An Overview of SLAM

Sufang Wang, Zheng Wu and Weicun Zhang

Abstract Simultaneous Localization and Mapping (SLAM) based on LIDAR and Visual SLAM (VSLAM) are key technologies for mobile robot navigation. In this paper, the SLAM algorithm based on these two types of sensors is described, and their advantages and disadvantages are comprehensively analyzed and compared. In order to better achieve active navigation and positioning, path planning and obstacle avoidance, the advantages of both should be brought into full play. In the end, the future development direction of mobile robot is discussed.

Keywords SLAM · Mobile robot · Navigation · Sensor

1 Introduction

For autonomous mobile robots, there are three aspects involved in navigation and positioning: (1) where am I? (2) where am I going? (3) How should I get there? [1]. These three issues correspond to three aspects respectively: (1) positioning, get the robot's current position. (2) Perception, get target location. (3) Path planning. SLAM stated that in an unknown environment, the mobile robot can perceive the surrounding environment according to its own sensors during the movement so as to perform autonomous positioning and incrementally construct an environmental map [2, 3].

S. Wang · W. Zhang (✉)
School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
e-mail: weicunzhang@263.net

S. Wang
e-mail: s20170610@xs.ustb.edu.cn

Z. Wu
Youer (Beijing) Robot Technology Ltd., Beijing 100081, China
e-mail: wuzheng@youerobot.com

Y. Jia et al. (eds.), *Proceedings of 2018 Chinese Intelligent Systems Conference*,
Lecture Notes in Electrical Engineering 528,
https://doi.org/10.1007/978-981-13-2288-4_64

Although the research on autonomous mobile robots has become a hot topic in the field of high-tech, we have seen and applied very few in our lives. Most of them are based on theoretical research. According to published papers at home and abroad, most of them are based on a single sensor, such as LIDAR SLAM and visual SLAM. SLAM using LIDAR has mature algorithms and solutions. With the follow-up of hardware devices, visual SLAM has also developed. In order to meet the requirements as much as possible, different products use different sensors.

The SLAM technology of mobile robot has very important theoretical significance and application value. For driverless cars, SLAM can build 3D environment models and position navigation by LIDAR. In terms of military, SLAM can allow mobile robots to reach many harsh environments that humans cannot reach. It can help realize intelligent reconnaissance and combat of robots. It can also be used to search for and remove hazardous explosives [4]. In life, SLAM can achieve household robots to walk autonomously and successfully avoid obstacles to complete high-quality tasks.

2 SLAM Classification

Intelligent robots gradually enter our lives. It makes our lives easier and more convenient. However, simple robots cannot walk autonomously. At this time, the auxiliary role of the LIDAR is needed to realize the robot's intelligence. It can acquire the information of the robot's environment in real time. However, robot cannot understand the information scanned by LIDAR. The powerful SLAM navigation algorithm is depended to achieve the intelligent walking of the robot. Common LIDAR, such as SICK, Velodyne and rplidar, can be used for SLAM.

VSLAM is mainly implemented with cameras. According to different working methods, VSLAM is classified into Monocular, Stereo Vision and RGB-D.

Using only one camera for SLAM is called monocular SLAM. Using multiple cameras as sensors is called stereo vision SLAM, the most widely used is the stereo camera. The combination of a monocular camera and an infrared sensor to form a sensor is called RGB-D SLAM.

3 LIDAR-SLAM

The key issue in SLAM is positioning [5]. The methods to solve positioning problems are divided into probabilistic and non-probabilistic method. The method based on probability is the mainstream method, and the method based on Bayesian estimation is the basis of probability estimation method. There are mainly Kalman filter (KF) method and particle filter (PF) method.

(1) Kalman Filter Method

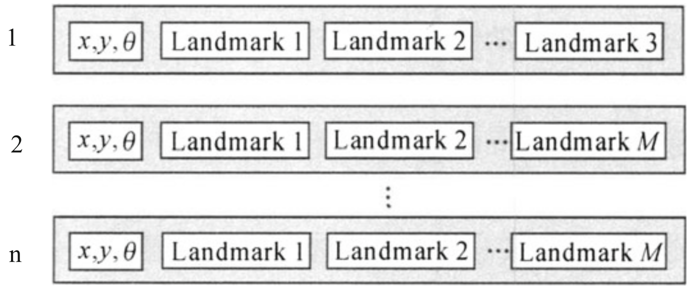

Fig. 1 RBPF-SLAM algorithm

In the case that the state noise and observation noise are gaussian distributions, data is observed by system input and output. Then KF makes the optimal estimate of the system state [6–9]. The premise of using KF is that the system is a linear system. However, the actual system is often a nonlinear system. Extended Kalman Filter (EKF) linearizes the nonlinear system through first-order Taylor expansion [10]. Unscented Kalman Filter(UKF) is a method of approximating nonlinear distribution using a sampling strategy [11]. It does not need to calculate the Jacobian, and has a higher linearization accuracy, and its performance is better than that of EKF.

