repeating the above actions, determine whether to reach the target point, and finish the work after completing the navigation task [15].

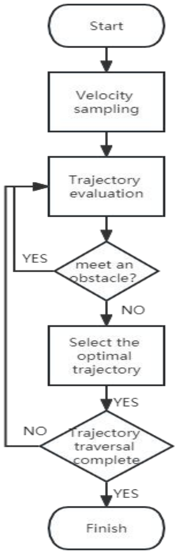


Fig. 4. The global path planning [2].

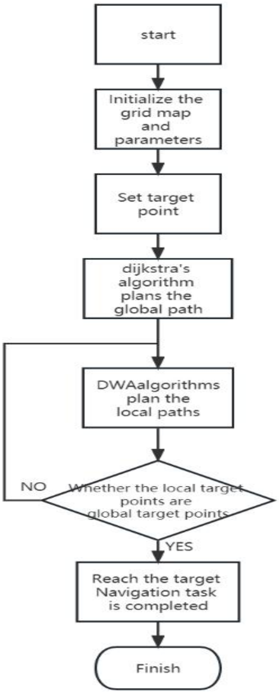


Fig. 5. The local path planning [2].

The planning of the local path is shown in Fig.5.After the beginning, the speed sensor and the engine crankshaft speed sensor are used to sample the speed, and then the global path is evaluated. If the obstacle is encountered, the planning is re-planned. If there is no obstacle, the optimal trajectory is selected. After that, it is judged whether the trajectory traversal is completed. If it is not completed, the above process will be repeated until the end [16].

6 Discussion

The image-based 3D detection method relies on the accumulation of deep learning technology in the field of image research for many years. However, due to the inherent defects of the image for the target depth perception, the detection performance of the image-based method is significantly worse than that of the point cloud-based method. How to effectively use the advantages of low image cost and fast calculation to improve the depth information capture ability of the image still needs more exploration.

LiDAR has a long detection distance and accurate perception of the depth information of the target, which can form a reliable data source for 3D target detection. However, the hardware cost of LiDAR is high, the calculation cost is large, and the point cloud is three-dimensional, sparse, disordered and irregular, which makes the 3D target detection method based on point cloud more expensive and more complex than the image-based method. How to reduce the cost of lidar, direct modeling of point cloud data, and more efficient extraction of information features of irregular point clouds are all areas where such methods need to be further improved. The deep SLAM currently used has the following deficiencies as shown in Table 2.

Table 2. Deficiencies of the current deep SLAM.

Data volume and annotation	Deep learning requires large-scale data and accurate annotation, but obtaining large-scale SLAM data sets is very difficult
Low real-time	Visual SLAM usually needs to be completed under the real-time requirements, even low frame rate, low resolution camera input will produce a large amount of data, requiring efficient processing and reasoning algorithms.
Generalization ability	Whether the model can accurately locate and map in new environments or scenes that have not been seen before.

7 Conclusion

At present, all kinds of existing methods have defects and shortcomings. The method based on feature matching can't be used in unknown environment. The vehicle must first obtain the garage's three-dimensional model and then extract plane features. The problem that all high-precision map research will face is that when the original point cloud structure is outdated, the high-precision map must be updated regularly. Otherwise, if the structure changes too much, the matching error will increase or the matching will fail, affecting the positioning accuracy. The inertial navigation system can provide continuous vehicle attitude and position information for the control system with higher frequency and times, and the output vehicle motion parameters also play a smoothing effect on other positions. In the future work, the inertial navigation system INS can be introduced as an auxiliary positioning technology to accumulate the vehicle motion parameters, so as to try to improve the accuracy and speed of the positioning method.

Future deep SLAM methods will be more and more close to human perception and cognitive models, and progress has been made in high-level map construction, human-like perception and localization, active SLAM methods, integration with task requirements, and memory storage and extraction. These developments will help robots achieve diverse tasks and self-navigation capabilities. The end-to-end training mode and information processing process are in line with human cognitive processes and have great potential.

References

1. Jin, K L. Li, F Y. Yang, T.: Research on autonomous navigation optimization of mobile robot under laser SLAM technology. Laser Journal **1**(6) (2024)
2. Huang, G W. Chen, M T. Liao, W T.: Autonomous mobile robot design based on visual SLAM. Mechanical engineer, 15-17 + 22 (2023)
3. Kovalenko, D. Korobkin, M. Minin, A.: Sensor aware LiDAR odometry. European Conference on Mobile Robots 1-6 (2019)
4. Deschaud, J-E.: IMLS-SLAM: scan-to-model matching based on 3D data.2. International Conference on Robotics and Automation (ICRA).2480-2485 (2018)
5. Yoneda, K. Tehrani, H. Ogawa, T. et al.: LiDAR scan feature for localization with highly precise 3-D map. Intelligent Vehicles Symposium Proceedings 1345-1350 (2014)
6. Lin, LL.: Location method of autonomous vehicles in underground parking lot based on point cloud registration. Xiamen University (2021)
7. Huang, G. Hu, Z. Mu, M. et al.: Multi-view and multi-scale localization for intelligent vehicles in underground parking lots. Transportation Research Record Journal of the Transportation Research Board **2673**(11), 791-800 (2016)
8. Liang, Z M, Huang, YP. Song, ZH. et al.: A review of 3D object detection methods based on deep learning in automatic driving. Journal of University of Shanghai for Science and Technology, 1-17 (2023）
9. Patel, M I. Thakar, V K. Shah, S K.: Image registration of satellite images with varying illumination level using HOG descriptor-based SURF. Procedia Computer Science **93**,382-388 (2019)
10. Lu, X. Ai, Y. Tian, B.: Real-time mine road boundary detection and tracking for autonomous truck. Sensors **20**(4),1121 (2020)
11. Chen, X. Ma, H. Wan, J. et al.: Multi-view 3d object detection network for autonomous driving. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 1907-1915 (2017)
12. Wu, Z. Song, S. Kholsa, A. et al.:3D shapenets: A deep representation for volumetric shapes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1912-1920 (2015)
13. Qi, C R. Su, H. Mo, K. et al.: Pointnet: Deep learning on point sets for 3d classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 652-660 (2017)
14. Zheng, H. Feng G L W. Pu, W J. et al.: Global Path planning in Closed environment based on Dijkstra Algorithm. Automotive Practical Technology **48**(16), 7-11(2019)
15. Aldibaja, M. Yanase, R. Suganuma, N.: Waypoint Transfer Module between Autonomous Driving Maps Based on LiDAR Directional Sub-Images. Sensors **24**(3) (2024)
16. Szrek, A. Romańczukiewicz, K. Kujawa, P. et al.: Comparison of TLS and SLAM technologies for 3D reconstruction of objects with different geometries. IOP Conference Series: Earth and Environmental Science **1295**(1), (2024)