Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5581788, 10 pages
https://doi.org/10.1155/2021/5581788

Research Article

Semantic Optimization of Feature-Based SLAM

Peng Li, Lili Yin, Jiali Gao, and Yuezhongyi Sun

School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, Heilongjiang, China

Correspondence should be addressed to Peng Li; printing3d@126.com

Received 30 January 2021; Revised 10 March 2021; Accepted 24 March 2021; Published 13 April 2021

Academic Editor: Zain Anwar Ali

The purpose of this paper is to provide reasonable recommendation and removal of inappropriate information for SLAM (Simultaneous Localization and Mapping) technology based on feature method. The methodology is to propose a semantic recognition of environment objects in the natural scene through object detection, which is a kind of bag of word method in SLAM problem between the key frames and object level, the method of establishing key frames, and the relationship between the target object levels, through the practical significance of the target object level to judge the merits of the target object level information, and then combined with key frames in the visual SLAM relations with relevant information, so as to get object level targets in each key frame and the relationship between the relevant information, so as to achieve through the object level semantic information to judge the merits of the key frames and screening, as well as to the key frames to judge the merits of the relevant information and screening purpose. The finding of the study is the above method can retain the information of high reliability and good stability for visual SLAM and process the key frames with poor reliability or low stability and the information related to key frames.

1. Introduction

For visual SLAM, in order to ensure the stability of positioning and mapping, key frames with good stability and relevant information should be retained as far as possible [1]. The relevant information here refers to the information used for mapping and map-ping correlation calculation in visual SLAM, since the front-end of visual SLAM (visual odometry) is divided into direct front-end and indirect front-end. For the former, this relevant information, namely, feature points, such as SIFT, SURF, and ORB, are effective feature points to be extracted. For indirect visual SLAM, the relevant information is the bright ness represented by a single pixel. In these SLAM methods, it is assumed that all motion estimation is carried out under a relatively ideal premise; that is, the information has good invariability with the change of time and space. However, in the actual production or living environment, neither the object target at the macro level nor the brightness of a single pixel at the micro level is constant. For the direct method, some researchers have proposed that the automatic calibration of luminosity can provide good pretreatment results for the visual odometer. However, in the indirect method, since the feature points have good rotation and scale invariance, and the brightness change in a certain range will not have a great impact on them, there is no need to preprocess for their constant luminosity. However, in actual activities, any object is not immutable, such as pedestrians on the roadside and books on the desk, which are possible moving visual targets. When the feature points collected in the process of positioning fall on these targets, the results calculated by the system will be abnormal.

The method proposed in this paper hopes to carry out semantic understanding of relevant information in the image through the object detection method and filter the key frames and relevant information obtained to eliminate highly dynamic objects and unstable objects in semantic concept, so as to improve the positioning robustness of SLAM system.

2. Related Work

In WACV2018, Zhong et al. used the method of target detection to SLAM remove dynamic points. In order to synchronize the target detection process with SLAM process, only the key frames were detected, and then the method of feature matching probability was extended to the moving

process to remove the influence of dynamic points. In the SLAM process, only the static points were tracked, and the slam mapping was projected onto the image. In order to improve the effect of target detection, the samples which are difficult to mine are used as training data [2]. In the research results of IROS2018, Yu et al. proposed a method of realizing semantic SLAM for dynamic environment, ds-slam, which combined semantic segmentation with motion consistency detection to remove dynamic points in the environment. In this method, the boundary of the object is obtained by semantic segmentation. If the dynamic points are located in the object according to the motion consistency detection, all the points in the object are denoted as dynamic points [3].

ECVV2018 port of articles, Shen Shao jieren team was proposed based on stereo vision the semantics of the 3D objects and track the autopilot, under the background of autonomous study the SLAM problem in dynamic environment, the Faster-R-CNN method for object detection using binocular camera, and the semantic information fusion to the solution of the unified optimization framework. ORB feature points can be divided into the background and object, first, use the background to estimate the maximum likelihood of the position of the camera and the position of the landmark point. After entering the posture tracking task of the camera, the trajectory is transformed into a priori object size information and semantic information; it is guided by the prior information. Next, the maximum posterior probability estimation method is used to estimate the position of the target, and finally the maximum posterior probability is transformed into a least squares problem to solve [4]. InIEEE2019, Wang et al. proposed SalientDSO, which changed the strategy of DSO [5] on the uniform selection of tracking points. First, the significance graph of the image was extracted through SalGAN [6] network, which was related to the attention of each pixel. Then, the image was semantically segmented by PSPNet [7], and the significance score of each pixel was adjusted by the semantic segmentation results to reduce the significance score of the region without information. Finally, the image was divided into $k \times k$ grids, and the median value of significance was calculated for each grid as the basis for each grid to be selected, and then the points of concern were further selected according to DSO in each grid. After the improvement, the change of point to light and angle of view in the significance region is more robust [8]. In 2018, on issues related to vehicle navigation, Ganti et al. proposed visual SLAM, where feature selection is determined by network uncertainty. In this paper, information entropy is calculated to determine whether the observed data is used to update the estimated state quantity, and the uncertainty of semantic segmentation is integrated into the information entropy and calculation. While calculating the entropy change of pose, if the category uncertainty of a feature point in semantic segmentation is higher, the entropy change of pose is lower and the feature is less easy to be selected. Through this method, the feature points are screened, the feature points with less information are removed, and the scale of the map is reduced without losing much accuracy [9]. In the research of ITSC2017, Murali et al. proposed a method of semantic landmarks to assist accurate vehicle navigation. By adding threshold signals to the factor graph, object categories were distinguished, and corresponding observation data were added to determine whether the corresponding observation data should be added, so as to remove relevant feature points of fixed objects [10].

