Research on Laser-Visual Fusion-based Simultaneous Localization and Mapping

Jianjie Zhenga, Haitao Zhang*, Kai Tang, Weidi Kong

College of Defense Engineering, PLA Army Engineering University, Nanjing 210007, China
a14291010@bjtu.edu.cn
907479323@qq.com

Abstract: In consideration of the restrictions of the single-sensor SLAM system in different application scenarios, this paper proposes a laser-visual sensor fusion strategy on the basis of analyzing the strengths and weaknesses of laser SLAM and visual SLAM. In the local mapping stage, the laser stability is used to facilitate the visual SLAM system while in the global map generation stage, the point cloud map with rich texture information as constructed by the visual SLAM system is used to improve the missing map information of the laser SLAM system and finally obtain a globally consistent 2D grid map. Through indoor real scene experiments, the fused SLAM system is verified so as to improve the accuracy of localization and mapping and yield a favorable robustness by combining the strengths of both effectively.

1. Introduction

Simultaneous Localization and Mapping (SLAM) refers to a key technology by which a robot determines its own location and constructs a map of the surrounding environment incrementally when moving in a strange environment through the sensor it carries [1]. The concept of SLAM was first proposed by Smith and Cheesman et al. [2][3], and the probability estimation method was also applied to robot localization and mapping at the same time. Nowadays SLAM technology has not only become one of the core technologies of robot intelligence and autonomization, but also boasts an extensive application prospect in such fields as 3D mapping, augmented reality, virtual reality (VR) and automatic driving. In particular, it has a huge potential of applying to the military field.

Based on the current sensor varieties, there are mainly laser SLAM and visual SLAM, both of which have their defects or weaknesses certainly due to the limited characteristics of sensors, regardless of the remarkable progress made with the efforts of many researchers [4]. Therefore, this paper proposes a laser-visual fusion scheme, which combines the strengths of both to complement their functions and enhance mapping accuracy and system robustness.

2. Performance Analysis of SLAM System Framework and Mainstream Algorithm

2.1. SLAM System Framework

With the rapid development of SLAM technology, a relatively mature system framework has been built at present, as shown in Fig.1. The framework mainly includes the following five parts:
Sensor data acquisition mainly includes reading and preprocessing the data and information for camera or laser, as well as the acquiring sensor information such as inertial navigation and code disk in the robot system.

The front-end odometer undertakes the main task to estimate the sensor position and motion and to build a local map.

The back-end optimization is aimed to build a globally consistent trajectory and map through optimizing the sensor positions estimated by the front-end odometer at different times and the loopback detection results.

Loopback detection refers to the process in which the robot judges whether it has returned to the place that it ever reached in course of moving. If loopback is detected, the result will be passed back to the back-end for global optimization.

Mapping means building the globally consistent map according to the task requirements.

![SLAM System Framework](image)

**Fig.1 SLAM System Framework**

### 2.2. Mainstream Laser SLAM Algorithm

The mainstream 2D laser SLAM algorithms available at present include "Hector-SLAM" [5], "KartoSLAM" [6] and "Gmapping" [7].

Hector-SLAM tries to solve the scanning matching problem by using Gauss-Newton's method. This algorithm does not need odometer information and less relies on sensors other than laser, making it feasible to map with aerial UAV and robots with uneven road conditions. Therefore, it can be applied in more scenarios. However, this algorithm poses very high requirements on the performance of lidar which must have a high update frequency and low measurement noise. In addition, the Hector-SLAM algorithm does not have a closed-loop detection function. All of the factors above restrict the robot from moving too fast; otherwise, the mapping effect will be dissatisfactory.

KartoSLAM means representing map using graph optimization method in the form of the mean value of graph. This algorithm takes into account the system sparsity and optimizes the accumulated error by using the nonlinear least square method, which certainly replaces the filter-based laser SLAM scheme. However, it is time consuming as a submap must be established before using local submap matching.

