Visual SLAM Location Methods Based on Complex Scenes: A Review

Hanxiao Zhang[1] and Jiansheng Peng[1,2](✉)

[1] School of Electrical and Information Engineering,
Guangxi University of Science and Technology, Liuzhou 545000, Guangxi, China
sheng120410@163.com

[2] School of Physics and Mechanical and Electronic Engineering, Hechi University,
Yizhou 546300, Guangxi, China

Abstract. In recent years, positioning for simple static scenes has been unable to meet the requirements of people's production and life. People want to achieve accurate positioning in practical scenarios such as airports, exhibition halls and stations. Therefore, the research on visual SLAM positioning in complex dynamic scenes is increasing day by day. This article reviews the research results of SLAM positioning methods and visual SLAM positioning methods for complex scenes in recent years. Firstly, the development process of laser SLAM, visual SLAM, semantic SLAM and multi-sensor fusion is introduced, but the focus is on visual SLAM. Secondly, the paper summarizes the methods of moving object detection and visual SLAM localization in complex dynamic scenes. Then the paper describes the development of deep learning and multi-sensor fusion in visual SLAM positioning based on complex dynamic scenes. Finally, the shortcomings of visual SLAM positioning methods based on complex scenes are summarized and the research prospects are prospected.

Keywords: Complex scene · Visual SLAM · Positioning

1 Introduction

Indoor positioning and navigation technology has always been the key technology for intelligent service robots. As the most important link for robots to achieve autonomous navigation, positioning has always been the focus of scientists' research. The current positioning technology is mainly realized through a variety of sensor information sensing environments such as laser radar, code disks, cameras and ultrasonic sensors. However, the positioning effect is often related to the complexity of the indoor environment and the cost of the sensor. Most of the visible on the market is based on laser to achieve positioning and navigation, but the cost of laser radar is high and the data is single. In comparison, vision-based positioning technology has more development potential than laser radar. Visual positioning and navigation technology has developed rapidly in recent years, with the advantages of low cost, rich information, and strong scalability. Visual positioning navigation is a popular direction of current positioning navigation technology.

X. Sun et al. (Eds.): ICAIS 2020, LNCS 12240, pp. 487–498, 2020.
https://doi.org/10.1007/978-3-030-57881-7_43

The current visual SLAM positioning and navigation technology is unstable and is more commonly used in simple scenes. The application scenarios of indoor positioning technology include airports, restaurants, hotels, exhibitions, and home environments. The application scenarios of indoor positioning technology are all complex dynamic scenarios. The study of simple scene positioning is not enough to meet the requirements of production and life, so the positioning of complex scenes is more practical and has become the focus of research in recent years. The core problem of the existing visual positioning technology is that it has poor robustness to complex dynamic scenes, and it is easy to lose positioning. How to achieve high robust positioning in complex dynamic scenes is the key issue of visual positioning technology. In order to improve the recognition accuracy, many scholars have combined deep learning with visual SLAM of complex scenes, which has improved the recognition accuracy of images. However, it often takes more time to process the data of complex scenes after adding deep learning, resulting in poor real-time positioning of complex scenes and poor positioning results, which makes them widely used in complex scenes with slow backgrounds.

The second chapter of this paper summarizes the current research status of SLAM positioning technology. The second chapter mainly introduces positioning methods from laser SLAM, visual SLAM, semantic SLAM and multi-sensor fusion, but mainly focuses on visual SLAM positioning research. The third chapter summarizes the research on moving object detection methods and visual SLAM positioning methods based on complex dynamic scenes. The purpose of moving object detection is to eliminate the impact of dynamic obstacles in complex environments and improve the positioning accuracy of robots in complex environments. The fourth chapter introduces the development direction of visual SLAM positioning based on complex dynamic environment around the future of deep learning. Finally, the shortcomings of the existing methods are further summarized and prospects are discussed.

