Recent Advances in Visual SLAM Algorithms in Dynamic Environments

Mingyang Li

Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, China

202164010288@mail.scut.edu.cn

Abstract. This paper reviews the latest research progress of visual SLAM in dynamic scenes. The improved algorithms dependent on deep learning and multi-sensor fusion are discussed. First of all, the challenges of traditional visual SLAM in dynamic environments are stated, especially the interference of high dynamic objects on system feature extraction, pose estimation and mapping. Then, the application of several object detection technologies such as YOLOv5 and YOLOv7 in dynamic visual SLAM was analyzed in detail. These methods significantly improve the positioning precision and robustness of the system by identifying and eliminating dynamic feature points in the scene. In addition, this paper explores the visual SLAM system combining semantic segmentation and geometric constraint optimization, and finally introduces the multi-sensor SLAM method combining IMU data, point cloud data and visual information. Experimental results show that these improved algorithms have a significant improvement in positioning accuracy and system robustness compared with traditional visual SLAM algorithms under TUM and other dynamic data sets.

Keywords: Visual SLAM, dynamic scenes, object detection, semantic segmentation, multi-sensor fusion.

1. Introduction

Visual Simultaneous Localization and Mapping (SLAM) is a technology in which cameras are used as external sensors for self-positioning and map construction [1]. For the past few years, the application of visual SLAM in dynamic environments has received wide attention. With the continuous improvement of computer vision and deep learning technology, the ability of visual SLAM algorithms to locate and construct maps in complex environments is also being strengthened. Most existing visual SLAM algorithms are based on static scenes, which fall roughly into two categories: the feature point method is represented by Parallel Tracking and Mapping (PTAM) and ORB-SLAM2, and the direct method is represented by Direct Sparse Odometry (DSO) and Deep Direct Dense Visual Odometry(D3VO) [2]. Based on the assumption of luminosity consistency, the direct method utilize the luminosity information of pixels in the input image as the luminosity error and is used to construct the dense map. Feature point method usually uses feature extractors and descriptors to extract feature points, and relies on feature points to match and build sparse maps [1].

However, most visual SLAM algorithms are dependent on static environments, but in daily actual scenes, there will be high dynamic objects (like vehicles, people, and animals) and some potential low dynamic objects (like temporarily moved objects). The current mainstream visual SALM algorithms such as ORB-SLAM2 perform well in static scenes and environments with a few dynamic factors, and they can achieve extremely high accuracy and robustness. However, when they deal with highly dynamic objects in the environment, they are prone to errors and uncertainties, and the algorithm performance will be significantly reduced.

Therefore, in order to meet the challenges in dynamic scenarios, researchers have proposed various of improvement strategies to improve the accuracy and robustness of visual SLAM algorithms. Some researchers combine deep learning techniques (semantic segmentation, object detection model) with visual SLAM. Wang Yan and her partners proposed an indoor dynamic visual SLAM algorithm combined with semantic segmentation [3]. The algorithm uses Seg-T-Mask network for dynamic point filtering, combined with geometric constraints to improve the localization accuracy of the SLAM system in a dynamic environment. The similarity matching of loop detection is improved by processing the semantic information of image frames. The semantic point cloud map building module is added, the key frame selection is optimized, redundancy and dynamic ghosting are reduced, and the map clarity and expressiveness are improved. Other researchers use multi-sensor fusion methods to optimize visual SLAM. Wang Hongxing and his colleagues improved the SLAM system by integrating Inertial Measurement Unit (IMU) and visual information [4]. Firstly, the visual odometer is combined with IMU data to optimize the pose of the sensor carrier, solving the trajectory drift faced by the pure inertial navigation system. Secondly, in order to solve the frequency difference between IMU and visual data, the study introduced pre-integral processing of IMU data, which significantly improved the robustness of the system. In the vision /IMU joint pose optimization, rotation and shift constraints between the camera and the IMU are applied to initialize the system jointly. Then, the tightly coupled nonlinear optimization framework is constructed to optimize the state by obtaining the residual function. Finally, the real-time performance of the system is optimized by the sliding window strategy.

Based on the above research status, this paper will review the latest research progress of visual SLAM, focusing on algorithm improvement based on deep learning, geometric constraints and multi-sensor fusion. The main contribution of this thesis is to summarize and analyze the advantages and disadvantages of these different methods, so as to serve as a reference for future research and improvement.