In IEEE2018 robotics and automation journal, Bescos et al. proposed dylam, which meticulously processed the dynamic points of rgb-d input images, not only removing dynamic objects, but also restoring the background blocked by dynamic objects. For objects with motion characteristics, the method of no motion problem is adopted. The overlapping degree selected is the first five key frames of each current frame. When the difference between the projection point and depth value of the feature point in the current frame is calculated, when the difference between the feature point and the corresponding position in the depth map exceeds a certain threshold, the characteristic point will be determined as the dynamic point. In addition, the pixels with the same depth value around the dynamic pixel are also set as dynamic pixels. Finally, in order to prevent the accuracy of edge segmentation from causing the classification of points in the background as dynamic points, points with large variance in depth around dynamic pixels were set as static points [11]. Brasch IROS2018 article puts forward the monocular semantic SLAM for high dynamic environment, considering the high dynamic environment with a large number of dynamic objects, so this article does not directly point out the potential motion feature points but uses the idea of SVO depth filtering to estimate the dynamic rate of change of punctuation [12], constantly adding new observation data, and the depth of the right punctuation is updated. On the premise of the depth convergence of landmarks, the map is added. On this basis, a priori value is assigned to the static rate according to the output of the semantic segmentation network, and then the static rate of the landmark points is updated when new observation data is introduced to realize the smooth transition of the landmark points between dynamic and static [13]. In ICRA2018, Stenborg et al. explored the stability of long-term positioning using semantic segmentation while participating in vehicle positioning projects. It solves the problem that the time span of the current detected features and the saved map features in the application scene of automatic driving is large, and the robustness of the features is high. A localization algorithm based on semantic tag and 3D position is proposed. A unified observation model based on SIFT features and semantic features is defined, and the pixel category and corresponding landmark category in the image are consistent as much as possible by adjusting the pose [14]. In ICRA2017, Sean et al. proposed the probabilistic data association of semantic SLAM, unified optimization of geometric information, semantic information, and IMU data in an optimization framework, and solved the problem through EM algorithm, realizing a SLAM system with higher accuracy [15, 16]. Suwoyo et al. proposed the adaptive development of svsf for a feature-based slam algorithm using maximum likelihood estimation and expectation maximization, which is designed to solve the online problem of Simultaneous Localization and Mapping (SLAM) [17].

The existing research in the field cannot provide reasonable recommendation and removal of inappropriate information for SLAM technology based on feature method. The overall idea of this project is to combine semantic information with ORB-SLAM2 and improve the traditional visual SLAM scheme through acquisition of object level semantic information, for higher efficiency of the loop closure testing provides a good database. So this paper proposed a semantic recognition of environment objects in the natural scene through object detection, and specific content is a key frame and target layer between the SLAM problem, establish a relation between key frames and target layer method, through the practical significance of the target layer to judge the merits of the target layer information, and combining with the SLAM relationship of key frames on the vision and the related information, the relationship between the object level targets in each key frames and related information are obtained, so as to realize through the object-level semantic information to judge the pros and cons of key frames. The above methods can retain the information of high reliability and good stability for visual SLAM and process the key frames with poor reliability or low stability and the information related to key frames.

The paper is structured as follows: Section 3 addresses the semantic optimization, which includes four parts: local mapping based on object level information filtering, semantic bags of words, and key frame filtering and landmark point filtering. Section 4 demonstrates the benefits of the proposed SR in a real scenario. This paper ends in Section 5 with the conclusions.

3. Semantic Optimization

In this paper, the object detection algorithm YOLO (You Only Look Once) is used to process the RGB (Red, Green, and Blue) images corresponding to the key frames obtained by the tracking thread in the local map construction thread, obtain the category and position information of objects in these RGB images, and establish the data association between the key frames and objects through the word bag model. The probability that an object is a dynamic object is calculated according to the category and probability of the object output by YOLO. Then for each key frame corresponding to the RGB image through dynamic object corresponding to the influential factors and dynamic map point within the region accounted for calculating the factors influencing the dynamic score, thus to screening of key frames, the dynamic road signs point to filter, improve the stability of the process of building a local punctuation, and improve the matching relation between the visual framework.

3.1. Local Mapping Based on Object Level Information Filtering. The tracking thread only determines whether the current frame needs to be added with a key frame. It does not really add the map. The main function is local positioning. The local build thread of ORB-SLAM2 is primarily tasked with maintaining the local map and managing key frames. The local mapping thread screens the key frames obtained by the tracking thread, fuses the landmark points in the key frames, eliminates redundant key frames and landmark points, maintains a stable global map, and provides the filtered key frames for loop detection. The thread of the local graph modifies the position of the key frame and the location of the signpost by determining more constraint relations. Specific tasks include the addition of key frames, elimination of landmark points, creation and fusion of landmark points, local BA, and local key frame elimination.