2 Research Status of SLAM Positioning Technology

According to the different sensors carried by the robot, SLAM is mainly divided into laser SLAM and visual SLAM. Semantic information extracted based on visual information is the key to intelligent robots to perform advanced tasks, so semantic SLAM is derived from the latest research directions. Researchers fuse these different sensors, and a multi-sensor fusion method has emerged. This chapter will focus on four areas: laser SLAM, visual SLAM, semantic SLAM and multi-sensor fusion, but we mainly focus on visual SLAM.

2.1 Laser SLAM

After SLAM was proposed, it has gone through several research stages. The early laser SLAM used extended Kalman filter [1], and then estimated the pose of the robot through the information provided by the odometer and laser, but the effect was not good. For some strong nonlinear systems, this method will bring more The truncation error makes it impossible to accurately locate and map. Then the particle filter-based laser SLAM gradually became the mainstream, and it was widely used in non-linear systems, and it is still a popular method until today [2–4]. However, particle filtering consumes memory and is not suitable for detecting large scenes. Grissett et al. [5] proposed an improved particle filtering method and proposed an adaptive resampling method to reduce the computational load of particle filtering. Then someone proposed a method of graph optimization instead of filtering [6–10]. It is generally considered that with the same amount of calculation, nonlinear optimization can achieve better results than Kalman filtering [11]. One of the major problems with the above method is the laser closed-loop detection problem. For the difficult closed-loop phenomenon caused by accumulated errors, Google proposed a cartographer [12], a solution that uses laser data and maps to identify closed-loops. This scheme can guarantee the accuracy of the map structure, but this scheme only relies on lasers and the closed loop is prone to errors.

2.2 Visual SLAM

Visual SLAM has become a research hotspot of SLAM due to low cost of sensors and rich image information. However, visual SLAM is more complicated than laser SLAM. Visual SLAM extracts feature points to achieve positioning and mapping. At first, Kalman filter [13] and particle filter [14] were used to extract feature points for matching to calculate the camera pose. The feature points are extracted to match the camera pose. The commonly used image features are SIFT [15], SURF [16] and ORB [17], etc. However, processing all data directly will cause excessive calculation. In order to achieve better real-time performance, the subsequent visual SLAM adopts the form of multi-threading. SLAM is divided into two parts: front-end and back-end. Nonlinear optimization method is used [6] and ORB feature is used to improve real-time performance and use the bag of words model Implement closed-loop detection [7].

Visual SLAM can also be combined with direct methods to achieve positioning and mapping. Compared with extracting feature points, the direct method is more practical for weakly textured scenes. Stuhmer et al. [8] proposed pixel calibration of dense images to construct dense three-dimensional maps. Engel et al. [18, 19] proposed dense image registration method to calculate camera pose and LSD-SLAM. LSD-SLAM is a successful example of direct method in semi-dense monocular SLAM. Subsequent researchers have successively proposed Kinect-based fusion methods [20], direct RGB-D SLAM method [21] and multi-camera realization of accurate positioning [22], etc. to perform positioning mapping work.

2.3 Semantic SLAM

Semantic SLAM has only been developed in recent years in order to make robots understand the environment closer to humans. Bao et al. [23] used image semantics to improve the robot's ability to recognize objects in the environment. Salas-Moreno et al. [24] generated dense, accurate problem-level semantic maps in order to be able to identify pre-modeled objects. Vineet et al. [25] for the first time realized a system capable of simultaneous semantic segmentation and mapping. Bowman et al. [26] and Schönberger et al. [27] implemented SLAM positioning using semantic information. The RGB-D SLAM system proposed by Rünz et al. [28] can realize scene multi-target recognition.

2.4 Multi-sensor Fusion

In terms of multi-sensor fusion, there are currently various sensors such as vision sensors, lidar, and inertial measurement unit (IMU). Currently, the mainstream approach is to fuse laser radar with vision sensors and IMU with vision sensors. Zhang Jie and Zhou Jun [29] proposed a SLAM method combining lidar and vision, which uses laser map navigation and visual map to restore the detection scene. Qin et al. [30] proposed a tight coupling scheme between vision and IMU, combining the visual structure and the residuals of the IMU structure to form a joint optimization problem. Yin Lei et al. [31, 32] proposed a method of positioning and mapping combining laser and vision. Combining laser data and image information effectively improved the accuracy of positioning and mapping.