2. Challenges and requirements of visual SLAM in dynamic scenarios

Traditional Visual SLAM in dynamic scenarios has many shortcomings. Various from the static environment, there are many high-dynamic moving objects and low-dynamic potential moving objects in dynamic scenarios, which will affect the feature extraction process to some extent, thus interfering with the subsequent pose estimation and map construction process. To ensure both the precision and real-time capability of visual SLAM in a dynamic environment, the researchers have adopted different improvement measures to improve the performance of the system.

2.1. Features of dynamic scenarios

The main feature of a dynamic scene is the existing moving objects in the environment (such as pedestrians, animals, vehicles, etc.), and the movement of these objects will interfere with extracting the feature and matching in the system. In systems such as ORB-SLAM2, the extraction and matching of feature points rely on the static background in the received images. Therefore, the movement of objects will cause the instability of the extracted feature points, generating the wrong matching of feature points, which will interfere with the subsequent pose estimation and global map construction [2]. At present, the main problems faced by SLAM systems in dynamic environments are the drift and inconsistency of feature points, which make it difficult for the system to distinguish between dynamic objects and static backgrounds, resulting in the extracted feature points containing a lot of error information.

2.2. The influence of dynamic scenes on SLAM

Moving objects in dynamic scenes will not only affect the extraction of feature points, but also affect pose estimation and map construction, and even reduce the positioning precision of the system. In terms of pose estimation and map construction, the SLAM system estimates the pose change of the camera by matching multiple feature points of different frames at different times. However, the drift of feature points caused by dynamic objects will introduce noise, resulting in a deviation of pose estimation [5]. In addition, moving objects may leave "ghosting" or error messages in the map, which affects the precision of the map constructed and degrades the performance of the whole system. For the localization precision and robustness of the SLAM system, the uncertainties in dynamic environments greatly increase the localization complexity of the SLAM system. Experiments demonstrate that the Absolute Trajectory Error (ATE) and Relative Pose Error (RPE) of traditional visual SLAM increase significantly in dynamic environments, which means that the localization precision of the system decreases significantly. For example, Shi Tao and his partners showed that the localization error of the traditional ORB-SLAM2 algorithm would increase significantly under highly dynamic sequences in the TUM dataset [5]. Their results evinced that under the static sequence "Sitting_static", the ATE root mean square error of ORB-SLAM2 is small and remains at about 0.0086 meters, while under the dynamic sequence "Walking_xyz", the Root Mean Square Error (RMSE) increases significantly and reaches about 0.7037 meters [5].

2.3. Requirements to handle dynamic scenarios

In dynamic scenes, visual SLAM systems not only need to efficiently detect dynamic objects and filter dynamic feature points, but also need to deal with complex computing tasks on the basis of ensuring real-time performance. However, although many deep learning models such as Mask Region-based Convolutional Neural Networks (Mask R-CNN) can effectively segment dynamic objects through semantic segmentation, their computational overhead is too large to meet the needs of real-time SLAM [2]. Therefore, how to improve the positioning accuracy while reducing the amount of system computation to optimize the real-time performance of the system has become an urgent problem to be solved in SLAM in dynamic scenes. To sum up, in dynamic scenarios, visual SLAM systems must find a balance between precision, real-time performance, and computational resources. This requires researchers not only to consider how to improve the accuracy of object detection and pose estimation, but also to optimize the complexity of the algorithm to adapt to the needs of various hardware platforms when developing new algorithms.

3. Improved method of visual SLAM based on deep learning

3.1. Application of YOLOv5, YOLOv7 and other object detection technologies in dynamic SLAM

In the visual SLAM system in dynamic scenes, the object detection technology based on the YOLO series algorithm can accurately and quickly identify the dynamic object and remove the dynamic feature points on it, so as to improve the positioning accuracy and robustness of the system. As popular object detection algorithms, YOLOv5 and YOLOv7 show excellent application effects in different SLAM frameworks. This section will introduce the improved visual SLAM scheme based on the YOLO series object detection algorithm in detail.

