Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/41/2024CH0171

Advances in monocular ORB-SLAM system: A review

Ziyi Yuan

School of Advanced Manufacturing, Guangdong University of Technology, Guangzhou, China

3121009096@mail2.gdut.edu.cn

Abstract. Perception and localization are the main factors to determine the success of unmanned vehicles. Therefore, researchers have conducted substantial studies, which made unmanned driving not only to perceive and comprehend the around environments but also refer to the detail about the environments by constructing 3D map. While there is still a lack of uniform explanation of Oriented Fast and Rotated Brief - Simultaneous Localization and Mapping (ORB-SLAM) for monoculars. By selecting and collecting the combination and application of the recent four types of monocular ORB-SLAM in unmanned driving scenarios, this paper discusses the question of how to decrease cumulative error and ensure accuracy and robustness in dynamic environments. It is revealed that after comparing the recent four types of ORB-SLAM systems with conventional ORB-SLAM systems, the fusion system's robustness and accuracy have been improved. Combining visual SLAM sensors with different algorithms and studying in different complex environments will be mainstream in future research.

Keywords: Localization, monocular vision, simultaneous localization and mapping.

1. Introduction

With the fast development of technology, Simultaneous Localization and Mapping (SLAM) widely be used in high-tech industries, such as the robot industry, construction industry, and unmanned vehicles. While it has developed from the traditional SLAM in recent years, SLAM has been divided into two categories, one is based on the laser sensors, which use laser to measure, while the other one is based on the visual sensors. The visual SLAM is mostly using cameras to make measurements, and it can be divided into three categories by their way of working, including monocular, multiocular, and RGB-D cameras.

In recent years, the SLAM has been used to understand the surrounding environments, map around environments and determine the location within the area. By detecting the objective, utilizing the deep estimation and visual SLAM, the perception and measurement of the surrounding environment are realized. There are also examples of applying SLAM techniques to microrobots for minimally invasive surgery. However, there is still a lack of a uniform explanation for the monocular ORB-SLAM system under the visual SLAM system. Through investigating four recent monocular ORB-SLAM systems, and summarizing the research of the current ORB-SLAM system for monoculars, this research discusses how to improve accuracy and robustness to reduce cumulative error in different environments. It is concluded that combining the sensor with different innovational algorithms will effectively decrease cumulative error and guarantee accuracy and robustness in complex environments.

DOI: 10.54254/2753-8818/41/2024CH0171

2. Application of the ORB-SLAM for monocular

ORB-SLAM, is an open-source visual system, about the application of the monocular ORB-SLAM. In recent years, researchers also studied how to use ORB-SLAM in unmanned vehicles, the construction industry, the robot industry and so on. In unmanned vehicles, the ORB-SLAM is used to recognize pavement information and detect road obstacles. ORB-SLAM is also used to investigate and improve drivers' driving habits. In a survey, the author mentioned a method using ORB-SLAM to realize the track of head scanning movement when driving, which aimed at gaining awareness of driving safety technology applications [1].

In the construction industry, there is research about using ORB-SLAM to enhance robot localization in dynamic construction environments, where the construction robots have to do precise positioning work. However, it was difficult to recognize the dynamic objects in previous research and they mainly investigate static objects. With the deepening of SLAM technology studies, the ORB-SLAM system has a breakthrough in accurately segmenting dynamic objects and improving localization accuracy, which justifies the ORB-SLAM potential for applications in complex construction environments.

In the robot industry, the ORB-SLAM system has been proposed for the surgical treatment of microrobots, such as minimally invasive intestinal surgery. there is a survey introduced that minimally invasive surgery has a series of problems in microrobot applications such as low reconstruction accuracy, small surgical field, and low computational efficiency, a framework based on the ORB-SLAM system for real-time dense reconstruction in binocular endoscopy scenes to solve these problems [2].

3. Algorithm based on SLAM for monocular

3.1. Conventional algorithm

The conventional SLAM systems include two main threads to be executed in parallel, which are called tracking and mapping. However, the visual SLAM framework needs to include the following parts: sensor information reading, front-end, back-end, map construction, and closed-loop detection [3]. Sensor information reading recognizes and preprocesses the image information. The front-end is known as visual odometry which is in charge of processing the input images of the previous step and estimating the camera posture at different times. The back-end is called nonlinear optimization, it can receive the camera posture at different times returned by visual odometry and optimize the posture. In addition, the back-end also receives closed-loop detection information and executes the global optimization to obtain globally consistent tracks and maps. The last part is closed-loop detection which is used to certain whether the mobile robot has passed through a previously visited location. The feature of pure visual SLAM tracks the movement of key points through successive camera frames to infer the posture of the camera.