KF-SLAM occupies a dominant position in many solutions, because it has greater advantages in convergence and implementation complexity. But it has the problem of lack of self-closed loop capability and associated fragility. When an error occurs in the data association, it will eventually be brought into the entire SLAM state estimation, sometimes even causing the entire prediction process to diverge. Therefore, a robust data association method is very important.

(2) Particle Filter Method

As a new type of filter, the particle filter can get rid of the linear assumption of the system and the constraints of the Gaussian noise assumption of the sensor. It can approximate any probability distribution and the calculation is simple and convenient. It can effectively solve the robot positioning problem [12]. Murphy [13], Murphy and Russell [14] found that if a robot's motion trajectory is known, then the probability between landmark positions is conditionally independent. Therefore, Rao-Blackwellised decomposition is proposed and implemented, which provides a theoretical basis for the particle filter to solve the SLAM problem. Based on this, Montemerlo et al. [15] demonstrated the feasibility of using Rao-Blackwellised Particle Filter (RBPF) to solve the SLAM problem, and proposed the FastSLAM algorithm.

The FastSLAM uses a improved particle filter to estimate the posterior distribution of the robot path. In Fig. 1 [16], each particle maintains a state estimate and a set of individual feature location information. Each particle represents a path traveled by a robot. Each feature is individually estimated using an EKF. Each particle maintains M EKFs and there are a total of N particles to predict the robot state [17].

Compared with other methods, the FastSLAM reduces the sampling space, thus greatly reducing the complexity and improving the calculation speed; it has high precision and better robustness; it can be applied to non-Gaussian, nonlinear, unknown posterior density function. However, it was found in the experiments that it requires more particles to avoid estimation divergence, and that divergence is closely related to data association. In addition, as a new algorithm for solving SLAM problem, particle filters also have many areas for improvement and optimization.

4 Visual SLAM Algorithm

Visual SLAM is mainly divided into visual front-end and optimized back-end. The front-end is also called visual odometry (VO). It estimates rough camera movement based on the information of adjacent images and provides a good initial value for the back-end. The implementation method of VO is divided into feature point method and direct method according to whether or not features need to be extracted. Feature-based method is stable in operation and insensitive to light and dynamic objects. So it is considered as the mainstream method of VO [18].

(1) Feature-based method

Davision et al. [19] first proposed a monocular visual SLAM (MonoSLAM) system that uses EKF as the back-end to track sparse feature points on the front-end. After the keyframe-based monocular visual SLAM gradually developed [20–23]. The most representative of these is parallel tracking and mapping (PTAM). Klein and Murray [21] proposed a simple and effective method for extracting keyframes, parallelizing the tracking and mapping. And for the first time PTAM used nonlinear optimization as the back-end. Mur-Artal et al. [23] inherited and improved PTAM and innovatively proposed the three thread to implement the monocular visual SLAM system based on PTAM of the dual thread. The entire system is implemented around the ORB feature, namely ORB-SLAM. Subsequent studies have shown that the ORB-SLAM system is equally applicable to monocular, stereo, and RGB-D models, and has good general-purpose use.

Because a monocular camera captures a two-dimensional projection of a three-dimensional space object, which is a single image, and there is uncertainty in the motion and trajectory estimated by moving the camera, the true depth of the object cannot be determined. Therefore, there are enormous difficulties in the three-dimensional reconstruction work. At this time, stereo cameras and RGB-D cameras appear, but stereo cameras have a major problem in terms of calculations under the existing conditions. However, SLAM based on RGB-D data simplifies the complexity of 3D reconstruction. Henry et al. [24] first proposed method of 3D reconstruction of indoor environment using RGB-D camera, extracting the SIFT features in the color image and finding the corresponding depth information on the depth image. Then Random Sample Consensus (RANSAC) method is used to match the 3D feature points and the corresponding rigid motion transformation is calculated. RANSAC is applied to many cases with incorrect data, which can handle data with error matching and use it as the initial value of the iterative closest point (ICP) to find more accurate position and pose. RGB-D SLAM usually uses ICP algorithm to estimate pose and optimize the camera's motion transfer matrix. The basic process of ICP algorithm in reference [25] is as follows:

Step 1: Read the point sets P1, P2.
Step 2: Select the pair of points. Search P2 for the closest point to P1 to form a point pair. Find out all the pairs of points in the two point sets.
Step 3: Calculate two barycentric coordinates based on the pairs of points in the two point sets.
Step 4: From the new point set, calculate the rotation matrix R and the translation matrix t.
Step 5: From the obtained R and t, calculate the new point set $P2'$ after the rigid transformation of the point set P2.
Step 6: If the absolute value of the difference between the sum of squares of the distances from P2 to $P2'$ is less than the threshold value for two consecutive times, then it is converged and the iteration is stopped. Otherwise, the steps 1–6 are repeated until convergence.