This article fuses semantic information with ORB-SLAM2 to help ORB-SLAM2 better complete the task of locating and mapping. After receiving the key frame from the tracking thread, the local map is processed with YOLOv3 to obtain the semantic category label of the object in the image. Then, according to the dynamic probability of the given category label, the dynamic probability of the object in the image is presented. After obtaining accurate semantic labels and dynamic probabilities [18–20], combining with the position of objects output by YOLO, feature filtering is carried out for key frames and landmark points, so as to minimize the impact of dynamic objects on map construction and location.

In order to better complete the above process, this paper constructs the relationship between key frames and objects by analogy with the association between key frames and landmark points in ORB-SLAM2, so as to further analyze whether a key frame is redundant and to eliminate landmark points that are dynamic objects.

3.2. Semantic Bags of Words. In the process of inserting a key frame by the thread of local map construction, the common view needs to be updated for each key frame that is added; that is, a new node is added for the new key frame k_i, and the edge between the key frames that have a co-visual relationship with it is updated. The co-visual relationship is judged by the number of common landmarks. The spanning tree is then updated to link k_i to the key frame with which it views the most, and finally to calculate the word bag representation of the key frame k_i.

In order to find the corresponding relationship between the key frame and the object and facilitate the processing of dynamic feature points in the feature filtering stage, this paper USES yolov3-tiny to process the RGB image corresponding to the key frame passed in by the local map construction thread to obtain the category and position information of the object. After the classification and position information of the object is confirmed, the classifiers marked object category and the probability that the object is a dynamic object are used to classify the object. The information obtained after classification includes the position information, size information, category information, and the probability that the object belongs to a dynamic object.

According to the requirements of this paper, after obtaining object level information through YOLO, the relationship between key frames and objects is built through

the word bag model. During the construction of the relationship between key frames and objects, O_i is stored for each object:

(1) Coordinates of landmark points located on the object

(2) Class label of object

(3) Object location and size information

K_i is stored for each key frame:

(1) The corresponding RGB image is used for detecting the object

(2) Corresponding depth image for generation point cloud

(3) The object observed in this key frame is recorded with a BoW, which is marked as BoW_O in this article to distinguish it from the BoW native to ORB-SLAM2

From the above description, the relationship between the key frame and the object can be established so that the key frame can find its related object, and vice versa.

The previous chapter of this paper has made a simple introduction to the word bag model and the special matching based on the word bag model, the generation of ORB-SLAM2 dictionary using the classical k-means algorithm for feature clustering [21, 22]; clustering problems often use the method of unsupervised machine learning, through the machine to find the law of data to complete the clustering. For the problem of establishing a dictionary of k words with N feature points in the image, each word can be regarded as a set of local feature points. In the following studies, on the basis of k-means, the hierarchical clustering method and k-means++ [23, 24] were proposed to improve k-means. Considering the efficiency of search, the K fork tree is used to express the dictionary on the basis of k-means. There are N feature points. The algorithm steps to construct a K fork tree with a depth of d and bifurcation into K are as follows:

(1) At the node of root, to ensure uniform clustering, k-means++ is used to cluster all samples into k class.

(2) Cross k to each node in the first layer of the tree. The samples belonging to this node are regrouped into k class, and the second layer is obtained.

(3) The leaf layer is the word by analogy.

The structure of the k fork tree is a k branch, and the tree with a depth of d can contain k^d words. Nodes outside the leaf layer are used for quick search. The leaf layer is used to construct words. The dictionary schematic diagram of k-fork tree is shown in Figure 1. ORB-SLAM2 puts the extracted ORB features of the image into a container and then mobilizes DBoW2's build interface to generate the word bag model.

In this paper, a word bag model between key frames and objects is established. Due to the limited number of object level objects and the classification basis of category labels, clustering is also unnecessary. For key frame A, use object level objects as words to build A word bag model such as 1. η represents the weight of the corresponding word, which is represented by the dynamic probability of the object corresponding to the word, and N represents the sum of the number of detected objects. Then, the corresponding depth of the k-fork tree is 1, and the value of k is N. Then, the dictionary schematic diagram of the object level word bag model is shown in Figure 2.

$$A = \{(w_1, \eta_1), (w_2, \eta_2), \ldots, (w_N, \eta_N)\} \triangleq v_A. \quad (1)$$

3.3. Key Frame Filtering. Considering that there are two kinds of key frames containing dynamic objects, one is that the picture area occupied by dynamic objects is small, and the number of dynamic landmark points corresponding to dynamic objects is small. This paper considers to keep the key frames in this case and only remove the landmark points related to dynamic objects. The other case is that dynamic objects occupy a larger area of the picture, and there are more relevant dynamic landmark points. Such key frames contain limited information that can be used as effective landmark points. This paper considers eliminating such key frames. The object corresponding to each frame and the probability that the object is a dynamic object can be corresponding through the word bag model. The size, position, and relevant landmark points of the object can be searched through the corresponding dictionary.