3 Visual SLAM Positioning Method Based on Complex Scene

3.1 Moving Object Detection

Moving object detection refers to the process of detecting moving objects from complex background images. Usually, the common methods of detecting moving objects are divided into the following categories: frame-based method, background-based method and light flow-based method, etc.

Frame Difference Method. This method detects moving objects by comparing the gray value of the pixel corresponding to the current frame and the adjacent frame in the video image to find the difference. Cheng et al. [33] proposed a difference method for three consecutive frames. The basic principle is to use the difference between two adjacent frames of three consecutive frames to obtain the result. This method can effectively remove the background due to motion occlusion. The detection error caused. Min et al. [34] are prone to some loopholes in the frame difference method. The background model method is difficult to establish a model. The improved frame difference method of motion history image is combined with the background difference method based on improved Gaussian mixture model to perform moving targets detection.

Background Difference Method. The video image is divided into foreground and background images. And the current frame and the background frame are subjected to difference operation. Stauffer et al. [35] further proposed a mixed Gaussian model in order to improve the detection capability of the single Gaussian model in complex environments. Zhou Jianying et al. [36] to solve the problem that the parameters of the traditional gaussian mixture model converge slowly and it is difficult to adapt to the real-time change of the real background in the scene with time. And the traditional method leads to the increase of error detection rate of moving target. They proposed a moving target detection method based on gaussian mixture model.

Optical Flow Method. The optical flow method extracts the velocity field from the image by using the change of the gray scale information of the image, and derives the motion parameter and the object structure of the moving object due to the existence of the constraint condition.

