SLAM closed-loop detection method of 2D laser fusion vision in a larger space

Yong Cai Tong, Wei Cui* and Hua Ye Fu
School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun, 130022, China
2019100597@mails.cust.edu.cn

Abstract. In the process of implementing SLAM (Simultaneous-Localization-and-Mapping) for mobile robots, using a separate 2D lidar to build a map in a larger space will lead to inaccurate mapping and missed closed-loop detection. Therefore, an improved 2D laser fusion vision closed-loop detection algorithm is proposed. The algorithm first detects the loop through the visual information obtained by the camera, compares the similarity between the current frame and the key frame, and continuously selects a series of closed-loop frames to be selected near the current frame, and then filters the final closed-loop frame through geometric verification. More matching points can be obtained through transformation matrix projection, and the matching points can optimize the transformation matrix. If the current frame and the closed-loop candidate frame get enough common-view space points, then the closed-loop detection is determined to be successful. Finally, the visual loop information is fed back to the laser, and the robot's pose is adjusted to achieve global optimization. In the experimental environment, it is verified that the laser fusion vision effectively solves the problems such as the lack of robustness of the single laser closed loop detection.

1. Introduction
Closed-loop detection is also called loop-back detection. In the SLAM process, it judges whether to loop-back by matching the current position information with the constructed map. Successful loopback detection can largely avoid accumulated errors and improve the accuracy of mapping. Adding loop detection to the SLAM system is a necessary link to reduce the pose drift and build a complete map. On the laser loop, Renaud et al. [1] proposed a matching and closed-loop detection algorithm through the source point cloud segment and the target point cloud segment, but the laser real-time matching speed is only 1Hz, and the scanning matching speed is too low to meet the actual work of mobile robots. The cartographer algorithm proposed by Hess et al. [2] searches for loop frames through the branch and bound method, and adds them to pose optimization as loop detection constraints. However, this algorithm has relatively high requirements for the initial frame pose. If the initial value error is large, the composition will have a large deviation. In the visual loop, Gao Xiang [3] of Tsinghua University and others proposed a multi-layer neural network deep learning method for closed-loop detection in the simultaneous positioning and map construction system based on visual information. The bag-of-words method can learn more complex information in the image than the deep learning method. ORN-SLAM2 [4] includes three threads of tracking, local composition, and loop detection. It optimizes the sparse sub-map, which can improve the positioning accuracy and real-time performance of the system.

Although laser SLAM has better mapping and positioning effects than VSLAM, because the lidar sensor has less information, it cannot quickly determine the closed-loop point, which further causes the
environment map to be deformed and incomplete. The visual image contains a wealth of visual information. Although its positioning accuracy is lower than that of laser SLAM, it has inherent advantages in closed-loop detection and can quickly determine the closed-loop point. Therefore, this paper proposes an improved 2D laser fusion vision closed-loop detection algorithm, which can effectively reduce the cumulative error in the SLAM process and improve the accuracy of mapping.

2. Algorithm theory and foundation

2.1. Transformation relationship between sensor data and world coordinates

The world coordinate system is usually used to describe the movement and positioning of mobile robots. The mobile robot mainly adopts the robot coordinate system and the sensor coordinate system when working. The working process is: the environmental information collected by the external sensor in the sensor coordinate system is converted into the robot coordinate system for calculation, and finally converted to the world coordinate system. Representation output. Assuming that the coordinates of the lidar in the world coordinate system are \((x, y, \theta)\), the conversion matrix from the robot coordinate system to the world coordinate system is:

\[
T_{WR} = \begin{bmatrix}
\cos \theta & -\sin \theta & x \\
\sin \theta & \cos \theta & y \\
0 & 0 & 1
\end{bmatrix}
\]  

2.2. Related data association algorithm

In the data association problem, ICP Iterative Closet Point (ICP) is a common method to solve the matching problem, and it is a typical matching method between point sets. Since the original data is used, the matching between points has better robustness.

2.3. Multi-information joint error function

In order to further optimize the pose of the robot, a multi-information joint error function is required, which also considers the error of lidar data matching and the error of image matching.