First, according to the object's dynamic influence factors to selection of key frames, this paper designs a screening function key frames for screening, screening function is shown in formula (2) select, where *awh* for the area of the influence factors of dynamic object, a is the weight of the parameter, w, h by YOLO output calculation area factor, the second dfrac$M_s M_h$ as belonging to the influence factors of dynamic number of feature points of objects, including M_s for this belongs to the dynamic object in the key frames of mappoints number, M_h is the total number of key frame middle punctuation, d is the weight parameter of the item, C_i and D_i are, respectively, the probability that the object belongs to the ith category, and the probability that the object of the ith category is a dynamic object, $C_i D_i$ is the probability product, and $\text{sum}_{I=1}^{N} C_i D_i$ is the probability product sum of the object category of N.

Above take into account the dynamic object and the area of dynamic relative map points which accounted for two factors, the type of calculation on key frames of an object selection function, set in each key frames can identify an object, a total of count, key frame selection function calculating formula is shown in formula (3), f parameter, used to adjust the output in the range of [0, 1], setting out the key frames of threshold value is S_{com}, calculated judgment S_{sum} is greater than S_{com}, if formed, delete the corresponding key frames, and make the local graph thread receive the new key frame.

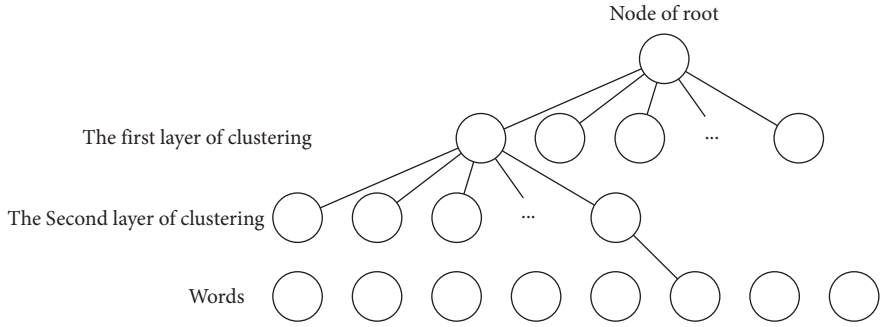

Figure 1: Semantic diagram of k-tree dictionary.

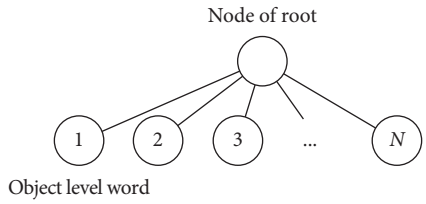

Figure 2: Dictionary diagram.

$$S = \sum_{i=1}^{N} C_i D_i \left(awh + d \frac{M_s}{M_h} \right), \tag{2}$$

$$S_{\text{sum}} = f \sum_{j=1}^{\text{count}} S_j. \tag{3}$$

3.4. Landmark Point Filtering. After adding key frames, the first thing you need to do is get rid of the bad signposts. A signpost in ORB-SLAM2 that wants to stay on the map must pass a rigorous test to determine if it can be traced to three consecutive key frames after it has been created. YOLO, this paper provides semantic information and adds a filtering wheel combining semantic information before ORB- SLAM2 original map points to filter the correct markers, filtering the features and placing key frames on the map points relative to the dynamic objects after filtering the step to calculate each object corresponding to the four formulas of key frames, each product of the probability of the corresponding N object category, and comparing with the corresponding threshold, in order to exceed the threshold map points associated with this object, optimize the total view of the key frames. Auxiliary key frames and landmark points are free from the interference of dynamic objects. The most important thing is that the landmark points in the local map are removed and the retained information is more meaningful.

$$\sum_{i=1}^{N} C_i D_i. \tag{4}$$

After the first round of screening, the landmarks associated with dynamic objects have been removed. While maintaining ORB-SLAM2's original screening process, in order to ensure a good relationship between the signposts and key frames, the following two conditions are required for good signposts to be screened:

(1) More than 25% frames can be observed in theory

(2) After the punctuation is created, it can be tracked by at least three consecutive key frames

According to the above conditions, the map points do not meet the conditions. Even if the road marking points meet the above conditions, the map points cannot be guaranteed not to be deleted. When the corresponding key frame is deleted, the map points are regarded as local BA external points, and the map points still belong to dynamic objects and will continue to be filtered.

The above two rounds of operations have preserved the good signpost points, and based on this, some new signpost points have been recovered by solving the ORB feature points connected to the key frames in the common view using PnP. Similarly, traverse the current key frame corresponding to the level of adjacent and secondary adjacent key frames, the current frame with the corresponding map points fusion, and finally, the current key frame corresponding to the property of the map points updates; these properties include the average direction of observation, observation distance, and best descriptor, such as the new signs point of view, the spanning tree.

4. Experiments

After the key frame is inserted, firstly, the object detection task is performed on the RGB image corresponding to each inserted frame through YOLOv3 to obtain the corresponding object category and size information in the frame, and the dynamic rate of the object is added according to the object category, and then according to the above information according to the method provided in this article to the key frame and landmark point selective filtering. This section conducts experiments for these two parts.