3.2 Research on Visual SLAM Location Method Based on Complex Scenes

In the study of visual SLAM based on complex dynamic scenes, the SLAMIDE algorithm [37] uses the expectation maximization algorithm to update the feature point motion model in the scene. After updating the motion model, the dynamic object is introduced into SLAM through a reversible model selection mechanism. The disadvantage of this algorithm is that the continuous introduction of dynamic map points in the map will increase memory consumption and reduce the search speed of map points. Tan et al. [38] proposed an adaptive RANSAC algorithm based on existing knowledge in order that the system can run stably in a dynamic environment. Tan et al. Iteratively screens static feature points in the image and removes external points introduced by dynamic objects, and according to the static feature points in the scene, obtains a more uniform distribution, a larger number, and a more consistent camera motion model. The algorithm adaptively models the dynamic environment and can effectively detect and handle appearance or structural changes. Oh [39] proposed a SLAM method based on dynamic extended Kalman filter (EKF) in order to solve the robot pose estimation and environment mapping errors caused by the change of landmark position. This algorithm divides SLAM into traditional static SLAM part and single dynamic SLAM part to reduce the impact of dynamic environment on SLAM and improve the positioning accuracy of robots in dynamic landmarks. Newcombe et al. [40] proposed that the Dynamic Fusion algorithm can reconstruct the dynamic changes of the environment and is suitable for various moving objects and scenes. However, the algorithm can only be used in smaller environments and requires the GPU to complete the composition, and its deployment is not easy. Kumar [41] proposed an online spatiotemporal joint model in order to obtain motion information. This model is used to estimate joint structure. The model predicts the future motion of joint objects by integrating spatial and temporal structures and adds it to the SLAM algorithm. This model enables the algorithm to include the motion of dynamic objects in the real-time SLAM framework, improving detection accuracy. Sun [42] proposed a motion culling method based on RGB-D data and integrated it into RGB-D SLAM. Sun first performs rough detection of moving targets through image difference, then uses particle filtering to track motion, and finally determines moving objects by estimating the maximum posterior probability on depth images. Experimental results show that this method effectively improves all the performance of RGB-D SLAM, but there are still some defects. For example, when the parallax between consecutive frames is large, the homography estimation will decrease and when the moving object becomes stationary, the tracking will fail. Sun [43] proposed a method of removing moving objects in order to effectively eliminate moving objects and assist the scene modeling algorithm to establish a clear scene model. This method mainly relies on optical flow method and depth information to achieve moving object culling in RGB-D SLAM. The disadvantage of this method is that the camera is assumed to be static during scene modeling. Barsan et al. [44] proposed a three-dimensional large-scale dynamic urban environment density mapping algorithm. The algorithm uses instance-perceived semantic segmentation and sparse scene flow to divide the object into background and movement and reconstruct the background and moving objects through the depth information calculated by visual odometry and stereo vision. The algorithm outputs high-quality static backgrounds and dense model dynamic objects of backgrounds, which prevents the impact of dynamic scenes on positioning. In order to improve the performance of RGB-D SLAM in high dynamic scenes, Yu Xiang [45] proposed to build and update the motion foreground model by learning. This model can implement moving object culling and apply it to visual SLAM, but this method still has some limitations. First, the parallax between consecutive frames is required to be small. Second, requiring static objects to dominate the detection environment reduces performance in low dynamic environments. Finally, this method is very time-consuming and has poor real-time performance. Berta et al. [46] proposed a dynamic object detection system DynaSLAM based on ORB-SLAM2 in order to track and reuse scene maps more accurately. DynaSLAM can detect moving objects through multi-view geometry method and deep learning method. DynaSLAM uses MASK-RCNN [47] for instance segmentation and segmentation of objects with mobility. In highly dynamic situations, DynaSLAM is superior to standard visual SLAM in accuracy, and it can also estimate the map of the static part of the scene. However, the performance of DynaSLAM on low dynamic sequences is not obvious. Yu et al. [48] proposed DS-SLAM in order to allow the robot to complete advanced tasks and reduce the impact of dynamic objects on pose estimation. The algorithm combines semantic segmentation network with optical flow method and provides semantic representation of octree maps. The algorithm focuses on reducing the impact of dynamic objects in vision-based SLAM and improving the robot's positioning and mapping performance in complex dynamic environments. The algorithm improves the robustness and accuracy of positioning and mapping. Xu [49] combined instance segmentation into RGB-D SLAM to provide stable motion estimation. However, this combination runs at a slower speed, which can only run at a speed of 2–3 Hz and does not include instance segmentation. Zhenlong Du et al. [50] proposes an improved SLAM algorithm, which mainly improves the real-time performance of classical SLAM algorithm, applies KDtree for efficient organizing feature points, and accelerates the feature points correspondence building. Moreover, the background map reconstruction thread is optimized, the SLAM parallel computation ability is increased. The improved SLAM algorithm holds better real-time performance than the classical SLAM. Xianyu Wu et al. [51] in order to solve the natural scene image has more interference and complexity than text. They propose a new text detection and recognition method based on depth convolution neural network is proposed for natural scene image. In text detection, this method obtains high-level visual features from the bottom pixels by ResNet network, and extracts the context features from character sequences by BLSTM layer, then introduce to the idea of faster R-CNN vertical anchor point to find the bounding box of the detected text, which effectively improves the effect of text object detection. The method can replace the artificially defined features with automatic learning and context-based features. It improves the efficiency and accuracy of recognition. Zhe Liu et al. [52] In order to avoid most image segmentation methods based on clustering algorithm, single objective function is used to realize image segmentation. An image segmentation method based on multi-objective particle swarm optimization (PSO) clustering algorithm is proposed. The unsupervised algorithm not only provides a new similarity calculation method based on electromagnetic force, but also obtains the appropriate number of clusters determined by scale space theory. Zhang Jinfeng et al. [53] proposed a SLAM method based on visual features in dynamic scenes in order to reduce the impact of dynamic objects on tracking and positioning. This algorithm introduces a deep learning-based object detection algorithm into the classic ORB_SLAM2 method. The algorithm divides feature points into latent dynamic features and non-latent dynamic features. The algorithm will calculate the motion model based on the non-potential dynamic feature points and then filter out the static feature points in the scene for pose tracking. The algorithm uses static feature points in non-latent dynamic features for mapping. Compared with ORB_SLAM2, the performance of the system has been significantly improved and the running speed of the system can meet the real-time requirements. Gao Chengqiang et al. [54] proposed a semi-direct RGB-D SLAM algorithm for indoor dynamic environment. The algorithm uses a sparse image alignment algorithm to make a preliminary estimation of the camera pose. The algorithm uses the pose estimation of the visual odometer to compensate the motion of the image. The algorithm establishes a Gaussian model based on real-time update of image blocks. The algorithm divides the moving target in the image according to the variance to eliminate the local map points projected on the moving area of the image. The algorithm completes the real-time update of the map in a dynamic environment and improves the camera pose accuracy.