Among many matching algorithms for depth images, such as Generic Algorithm, ANSAC, and ICP. ICP is the most widely used one. It does three-dimensional data processing directly to depth images, and does not need to assume and divide of object features. After selecting the initial value, the algorithm has good convergence, thus getting the global optimal value and obtaining more accurate matching results. So it quickly becomes a mainstream algorithm for depth image matching [26].

Although feature-based SLAM method has many advantages, the extraction of key points and the calculation of feature points are very time-consuming, and some useful image information may be discarded because only feature points are used. Such as a white wall, where there are few feature points. We will not be able to accurately calculate the motion of the camera.

(2) Direct method

According to the classification of the number of pixels used, direct method can be divided into sparse, dense and semi-dense. Assume that P is a spatial point of a known location. When P is derived from a sparse key point, it is called a sparse direct method. When P comes from a portion of pixels, it is called semi-dense direct method. If all pixels are used, it is called dense direct method, which can build a complete map.

Many people have been devoted to the study of direct method [27–33]. Irani and Anandan [27] gave a detailed and in-depth description of direct method. Silveira et al. [28] applied direct method to the visual SLAM and described main advantages and limitations. Subsequently, a sparse-based semi-direct monocular visual odometry (SVO) was proposed [29]. This algorithm has high accuracy, good robustness, and is faster than the most advanced methods currently available. This sparse method eliminates the needs of feature extraction and robust matching techniques for motion estimation, and is suitable for estimating the state of microcars in GPS denied environments. Next, Engel et al. [30] proposed a large-scale direct monocular SLAM (LSD-SLAM). Compared to other direct methods, it reconstructs the keyframe's pose map and the environment's semi-dense and highly accurate three-dimensional map in real time. Usenko et al. [33] proposed a novel direct visual-inertial odometry method for stereo cameras that use the complementarity of visual and inertial data to improve the accuracy of 3D reconstruction maps.

Newcombe et al. [34] integrate all the deep data and image information from Kinect into the observation scene, and reconstruct the 3D model, so as to get the global map. Particularly, it allows reconstruction of dense maps in real time, making a big step towards augmented reality (AR). Henry et al. [35] use a joint optimization algorithm to apply RGB-D cameras to the robot field in indoor environments. Kerl et al. [36] proposed a visual SLAM method based on a direct dense RGB-D camera. The error terms of this method are the photometric and the depth error. The optimal camera pose is solved using the g2o optimization library. And a keyframe selection and loop closure detection method entropy-based is proposed. Thus, the path error is greatly reduced.

Compared to feature method, direct method does not need to extract image features. It has a fast speed of execution, high robustness to the photometric error of the image, but a high requirement for the camera internal reference. When there is geometric noise, the algorithm performance decreases quickly. In the event of image motion blur, camera position can still be achieved. But direct method has poor robustness to large baseline motion.

5 Scheme Comparison

The advantages of LIDAR is that it has a wide range of visibility, and can detect the angle and distance of the obstacle points with high accuracy to achieve obstacle avoidance. But it's expensive. However, the camera has the advantages of small size, light weight, easy installation, abundant information extraction, convenient and flexible, and low price. Therefore, VSLAM has become a hot topic in the research of SLAM algorithm in recent years [37].

The performance of visual SLAM depends on the environment in which it is operating. Ideal conditions are as follows:

(1) Light is sufficient, without large changes in lighting. The camera must be able to identify features in the scene. Generally, a more complex scene, with lots of objects or geometry is best. Blank walls, floors, or ceilings are the worst cases. Many reflective surfaces, such as glass or mirrors, can cause problems. Also, direct sunlight can interfere with the depth cameras, which affects the accuracy of occupancy mapping.

(2) When the scene is mostly motionless, Visual SLAM works best. If people or objects are moving, performance will be affected more or less. If the entire scene is moving, such as in an elevator, SLAM will not work at all.
(3) When the camera motion is primarily translation, not a rotation. Visual SLAM works best. When rotation is necessary, it is best to rotate slowly.
(4) When SLAM begins, the camera must be stationary, and there must be sufficient visual features. If the camera is aimed at a blank wall, floor, or ceiling that is enough to block the camera's view, SLAM may not initialize properly.