4.1. Filtering of Key Frames Based on Object Level Information. To filtering, key frames in the thread of local embedding key frame selection, based on YOLOv3 object categories, dynamic rate, size, and location information, get relevant dynamic waypoint information, comprehensive overall dynamic keyframe information judgment; the influencing factors of high dynamic key frames are: avoid the dynamic

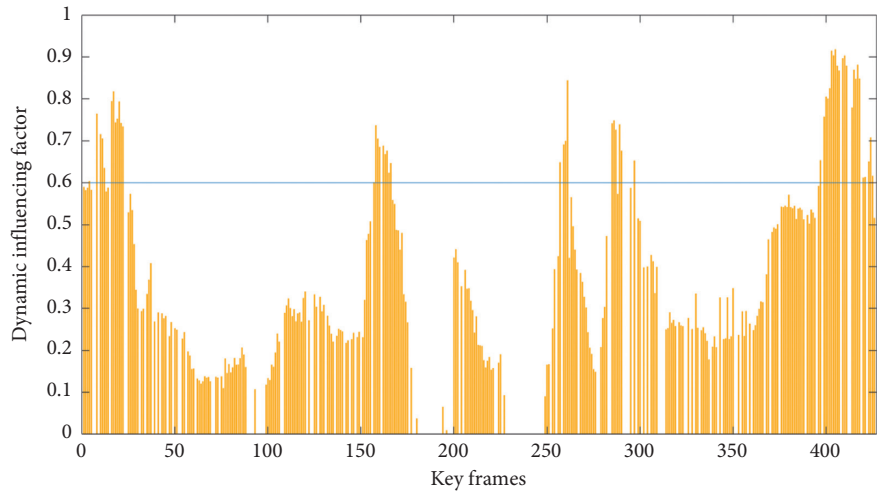

Figure 3: Key frame filtering.

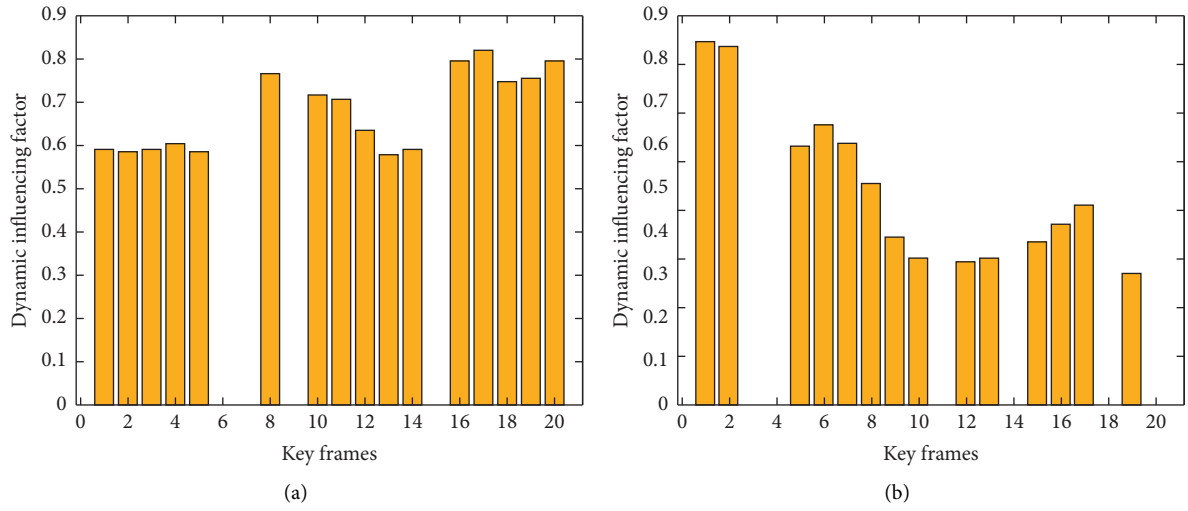

Figure 4: Key frame filtering of the first 40 key frames.

object, the more the greater the dynamic object, or the more dynamic mapping point retain key frames, key frames and the influence of the loopback detection obviously. In this paper, the dynamic influencing factors of key frames are calculated by formulas (2) and (3).

This section conducted experiments on the above process on the TUM dataset, which retained 426 key frames after tracking the thread. Each key frame was taken as the horizontal axis, and the dynamic influencing factors of the calculated key frames were taken as the vertical axis. The threshold value $S_{\text{com1}} = 0.6$ was taken to obtain the bar statistical graph as shown in Figure 3. The horizontal line of $y = 0.6$ in the figure is the cut-off line represented by the threshold value, and the corresponding key frame with the ordinate value higher than this line is removed.

