Simultaneous localisation and mapping of intelligent mobile robots

Sufang Wang*

Beijing Institute of Computer Technology and Application,
Beijing 100000, China
Email: 18232519562@163.com
* Corresponding author.

Tingyang Xie

Department of Computer Science and Engineering,
The Ohio State University,
Columbus 43210, Ohio, USA
Email: xie.963@buckeyemail.osu.edu

Peng Chen

School of Mechanical Engineering,
North China University of Science and Technology,
Tangshan 063210, Hebei, China
Email: 2529106687@qq.com

Abstract: Simultaneous Localisation and Mapping (SLAM) is the key technology of mobile robot navigation. In this field, visual SLAM (VSLAM) has become a research hotspot in recent years. This article discusses the Lidar and visual SLAM algorithms, including: introduction on commonly used solutions and improvements of the Lidar SLAM algorithm and exploring related difficulties, the characteristics of monocular, binocular and RGB-D cameras in VSLAM, the ORB-SLAM2 system based on the feature extraction method, the LSD-SLAM system based on the direct method, and comprehensively understanding and comparing the advantages and disadvantages between Lidar SLAM and visual SLAM, trying to fully utilise the two systems' advantages to implement better abilities of autonomous localisation, path planning and obstacle avoidance. Finally, the conclusion section discusses the development direction of multi-sensor fusion SLAM and the intelligent application of mobile robots in multiple fields.

Biographical notes: Sufang Wang is software development engineer, master, working in Beijing Institute of Computer Technology and Application. Her research direction is intelligent robot navigation algorithm, visual SLAM, multi-sensor information fusion.

Tingyang Xie is junior at The Ohio State University. His learning direction is computer artificial intelligence, computer vision.

Peng Chen is at School of Mechanical Engineering, North China University of Science and Technology. His research direction is intelligent robot navigation algorithm, visual SLAM, multi-sensor information fusion.

1   Introduction

In terms of autonomous navigation and localisation, mobile robots have three major problems: 1) Where am I?; 2) Where am I going?; 3) How should I get there? (Leonard et al., 1992; Durrant-Whyte, 2005). These problems generated three research directions, namely, the localisation problem, the perception problem of the target position and the path planning problem.

SLAM technology enables the robot to navigate its own location in an unknown environment. Achieved by installing sensors, robot during movement can read its surroundings information and construct navigational maps accordingly (Chen and Zhang, 2005; Csorba, 1997).

Although the research on autonomous mobile robots has become a research hot topic in the high-tech field, there are still very few autonomous mobile robots in our daily lives. Most recent research papers are based on one single type of sensor, such as Lidar SLAM or VSLAM algorithms. Algorithm based on Lidar SLAM is already well developed, but VSLAM is not.

SLAM algorithm has significant theoretical meanings and a wide range of applications. In self-driving field, SLAM achieves precise localising and navigation through three-dimension modelling with Lidar and Global Positioning System (GPS). In terms of military use, robots developed by SLAM are working in hazardous areas. In our daily lives, service robots powered by SLAM are capable of walking on their own without hitting any obstacles.

Based on different sensors, SLAM has different names. When the sensor is Lidar, we call it, Lidar SLAM. When the sensor changes to a camera, we call it VSLAM.

As intelligent robots gradually enter our lives, our lives become easier and more convenient; however, simple robot cannot achieve self-walking. At this time, Lidar can help robot higher intelligent level, which can access the robot's surrounding information. With powerful navigation algorithm SLAM, the information sensors scanned is passed on to robots and helps them to walk. There are many Lidar commonly used for SLAM, such as SICK, Velodyne, the SLAMTEC's RPLIDAR and Leishen.

VSLAM is majorly achieved by using cameras. According to different task, cameras are classified to three major branches: Monocular, Stereo Vision and RGB-D.

Exploiting a single camera to do SLAM algorithm is called monocular SLAM. Correspondingly, Stereo Vision SLAM is carrying multiple cameras, among them binocular camera is the most widely used sensor. SLAM combining monocular camera and infrared sensor is called RGB-D SLAM.