In order to better reflect the improvement results of different algorithms, this article selects some algorithms for comparison. The comparison results are shown in Table 1. Table 1 compares the absolute path error (ATE) of ORB-SLAM2, DS-SLAM [48] and Zhang Jinfeng's improved dynamic scene SLAM method [53]. Both DS-SLAM and Zhang Jinfeng's improved dynamic scene SLAM methods are improvements based on the original system ORB-SLAM2. It can be seen from Table 1 that the improved dynamic scene SLAM method proposed by DS-SLAM and Zhang Jinfeng has improved performance compared to the original system ORB-SLAM2.

Table 1. Absolute path error (ATE) comparison

Data		Walking_static	Walking_rpy	Walking_halfsphere
ORB-SLAM2	RMSE	0.3900	0.8705	0.4863
	Mean	0.3554	0.7425	0.4272
	S.D.	0.1602	0.4520	0.2290
DS-SLAM	RMSE	0.0081	0.4442	0.0303
	Mean	0.0073	0.3768	0.0258
	S.D.	0.0036	0.2350	0.0159
Zhang Jinfeng's improved method	RMSE	0.0088	0.0595	0.0394
	Mean	0.0079	0.0427	0.0326
	S.D.	0.0040	0.0362	0.0220

Visual SLAM started late and most of the research on visual SLAM positioning in complex dynamic scenes has only begun in recent years. Research on visual SLAM in complex scenes is relatively shallow and only applicable to simple dynamic scenes. The most common optical flow method often assumes that the largest connected area in the picture is a static background, but the opposite situation may exist in the actual scene. Using deep learning to consider semantic information into dynamic scene vision SLAM also has the problem of large amount of calculation and cannot be real-time. Moreover, the deep segmentation of the object semantics in the image through deep neural network can only obtain its own information, but cannot directly determine the movement.

4 Direction of Development

The following discussion will help to advance the development of visual SLAM positioning methods based on complex environments:

(1) Visual SLAM positioning and deep learning

Aiming at the impact of dynamic objects on the robot's current positioning in complex dynamic scenes, a lightweight dynamic semantic segmentation network architecture with real-time performance can be used. This architecture can optimize existing semantic segmentation network models and reduce the number of network model layers. Specific training samples are given for specific scenarios, for example, for airports, the environment contains a large number of people, luggage, walkers, and so on. The architecture specifically trains network models for these objects and extracts dynamic semantic information from images. The architecture redesigns a learning inference model that combines semantically segmented images with the original images for the detection and elimination of dynamic objects. The architecture reduces the variety of semantics through specialized training samples to reduce network complexity and achieve lightweight real-time semantic segmentation. At the same time, in the sample processing stage, artificially designed mobility tags are added for classified semantics. The tag can distinguish whether the object has mobility or not, and can distinguish whether the object is easily moved.