For more convenient explanation, the data of 1–20 frames and 21–40 frames are enlarged and displayed. See Figure 4 for details. The left figure shows the data of 1–20 frames and the right figure shows the data of 21–40 frames. In this figure, it can be seen that the dynamic influencing factors of frames 6, 7, 15, 23, 24, 31, 34, 38, and 40 are zero, indicating that there is no dynamic object in these key frames. To remove frames 8, 10, 11, 12, 16, 17, 18, 19, 20, 21, 22, etc., in which the dynamic factor is high, the removal of these key frames can eliminate the unreliability of the total view, depending on the relationship, largely avoiding the concentration of many dynamic feature points on dynamic objects which will provide wrong action orientation, from the entire data set key frames as shown in Figure 4, It can be seen from Figure 20 that 10 frames and around 170–160 frames can be observed. Frames 400–420 are the continuous key frames of the concentrated activity of dynamic objects, and the key frames of these positions are selectively eliminated by the judgment of dynamic influencing factors. The RGB images corresponding to these position key frames are selectively posted in Figure 5. Two people can be observed in

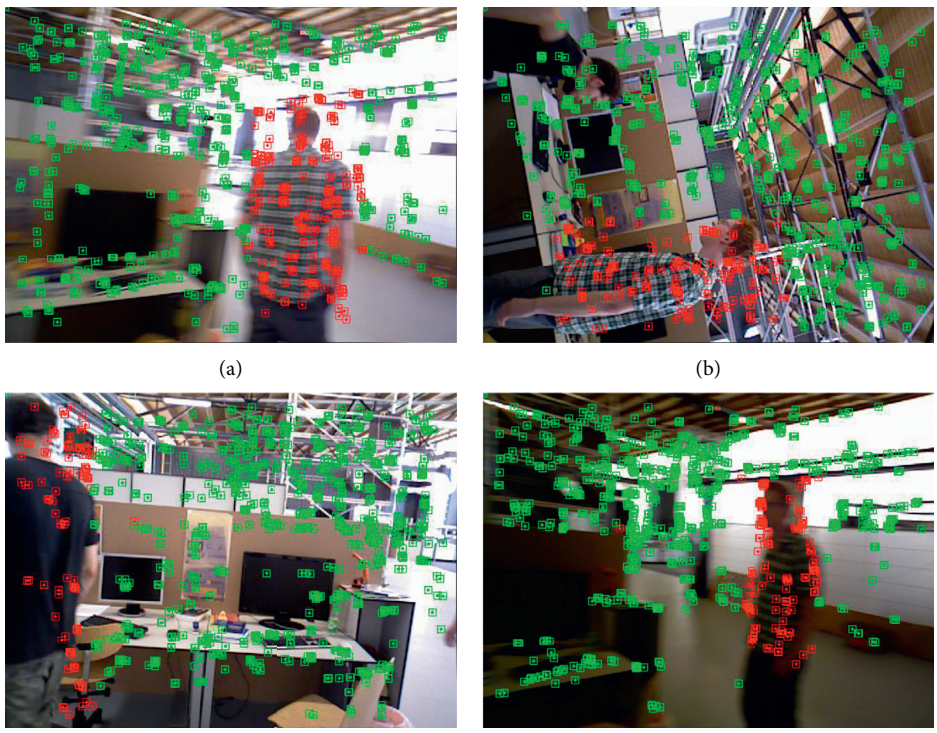

Figure 5: The RGB images corresponding to the key frames filtered out.

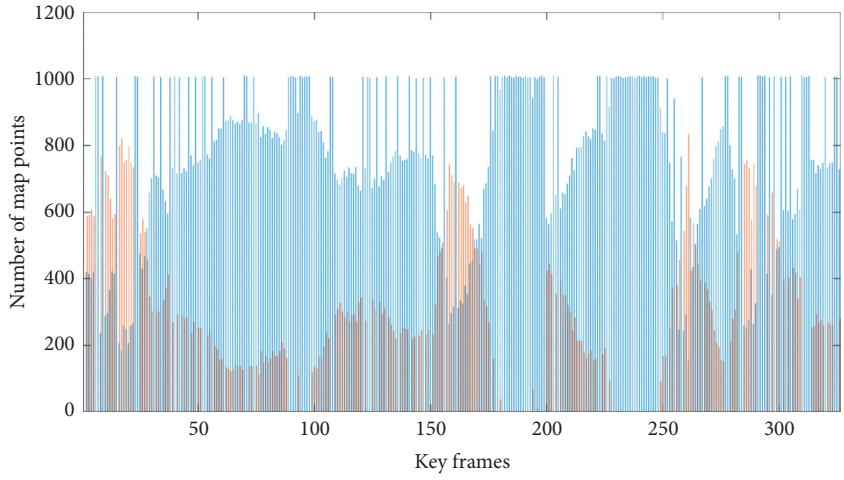

Figure 6: Map point filtering.

the key frames at these positions, which occupy a large area of the whole RGB image and have extremely high activity.

Through the above experiments, it can be seen that the improved ORB-SLAM2 has accurately filtered the extremely high key frames of dynamic influencing factors from the tracking thread in the local map construction thread, providing a good foundation for the stability of loop detection.