2   Key technology

2.1   Lidar SLAM algorithm

Lidar are divided into single-mode and multiple-mode, and they have their own merits of angular resolution and accuracy. Also, the implement method has its own advantages and disadvantages. The most important problem for SLAM is localisation, which can be solved by probability method and non-probability method. In current situation, the most occurred problems are solved by probability-based method whose basis is Bayesian estimation. It is a filter combined with Kalman filter and particle filter. For non-possibility method is based on graph-based optimisation.

1) *Kalman filter*: Kalman filter is the process where the system collects input and output data to achieve optimisation. The assumption of Kalman filter is that state noise and observation noise obey Gaussian distribution (Wu et al., 2007; Yavuz et al., 2009; Wei and Zuo, 2009; Kang et al., 2010; Wang et al., 2013). In addition, the assumption above is also saying that the system is described by linear state equations. However, non-linear systems widely exist in applications, so the extended Kalman filter, a method specifically built for non-linear system, was proposed: non-linear system can be transferred to linear system by applying first derivative based on Taylor's theorem in order to achieve an approximated system model. If system's non-linear degree is high, error of linearisation will increase, so errors may occur when constructing a map. Therefore, EKF is not suitable for SLAM research in large-scale surroundings.

    Unscented Kalman filter, UKF, is a method of approximating non-linear distribution with sampling method by applying conditional Gaussian distribution to test probability density. Its advantage is that it does not need to calculate Jacobian matrix, and it has better performance than EKF. Unscented Kalman filter uses sampling method to approximate non-linear distribution. Most of solutions use Kalman filter, since it has advantages on convergence and implementing difficulties. However, this method only considers the influence of previous frame. Once data association errors occur, whole SLAM system will be affected.

2) *Particle filter*: Particle filter, a novel filter, effectively removes constraints of linear assumption and the assumptions of sensor Gaussian noise. It plays an important role in mobile robots' localisation and mapping. It is good at approximating any probability distribution and calculate simply and efficiently (Arulampalam et al., 2002). However, it cannot achieve the desired effect, when it comes to high-dimension problems.

Murphy et al. found that under the condition where robot moving routine is known, probability between signpost locations is conditionally independent. Consequently, they pointed out Rao-Blackwellised which is the theoretical basis for particle filter to solve SLAM problems. In order to test the possibilities by applying Rao-Blackwellised Particle Filter (RBPF) to solve SLAM problems, Montemerlo et al. (2002) done some relative application practices, and pointed out a Fast-SLAM solution.

This method estimates robot routine by applying improved particle filter. Each particle maintains one approximation state and one individual feature location information set, as shown in Figure 1. Each trajectory of the robot corresponds to a particle. Robot's state is predicted by N particle, each feature is estimated by Extended Kalman Filter (EKF), and each particles maintains M EKF (Thrun et al., 2001). Comparing with other methods, this method reduces sampling space, which significantly reduces complexity and improves calculation speed. High precision means it has better accuracy and robustness. This method, therefore, can apply to situations where the posterior density function is unknown, non-linear or non-Gaussian conditions. However, this method also has some defects. In order to avoid divergence, more particles are needed in experiments, and divergence is closely related to data association. As a novel SLAM algorithm, particle filter still needs to be consummated and improved.

