Proceedings of the 2023 International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/12/20230291

Research on visual SLAM for dynamic scenes

Xincheng Xu

China Europe Institute of Engineering and Technology, Shanghai University, Shanghai, 200444, China

xincheng_xu@shu.edu.cn

Abstract. Simultaneous Localization and Mapping (SLAM) technology has many applications, such as intelligent robots, autonomous driving, and smart cities. It is a key technology for achieving autonomous robot navigation and environmental awareness. Robots need to perceive the surrounding environment in real-time and construct maps while conducting autonomous localization and path planning, which involves algorithms of SLAM technology. This article aims to introduce the research on visual SLAM algorithms in dynamic scenes. This article will first introduce the traditional SLAM algorithm, its research status, and its limitations. Next, introduce the research on the visual SLAM algorithm in dynamic scenes. Finally, the current visual SLAM algorithm's problems and future research directions were discussed. The research in this article will have significant value for the research and application of autonomous robot navigation, environmental perception, and smart cities.

Keywords: SLAM, dynamic scene, intelligent perception, smart city.

1. Introduction

Entering the 21st century, robots have gradually entered people's lives and have important applications in fields such as aerospace, medical treatment, and convenience of life. How robots complete autonomous positioning and navigation has always been a hot and difficult research topic. A common technology is GPS technology, but the positioning accuracy of GPS receivers is affected by satellite signal conditions and road environment, as well as many factors such as clock error, propagation error, receiver noise, etc. [1]. Moreover, there are ultrasonic navigation and location techniques. However, they are subject to the influence of the weather, the environment (reflection of the mirror or the limitation of the beam), and the obstacles [2]. Therefore, in most cases, obtaining prior environmental information in advance is difficult, requiring robots to calculate their position while creating environmental maps when moving in unfamiliar environments. Currently, SLAM technology combines robot perception and control, Excellent completion of autonomous positioning and map construction, with high autonomy and repeatability [3].

This article introduces the composition and static algorithms of traditional SLAM technology, points out its limitations for dynamic scenarios, introduces the development and application of dynamic SLAM algorithms, and finally points out its limitations and future development directions.

2. Introduction to SLAM technology

SLAM is used to locate and map simultaneously. Starting from an unknown position in an unfamiliar environment, the robot determines its position, posture, and trajectory with a sensor (e.g., radar or camera). It then incrementally builds a map based on its location to achieve the goals of both positioning and map building. SLAM is the culmination of a variety of technologies that have been used to achieve this goal.

SLAM systems are generally divided into five modules: sensor data, visual odometry, optimization, mapping, and loop closing. Its workflow is shown in Figure 1.

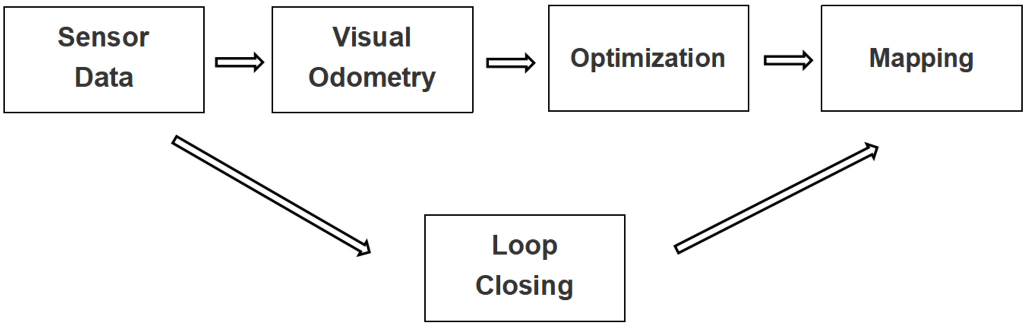

Figure 1. Overall visual SLAM flowchart.

After the sensor receives the data, the visual odometer estimates the relative motion (Ego-motion) of 2 seconds. The optimization processes the accumulated error of the visual odometer estimate, build a map on the basis of the estimated trajectory and make a loop to detect the image at different times in the same scene and use space constraint to remove the accumulated error.

3. Kalman filtering

Kalman filtering is a technology which makes use of the state equation of a linear system to obtain the optimal estimation of the state of the system. Due to the presence of noise and disturbance in the system, the optimal estimation can also be considered as a filtering process.

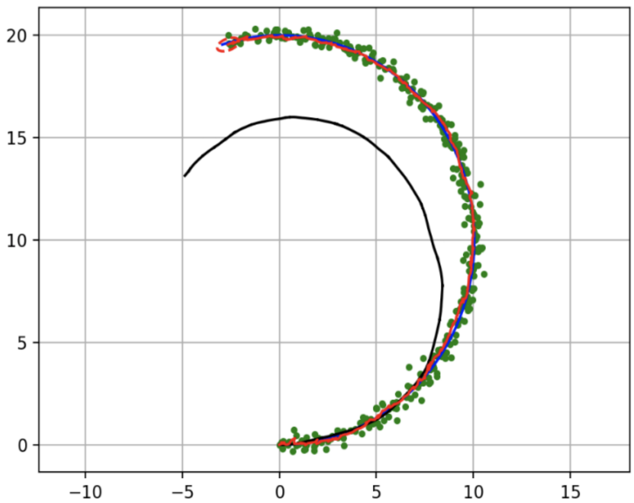

Figure 2. Kalman filter diagram (unit: meters).