As the most widely used 2D laser SLAM algorithm, Gmapping deals with optimization with RBPF particle filter to obtain the optimal particles. By introducing adaptive resampling technology, this algorithm is able to provide a selective resampling operation which could relieve the problem of particle dissipation very significantly. Plenty of studies show that Gmapping yields a good effect in the long corridors or the scenarios with less feature information. However, it is also disadvantageous as it is strongly dependent on odometer information, making it inapplicable to the scenarios where odometer cannot be used, such as UAV.

Through the research and analysis of the above three mainstream 2D laser SLAM algorithms, "Gmapping" algorithm proves to be favorable in performance and accuracy. Therefore, this algorithm is used in this paper for research and is fused with the visual SLAM algorithm.

### 2.3. Mainstream Laser SLAM Algorithm

The mainstream visual SLAM algorithms at present include LSD-SLAM algorithm, ORB-SLAM algorithm, and ORB-SLAM2 algorithm.
First proposed by Jakob Engel in 2014 [8], LSD-SLAM algorithm is the most mature monocular visual SLAM algorithm based on direct method and has been widely applied in robotics. It is able to apply direct method to semi-dense monocular SLAM directly, making it easy to realize semi-dense and large-scale map construction on CPU. However, the algorithm is sensitive to camera internal parameters and exposure, and in particular, loss easily happens when the camera is moving quickly. The system robustness is also not as high as expected.

ORB-SLAM algorithm, which was first proposed by Raul MurArtal in 2015, is one of the monocular SLAM systems with the best performance at present and uses ORB feature point detection method. It is superior in terms of accurate localization, good stability and excellent compatibility in a static environment, and supports multiple modes such as monocular, binocular, RGB-D, etc. At the same time, it adds the loopback detection thread of the SLAM back-end, which could solve such problems as error accumulation effectively. It also has some weaknesses, such as low running speed ascribed to the sparse constructed point cloud map and possibility of losing the dynamic objects that are followed up due to the limited feature point extraction speed.

Based on ORB-SLAM, the ORB-SLAM2 has been greatly improved and is actually a complete set of SLAM scheme based on monocular, binocular and RGB-D camera with such functions as re-localization and loopback detection. The beam adjustment optimization method is designed at the back end, contributing to high trajectory accuracy. In addition, ORB-SLAM2 includes a lightweight localization mode, with which, the system is able to follow up and conduct feature points matching for unmapped areas using visual odometers once zero drift occurs.

It was found through studying and analyzing the three mainstream visual SLAM algorithms above that the ORB-SLAM2 algorithm is high in performance and accuracy and strong in compatibility and applicability. Therefore, this algorithm is used in this paper for research and fused with the laser SLAM algorithm.

3. Laser-Visual Fusion SLAM Strategy

3.1. Lidar-based Map Construction

The Gmapping laser SLAM system is adopted in this paper and is realized by using the RBPF-SLAM (Rao-Blackwellized Particle Filter) algorithm. With non-parametric Monte Carlo, it is able to replace the Bayesian method to solve the problem of probability estimation in large and complex environments [9]. The problem to be settled by SALM is to find the joint distribution of position and map by controlling data and observation data . The Rao-Blackwellized transformation based on SLAM problem can be expressed as:

\[
P(x_t, m | z_{1:T}, u_{1:T-1}) = P(m | x_t, z_{1:T}) \cdot P(x_t | z_{1:T}, u_{1:T-1})
\]

Formula (1) describes the core idea of Rao-Blackwellized decomposition. is split using conditional probability to first solve the position and then build the map. RBPF uses particle filtering to estimate the position. The algorithm consists of four steps: sampling-calculating weight-resampling-map estimation.

Step 1: Sampling.

\[
x_t^{(i)} = P(x_t | x_{t-1}^j)
\]

Step 2: Calculating weight.

\[
w(x_t, z_{1:T}, u_{1:T-1}) = \frac{P(x_t | z_{1:T}, u_{1:T-1})}{\pi(x_t | z_{1:T}, u_{1:T-1})}
\]

Step 3: Resampling. Retain high-weight particles by quitting the low-weight ones. The posterior probability of the map can be obtained from the posterior distribution and posterior probability density of the particles and is expressed as:
\[
P_N = (m \mid z_{t_2}, u_{t_2}) = \sum_{i=1}^{N} \nu_{t_i}^{(i)} p(m \mid z_{t_2}, u_{t_2-1}, x_{t_2}^{(i)})
\]  

\text{(4)}

Step 4: Map estimation. Use Kalman filter to calculate the above formula.