LIDAR and camera have their own advantages and disadvantages. It is very easy for the camera to identify the same object, but it is difficult for the LIDAR. if the camera tells the LIDAR that the two frames are the same object, then it is possible to know what the speed and displacement of this object are between the two frames by the LIDAR. It can be seen that recognition and tracking are easy to achieve, resulting in more accurate maps.

6 Conclusions

After more than thirty years of research and efforts by predecessors, SLAM has made great progress. Because of the limitations of the indoor operating environment, GPS cannot be used to restrict positioning errors, and SLAM opens the door for the development of the indoor robot field. SLAM based on LIDAR has already a relatively mature scheme, but the high cost is still the primary problem. Therefore, the low-cost visual SLAM has become a research hotspot in recent years. However, no matter which sensor is used alone, there are some defects. The multi-sensor fusion technology based on LIDAR, vision sensor and inertial measurement unit [38, 39] not only can realize the cooperative operation among sensors, but also greatly enhance the robustness. It is believed that the research and application of multi-sensor fusion technology will bring wider space to driverless, robotics, augmented reality and virtual reality. In addition, SLAM is combined with deep learning to perform image processing [40], to generate semantic maps of the environment and improve the human-computer interaction techniques, so that intelligence can be better realized.

Acknowledgements This work was supported by National Natural Science Foundation of China (No. 61520106010; 61741302).

References

1. J.J. Leonard, H.F. Durrant-Whyte, I.J. Cox, Dynamic map building for an autonomous mobile robot, vol. 11, no. 4 (Sage Publications Inc., 1992), pp. 286–298
2. W. Chen, F. Zhang, Review on the achievements in simultaneous localization and map building for mobile robot. Control Theory Appl. **22**(3), 455–457 (2005)

3. M. Csorba, *Simultaneous Localization and Mapping* (University of Oxford, 1997), pp. 56–89
4. M. Montemerlo, S. Thrun, D. Koller, Fast SLAM (Simultaneous Localization And Mapping), in *Proceedings of the AAAI National Conference on Artificial Intelligence* (Menlo Park, CA, USA, 2002), pp. 593–598
5. H. Durrant-Whyte, Where am I? A tutorial on mobile vehicle localization. Ind. Robot **21**(2), 11–16 (1994)
6. S. Yavuz, Z. Kurt, M.S. Bicer, Simultaneous localization and mapping using extended Kalman filter. in *2009 IEEE 17th Signal Processing and Communications Applications Conference* (IEEE, Antalya, 2009), pp. 700–703
7. Y. Wei, Z. Zuo, Improvement of the simultaneous localization and map building algorithm applying scaled unscented transformation, in *International Conference on Industrial Mechatronics and Automation*, *2009*, *ICIMA 2009* (IEEE, Chengdu, 2009), pp. 371–374
8. J.G. Kang, W.S. Choi, S.Y. An et al., Augmented EKF based SLAM method for improving the accuracy of the feature map, in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, Taipei, 2010), pp. 3725–3731
9. D. Wang, H. Liang, T. Mei, Lidar scan matching EKF-SLAM using the differential model of vehicle motion, in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, vol. 36, no. 1 (IEEE, Gold Coast, QLD, 2013), pp. 908–912
10. S.J. Julier, J.K. Uhlmann, A counter example to the theory of simultaneous localization and map building. in *IEEE International Conference on Robotics and Automation* (Seoul, Korea: [s. n.], 2001), pp. 4238–4243
11. S.J. Julier, The spherical simplex unscented transformation, in *Proceedings of the American Control Conference* (Denver: [s. n.], 2003), pp. 2430–2434
12. M.S. Arulampalam, S. Maskell et al., A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process. **50**(2), 174–188 (2002)
13. K. Murphy, Bayesian map learning in dynamic environments. Adv. Neural Inform. Process. Syst. **12**, 1015–1021 (2008)
14. K. Murphy, S. Russell, Rao-Blackwellised particle filtering for dynamic bayesian networks, vol. 43, no. 2 (Springer New York, 2001), pp. 499–515
15. M. Montemerlo, S. Thrun, D. Koller, Fast SLAM (Simultaneous Localization And Mapping), in *Proceedings of the AAAI National Conference on Artificial Intelligence* (AAAI, Menlo Park, CA, USA, 2002), pp. 593–598
16. E. Wu, X. Zhiyu, M. Shen, J. Liu, Robot SLAM algorithm based on laser range finder for large scale environment. J. Zhejiang Univ. **41**(12), 1982–1986 (2007)
17. S. Thrun, D. Fox, W. Burgard, F. Dallaert, Robust Monte Carlo localization for mobile robots. Artif. Intell. **128**(1–2), 99–141 (2001)
18. X. Gao, T. Zhang, Y. Liu, Q. Yan, *Visual SLAM Fourteen Lectures: From Theory to Practice* (Publishing House of Electronics Industry, Beijing, 2017), pp. 132–204
19. A.J. Davison, I.D. Reid, N.D. Molton, O. Stasse, Monoslam: real-time single camera SLAM. IEEE Trans. Pattern Anal. Mach. Intell. **29**(6), 1052–1067 (2007)
20. M. Quan, S. Piao, L. Guo, An overview of visual SLAM. CAAI Trans. Intell. Syst. **11**(6), 768–776 (2016)
21. G. Klein, D. Murray, Parallel tracking and mapping for small AR workspaces, in *IEEE and ACM Intemational Symposium on Mixed and Augmented Reality* (Nara, Japan, 2007), pp. 225–234
22. R. Mur-Artal, J.D. Tardós, Fast relocalisation and loop closing in keyframe-based SLAM, in *IEEE International Conference on Robotics and Automation* (New Orleans, LA, 2014), pp. 846–853
23. R. Mur-Artal, J.M.M. Montiel, J.D. Tardós, ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. Rob. **31**(5), 1147–1163 (2015)
24. P. Henry, M. Krainin, E. Herbst et al., RGB-D mapping: using depth cameras for dense 3D modeling of indoor environments, vol. 31, no. 5 (Springer Berlin Heidelberg, 2014), pp. 647–663
25. K. Zhu, H. Liu, Q. Xia, Survey on monocular visual SLAM algorithms. Appl. Res. Comput. **35**(1), 1–6 (2018)