3.1.1. Improved algorithm based on YOLOv5. Wang Hongyu and his partners added an object detection thread and a module for eliminating dynamic feature points to the visual odometer of the original ORB-SLAM2 system [2]. In the object detection thread, the researchers chose YOLOv5s, which has the smallest model and the fastest running speed among the YOLOv5 series [2]. To enhance the real-time performance of the SLAM system, Wang and his partners substituted the Backbone layer of YOLOv5s with MobileNetV3-Small and modified the feature maps of each layer for better alignment [2]. Once the dynamic feature points are identified by the object detection thread, the system will eliminate the feature points within the vicinity of dynamic objects. However, the shape of moving objects in reality is mostly irregular, and the projection box produced by the object detection thread is a regular rectangular area, which will make the original static feature points around the dynamic object be wrongly removed. Therefore, the researchers decided to remove the feature points which are dynamics again by combining the epipolar geometry constraint and LK optical flow method to make sure as few static feature points as possible are wrongly removed. For the epipolar geometric constraint method, researchers evaluate the distance between the quasi-static feature point and the epipolar line against a predefined threshold range. If the distance surpasses the threshold range, it is regarded as a dynamic feature point and is removed. However, there are exceptions, that is, when the moving direction of the object is parallel to the camera, the conditions of the epipolar geometry constraint can also be satisfied. Therefore, the LK optical flow method should be utilized to further remove feature points which are dynamics.

In order to evaluate the performance of the improved algorithm, Wang and partners used the TUM RGB-D dataset and selected the "fr3_walking" high dynamic sequence and "fr3_sitting" low dynamic sequence in the dataset for testing. The ATE and RPE are used to evaluate the performance of the algorithm. In terms of ATE, the RMSE, Mean and STD of the four groups of highly dynamic sequences are greatly improved compared with the ORB-SLAM2 algorithm before the improvement, and the errors are reduced by more than 90%. Among them, the visual SLAM algorithm based on YOLOv5 and geometric constraints (the improved algorithm) performs best in the Walking_static sequence, with RMSE, Mean and STD increased by up to 97.79%, 97.88% and 97.35% respectively.

Huang Yourei and his colleagues proposed a dynamic visual SLAM algorithm combined with lightweight YOLOv5s, taking ORB-SLAM3 as the main framework, and mainly focusing the improvement on the tracking threads [1]. In their paper, a lightweight YOLOv5s object detection thread and a dynamic feature point elimination module are incorporated into the tracking thread of ORB-SLAM3. Firstly, the prior information obtained from the object detection model is used to detect the dynamic target and potential dynamic target in the image frame, and the corresponding anchor coordinates are output. Then, in the dynamic feature point elimination module, the LK optical flow method is used to judge the feature points in the dynamic target anchor frame, and the judged static feature points are used for the subsequent pose optimization. In the YOLOv5s object detection thread, the researchers combined the Ghost module to replace the ordinary convolution and C3 module in the network with the lightweight GhostConv and C3Ghost modules respectively, and the SimAM attention mechanism was integrated into the backbone network to improve the network's capability to capture key features. Thus, the accuracy of model checking after lightweight is improved.

Similar to Wang Hongyu, Huang Yourui and his colleagues selected five image sequences of sitting_static, sitting_xyz, walking_half, walking_static, and walking_xyz in the TUM dataset. The two sequences sitting_static and sitting_xyz are low dynamic sequence data sets, and the three sequences walking_half, walking_static and walking_xyz are high dynamic sequence data sets. The evaluation metrics of the functionality of the algorithm are also used in ATE and RPE. Among them, the improved algorithm proposed by Huang Yourui improves the ATE by 89.29%, 65.34%, and 94.43% respectively compared with ORB-SLAM3 algorithm in walking_half, walking_static and walking_xyz high dynamic sequences. Moreover, RPE increased by 39.58%, 51.41%, and 52.36% respectively. The experimental results demonstrate that the improved algorithm proposed by Huang Yourui has a significant improvement over the ORB-SLAM3 algorithm, and the positioning precision of the system is higher.

3.1.2. Improved algorithm based on YOLOv7. Shi Tao and his partners added a module which is utilized to remove dynamic feature points to the framework of ORB-SLAM2. The improved YOLOv7 lightweight object detection network was utilized to identify dynamic objects, integrating multi-view geometric constraints to eliminate dynamic feature points. Subsequently, the remaining static feature points were employed to estimate the camera's pose [5]. In the object detection procedure of the visual SLAM system proposed by Shi Tao, in order to reduce the computational cost of the system, the research team replaced the ordinary convolution layer with the more lightweight GhostConv convolution to reduce the model parameters and computational complexity. At the same time, it combined Batch Normalization (BN) layer and Sig-Moid Linear Unit (SiLU) activation function. Thus, new GBS modules are formed to perform feature extraction. After the improvement, the model parameters are reduced from the original 37.62 million to 36.01 million. Simultaneously, in order to improve the system's ability to capture dynamic information, Shi Tao and his colleagues combined the squeeze-and-excitation (SE) mechanism with the convolution operation, proposed the Conv_SE module, and applied it to the effective long-range aggregation network (ELAN) aggregation network in Backbone. Through these improvements, the detection accuracy of the model was increased from 95% to 96.8%. In the real environment, there may exist some potential dynamic points (such as chairs moved by people) that cannot be detected by the object detection model. To solve this problem, Shi Tao used epipolar geometric constraints to further identify the remaining potential dynamic feature points. Finally, for static objects and unrecognized surrounding environments, the original feature point extraction method of ORB-SLAM2 system was adopted, and the extracted feature points were the static feature points for subsequent pose estimation and local map construction.