3.2. Conventional algorithm of ORB-SLAM

The conventional Algorithm of ORB-SLAM divides SLAM system into three threads, feature points are attached to them. The ORB-SLAM algorithm is modified based on the Parallel Tracking and Mapping (PTAM) algorithm. The original PTAM algorithm has made a great breakthrough in the conventional visual SLAM, which first proposed the parallelization of the tracking and mapping process, and uses nonlinear optimization to replace the traditional filter as the back-end scheme, introducing a mechanism of keyframes in the PTAM algorithm [3].

The mechanism suggests that each image can be processed without fine processing, instead, it can proceed by connecting several images and then optimizing its tracks and maps. However, the closed-loop detection cannot be performed in the PTAM algorithm. So the scenario it applies in is small and the tracking is easy to lose. Compared to the ORB-SLAM proposed after PTAM, the ORB-SLAM algorithm uses the ORB feature points and its descriptors to detect and track the feature points in the image, and to estimate the camera pose through the resulting feature points. The ORB feature points are a very fast feature extraction method with rotational invariance. The use of uniform ORB features helps SLAM algorithms to have endogenous consistency in the steps of feature extraction and tracking, key

DOI: 10.54254/2753-8818/41/2024CH0171

frame selection, 3D reconstruction, and closed-loop detection [4]. The ORB-SLAM algorithms divide SLAM system into three threads, feature point tracking, spatial mapping, and loop detecting. The advantage of this algorithm is that ORB-SLAM can realize real-time tracking and it is easy to find back the lost keyframe when it returns to the original scene [5]. Besides, using the ORB-SLAM algorithms can effectively improve the positioning stability and track the object in a simple scenario. This algorithm compared with the PTAM algorithm provides more closed-loop detection parts than the PTAM algorithm and can effectively solve the cumulative error problem left by PTAM algorithm.

4. Optimization of ORB-SLAM for algorithm

With the continuous innovation and optimization of the ORB-SLAM algorithm in recent years, this section will introduce four derivative algorithms based on ORB-SLAM. By integrating monocular ORB-SLAM with different methods, its robustness and accuracy in different environments have been improved.

4.1. A graph recovery algorithm

Based on the ORB-SLAM, different progress has been made on monocular visual SLAM. Through the SLAM graph recovery algorithm based on subgraphs and undirected connection graphs, the system uses the mapping connection to re-initialize and reconstruct the individual parts of the map without tracking. The survey shows that by evaluation in drone image simulations and datasets of ground and indoor testing, it is concluded that in the situation of tracking failure, the SLAM graph recovery algorithm based on subgraph and undirected connection graph can make the integrity of the map better than other mainstream SLAM methods, ensuring the map integrity in the unmanned driving under the system of tracking failures [6].

The main breakthrough is that after creating tracking failures in unmanned driving, missing maps can be repaired by creating subgraphs. Then, the integrality of its subgraphs is guaranteed by a new selection method. Finally, the undirected connection graph is used to maintain the connection relationship between the subgraphs. The number of keyframes retained in the UAV environment, and the proposed system is about four times the keyframe retained based on the original ORB-SLAM 2. In an outdoor street environment, the proposed system can effectively reconstruct a more complete scene map [6].

4.2. A semi-direct monocular SLAM with three-level parallel optimization

The conventional visual SLAM method can be divided into the feature-based method and the direct method. The method of feature-base is to extract the feature points from the image data which is received by the camera and analyze the feature points to realize the estimation of the camera posture. However, the method of direct utilizes the photometric error to estimate the posture of the camera, for the reason to effectively combine the advantages of the two methods and achieve more accurate camera pose estimation. The survey proposed a semi-direct monocular SLAM with three-level parallel optimization [7]. In this study, a new framework for SLAM operation called DO-SLAM is explored [7]. The first half part of the DO-SLAM system, by using direct methods to quickly and robustly track the camera pose. While the second half part of the system uses a feature-based approach, to refine the pose of the keyframes, execute the loop, and construct the reusable globally consistent, long-term, and sparse feature maps. The survey, as demonstrated by its evaluation of two benchmark datasets, using this method has higher accuracy and robustness in motion estimation in unmanned driving.