On the basis of the RBPF algorithm, the system constructs the environment map by referencing the information of lidar and odometer, and estimates its own state. The algorithm has a stable performance which is not affected by the factors such as light in the environment, and has excellent anti-interference. However, it is able to obtain single scene information alone and unable to scan and identify the obstacles below the laser scanning plane, resulting in the lack of information on the constructed map [10]. As shown in Fig.2, obstacle 1 cannot be identified.

3.2. Visual-based Point Cloud Depth Information Acquisition

The RGB-D camera obtains the depth information of each spatial point through an active measurement method. This paper uses the D435i depth camera of the Microsoft Intel RealSense R200 series. The measurement principle of D435i is the infrared structured light method. As shown in Fig.3, by sending infrared light to the object to be measured, the distance between the camera itself and the object is calculated according to the returned structured light pattern received by the camera.

![Fig.3 Schematic Diagram of Infrared Mechanism Light Method](image)

It was concluded through research that the visual SLAM system based on RGB-D camera can construct a 3D point cloud map with rich texture, but the visual SLAM largely depends on the feature information in the environment. Therefore, in such environment with low-texture and low-light, matching easily fails, resulting in a lower accuracy. As shown in Fig.4, when encountering a smooth wall, matching fails and tracking is lost, which finally lowers map accuracy.
3.3. Process and Steps of Laser-visual Fusion

It can be found through the research above that it is difficult to build a high-precision map with a single sensor alone. Therefore, the laser and visual sensors are fused to improve the accuracy of robot localization and mapping through combining and supplementing the strengths of both. The algorithm fusion process is shown in Fig.5:

In the local mapping stage, laser stability is used to assist vision. When the visual tracking fails, the position information obtained by the laser at this time will be transmitted to the visual SLAM system, for the purpose of secondary initialization until the visual system tracking is successful.

The point cloud map with rich texture information constructed by visual SLAM is used in the global map generation stage to improve the map information lost by the laser, and finally obtain a globally consistent 2D grid map for autonomous localization and navigation of robot. The main steps are as follows:

Step 1: Convert the point cloud depth coordinate \( p(u,v,z) \) obtained by RGB-D into a coordinate \( P(x,y,z) \) in the camera coordinate system.

Step 2: Take the plane formed by the z-axis and x-axis of the camera coordinate system as the point cloud data projection plane, project the coordinate point \( P \) onto the plane, and represent it with the distance \( d \) between the projection point and the optical center, and the angle \( \theta \) between the projection line and the z-axis [11].

\[
d = \sqrt{x^2 + z^2} \quad (5)
\]

\[
\theta = \arctan\left(\frac{z}{x}\right) \quad (6)
\]

Step 3: Fuse the local 2D grid map formed by the projection and the local 2D grid map obtained by the lidar to improve the map information lost by the laser so as to finally obtain a global 2D grid map.
4. Experiment Construction and Result Analysis

4.1. Experimental Platform Construction

The core processor of the experimental platform is Intel-NUC-8i5BEK, which comprises the 8th generation of Intel Core i5-8259U processor, an 8G memory, an Intel Iris Plus graphics card 655, 2 HDMI interfaces, 2 front USB3.1 interfaces, 2 rear USB3.1 interfaces, 2 internal USB2.0 ports, 64G internal memory as well as the Intel dual-band Wireless-AC 9560 wireless network card. The lidar is RPLIDAR-A1 (improved version), which is a lidar fitted with triangular ranging technology, with the ranging scope of 0.15~12 m, the scanning angle of 0-360 degrees, and the measuring frequency of 4,000 Hz. The RGB-D camera is D435i, which is the next generation of Intel’s RealSense depth camera D435, with a global shutter and the capacity of handling fast-moving objects. It is applicable both indoors and outdoors with a depth distance of 0.1 m ~ 10 m, and the field of view angle is 85×58 degrees. The internal IMU adds a data with six degrees of freedom and the built-in inertial measurement unit combines various linear accelerometers and gyroscopes to detect the rotation and translation of three axes. The constructed experimental platform is shown in Fig.6:

4.2. Real Scene Experiment and Analysis

This paper selects the experimental building corridor as the experimental scene, where several obstacles of different heights were set randomly, including one that is lower than the lidar scanning plane, as shown in Fig.7:

The laser SLAM system, the visual SLAM system, and the laser-visual fusion SLAM system in this paper are used to conduct three sets of experiments. In Fig.8, (a) is a 2D grid map constructed using a single sensor lidar, (b) is a point cloud map constructed with a single visual sensor, and (c) is a 2D grid map constructed in virtue of a laser-visual fusion system. It can be concluded from the results of mapping that map information will be lost when using a single sensor, and obstacles below the laser scanning plane do not appear in the grid map. The fusion scheme proposed in this paper can remove the defect and improve mapping accuracy.
5. Conclusion
In light of the different restrictions and defects of a single sensor SLAM system in different scenarios, this paper proposes a laser-visual sensor fusion strategy. In the local mapping stage, the laser stability is used to facilitate visual SLAM system while the point cloud map with rich texture information as constructed by the visual SLAM system is used in the global map generation stage to improve the map information lost by the laser, so as to finally obtain a globally consistent 2D grid map. Through indoor real scene experiments, the fused SLAM system is further verified so as to improve the accuracy of localization and mapping, and yield good system robustness by combining the strengths of both effectively.

Acknowledgments
The authors would like to thank the anonymous reviewers for their valuable comments and helpful suggestions. This work was supported by National Natural Science Foundation of China (No.61602507).

References
[1] Du Chen. Research Review of Simultaneous Localization and Map Construction Technology Based on Mobile Robots [C]. Proceedings of the 22nd Annual Conference of New Network Technologies and Applications in 2018 of the Network Application Branch of China Computer Users Association, 2018:212-215+227(in Chinese).
[2] Randall C. Smith. On the Representation and Estimation of Spatial Uncertainty[J]. The International Journal of Robotics Research, 1986, 5(4):56-68.
[3] Lin Zhilin, Zhang Guoliang, Wang Feng, Yao Erliang, Jia Xiao. A VSLAM-based Indoor Navigation Map Preparation Method [J]. Electro-Optics and Control, 2018, 25(01): 98-103(in Chinese).
[4] Kang Lu. Research on Vision/Laser-based Mobile Robot Localization and Mapping [D]. Nanjing: Southeast University, 2019(in Chinese).
[5] Kohlbrecher S, Stryk O V, Meyer J, et al. A flexible and scalable SLAM system with full 3D motion estimation[A]. Kyoto: IEEE, 2011:155-160.
[6] Grisetti, Giorgio; Stachniss, Cyrill; Burgard, Wolfram. Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters[J]. IEEE Transactions on Robotics, 2007, 23(1): 34-46.
[7] Wolfgang Hess; Damon Kohler; Holger Rapp; Daniel Andor. Real-time loop closure in 2D LIDAR SLAM[A]. Robotics and Automation (ICRA), 2016 IEEE International Conference on[C], 2016:1271-1278.
[8] Jakob Engel; Thomas Schps; Daniel Cremers. LSD-SLAM: Large-Scale Direct Monocular SLAM[J]. Lecture Notes in Computer Science, 2014, Vol.8690(1): 834-849.
[9] Zhang Yanan, Sun Fengcai, Shi Xuhua. An Improved RBPF Laser SLAM Algorithm [J]. Wireless Communication Technology, 2017, 26(04): 16-20+25.
[10] Wang Guangting, Cao Kai, Liu Hao. SLAM Method Based on Lidar and Visual Information Fusion [J]. Journal of Shandong University of Technology (Natural Science Edition), 2019, 33 (01): 9-13+19(in Chinese).
[11] Li Yangyu. Research on SLAM of Mobile Robot Vision Based on Depth Sensor [D]. Chongqing: Chongqing University, 2017(in Chinese).