Figure 1   Robot SLAM algorithm based on RBPF
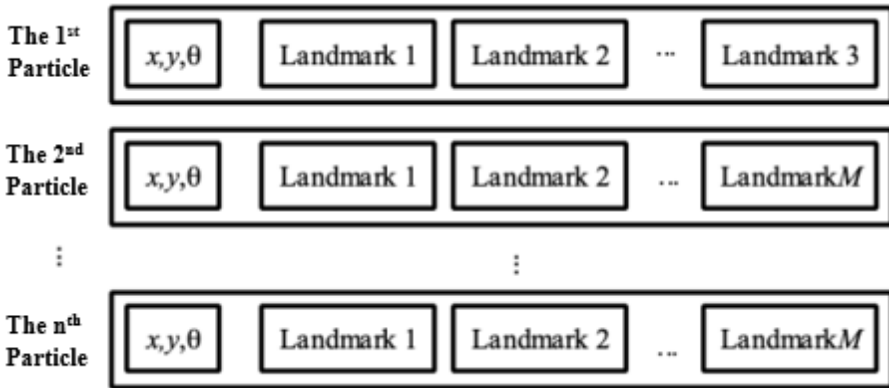
3) *Graph-based optimisation*: Lu and Milios (1997) first proposed graph-based optimisation. In their articles, two-dimensional map which was created by Lidar is taken as an example, and the influence of all frames is taken into consideration to form a spatial constraint relationships. Compared with filter method which only considers the influence of previous frame, if a random moment between x_0 and x_1 occurs an error, the later posture is difficult to correct. However, as surrounding's size increases, the posture error rate increases. Therefore, filter method is not suitable to estimate posture in big surroundings. But graph-based optimisation can optimise all frames to calculate correct postures. Graph-based optimisation estimates the trajectory of the entire robot, and continuously optimises the linear

points in order to minimise the errors. Since the precision after optimising multiple times is significantly higher than filter method, it is suitable for applying to estimate posture in big scenario. But the efficiency will be reduced accordingly. Graph-based optimisation generally optimises graphs which are constructed by SLAM front-end estimation and loop closure detection. So, graph-based optimisation is also called back-end optimisation.

Considering that most application situations of robots are non-linear systems, posture graph can be solved by non-linear least squares. Basic idea of nonlinear least squares is that given a non-linear system with nonlinear state equations we need to find optimal solution in order to minimise the error of estimates (observations). Also, in order to find the target function of nonlinear least squares, we assume that the error obeys Gaussian distribution.

2.2 *VSLAM Algorithm*

VSLAM is constituted by Visual Odometry (VO) and optimised back-end. Visual Odometry is also called front-end, and its main function is to approximately estimate camera's movement according to adjacent frames information. The movement trail can be used for back-end's initial values. Since feature extraction method operates stably and is not suitable for the change of surroundings, feature extraction method is the most often used solution in Visual Odometry (Gao et al., 2017). The back-end optimisation is similar to the laser optimisation algorithm, which will not be repeated in this article.

1) *Feature extraction method*: Monocular visual SLAM system, MonoSLAM, was first pointed out by Davision et al. (2007), which uses Extended Kalman Filter as back-end, and tracks the sparse feature points on the front end. After this monocular visual SLAM which is based on key frame is developed. The most representative key frame in visual SLAM is parallel tracking and mapping (PTAM) which is pointed out by Klein and Murray (2007). PTAM is a simple and effective method to extract key frame, which implements tracking and constructing graph simultaneously, and it is the first time using non-linear optimisation as back-end. Mur-Artal et al. (2015) extends and improves PTAM, creatively pointing out three threads monocular VSLAM system consisted of two threads PTAM. The entire system is completed around ORB features, so it is called ORB-SLAM. The entire SLAM processes can be divided into tracking thread which real-time tracks feature point, optimising thread for partial bundle adjustment and loop closure detecting and optimising thread of global pose graph. Also, applying identical ORB feature to the three threads above makes them interact with each other, which diminishes the accumulated errors when constructing the graph, and guarantees the global consistence between motion trajectory and map. What the ORB-SLAM2 system improved on this basis is that it becomes applicable to monocular, binocular, and RGB-D modes, and has good versatility. ORB-SLAM2 algorithm framework is shown in Figure 2.