To testify the functionality of the system, Shi Tao and his colleagues performed algorithm validation on the TUM dataset. The verification results show that under the dynamic Walk sequence, the RMSE of ATE of the improved SALM algorithm proposed by the researcher is reduced by 96.5% on average compared with ORB-SLAM2, and the RMSE of the translation part of the RPE is reduced by more than 90% on average. It can be seen that Shi Tao's improved algorithm exhibits superior performance compared to ORB-SLAM2 in dynamic scenes, and can run more stably in dynamic scenes.

Yinzhen Liu and his partners proposed a SLAM system based on the combination of Semantic and Geometric constraints (Dynamic Semantic Geometric SLAM) to deal with the challenges caused by Dynamic scenes [6]. DSG-SLAM adds a parallel semantic thread to obtain 2D semantic information based on ORB-SLAM2, and adds a rapid dynamic feature elimination algorithm in the tracking thread combining semantic and geometric constraints. Among them, the 2D semantic information is provided by the object detection network, and the researchers introduce the lightweight object detection network GhostNet-YOLOv7 into the DSG-SLAM system. Ghostnet-yolov7 replaces its backbone feature extraction network with a lightweight GhostNet network under the framework of YOLOv7, thereby reducing the complexity of the YOLOv7 model and the number of parameters of the model. After obtaining 2D semantic information, the tracking thread in the DSG-SLAM system will combine geometric constraints and 2D semantic information to eliminate dynamic feature points.

Liu evaluated DSG-SLAM in the TUM public dataset and the real environment. The results demonstrate that for the highly dynamic sequences in the dataset, the RMSE and standard deviation (SD) of ATE and RPE of DSG-SLAM are mostly improved by more than 90% compared with ORB-SLAM2 algorithm. Obviously, DSG-SLAM has good localization accuracy and robustness in highly dynamic scenes. However, for low dynamic sequences, the improvement of RMSE and SD of ATE and RPE of DSG-SLAM algorithm is only about 10%. This is because in low dynamic sequences, the majority of the objects are static, and moving objects account for a small proportion of the environment, so the amelioration of DSG-SLAM is not evident compared with ORB-SLAM2.In terms of real-time performance, the operating frequency of the system reaches 30 Hz, which shows that the system effectively improves positioning accuracy in dynamic scenes under the premise of ensuring real-time operation.

3.2. Application of semantic segmentation technology in dynamic SLAM

Fu Hao and his partners proposed a SLAM algorithm combining semantic segmentation and optical flow constraints, which added dynamic object detection and semantic map construction threads to the original thread of ORB-SLAM2 [7]. Firstly, Fu Hao employed the DeepLabv3 semantic segmentation network to carry out semantic segmentation on the input RGB-D image, so as to obtain the mask of the object in the scene [7]. Then, the dense optical flow between the current frame and 5 frames before is calculated by sliding window method. Finally, the dynamic probability of the object was obtained by combining semantic segmentation and optical flow, and the static feature points that could be used to predict the camera pose were obtained by filtering the feature points according to the dynamic probability. The researchers tested the algorithm on the TUM dataset and used two indicators of ATE and RPE to evaluate the performance of the algorithm in localization [7]. The experimental outcomes show that in high dynamic environments, Fu Tao's algorithm can achieve more than 95% improvement in ATE and RPE compared with ORB-SLAM2, and reduces the absolute trajectory error by 41% and 11% respectively compared with DS-SLAM and DynaSLAM. This indicates that the proposed algorithm offers better localization precision and robustness in highly dynamic environments. However, the real-time performance of the system needs to be improved. The semantic segmentation module has the greatest influence on the real-time performance of the algorithm, and its running time is 169.4 milliseconds. Unlike traditional object detection algorithms, the task of semantic segmentation is to classify each pixel, so the computational complexity is much higher than object detection algorithms like the YOLO family.