The reprojection error function between two matched images is:

\[
e_i = \frac{1}{2} \sum_{j=1}^{m} \left| \left| u_j - \frac{1}{d_j} K (R P_j + T) \right| \right|_2
\]

Among them, \(m\) is the number of pairs of feature points on the matching, \(K\) is the parameter of the camera, \(u_j\) is the position of the feature point, \(P_j\) is the corresponding position of the point matched by the feature point, and \(d_j\) is the depth value of the feature point.

The multi-information joint error function can be expressed in the following form:

\[
e = \beta \sum_{j=1}^{m} \left| \left| u_j - \frac{1}{d_j} K (R P_j + T) \right| \right|_2 + (1 - \beta) \sum_{i=1}^{n} \left| \left| P_j - (R Q_i + T) \right| \right|_2
\]

\(\beta\) is the balance coefficient, the value range is (0,1), it mainly determines the degree of dependence of the two observation methods of laser and vision. If the robot is working in a long corridor and other environments where laser matching is unstable but full of visual features, a larger value can be used; In order to facilitate subsequent optimization calculations, the robot's pose conversion relationship \((R, T)\) is expressed in the form of the following Lie algebra:

\[
RP + T = \exp(\hat{\xi})P
\]

Therefore, the error function can be rewritten as the following Lie algebra form, which is a function of the independent variable:
For the poses of M robots, the total error function can be written as:

$$E = \sum_{k=1}^{M} e_k = \sum_{k=1}^{M} F_k(\xi_k)$$  \hspace{1cm} (6)$$

By minimizing this error, the optimal robot pose sequence can be obtained.

3. Laser Fusion Vision Closed Loop Detection

3.1. Current status of laser loop

Due to the small amount of information of the lidar sensor, it takes too long to complete an effective closed-loop detection. The matching requires a large rotation and translation distance between the two frames, and the 2D laser loop can only be relied on, so the 2D laser closed-loop detection is easy to fail at times, which will further cause the environmental map constructed by it to appear deformed.

3.2. The overall framework of the algorithm

The overall algorithm framework of 2D laser fusion vision closed-loop detection is shown in Figure 1:

Laser data and image data are input to the system at the same time. The laser data adopts the frame-image matching method to obtain the pose estimation for back-end optimization; the image data is extracted by feature and used to construct the image feature dictionary for loop detection and loop detection in, when the visual SLAM system detects the loop, the loop information is transmitted to the laser, and the laser adds loop constraints to the pose map, and then uses the pose map optimization theory to optimize the robot's pose and complete the construction of the 2D environment map.

4. Experiment analysis

This article completes the experiment on the mobile robot platform shown in Figure 2, which is equipped with 2D lidar and RGB-D camera. The lidar uses single-line lidar RPLIDAR A1 with a measuring distance of 12m; the RGB-D camera uses a Microsoft XBox360 Kinect 1.0 depth camera with a measuring distance of 0.8m-4.0m. The effective angle is 57.5° horizontally and 43.5° vertically. Mini industrial computer is equipped with Linux (Ubuntu 18.04) operating system. The software environment used in this article is the open source robot operating system ROS (Robot Operating System, ROS), the version is Kinetic.
In order to verify the algorithm in this paper, it is mainly to test the accuracy of lidar single loop detection and lidar fusion visual loop detection in a 128m² teaching building lobby environment, and compare the loop pose and the current laser frame pose error.

Figure 2 Experimental hardware platform

![Experimental hardware platform](image)

Figure 3 Composition effect of 128m² experimental environment

![Composition effect](image)

Figure 3a is the actual environment where the mobile robot is located, Figure 3b is its separate laser mapping in this environment and the loopback fails, and Figure c is the laser fusion vision mapping and the loopback is successful. The robot's starting point and looping point are both in the red circle on the way.