Figure 2   ORB-SLAM2 algorithm framework
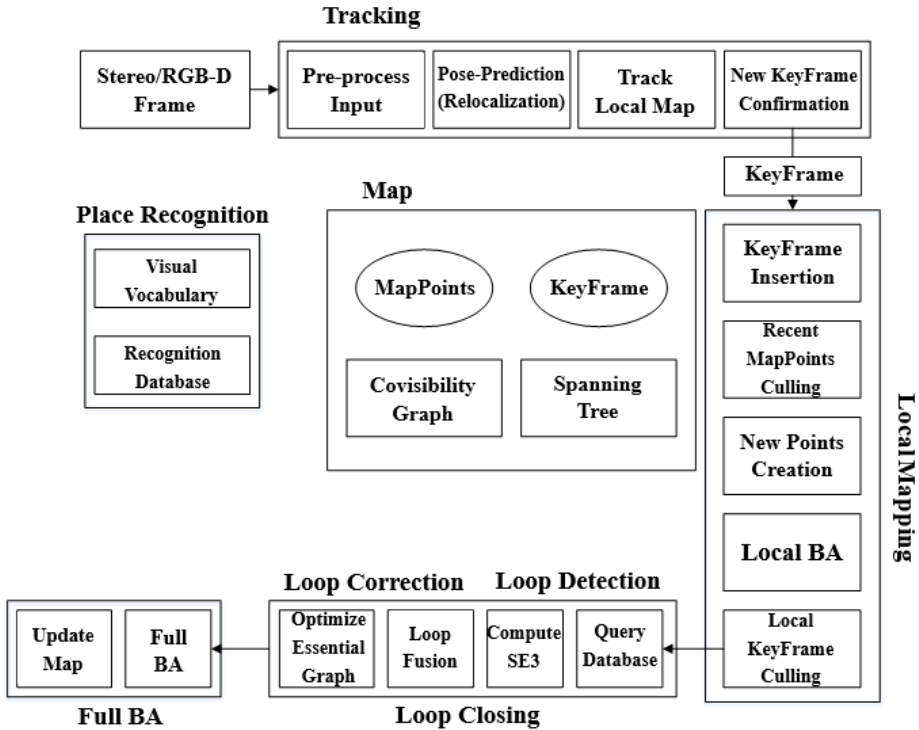
ORB-SLAM2 system is built with four parallel threads: tracking, local mapping, loop closure detecting and global BA optimising. Among them, the fourth thread executes after having the confirmation of the previous thread. Specific explanation for the four threads: 1) tracking: pre-processes the input to get features at the location of significant key point. 2) local mapping: executes partial BA to achieve local map's regulation and optimisation. 3) loop closure detecting: this thread is divided into two steps: first, cycle inspection and verification; second, cyclically correction and optimisation of the posture map. Compared with monocular ORB-SLAM camera which might result scale drift, stereo/depth information makes the scale observable. Geometry verification and target optimisation are no longer needed while system deals with scale drift issues. Also, based on rigid body transformation instead of similarity, the global BA optimisation of the fourth thread is executed after the posture graph. 4) global BA: after the optimising of posture graphs, optimised system structure and motion result can be calculated by executing global BA. This system embeds a DBoM2-based position recognition model used for relocation, which can effectively avoid tracking failures, like occlusion. This system applies scenes graph re-initialisation and loop closure detection, etc. This system applies ORB feature to localising and constructing map, also it has good scale invariance and ideal rotation invariance. Besides, it can promptly extract features to match, so that it can meet the requirements of real-time operation.

Since the monocular camera shoots the two-dimension projection of a three-dimension object, which is a single picture; however, the movement and trajectory estimated by moving camera have uncertainties, so it cannot have the real depth for all

objects. This may result in disadvantages of low accuracy for three-dimensional constructions. At this time, binocular camera and RGB-D become useful. However, based on binocular camera current situation, computing power becomes the major problem, but SLAM based on RGB-D camera can capture depth information directly. This significantly reduces the difficulty of three-dimensional constructions. Henry et al. (2014a) first pointed out a method to complete three-dimensional construction in in-door environment by applying RGB-D camera. First, SIFT feature can be extracted in RGB picture, and then depth information can be found in depth image according to the feature point. Second, Random Sample Consensus, RANSAC, is applied to match 3D feature points with samples to calculate camera's posture and rigid body motion transformation. RANSAC is suitable for situations with wrong data, which can deal with wrong matching data, and then Iterative Closest Point (ICP) algorithm is used to calculate a more precise posture applying initial values. RGB-D SLAM usually uses the ICP algorithm to estimate the postures and optimise the motion transferring matrix of the camera. Zhu and Liu (2018) gave the basic process of ICP algorithm as follows:

1) Read point sets P_1, P_2.