(2) Multi-sensor fusion positioning

Multiple sensor fusion positioning is also the focus of research. The laser sensor has high detection accuracy, but it is difficult to achieve a closed loop. The difficulty of closed loop greatly affects the establishment of maps by robots in indoor environments. Vision sensors are currently widely used and are easy to implement closed loops for maps. Can be visual sensor is detection accuracy is not enough. The inertial measurement unit (IMU) has large drift errors and accumulated errors. Semantic SLAM has only been developed in recent years and is mostly used in complex dynamic environments. But semantic SLAM detection accuracy is not enough to meet the requirements of production and life. At present, there are many researches on the combination of semantic segmentation and visual SLAM, but the current research is insufficient to meet people's positioning requirements for indoor robots in complex dynamic environments. Multi-sensor fusion also requires researchers to continuously improve their methods to improve robustness and real-time performance.

5 Conclusion

Visual SLAM positioning based on complex scenes has made great progress, driven by the continuous development of deep learning and computer vision. However, the visual SLAM based on complex scenes still fails to meet the requirements of people's production and life in terms of real-time, robustness and scalability. At present, the combination of SLAM and deep learning has improved the object detection of complex scenes to a certain extent. The combination of SLAM and deep learning also has an important impact on the rapid and accurate generation of high-level semantics and the construction of robot knowledge bases. However, after SLAM combined with deep learning, the detection and recognition time of complex scenes increases significantly, which makes it difficult to meet the system real-time requirements. And the training parameters greatly depend on the experience of adjusting the parameters. The detection result also greatly depends on the similarity of the current field application scenario. Therefore, the visual SLAM of complex scenes in the future needs to rely on the further development and integration of target recognition, semantic segmentation, and deep learning. And how to apply deep learning to the entire SLAM system instead of just positioning, closed-loop modules, etc. is still a huge challenge.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (Grant No. 61640305). This research was financially supported by the project of Thousands outstanding young teachers' training in higher education institutions of Guangxi, The Young and Middle-aged Teachers Research Fundamental Ability Enhancement of Guangxi University (ID: 2019KY0621), Natural Science Foundation of Guangxi Province (No. 2018GXNSFAA281164). Guangxi Colleges and Universities Key Laboratory Breeding Base of System Control and Information Processing, Hechi University research project start-up funds (XJ2015KQ004), Supported by Colleges and Universities Key Laboratory of Intelligent Integrated Automation (GXZDSY2016-04), Hechi City Science and Technology Project (1694-3-2), Research on multi robot cooperative system based on artificial fish swarm algorithm (2017CFC811).

References

1. Smith, R., Self, M., Cheeseman, P.: Estimating uncertain spatial relationships in robotics. In: Proceedings of IEEE International Conference on Robotics and Automation, Raleigh, NC, USA, 31 March–3 April 1987
2. Sileshi, B.G., Oliver, J., Toledo, R., Goncalves, J., Costa, P.: On the behaviour of low cost laser scanners in HW/SW particle filter SLAM applications. Robot. Auton. Syst. **80**(C), 11–23 (2016)
3. Thallas, A., Tsardoulias, E., Petrou, L.: Particle filter—scan matching hybrid SLAM employing topological information. In: 24th Mediterranean Conference on Control and Automation, Athens, Greece, 21–24 June 2016
4. Song, W., Yang, Y., Fu, M., Kornhauser, A., Wang, M.: Critical rays self-adaptive particle filtering SLAM. J. Intell. Robot. Syst. **92**(1), 107–124 (2017). https://doi.org/10.1007/s10846-017-0742-z
5. Grisetti, G., Stachniss, C., Burgard, W.: Improved techniques for grid mapping with Rao-Blackwellized particle filters. IEEE Trans. Rob. **23**(1), 34–46 (2007)

6. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, Japan, 13–16 November 2007
7. Mur-Artal, R., Montiel, J.M., Tardós, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. Rob. **31**(5), 1147–1163 (2015)
8. Stühmer, J., Gumhold, S., Cremers, D.: Real-time dense geometry from a handheld camera. In: Goesele, M., Roth, S., Kuijper, A., Schiele, B., Schindler, K. (eds.) DAGM 2010. LNCS, vol. 6376, pp. 11–20. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15986-2_2
9. Konolige, K., Grisetti, G., Kümmerle, R.: Efficient sparse pose adjustment for 2D mapping. In: International Conference on Intelligent Robots and Systems, pp. 22–29 (2010)
10. Engel, J., Sturm, J., Cremers, D.: Semi-dense visual odometry for a monocular camera. In: IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013
11. Strasdat, H., Montiel, J.M.M., Davison, A.J.: Visual SLAM: why filter? Image Vis. Comput. **30**(2), 65–77 (2012)
12. Hess, W., Kohler, D., Rapp, H., Andor, D.: Real-time loop closure in 2D LIDAR SLAM. In: IEEE International Conference on Robotics and Automation, Stockholm, Sweden, 16–21 May 2016
13. Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: MonoSLAM: real-time single camera SLAM. IEEE Trans. Pattern Anal. Mach. Intell. **29**(6), 1052 (2007)
14. Sim, R., Elinas, P., Griffin, M.: Vision-based SLAM using the Rao-Blackwellised particle filter. In: IJCAI Workshop on Reasoning with Uncertainty in Robotics, vol. 9(4), pp. 500–509 (2005)
15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
16. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_32
17. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF, vol. 58, no. 11, pp. 2564–2571 (2011)
18. Engel, J., Sturm, J., Cremers, D.: Semi-dense visual odometry for a monocular camera. In: IEEE International Conference on Computer Vision, pp. 1449–1456 (2013)
19. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: large-scale direct monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8690, pp. 834–849. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10605-2_54
20. Newcombe, R.A., Izadi, S., Hilliges, O., et al.: KinectFusion: real-time dense surface mapping and tracking. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 127–136 (2011)
21. Kerl, C., Sturm, J., Cremers, D.: Dense visual SLAM for RGB-D cameras. In: International Conference on Intelligent Robots and Systems, pp. 2100–2106 (2014)
22. Yang, S., Scherer, S.A., Yi, X., et al.: Multi-camera visual SLAM for autonomous navigation of micro aerial vehicles. Robot. Auton. Syst. **93**(1), 116–134 (2017)
23. Bao, S.Y., Bagra, M., Chao, Y.W., et al.: Semantic structure from motion with points, regions, and objects. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2703–2710 (2012)
24. Salas-Moreno, R.F., Newcombe, R.A., Strasdat, H., et al.: SLAM++: simultaneous localisation and mapping at the level of objects. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1352–1359 (2013)