4.2. Filtering of Landmark Points Based on Object Level Information. Through the above process, the total area distribution is larger, and there are more dynamic object keyframes, but some dynamic objects still occupy less area, and the related dynamic map points have fewer keyframes, so the corresponding dynamic map points are also retained. Some buildings, roads, and signs affect the accuracy. Therefore, in this paper, through formula (4), the dynamic

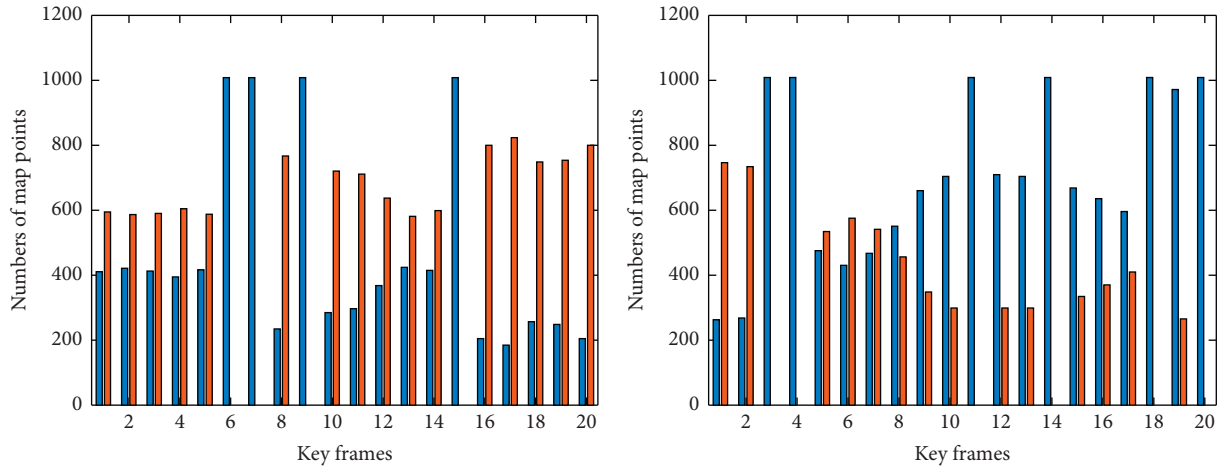

Figure 7: Map point filtering of the first 40 key frames.

influencing factors of dynamic objects related to dynamic landmark points are determined, and the corresponding key frames are retained, but the object-related landmark points with relatively large dynamic influencing factors are removed.

In this section, experiments were conducted on the TUM dataset on the above process. Each key frame was taken as the horizontal axis, and the number of signposts was taken as the vertical axis. The filtering of signposts is represented by Figure 6. In order to facilitate the explanation, the data of 1–20 frames and 21–40 frames in the remaining key frames after key frame filtering were enlarged and displayed. See Figure 7 for details. The left figure shows the data of 1–20 frames and the right figure shows the data of 21–40 frames. Among them, frames 7, 15, 23, 24, 31, 34, 38, and 40 have no corresponding orange bars, indicating that there is no dynamic object in these key frames, or the probability of dynamic object is very small. The landmark points are not considered to be eliminated, and the road punctuation is eliminated to different degrees in other frames. The landmarks correspond to the feature points obtained by ORB detection, and the elimination of the corresponding landmarks in 3D space can be expressed by the feature points, which can be represented by displaying the feature points on the RGB image corresponding to the key frame.

Extract four keys to weed out corresponding map points corresponding to RGB images, as shown in Figure 5. To weed out the 3D map points corresponding feature points shows red, not out of the 3D map points corresponding feature points according to green. People can be seen in the figure; the dynamic related feature point is marked in red.

As can be seen from the above experiments, this paper improved the selection of punctuation in the middle of ORB-SLAM2 by filtering signposts based on object level information. Dynamic signposts were removed correctly while static signposts with high stability were retained, which improved the stability of ORB-SLAM2.

5. Conclusions

The overall idea of this project is to combine semantic information with ORB-SLAM2 and improve the traditional visual SLAM scheme through acquisition of object level semantic information. YOLO aids the process of selecting keyframes and selecting map points through object detection algorithms for visual SLAM object-level semantic information, enabling ORB-SLAM2 to obtain more stable static map points and more stable keyframes for loopback, and to preword wrap objects through keyframes and object models, providing a good database for higher closed-loop testing efficiency.

Key frames received by the local build thread of ORB-SLAM2 are filtered through object level information. By detecting the RGB image corresponding to each key frame through YOLO, the category, position, and size information of the object corresponding to each frame can be obtained. After the dynamic probability of the object is increased according to the category of the object, the dynamic influencing factors are calculated for each key frame based on the above information, and the key frames whose dynamic influencing factors exceed the threshold are eliminated. On the one hand, this improves the reliability of the common view and prevents the co-vision relationship generated under the condition that there are many landmarks related to the dynamic objects in the adjacent key and the degree of co-vision is high. This co-vision relationship is unstable and may no longer exist with the movement of the dynamic objects. On the other hand, it increases the stability of loop detection and prevents the judgment of loop detection from being affected by the repeated detection of large objects with more relevant dynamic landmarks. Through experiments, it was found that the RGB images corresponding to the deleted key frames had larger dynamic objects and more dynamic landmarks, which was in line with the original intention of the design and achieved a better effect.