2) Find closest pair. Search for the point which is closest to P_1 in P_2 to form a closest pair; filter all closest pairs in the two-point sets.

3) According to the two-point pair sets, calculate two coordinates of centre of gravity.

4) According to the new point set, calculate the translation matrix t and rotation matrix R.

5) Using R, t, to calculate P_2' which is a new point set after the motion.

6) Calculate twice the absolute value of the difference between the sum of squared distances from P_2 to P_2'. If the absolute value is less than the threshold, it is convergence, and stops iterating; otherwise repeat 1–6 steps, until the error is convergence.

Among genetic algorithm, RANSAC and ICP, etc., since ICP can directly deal with images depth information, and does not need to make any hypotheses and segments for the features of the object, ICP is the most widely used algorithm. After choosing the initial values, the algorithm has better convergence so that it can get global optimal values and achieve relatively precise results. That is the reason it becomes the main depth image matching algorithm (Li et al., 2009).

Although feature extraction method has many advantages, the process of extracting the feature points is time consuming. Also only applying feature extraction method under real environment where there are few feature points, like a completely white wall, it will ignore useful image information, resulting in imprecise calculations of camera's movements.

1) *Direct method*: According to the number of pixels, direct method can be classified into three kinds: sparse, dense and semi-dense. For example, there is a space point of a known position. When P is from sparse key point, it is called sparse direct method, which can quickly calculate camera's posture. Since this method only uses hundreds of pixels instead of calculating descriptors, it can only calculate the sparse reconstruction. When P is from parts of pixels, it is called semi-dense direct method. This method changes previous key points to pixels with obvious gradients ignoring

the position with little gradient changes. In this way, it can reconstruct a semi-dense map. When P is from whole pixels, it is called dense direct method which can reconstruct the complete map. However, dense direct method needs to use GPU programming, so it has relatively high demands for the hardware.

Many people have been committed to the study of the direct method. Irani and Anandan (1999) gave a detailed and in-depth description of the direct method. Silveira et al. (2008) applied the direct method to visual SLAM and described the main advantages and limitations of the method. Subsequently, Forster et al. (2014) proposed the visual mileage calculation method based on the sparse monocular camera method (SVO for short). The precision of this algorithm has good robustness, is improved, and faster than current methods. This sparse method diminishes the demand of movement estimation feature extraction and robustness matching technology, and it is suitable for the condition that state estimation of mini car in GPS refusal environment. Right after this, Engel et al. (2014) proposed a large-scale direct SLAM, LSD-SLAM. This algorithm is composed by three parts: tracking link, depth map estimation and map optimisation, shown in Figure 3. In the algorithm, tracking link continuously tracks new camera images using image data from previous moment in initial process, so that it can calculate camera's posture at current moment. In depth estimation process, tracking frame refreshes the used key frame. If the algorithm chooses not to adopt, then optimise current frame. By filtering multiple unit pixels, doing small baseline stereo contrast, and applying staggered space regularisation method, the depth information can be optimised. If camera's movement speed is fast which makes distance of the movement far, then the current key frame will be optimised by the projection points of historical key frames. The relative distance and angle of the current key frame are weighted and combined, the depth map data is scaled, and the average inverse depth is normalised to 1. Based on this, the new generated frame will replace current key frame. When the reference tracking key is replaced, it will add to the global map in the process of map optimisation. For the not adopted frame, they will be used to deal with current frame's data. In order to detect loop closure and scale drift, a method will be applied, which will use two images with different scale in algorithm (function) sim (3) to align, so that scale drift will be detected, and make estimation to the loop closure similarity transformation (including to the previews key frame). Comparing to the other methods, it can reconstruct the pose map of the key frame and the semi-dense and highly accurate 3D map of the environment in real time. Usenko et al. (2016) proposed a noble stereo camera direct vision-inertial odometer localisation method using vision and inertial data's complementarity to improve the precise of three-dimensional map reconstructions.