25. Vineet, V., Miksik, O., Lidegaard, M., et al.: Incremental dense semantic stereo fusion for large- scale semantic scene reconstruction. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 75–82 (2015)
26. Bowman, S.L., Atanasov, N., Daniilidis, K., et al.: Probabilistic data association for semantic slam. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 1722–1729 (2017)
27. Schönberger, J.L., Pollefeys, M., Geiger, A., et al.: Semantic visual localization. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–21 (2008)
28. Rünz, M., Agapito, L.: MaskFusion: real-time recognition, tracking and reconstruction of multiple moving objects. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality, pp. 1–10 (2018)
29. Jie, Z., Jun, Z.: A study on laser and visual mapping of mobile robots with improved ICP algorithm. Mech. Electr. Eng. (12) 1480–1484 (2017)
30. Qin, T., Li, P., Shen, S.: VINS-Mono: a robust and versatile monocular visual-inertial state estimator. IEEE Trans. Rob. **3**(4), 1–17 (2017)
31. Lei, Y.: Research on simultaneous positioning and mapping of indoor robots. Guangxi University of Science and Technology (2019)
32. Yin, L., Peng, J., Jiang, G., Ou, Y.: Research on synchronous positioning and mapping of low-cost laser and vision. Integr. Tech. (2) (2019)
33. Cheng, Y.H., Wang, J.: A motion image detection method based on the inter-frame difference method. Appl. Mech. Mater. **490–491**, 1283–1286 (2014)
34. Min, H., Shu, H., Liu, Q., Xia, Y., Gang, C.: Moving object detection method based on NMI features motion detection frame difference. Adv. Sci. Lett. **6**(1), 477–480 (2012)
35. Stauffer, C., Grimson, E.: Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 747–757 (2000)
36. Zhou, J., Wu, X., Zhang, C., Lu, W.: A moving object detection method based on hybrid Gaussian model of sliding window. J. Electron. Inf. Tech. **35**(07), 1650–1656 (2013)
37. Hahnel, D., Triebel, R., Burgard, W., et al.: Map building with mobile robots in dynamic environments. In: IEEE International Conference on Robotics and Automation, pp. 1557–1563. IEEE, Piscataway (2003)
38. Tan, W., Liu, H.M., Dong, Z.L., et al.: Robust monocular SLAM in dynamic environments. In: 12th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 209–218. IEEE Piscataway (2013)
39. Oh, S., Hahn, M., Kim, J.: Dynamic EKF-based SLAM for autonomous mobile convergence platforms. Multimedia Tools Appl. **74**(16), 6413–6430 (2014). https://doi.org/10.1007/s11042-014-2093-0
40. Newcombe, R.A., Fox, D., Seitz, S.M.: DynamicFusion: reconstruction and tracking of non-rigid scenes in real-time. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 343–352. IEEE, Piscataway (2015)
41. Kumar, S., Dhiman, V., Ganesh, M.R., Corso, J.: Spatiotemporal Articulated Models for Dynamic SLAM (2016)
42. Sun, Y., Liu, M., Meng, Q.H.: Improving RGB-D SLAM in dynamic environments: a motion removal approach. Robot. Auton. Syst. **89**, 110–122 (2017)
43. Sun, Y., Liu, M., Meng, Q.H.: Invisibility: a moving-object removal approach for dynamic scene modelling using RGB-D camera. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), Macau, China, 5–8 December 2017
44. Barsan, I.A., Liu, P., Pollefeys, M., Geiger, A.: Robust dense mapping for large-scale dynamic environments. In: International Conference on Robotics and Automation (ICRA), pp. 7510–7517 (2018)

45. Sun, Y., Liu, M., Meng, M.: Motion removal for reliable RGB-D SLAM in dynamic environments. Robot. Auton. Syst. **108**, 115–128 (2018)
46. Berta, B., Facil, J.M., Javier, C., Jose, N.: DynaSLAM: tracking, mapping and inpainting in dynamic scenes. IEEE Robot. Autom. Lett. **3**(4), 4076–4083 (2018)
47. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask R-CNN. IEEE Trans. Pattern Anal. Mach. Intell. (2018)
48. Yu, C. et al.: DS-SLAM: a semantic visual SLAM towards dynamic environments. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1168–1174 (2018)
49. Xu, B., Li, W., Tzoumanikas, D., Bloesch, M., Davison, A., Leutenegger, S.: MID-Fusion: octree-based object-level multi-instance dynamic SLAM. In: International Conference on Robotics and Automation, Montreal, Canada, 20–24 May 2019
50. Du, Z., Ma, Y., Li, X., Lu, H.: Fast scene reconstruction based on improved SLAM. Comput. Mater. Continua **61**(1), 243–254 (2019)
51. Wu, X., Luo, C., Zhang, Q., Zhou, J., Yang, H., Li, Y.: Text detection and recognition for natural scene images using deep convolutional neural networks. Comput. Mater. Continua **61**(1), 289–300 (2019)
52. Liu, Z., Xiang, B., Song, Y., Lu, H., Liu, Q.: An improved unsupervised image segmentation method based on multi-objective particle, swarm optimization clustering algorithm. Comput. Mater. Continua **58**(2), 451–461 (2019)
53. Zhang, J., Shi, C., Wang, Y.: SLAM method based on visual features in dynamic scenes. Computer program, pp. 1–8, 04 November 2019.http://kns.cnki.net/kcms/Detail/31.1289.tp.20191025.1559.006.html
54. Gao, C., Zhang, Y., Wang, X., Deng, Y., Jiang, H.: Semi-direct method RGB-D SLAM algorithm for indoor dynamic environment. Robot **41**(03), 372–383 (2019)