26. L. Shifei, P. Wang, S. Zhenkang, A survey of iterative closest point algorithm. Sig. Process. **25**(10), 1582–1588 (2009)
27. M. Irani, P. Anandan, About direct methods, in *International Workshop on Vision Algorithms*: *Theory & Practice* (1999), pp. 267–277
28. G. Silveira, E. Malis, P. Rives, An efficient direct approach to visual slam. IEEE Trans. Rob. (IEEE Press) **24**(5), 969–979 (2008)
29. C. Forster, M. Pizzoli, D. Scaramuzza, SVO: fast semi-direct monocular visual odometry. in *IEEE International Conference on Robotics and Automation (ICRA)* (2014), pp. 15–22
30. J. Engel, T. Schöps, D. Cremers, LSD-SLAM: large-scale direct monocular SLAM, in *Computer Vision-ECCV 2014* (Springer International Publishing, 2014), pp. 834–849
31. J. Engel, J. Sturm, D. Cremers, Semi-dense visual odometry for a monocular camera, in *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 1449–1456
32. J. Engel, V. Koltun, D. Cremers, Direct sparse odometry. IEEE Trans. Pattern Anal. Mach. Intell. **PP**(99), 1 (2017)
33. V. Usenko, J. Engel, J. Stückler, D. Cremers, Direct visual-inertial odometry with stereo cameras, in *IEEE International Conference on Robotics and Automation (ICRA)* (2016), pp. 1885–1892
34. R.A. Newcombe, S. Izadi, O. Hilliges et al., KinectFusion: real-time dense surface mapping and tracking, in *IEEE International Symposium on Mixed and Augmented Reality* (2011), pp. 127–136
35. P. Henry, M. Krainin, E. Herbst et al., RGB-D mapping: using kinect-style depth cameras for dense 3D modeling of indoor environments. Int. J. Rob. Res. **31**(5), 647–663 (2012)
36. C. Kerl, J. Sturm, D. Cremers, Dense visual SLAM for RGB-D cameras, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 8215, no. 2 (2014), pp. 2100–2106
37. M. Wu, J. Sun, Extended Kalman filter based moving object tracking by mobile robot in unknown environment. ROBOT **32**(3), 334–343 (2010)
38. W. Zhang, *Research on Autonomous Navigation Method for Indoor Robots Based on Multi-sensor Fusion* (University of Science and Technology of China, 2017)
39. Q. Lai, S. Yu, S. Ding, Mobile robot SLAM and path planning system. Comput. Knowl. Technol. **33**(13), 24–37 (2017)
40. Y. Zhao, G. Liu, G. Tian et al., A survey of visual SLAM based on deep learning. ROBOT **39**(6), 889–896 (2017)