Newcombe et al. (2011) reconstructed a three-dimensional model by combining all depth information and image information which are all from Kinect into observation sense in real time, so that it can obtain the global maps. Specially, it allows to reconstruct the dense maps in real time, which makes a huge progress in AR field. Henry et al. (2012) adopted Joint Optimisation Algorithm. This algorithm makes RGB-D camera can be used in in-door mobile robot field. Kerl et al. (2014) proposed a VSLAM method which is based on direct dense RGB-D camera. This method has two kinds of errors: luminosity error and depth error, and it uses G2O optimisation library to calculate the best camera posture and uses entropy method to choose key frame for loop closure detection. This significantly reduces the path error.

Figure 3 LSD-SLAM algorithm framework
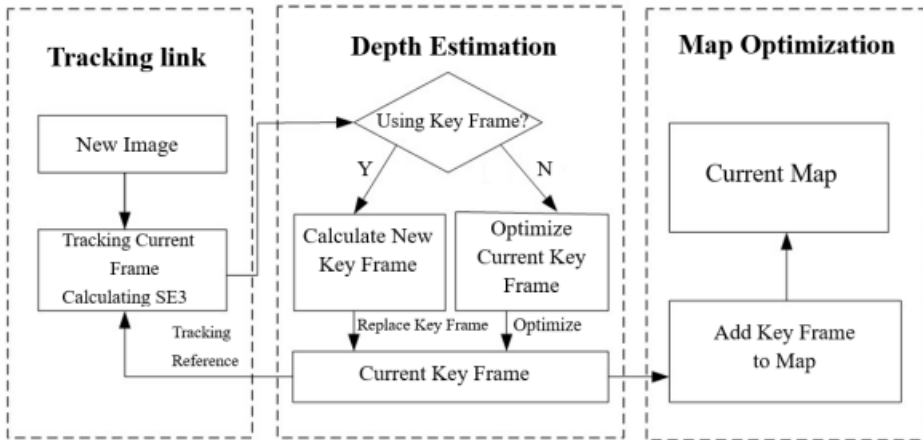
Direct Method is faster than feature extraction method in executing time perspective, and it does not need to extract image features. However, direct method is easy to make error because of lighting environment, when grey level is constant. Feature extraction method will not reconstruct the map successfully, when camera doing rapid motion makes images blur so that the feature points loses. But Direct Method does not need feature point to obtain useful data, it uses prime gradient data to solve the camera rotation matrix (Henry et al., 2014b). Also, the prime gradient leads the optimisation's direction, which requires the camera's movement cannot be too large, so that the right posture can be calculated by most of the prime gradients.

3 Perspectives

After more than 30 years of research and hard work on SLAM system (including VSLAM) algorithms by predecessors, the SLAM (including VSLAM) system with Lidar and camera as the main sensor has been developed by leaps and bounds. Because of the limitations of indoor operation environment, GPS cannot be used to limit the localisation error, but SLAM has opened a door for the development of indoor robotics. Lidar SLAM already have a relatively mature method, but the high cost is always the primary issue (Wu and Sun, 2010). Low-priced VSLAM has become a popular research topic in recent years. No matter what kind of sensors is used alone, there are some certain defects. Therefore, multiple sensor fusion technology combining Lidar sensor, vision sensor and Inertial Measurement Unit (IMU) (Zhang, 2017; Lai et al., 2017) has been developed, which not only can achieve the cooperative operation between various sensors, but also can improve the robustness. The researches and applications of multi-sensor fusion technology will bring a broader space to the fields of self-driving, robotics, augmented reality and virtual reality. Besides, SLAM also can be combined with deep learning to process images (Zhao et al., 2017), generate environmental semantic map, and improve the robot's human-computer interaction capabilities. This can boost intelligence level to a higher ground.

4   Conclusions

Lidar sensors' advantages are wide visual range, high precision which can measure the angle and distance of surrounding obstacles precisely, so that it can help robots to avoid obstacle, but the price of Lidar is expensive. In contrast, camera has advantages, like lower cost, lightweight, easy to install, convenient and flexible, can extract plenty of scenes information; therefore, visual SLAM becomes popular for SLAM researchers these years. Visual SLAM and density reconstruction use little CPU. It provides enough capability for application's logic and processing of other sensors. By providing the situational awareness of robots or drones, depth camera can help solving SLAM problems better.