In the experiment, the calculation of the loop frame pose and the current frame laser frame pose of the robot in the two cases of single laser loop positioning and vision fusion laser loop positioning are recorded. The relative position relationship of the two poses can be expressed through Lie group Li algebra. The smaller the error, the more accurate the positioning of the robot during loopback:

$$\Delta c_i = c_i^{-1}c_j$$

Error

$$e_i = \ln (\Delta c_i^{-1}c_i^{-1}c_j)$$

After many experimental verifications, due to space issues, five experiments that can represent the overall experimental effect are selected to solve the calculation of the relative position relationship between the looped laser frame pose and the current laser frame pose. As shown in Table 1 (“n” in the table means that the robot cannot detect the loop).
Table 1  Relative error of initial pose and looping pose of 128m² environmental robot

| Laser | Laser fusion vision |
|-------|---------------------|
| Looping pose | Optimize pose | Error | Looping pose | Optimize pose | Error |
| $c_i(x, y, \theta)$ | $c_i(x, y, \theta)$ | $(10^{-16})$ | $c_i(x, y, \theta)$ | $c_i(x, y, \theta)$ | $(10^{-16})$ |
| 0 | (3.31, 0.20, -0.12) | (3.34, 0.32, 0.12) | (-3.86, -0.38, -0.10) | (0.01, 0.01, 0.01) | (-0.02, 0.03, 0.03) | (-0.12, 0.0, 0.0) |
| 1 | (0.00, 0.00, 0.00) | $\alpha$ | $\alpha$ | (0.58, -0.02, -0.04) | (0.74, 0.06, -0.01) | (-1.10, 0, 0) |
| 2 | (0.6, 0.6, 0.02) | (0.02, 0.16, 0.18) | (-0.12, 0.20, 0.26) | (3.24, -0.42, -1.82) | (3.22, -0.48, -1.62) | (1.10, 3.34, 0.01) |
| 3 | (0.48, 0.19, 0.18) | $\alpha$ | $\alpha$ | (2.56, -0.38, -1.98) | (2.48, -0.49, -2.34) | (-2.42, 1.14x10^{-6}, 1.12) |
| 4 | (0.86, 0.20, 0.12) | (1.38, 0.36, 0.26) | (-0.32, -0.12, 0) | (0.02, 0.07, 0.07) | (0.34, 0.26, -0.16) |

The loop detection rate of the single laser SLAM composition is 60%, and the loop accuracy rate of the visual fusion laser composition is 100%. The relative pose errors of the single laser SLAM composition and the visual fusion laser SLAM composition are both 10-16 orders of magnitude, and there is not much difference in the pose errors between the two. However, the loop-back missed detection of the single laser SLAM in this environment composition causes a relatively large drift error in the map construction. Therefore, experiments can verify that the proposed closed-loop detection method of laser fusion vision can effectively improve the success rate of closed-loop detection.

5. Concluding remarks

In this paper, laser fusion visual inspection closed loop is used. Through appearance verification, all common view key frames of the current frame and the word bag model BOW score of the current frame are selected to select the closed-loop candidate frame of the current frame; for geometric verification, the current frame and the closed-loop candidate frame are calculated the transformation relationship and the transformation matrix reprojection to find the common view space matching points, complete the closed-loop detection and optimize the robot pose, and verify it in the actual environment. The experimental results show that, compared with the single laser loop detection algorithm, the algorithm in this paper solves the problem of a single 2D laser in an environment lacking obvious characteristics and a large cumulative error of the odometer, which leads to loop detection and poor robustness, and improves the robot loop. Optimized positioning accuracy. With the development of vision and laser multi-sensor fusion, how to optimize the laser composition and improve the accuracy of laser composition will be the next research problem.

References

[1] Dube R, Dugas D, Stumm E, et al. Seg Match: Segment based place recognition in 3D point clouds[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ: IEEE Press, 2017: 5266-5272.
[2] 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). Piscataway, NJ:IEEE. Press, 2016: 1271-1278.
[3] Gao X, Zhang T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system[J]. Autonomous Robots, 2017, 41(1):1-18.
[4] Mur-Artal R, Tardos J D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras[J]. IEEE Transactions on Robotics, 2017, 33 (5): 1255-1262.
[5] Sun Longlong. Research on closed-loop detection of mobile robot visual SLAM [D]. An-Hui Engineering University, 2020.
[6] Hou Xueyong. Road feasible area extraction for mobile robot based on two-dimensional laser ranging [D]. Hangzhou University of Electronic Science and Technology, 2012.
[7] Olson E B. Real-time correlative scan matching[C]/ /2009 IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE Press, 2009: 4387 -4393.

[8] Kummerle R, Grisetti G, Strasdat Het al. G2o: A general framework for graph optimization[C]/ /2011 IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE Press, 2011: 3607 – 3613.