4.3. Optimization of 3D points

Due to the limitations of monocular cameras, the scale of the monocular camera is fuzzy and limited in the system and the environment. It is difficult to accurately measure the depth of the target scene and the distance from the camera, which can reduce the impact on measurement accuracy during unmanned driving. A study proposed a scaling estimation method [8]. By using the method for monocular visual odometers in unmanned driving scenarios, the innovative approach is to use two consecutive keyframes

DOI: 10.54254/2753-8818/41/2024CH0171

to reconstruct the 3D ground points, then use more processing frames to increase the number of 3D ground points, to estimate more precise camera height [8]. This method adopts the mean ORB-SLAM movement error of 1.19% on the KITTI dataset, compared to the state-of-the-art traditional monocular SLAM method.

4.4. HFNet-SLAM

ORB-SLAM 3 is a SLAM system in visual SLAM with feature-based methods and higher robustness and accuracy. A study replaced the vulnerability of traditional algorithms in complex environments, proposed the HFNet-SLAM system [9]. This is an accurate real-time monocular SLAM system based on ORB-SLAM 3. The system is combined with a deep convolutional neural network (CNN). The difference between HFNet-SLAM and conventional ORB-SLAM 3 is that its local and global features are extracted from deep CNN, HF-Net system, while experiments show that even with highly reproducible local features of a deep CNN in complex environments, this is better than the traditional feature extraction. The performance of this system has been validated on public data sets against other state-of-the-art algorithms, the results show that HFNet-SLAM achieves the lowest error among the systems available in the literature.

4.5. Comparison between optimized ORB SLAM algorithms and conventional SLAM

The four different fusion systems that are based on ORB-SLAM introduced above are compared to the traditional visual SLAM system in this section. Firstly, a graph recovery algorithm based on the subgraph and the undirected connection graph is compared with the traditional visual SLAM algorithm in the case of failed system tracking, the fusion algorithm can ensure the integrity of the map in its lost situation and is able to recover previously missing partial maps under the method of creating subgraphs. While in the method based on the three-level parallel optimization, the direct approach with the advantages of the feature approach was combined. In terms of the previous visual SLAM method, which can obtain a more accurate and more robust camera posture estimation method. The approach uses the 3D ground feature points, collects 3D ground points from multiple processing frames, and then utilizes robust scale estimation. Comparing it with traditional visual SLAM effectively reduces its scale drift. Finally, by integrating the deep convolutional neural network (CNN) algorithm with the ORB-SLAM 3 system, compared with the conventional ORB-SLAM system, it shows that the resulting proposed system has higher robustness and accuracy than the previous systems. The fusion system is twice as accurate as the ORB-SLAM3 system in medium and large environments in the TUM-VI dataset [9].

5. Limitations of ORB-SLAM

Monocular visual sensor, compared with multiocular visual sensor and depth camera, the cost is cheaper. However, in the application scenario, it often cannot exactly get the absolute depth of the environment. For a monocular camera, it cannot get the true value of the trajectory and the map size leading to the measured results of certain deviation values. The multiocular camera and depth camera can measure the scene depth, so the aspect of the sensor monocular camera in the application scene still has certain limitations. For example, Kinect Fusion proposed the use of Kinect cameras for 3D reconstruction, and the RGB-D camera is used in ORB-SLAM2 system [10].

In the current experimental survey, most studies concentrated on both static and low-speed environments. While studies in highly dynamic environments are still scarce. The robustness and accuracy of the monocular vision system in highly dynamic or complex environments still cannot be guaranteed [10], such as the method of scale estimation by acquiring 3D ground points [8]. The researchers found that in a curved road or slope, only a less of 3D ground points can be collected, which results in introducing new scale factors for the purpose of correcting the camera posture. This will lead to a measurement error. Therefore, the lack of research on high-speed dynamic and complex scenarios is one of the reasons for the limitation of the application of monocular vision systems in different scenarios

DOI: 10.54254/2753-8818/41/2024CH0171

Although a lot of research in the direction of feature extraction, the monocular visual system still has some limitations in the extraction of feature points, such as integrating the deep convolutional neural network algorithm with the ORB-SLAM [9]. It utilizes the deep learning method but still found in the process of the experiment, the system will break in extreme rotation of the situation, and the use of neural network method needs to use the support of external equipment, which undoubtedly increases the burden of mobile robots.