The performance of visual SLAM depends on the environment it is situated in. The ideal conditions are:

1)   The light is enough and remains unchanged. Cameras must identify some features in the environment. Generally, a much more characteristic field of vision, like plenty of objects or geometric figures, can help to achieve the desired results. The worst scenarios include blank walls, floors or ceilings. If there are reflective surfaces, like mirrors or glass, errors probably will occur. Also, direct sunlight will affect cameras' determination, which will influence the precision of the mapping.

2)   When most parts in the environment remain unchanged, VSLAM works the best. If people or objects are moving, the performance of VSLAM will be influenced less or more. If the entire environment is moving, then VSLAM will not work at all.

3)   When the major work of cameras is transforming, not rotation, VSLAM works the best. When cameras have to rotate, VSLAM works better for the slow rotation.

4)   When a VSLAM system starts, cameras must stand still, and cameras need to have sufficient visual features. Sometimes, blank wall, floor, or celling may block cameras' sight. This may mean the VSLAM system cannot be successfully initialised.

Lidar and cameras have their advantages and disadvantages. It is easy for cameras to identify the identical objects. If cameras tell Lidar that the object in two frames before and after is the same object, the Lidar can calculate the object's velocity and displacement between the two frames. Therefore, identifying and tracking can be easily implemented, so that more applicable and more precise maps can be achieved.

References

Arulampalam, S.M., Maskell, S. and Gordon, N. (2002) 'A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking', *IEEE Transaction Signal Processing*, Vol. 50, No. 2, pp.174–188.

Chen, W. and Zhang, F. (2005) 'Review on the achievements in simultaneous localization and map building for mobile robot', *Control Theory and Applications*, Vol. 22, No. 3, pp.455–460.

Csorba, M. (1997) *Simultaneous Localization and Map Building*, University of Oxford, Oxford.

Davison, A.J., Reid, I.D. and Molton, N.D. et al. (2007) 'Monoslam: real-time single camera SLAM', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6, pp.1052–1067.

Durrant-Whyte, H. (2005) 'Where am I? A tutorial on mobile vehicle localization', *Industrial Robot*, Vol. 21, No. 2, pp.11–16.

Engel, J., Schöps, T. and Cremers, D. (2014) 'LSD-SLAM: large-scale direct monocular SLAM', *European Conference on Computer Vision*, Springer International Publishing, pp.834–849.

Forster, C., Pizzoli, M. and Scaramuzza, D. (2014) 'SVO: fast semi-direct monocular visual odometry', *IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, China.

Gao, X., Zhang, T. and Liu, Y. et al. (2017) *Visual SLAM Fourteen Lectures: From Theory to Practice* (in Chinses)*.*, Publishing House of Electronics Industry, Beijing.

Henry, P. Krainin, M. and Herbst, E. et al. (2014b) 'RGB-D mapping: using depth cameras for dense 3D modeling of indoor environments', *Springer Berlin Heidelberg*, Vol. 31, No. 5, pp.647–663.

Henry, P., Krainin, M. and Herbst, E. et al. (2012) 'RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments', *International Journal of Robotics Research*, Vol. 31, No. 5, pp.647–663.

Henry, P., Krainin, M. and Herbst, E. et al. (2014a) 'RGB-D mapping: using depth cameras for dense 3D modeling of indoor environments', *Springer Berlin Heidelberg*, Vol. 31, No. 5, pp.647–663.

Irani, M. and Anandan, P. (1999) 'About direct methods', *International Workshop on Vision Algorithms: Theory and Practice*, pp.267–277.

Kang, J.G, Choi, W.S. and An, S.Y. et al. (2010) 'Augmented EKF based SLAM method for improving the accuracy of the feature map', *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Taipei, Taiwan.

Kerl, C., Sturm, J. and Cremers, D. (2014) 'Dense visual SLAM for RGB-D cameras', *IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Japan.