On the basis of key frame filtering, the dynamic feature points corresponding to the reserved key frames were removed to screen out more reliable signpost points for ORB-SLAM2, which is conducive to the accuracy of positioning and mapping. Through the experiment, it is found that the dynamic landmark points retained by the key frame filter are removed accurately and the desired effect is achieved.

This paper provides reasonable recommendation and removal of inappropriate information for SLAM technology based on feature method. The methodology is to propose a semantic recognition of environment objects in the natural scene through object detection. The method can retain the information of high reliability and good stability for visual SLAM and process the key frames with poor reliability or low stability and the information related to key frames.

Data Availability

This publication was supported by TUM RGB-D datasets, which are openly available at location cited in [25].

Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

This research was funded by the Fundamental Research Foundation for Universities of Heilongjiang Province (LGYC2018JQ003) and University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (No. UNPYSCT-2018208).

References

[1] K. Tateno, F. Tombari, I. Laina, and N. Navab, "Cnn-slam: real-time dense monocular slam with learned depth prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6243–6252, Seattle, WA, USA, June 2017.

[2] F. Zhong, S. Wang, Z. Zhang, and Y. Wang, "Detect-slam: making object detection and slam mutually beneficial," in *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1001–1010, IEEE, Lake Tahoe, NV, USA, March 2018.

[3] C. Yu, Z. Liu, X.-J. Liu et al., "Ds-slam: a semantic visual slam towards dynamic environments," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1168–1174, IEEE, Madrid, Spain, October 2018.

[4] P. Li and Q. Tong, "Stereo vision-based semantic 3d object and ego-motion tracking for autonomous driving," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 646–661, Glasgow, UK, August 2018.

[5] R. Wang, M. Schworer, and D. Cremers, "Stereo dso: large-scale direct sparse visual odometry with stereo cameras," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3903–3911, Venice, Italy, October 2017.

[6] J. Pan, C. F. Cristian, and K. McGuinness, "Salgan: visual saliency prediction with generative adversarial networks," 2017, http://arxiv.org/abs/1701.01081.

[7] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890, Seattle, WA, USA, June 2017.

[8] H.-J. Liang, J. Nitin, C. Fermüller, and Y. Aloimonos, "Salientdso: bringing attention to direct sparse odometry," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1619–1626, 2019.

[9] P. Ganti and S. L. Waslander, "Visual slam with network uncertainty informed feature selection," 2018, http://arxiv.org/abs/1811.11946.

[10] V. Murali, H.-P. Chiu, S. Samarasekera, and R. Teddy Kumar, "Utilizing semantic visual landmarks for precise vehicle navigation," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–8, IEEE, Yokohama, Japan, October 2017.

[11] B. Bescos, J. M. Facil, J. J. Civera, and J. Neira, "Dynaslam: tracking, mapping, and inpainting in dynamic scenes," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4076–4083, 2018.

[12] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "Svo: semidirect visual odometry for monocular and multicamera systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2016.

[13] N. Brasch, Aljaz Bozic, L. Joe, and F. Tombari, "Semantic monocular slam for highly dynamic environments," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 393–400, IEEE, Madrid, Spain, October 2018.

[14] E Stenborg, C. Toft, and L. Hammarstrand, "Longterm visual localization using semantically segmented images," in *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6484–6490, IEEE, Brisbane, Australia, May 2018.

[15] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, Hoboken, NJ, USA, 2007.

[16] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas, "Probabilistic data association for semantic slam," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1722–1729, IEEE, Marina Bay Sands, Singapore, May 2017.

[17] H. Suwoyo, W. Y. Tian, A. L. AdriansyahLi, and G. Yuan, "Adaptive development of svsf for a feature-based slam algorithm using maximum likelihood estimation and expectation maximization," *IIUM Engineering Journal*, vol. 22, no. 1, pp. 269–286, 2021.

[18] Ji-H. Xi, K.-M. L. Dong-Hyun Lee, and C.Ho Lin, "An improved yolov3-based neural network for de-identification technology," in *Proceedings of the 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Korea, March 2019.

[19] F. Zeng and C. Wang, "Visual navigation with asynchronous proximal policy optimization in artificial agents," *Journal of Robotics*, vol. 2020, Article ID 8702962, 7 pages, 2020.

[20] J. Ni, T. Gong, Y. Gu, J. Zhu, and X. Fan, "An improved deep residual network-based semantic simultaneous localization and mapping method for monocular vision robot," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 7490840, 14 pages, 2020.

[21] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[22] J. Song and L. B. Kish, "On the theory and design of cold resistors," *Fluctuation and Noise Letters*, vol. 20, no. 1, Article ID 2150001, 2020.

[23] D. Arthur and V. Sergei, "k-means++: the advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, LA, USA, January 2007.

[24] Z. A. Ali, Z. Han, and Bo H. Wang, "Cooperative path planning of multiple UAVs by using max–min ant colony optimization along with cauchy mutant operator," *Fluctuation and Noise Letters*, vol. 20, no. 1, Article ID 2150002, 2021.

[25] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proceedings of the International Conference on Intelligent Robot Systems (IROS)*, Vancouver, Canada, October 2012.