6. Tendency and improvement

Through the above research and analysis, the future development trend and research direction of the monocular visual system can be seen. One direction proposed in this survey is that it can be expanded in the existing ORB-SLAM algorithm to support multiocular visual SLAM and RGB-D SLAM, because both are more robust and accurate than monocular visual SLAM. The deep convolutional neural network can be used as its feature extraction method, and then similar methods to solve the drift problem, to ensure the stability of the whole system and reduce its error. This method combines the above methods and it combines some of their advantages.

Experimental studies in highly dynamic and complex scenarios are scarce. So for future research, a visual SLAM system in various dynamic scenarios is needed. Through the exploration and study of complex scenes and highly dynamic environments, it is noticed that robustness can be achieved by repeating the subgraph or improving the algorithm while improving the accuracy and robustness of the visual SLAM system. By combining the visual SLAM system with the emerging algorithm, achieving a much lower systematic error will be the future research direction.

7. Conclusion

This paper discusses and analyzes four recent ORB-SLAM systems combined with other innovational algorithms, based on their differences from the conventional visual SLAM system. It is proposed that in complex conditions, different algorithms with sensors should integrate with the monocular ORB-SLAM, such as graph recovery algorithm for subgraph and undirected connection graph, three-level parallel optimization method, and feature extraction method for constructing 3D ground points. An emerging system combining the ORB-SLAM system with a deep convolutional neural network can make visual SLAM have a prominent improvement in accuracy and robustness but also decrease the cumulative error. Investigating during the study, current research on ORB-SLAM systems in highly dynamic environments is scarce, and there is no study for monocular ORB-SLAM systems to apply in complex environments. In the subsequent scientific studies, researchers can focus on studying the visual SLAM system in highly dynamic environments and successively improve the ability to extract feature points in the complex environment, to improve the accuracy and robustness of the monocular ORB-SLAM system and reduce the cumulative error.

References

[1] Wang, S., Li, J., Yang, P., Gao, T., Bowers, A. R., & Luo, G. (2020). Towards Wide Range Tracking of Head Scanning Movement in Driving. International journal of pattern recognition and artificial intelligence, 34(13), 2050033. https://doi.org/10.1142/s0218001420500330

[2] Huo, J., Zhou, C., Yuan, B., Yang, Q., & Wang, L. (2023). Real-Time Dense Reconstruction with Binocular Endoscopy Based on Stereo Net and ORB-SLAM. Sensors (Basel, Switzerland), 23(4), 2074. https://doi.org/10.3390/s23042074

[3] Zhu P, Zhou H, Zhang H, Lu S& Wei R. Visual simultaneous localization and mapping method for a mobile robot. JOURNAL OF TIANJIN UNIVERSITY OF TECHNOLOGY1-10.

[4] Tourani, A., Bavle, H., Sanchez-Lopez, J. L., & Voos, H. (2022). Visual SLAM: What Are the Current Trends and What to Expect? Sensors (Basel, Switzerland), 22(23), 9297. https://doi.org/10.3390/s22239297

DOI: 10.54254/2753-8818/41/2024CH0171

[5] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," in IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, Oct. 2015, doi: 10.1109/TRO.2015.2463671.

[6] Z. Zhan, W. Jian, Y. Li and Y. Yue, "A SLAM Map Restoration Algorithm Based on Submaps and an Undirected Connected Graph," in IEEE Access, vol. 9, pp. 12657-12674, 2021, doi: 10.1109/ACCESS.2021.3049864

[7] S. Lu, Y. Zhi, S. Zhang, R. He and Z. Bao, "Semi-Direct Monocular SLAM With Three Levels of Parallel Optimizations, " in IEEE Access, vol. 9, pp. 86801-86810, 2021, doi: 10.1109/ACCESS.2021.3071921

[8] M. Fan, S. -W. Kim, S. -T. Kim, J. -Y. Sun and S. -J. Ko, "Simple But Effective Scale Estimation for Monocular Visual Odometry in Road Driving Scenarios," in IEEE Access, vol. 8, pp. 175891-175903, 2020, doi: 10.1109/ACCESS.2020.3026347

[9] Liu, L., & Aitken, J. M. (2023). HFNet-SLAM: An Accurate and Real-Time Monocular SLAM System with Deep Features. Sensors (Basel, Switzerland), 23(4), 2113. https://doi.org/10.3390/s23042113

[10] Bala, J. A., Adeshina, S. A., & Aibinu, A. M. (2022). Advances in Visual Simultaneous Localisation and Mapping Techniques for Autonomous Vehicles: A Review. Sensors (Basel, Switzerland), 22(22), 8943. https://doi.org/10.3390/s22228943