Klein, G. and Murray, D. (2007) 'Parallel tracking and mapping for small AR workspaces', *IEEE and ACM International Symposium on Mixed and Augmented Reality*, IEEE, Nara, Japan, pp.225–234.

Lai, Q., Yu, S. and Ding, S. (2017) 'Mobile robot SLAM and path planning system', *Computer Knowledge and Technology* (in Chinses), Vol. 33, No. 13, p.24.

Leonard, J.J., Durrant-Whyte, H.F. and Cox, I.J. (1992) 'Dynamic map building for an autonomous mobile robot', *The International Journal of Robotics Research*, *Sage Publications*, Inc., Vol. 11, No. 4, pp.286–298.

Li, S., Wang, P. and Shen, Z. (2009) 'A survey of iterative closest point algorithm', *Signal Processing*, Vol. 25, No. 10, p.1582

Lu, F. and Milios, E. (1997) 'Globally consistent range scan alignment for environment mapping', *Autonomous Robots*, Vol. 4, No. 4, pp.333–349.

Montemerlo, M., Thrun, S. and Koller, D. (2002) 'Fast SLAM (simultaneous localization and mapping)', *Proceedings of the AAAI National Conference on Artificial Intelligence*, Menlo Park, CA, USA, pp.593–598.

Mur-Artal, R., Montiel, J.M.M., and Tardós, J.D. (2015) 'ORB-SLAM: A Versatile and Accurate Monocular SLAM System', *IEEE Transactions on Robotics,* Vol. 31, No 5, p.1147.

Newcombe, R.A., Izadi, S. and Hilliges, O. et al. (2011) 'KinectFusion: real-time dense surface mapping and tracking', *IEEE International Symposium on Mixed and Augmented Reality*, pp.127–136.

Silveira, G., Malis, E. and Rives, P. (2008) 'An efficient direct approach to visual SLAM', *IEEE Transactions on Robotics*, Vol. 24, No. 5, pp.969–979.

Thrun, S., Fox, D. and Burgard, W. (2001) 'Robust Monte Carlo localization for mobile robots', *Artificial Intelligence*, Vol. 128, No. 2, pp.99–141.

Usenko, V., Engel, J. and Stückler, J. et al. (2016) 'Direct visual-inertial odometry with stereo cameras', *IEEE International Conference on Robotics and Automation (ICRA)*, pp.1–8.

Wang, D., Liang, H. and Mei, T. (2013) 'Lidar scan matching EKF-SLAM using the differential model of vehicle motion', *Intelligent Vehicles Symposium (IV)*, IEEE, Gold Coast, QLD, Vol. 36, No. 1.

Wei, Y-W. and Zuo, Z-Y. (2009) 'Improvement of the simultaneous localization and map building algorithm applying scaled unscented transformation', *International Conference on Industrial Mechatronics and Automation (ICIMA'09)*, IEEE, Chengdu.

Wu, E., Xiang, Z. and Shen, M. et al. (2007) 'Robot SLAM algorithm based on laser range finder for large scale environment', *Journal of Zhejiang University* (in Chinses), Vol. 41, No. 12.

Wu, M. and Sun, J. (2010) 'Extended Kalman filter based moving object tracking by mobile robot in unknown environment', *ROBOT*, Vol. 32, No. 3, p.334.

Yavuz, S., Kurt, Z. and Bicer, M.S. (2009) 'Simultaneous localization and mapping using extended Kalman filter', *Proceedings of the IEEE 17th Signal Processing and Communications Applications Conference*, IEEE, Antalya.

Zhang, W. (2017) *Research on Autonomous Navigation Method for Indoor Robots Based on Multisensor Fusion* (in Chinses), University of Science and Technology of China, Beijing.

Zhao, Y., Liu, G. and Tian, G. et al. (2017) 'A survey of visual SLAM based on deep learning', *ROBOT*, Vol. 39, No. 6, pp.889–896.

Zhu, K. and Liu, H. (2018) 'Qingyuan Xia: survey on monocular visual SLAM algorithms', *Application Research of Computers* (in Chinses), Vol. 35, No. 1.