Haochen Jiang and his colleagues proposed a semantic prior-based motion consistency detection algorithm with weighted epipolar lines and depth constraints [8]. The algorithm added a semantic segmentation thread dependent on the original framework of ORB-SLAM2, and added operations such as initial pose calculation, motion consistency detection and motion point elimination to the original tracking thread of ORB-SLAM2. The semantic segmentation thread uses the Light Weight RefineNet as the semantic segmentation network of the system, which performs semantic segmentation on the objects in the image to obtain the image mask of the potential moving object, so as to estimate the camera's initial pose. Then, the weighted epipolar line that is dependent on semantic prior and the motion consistency detection algorithm based on depth constraint is utilized to judge the "candidate feature points" on the potential motion object, and the feature points that meet the conditions are put back into the original static feature point set. Finally, the updated "exact pose" and static feature points are handed over to the back-end optimization. Haocen Jiang conducted multiple comparative tests on 9 dynamic scene sequences from the TUM dataset and 3 image sequences from Bonn complex dynamic environment dataset. Compared with the existing advanced dynamic SLAM system DS-SLAM, the RMSE of ATE of the improved algorithm is reduced by 10.53% to 93.75%. For the translation and rotation RPE, the RMSE of the improved algorithm achieves a decrease of 73.44% and 68.73% at most [8]. The experimental outcomes indicate that the improved method can dramatically improve the pose estimation precision and robustness of the visual SLAM system.

4. Improved visual SLAM method based on multi-sensor fusion

Bao and his partners proposed a radar-camera fusion SLAM algorithm in order to solve the problem of low mapping and positioning precision of traditional lidar-based SLAM systems in dynamic scenes [9]. The main ideas of the algorithm are as follows: data acquisition is performed first, where the lidar provides 360° point cloud data, while the camera is used to obtain image information. Second, data preprocessing is performed, which consists of IMU data pre-integration and laser point cloud distortion correction. The curvature of each laser point is calculated to determine if the point cloud belongs to a corner feature point or a plane point feature point. At the same time, YOLOv5 is used to detect and segment the camera image. Then, the point cloud is projected to the pixel coordinate system to combine the segmentation results of YOLOv5 to eliminate the point cloud data of dynamic objects. Finally, according to the original process of LIO-SLAM, feature extraction, factor graph optimization, map construction and loop detection are carried out to obtain the final mapping result after removing the dynamic point cloud. In general, Bao and his partners' paper uses camera-based vision processing techniques based on LIO-SAM to help lidar SLAM algorithms acquire semantic information of the environment. Bao used the open-source dataset KITTI to verify the improved algorithm and found that the mean APE of the proposed algorithm was 3.48% lower than LIO-SAM, the median was 4.85% lower than LIO-SAM, and the root Mean Square error was 2.86% lower than LIO-SAM. In conclusion, the radar-camera fusion SLAM algorithm proposed by Bao Baizhong and his partners has better trajectory localization accuracy than LIO-SAM.

Wang and his colleagues proposed a monocular vision /IMU fusion localization method based on dynamic target filtering. Firstly, the semantic segmentation method was utilized to filter the dynamic target, and then the rest static points in the image and the information collected by the IMU were input into the monocular /IMU fusion SLAM to calculate the camera pose and realize the map construction [10]. Given the outstanding feature extraction ability of the Mask R-CNN network and the good target detection effect, Wang decided to use this network to realize semantic segmentation and then complete the task of dynamic target elimination. Semantic segmentation is to process the image in pixels and segment the image into regions with semantic meaning. Then the Mask of the dynamic target is generated according to the semantic information and the mask is compared with the original image to filter the dynamic target. To keep costs down, the researchers used a monocular camera for data acquisition. However, the monocular camera cannot obtain the depth information directly or indirectly through calculation, and the lack of depth information will lead to the wrong judgment of the size of the object in the environment, which greatly reduces the precision of the pose estimation of the SLAM system. Therefore, Wang and his partners exploit a tightly coupled method to realize monocular /IMU fusion localization, that is, the state of both camera and IMU is combined, so that they jointly participate in SLAM observation and the construction of motion equation. Among them, IMU is an inertial measurement unit, and the scale information obtained by its integration can assist visual SLAM to estimate the scale of the object in the image, so as to improve the positioning accuracy of the system. In order to testify the functionality of the improved algorithm, Wang Zhe used the Euroc Machine Hall public data set to simulate and verify the algorithm. The experimental outcomes are as follows: in the dynamic environment, the maximum absolute pose error (APE) of monocular SLAM is 5.77, and the error drops to 1.43 after introducing the dynamic filtering method. After further fusing the IMU with the monocular camera, the absolute pose error decreases from 1.43 to 0.84. This shows that the monocular vision/inertial fusion positioning method based on dynamic target filtering proposed in this paper reduces the system drift to a certain extent and improves the overall positioning accuracy of the system.