In the Figure 2, the blue line is the true value of the track. The black line is the track derived from "track guess" (track guess refers to the method of inferring position solely by speed information, which includes the error of speed information in the position), the green point is the observation value

DOI: 10.54254/2755-2721/12/20230291

(ex. GPS), and the red line is the track filtered by EKF. The red ellipse represents the real-time covariance of the EKF output robot state.

The filter cycle consists of the following steps: motion prediction, measurement prediction, measurement, data association, and update.

4. Limitations of traditional SLAM algorithms

In recent years, significant progress has been made in the field of visual SLAM, but there are still technical difficulties. For example, dynamic objects can cause considerable inconsistencies in adjacent frames obtained by cameras, resulting in tracking loss, repositioning failure, track drift, and other issues which seriously affect the accuracy and robustness. Most current classical visual SLAM open-source systems assume that the environment is static, ignoring the impact of dynamic objects. However, there are a large number of dynamic objects in the natural environment. Therefore, it is of great significance and practical value to study how to improve the accuracy and robustness of visual SLAM systems in dynamic environments in the fields of autonomous navigation, unmanned navigation, AR/VR, etc [4].

5. Development and application of dynamic SLAM algorithm

In 2022, using the TUM RGB-D dataset and RGB-D depth camera, Li Yue proposed a pose estimation algorithm based on the DS-SLAM framework for a complex environment. It extracts ORB feature points by the adaptive threshold, assigns them reasonably using the quadtree algorithm, tracks corner motion using Pyramid LK optical flow method, and filters out dynamic feature points using geometric constraints combined with Segment semantics based on Cafe to preserve high-quality static feature points. At the same time, PnP is used to estimate camera motion and build a dense indoor semantic octree map.

Its future work is to use instance segmentation for detection and to increase the system's robustness. Train own dataset to increase the impact of different environments such as object class, distance, and light, and increase the mapping effect's richness. Choose lighter networks to reduce dependency on hardware conditions.

Its advantage is that the algorithm not only maintains a good posture effect in a static environment but also improves the performance in a dynamic environment and improves the positioning accuracy and real-time performance of the robot system in a complex environment [5].

In 2022, Zhang Jianhua, Zhang Tianjing, Zhao Yan, and others proposed a multi-sensor SLAM cooperative scheme combining a thermal imager and depth camera using a thermal imager and RGB-D depth camera from the point of view of "temperature". First, the joint calibration and multimodal image registration of the multi-sensor visual system consisting of depth camera and thermal imaging are implemented. Then, a visual millimeter is built based on the RDH three-mode image, and a static map is created. Finally, a multi-sensor visual SLAM algorithm is proposed for applying the visual system to the SLAM system.

Its future work will be to use better RGB-D cameras and thermal imaging sensors; In future research, deep learning can be used to process thermal images to better extract facial mask images.. The following research will study different dynamic scene types, complexity, and density parameters to explore their impact on the algorithm.

The advantage of the proposed algorithm is that it can effectively remove the interference of dynamic object feature points and improve the drawing effect compared with the traditional visual SLAM algorithm [6].

In 2021, Naigong Yu, Mengzhe Gan, Hejie Yu, and Kang Yang presented a dynamic RGB-D SLAM (DRSO-SLAM). Based on semantic information, the algorithm makes use of the pixel-level segmentation network to estimate the rough self-motion of the system and then extracts the moving target from the dynamic scene with the light stream and semantic information. Then, the position of the SLAM system is estimated by the static characteristic point [7].

DOI: 10.54254/2755-2721/12/20230291

In 2023, Xiaomin Ma and Ye Yang proposed a real-time and robust DIR (DIR) algorithm based on prior knowledge and geometry. In this paper, a semantic label CNN network based on descriptor correlation is proposed. To remove dynamic interference, a Geometric Consistency Check Module is designed that computes Bundle Adjustments to Predefine Fixed Key Points and then uses Semantic Weighted Polar Constraints to identify Dynamic Outliers. The method is incorporated into optimizing ORB-SLAM2, and the dynamic keywords are filtered out.

In the future, they will try to find an affine invariant feature extractor that is more suited to the motion of the camera. Furthermore, they aim to integrate semantic mapping and background restoration in DIR-SLAM, which is superior to low-dynamic scenes and robust to unknown environments [8].

In November 2022, Rushmian Annoy Wadud, Wei Sun, et al. proposed the DyOb-SLAM algorithm using the KITTI tracking dataset, a SLAM system that contains dynamic objects in its mapping and tracks objects in each frame. With advanced technologies such as convolution neural networks and dense optical flow algorithms, dynamic content can be accurately detected and used for further processing. The back-end of the DyOb-SLAM system can provide more optimized data to obtain an optimized sparse graph of static key points in the environment and a global map of camera and object tracks, and subsequently process each frame.

The advantage is that the DyOb SLAM system has very low average attitude error and superior object speed performance for VDO-SLAM. The system uses the Mask RCNN network to segment and process dynamic objects, which makes them more accurate in detecting and segmenting objects. In addition, the system can measure the speed of detected objects by using object ID tags and computing speed.

Its future work is to get a better SLAM processing system that can use a cloud environment to process both the primary processing unit and the data in the cloud. This way, the attitude estimation error can be reduced, and the drive output faster [9, 10].

6. Prospects of SLAM technology

6.1. Real-time and accuracy will be further improved

As computer computing power and sensor technology improves, SLAM algorithms will become more efficient and accurate. Future SLAM systems can build and update high-quality maps in real-time and handle more complex environments.

6.2. Multisensor fusion will become the main development direction of SLAM

Future SLAM systems will continue to explore the possibility of multisensor fusion, such as depth cameras, inertial measurement units (IMUs), and lidar. Data from these sensors can be complemented to improve the accuracy and reliability of SLAM systems.

6.3. SLAM will gradually be used in more areas

Currently, SLAM technology is mainly used in robots, auto-driving, etc. With the development and maturity of SLAM technology, SLAM will gradually be applied to a wider range of fields, such as augmented reality, virtual reality, smart home, medical and other fields.

7. Conclusion

Future SLAM systems will be integrated with other technologies, including computer vision, deep learning, and natural language processing. The integration of these techniques will make the SLAM system more accurate and intelligent, which will promote the development and application of SLAM.

SLAM technology is widely used in uncrewed vehicles, robot cleaners, and robot navigation.

SLAM technology still has challenges, such as positioning errors, sensor noise, and motion model errors in complex environments. These challenges require continuous improvement and optimization of SLAM algorithms to accommodate more complex application scenarios.

DOI: 10.54254/2755-2721/12/20230291

In addition, with the development of in-depth learning technology, the combination of in-depth learning and SLAM has become a current research hotspot. It can improve the robustness and accuracy of the SLAM algorithm by using a neural network to process and extract features from sensor data.

In general, SLAM is an important issue in the field of robots. The development of SLAM technology will help to achieve autonomous navigation and control of robots and provide a broad space for future applications of robots.

References

[1] Z. Yu,"Research on Visual SLAM Algorithm in Dynamic Scenes", Master's degree, no. 01, p.64, doi:10.27147/d.cnki.ghdju.2021.000300.

[2] W. Tao,"Overview of Visual SLAM for Mobile Robots",Data Communication, vol. 206, no. 01. pp.48-51, Feb. 2022.

[3] L. Yanzhen, S. Liguo, et al."Overview of Research on Mobile Robot Vision SLAM,"Intelligent Computer and Application, vol. 12, no. 07, pp. 40–45, Jul. 2022.

[4] H. Cheng , “Research on Visual SLAM System for Dynamic Scene,”Master's degree, no. 02, p.75, doi:10.27251/d.cnki.gnjdc.2022.001719.

[5] L. Yue and L. Zhenyu, “Visual SLAM System Based on Optical Flow Method and Semantic Segmentation,” Master's degree, no. 02, p. 62, Jun. 2022, doi ： 10.27322/d.cnki.gsgyu.2022.001147

[6] Z. Jianhua, crystal, Zhaoyan, Z. Ling, and H. U. of T. Zhou Hao, “Multi-Sensor Visual SLAM Method for Indoor Dynamic Scene,” Information and Control, vol. 51, no. 06. pp. 641-650+661, 2022.

[7] N. Yu, M. Gan, H. Yu and K. Yang, "DRSO-SLAM: A Dynamic RGB-D SLAM Algorithm for Indoor Dynamic Scenes," 2021 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 2021, pp. 1052-1058, doi: 10.1109/CCDC52312.2021.9602705.

[8] X. Ma et al., “DIR-SLAM: Dynamic Interference Removal for Real-Time VSLAM in Dynamic Environments,” Mobile Information Systems, vol. 2023, p. 1145346, Feb. 2023, doi: 10.1155/2023/1145346.

[9] Y. Sun, M. Liu, and M. Q.-H. Meng, “Motion removal for reliable RGB-D SLAM in dynamic environments,” Robotics and Autonomous Systems, vol. 108, pp. 115–128, Oct. 2018, doi: 10.1016/j.robot.2018.07.002.

[10] W. Jianqing, S. Xiuguang,“Overview of the Development of Synchronous”,Qilu School of Transportation, Shandong University, vol. 51, no. 05, pp. 16–31, Sept. 2021.