5. Conclusion

Addressing the challenges encountered by visual SLAM in dynamic scenes, this paper combines a number of the latest research literature and analyzes the improvement effect of deep learning technology and multi-sensor fusion method on visual SLAM. This paper first discusses the improved visual SLAM methods based on object detection models (YOLOv5, YOLOv7, etc.). Many existing literature results show that the lightweight YOLO-based object detection model brings out good accuracy and real-time performance in dealing with high dynamic circumstances, especially in the case of large target motion range and severe illumination change. Secondly, this paper discusses the improved method of visual SLAM which is dependent on semantic segmentation. By comparing multiple sets of experimental data, researchers find that the SLAM system combined with semantic segmentation shows significant advantages in the filtering of dynamic targets, which can effectively improve the accuracy and robustness of system positioning. Finally, the improved SLAM method based on multi-sensor fusion is analyzed. By combining visual information with the data of IMU, lidar and other sensors, the improved SLAM system can perform more stability in complex dynamic environments, and significantly reduce the error accumulation of the system due to visual occlusion or short-term failure.

Although the existing improved visual SLAM has made remarkable progress, there are still many areas for further research in the future. Firstly, the deep learning model can be improved to prevent the decline of detection accuracy to the greatest extent on the basis of ensuring lightweight, so as to adapt to the resource-constrained embedded platform. Secondly, the adaptability of the SLAM system should be strengthened, so that it can work stably in a more complex dynamic environment. Finally, it is proposed to try to integrate more environmental perception information (such as point cloud data, IMU data, and image processing data) into the SLAM system to build a more robust dynamic visual SLAM solution.

References

[1] Huang Y. R., Wang Z. F., Han T., & Song H. P. (2024). Dynamic visual SLAM algorithm combining lightweight YOLOv5s. Electronic Measurement Technology, 1-12.

[2] Wang H. Y., Wu Y. Z., Chen L. J., & Chen X. (2024). Visual SLAM algorithm based on YOLOv5 and geometric constraints in dynamic scenes. Packaging Engineering, (03), 208-217. https://doi.org/10.19554/j.cnki.1001-3563.2024.03.024.

[3] Wang Y. (2023). Research on indoor dynamic visual SLAM and loop detection based on semantic segmentation (Master’s thesis, Xi'an University of Technology). https://doi.org/10.27398/d.cnki.gxalu.2023.000286.

[4] Wang H. X. (2021). Research on visual SLAM-IMU integrated navigation system (Master’s thesis, Harbin Engineering University). https://doi.org/10.27060/d.cnki.ghbcu.2021.000594.

[5] Shi T., Xiao N. Z., Ding Y., & Xu J. D. (2024). Visual SLAM algorithm integrating improved YOLOv7 in dynamic scenes. Foreign Electronic Measurement Technology, (07), 90-96. https://doi.org/10.19652/j.cnki.femt.2406101.

[6] Liu Y. Z., Xu X. R., Zhang H., & Yu Q. S. (2024). Visual SLAM algorithm based on semantic and geometric constraints in dynamic scenes. Information and Control, (03), 388-399. https://doi.org/10.13976/j.cnki.xk.2024.3089.

[7] Fu H., Xu H. G., Zhang Z. M., & Qi S. H. (2021). Visual simultaneous localization and mapping in dynamic scenes based on semantic and optical flow constraints. Computer Applications, (11), 3337-3344.

[8] Jiang H. C., Liu Y. Q., Peng J. Q., Li J. M., Zhu D. C., & Zhang X. L. (2021). RGB-D SLAM algorithm based on semantic prior and depth constraints in indoor dynamic scenes. Information and Control, (03), 275-286. https://doi.org/10.13976/j.cnki.xk.2021.0167.

[9] Bao B. Z., Zhan X. B., Yu D., Si Y., Duan J., & Shi T. L. (2024). Dynamic SLAM algorithm based on radar and camera fusion. Instrumentation Technology and Sensors, (07), 105-109.

[10] Wang Z., & Xiu C. D. (2023). A visual/IMU fusion localization method in dynamic environments. In Proceedings of the 17th National Conference on Signal and Intelligent Information Processing and Applications (pp. 180-184). Beijing University of Aeronautics and Astronautics.