Multi-session Map Construction in Outdoor Dynamic Environment*

Xiaqing Ding[1], Yue Wang[1], Huan Yin[1], Li Tang[1], Rong Xiong[1]

Abstract— Map construction in large scale outdoor environment is of importance for robots to robustly fulfill their tasks. Massive sessions of data should be merged to distinguish low dynamics in the map, which otherwise might debase the performance of localization and navigation algorithms. In this paper we propose a method for multi-session map construction in large scale outdoor environment using 3D LiDAR. To efficiently align the maps from different sessions, a laser-based loop closure detection method is integrated and the sequential information within the submaps is utilized for higher robustness. Furthermore, a dynamic detection method is proposed to detect dynamics in the overlapping areas among sessions of maps. We test the method in the real-world environment with a VLP-16 Velodyne LiDAR and the experimental results prove the validity and robustness of the proposed method.

I. INTRODUCTION

Substantial progresses in SLAM enable robots to construct an accurate and even a dynamic-free map within a single session using 3D LiDAR in outdoor environment [1]. When the environment is large, the map is usually represented as a conjunction of linked submaps, in which way the computation burden could be restricted under a sustainable range, which benefits the performance of SLAM and localization algorithms. However, a single-session map is not adequate for long-term operation tasks, especially in outdoor environment where the dynamics can not be suppressed. Besides the moving objects, there are many low-dynamic objects in outdoor environment, such as the parking cars, which keep still throughout a whole session but are potential to change positions across sessions. If no priori information is provided, they can not be distinguished within a single session and are remained in the constructed map. Those low dynamic objects in the map sometimes would introduce wrong data association results for localization algorithms as some part of the observation can not find correct correspondences in the map. What's more, they also deteriorate the performance of navigation algorithms as some free space is explicitly occupied in the map. To improve the performance of outdoor tasks, multi-session of data is required to remove the dynamics in the map.

[1]Xiaqing Ding, Yue Wang, Huan Yin, Li Tang, Rong Xiong are with the State Key Laboratory of Industrial Control and Technology, Zhejiang University, Hangzhou, P.R. China. Yue Wang is with iPlus Robotics Hangzhou, P.R. China. Yue Wang is the corresponding author wangyue@iipc.zju.edu.cn. Rong Xiong is the co-corresponding author rxiong@zju.edu.cn.

As for multi-session map maintenance, one of the most important issues is how to efficiently align maps across sessions. Maps in different sessions need to be conjuncted into a whole map after alignment and a global optimization could be applied to improve the consistency of the map. This can also be regarded as a broader loop closure detection problem, while in terms of laser-based algorithms, it still remains to be a difficult problem due to the high ambiguity of 3D point clouds.
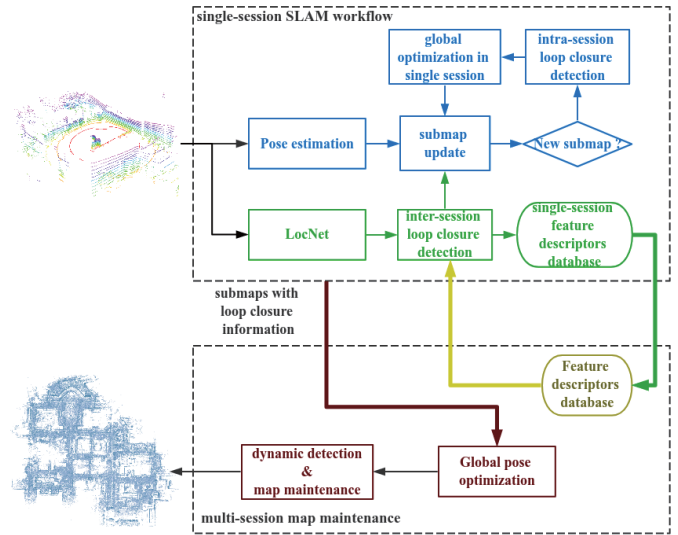

Fig. 1: The framework of the multi-session laser-based map maintenance method.

In this paper, a multi-session map maintenance method is developed in large scale outdoor environment using the 3D LiDAR. The map is conjuncted from linked submaps and the intra-session loop closure detection is performed based on the poses of submaps following an overlap-based outlier detection. Extended from our previous work, information from the history sessions is converted into feature vectors and organized into a database for quick index. The inter-session loop closure is detected for each scan based on the database while performing laser-based SLAM method in a new session and sequence information is used in order to improve the robustness. What's more, a new dynamic detection method designed for detecting in sparse point clouds is proposed to remove the dynamics out of the map. The main contributions of this paper are as follows:

- Introduce a multi-session laser-based map construction framework which is designed for serving localization and navigation tasks in large-scale outdoor environment.

- Integrate the laser-based loop closure detection method in our previous work into the multi-session map construction framework for map alignment based on submaps and introduce sequential information to improve the robustness.
- Propose a low-dynamic detection method between sparse point clouds from different sessions to maintain only the static parts in the map, which aims at suppressing the influence from the dynamics in the environment while localizing and navigating against the priori map.

The remain of the paper is organized as follows: Section II gives a review of the related works and the followed Section III introduces the system overview of the proposed multi-session map construction method. Section IV, V, VI separately demonstrate the single-session laser-based SLAM method, inter-session loop closure detection method and the multi-session map construction method implemented in the proposed system. And in Section VII the experimental results are presented. Conclusion and some discussions are showed in Section VIII.

II. RELATED WORK

To bring the SLAM algorithms into practice, many researches focus on how to improve the robustness in terms of long term operation. Two of the relevant issues are addressed in this paper: one is how to efficiently align the multi-session results into a consistent map and the other is how to extract the stable parts of the environment according to massive sessions of data.

To realize multi-session map alignment, robots should be capable of detecting loop closure globally and across sessions. With the development of computer vision and machine learning, vision-based loop detection method is gradually mature in recent years. McDonald et al. [2] introduce anchor nodes to combine the pose graphs from massive sessions based on the loop closures detected from the visual-based place recognition method. Recently Schneider et al [3] propose an impressive visual-inertial mapping and localization framework named maplab. They use an efficient binary description for loop detection then add the information as edges into the posegraph for optimization. While in context of laser-based methods, there still exists large difficulties. Yang et al. [4] propose a global point cloud registration method expanding from ICP algorithm [5] with Branch and Bound [6] scheme. Though efficiently, it's still hard to be applied in large scale outdoor environment. SegMatch [7] extracts segments from 3D point cloud and tries to match them with segments extracted from history trajectories for loop candidate detection. Being different from SegMatch that utilizes features in the environment for detection, Yin et al. [8] introduce LocNet to describe the whole range information of a 3D point cloud for matching, which reveals good performance in unstructured outdoor environment. In this paper, we also introduce LocNet for point cloud representation and detect loop closure for each laser scan against the history point cloud information indexed in the database.

In the context of long-term localization and navigation, it's essential that robots should be capable of distinguishing dynamics out of the static environment. Walcott-Bryant et al. [9] propose the 2D dynamic pose graph SLAM and detect dynamics across time on the occupancy grid map. While in 3D space, Fehr et al. [10] represent the environment using tsdf-based structure, which could inherently distinguish the empty voxel out of the unobserved one. Pomerleau et al. [1] update the dynamic probabilities of the point cloud based on the observations from each new scan while incrementally mapping, they can also to large extend suppress dynamic points from adding into the maps. In this paper we try to detect the dynamics between two sparse submaps in outdoor environment. Instead of directly converting the whole submap into a tsdf-based map, we propose a new detection method which could largely decrease the computation burden of building the voxel map and designed especially for dynamic detection in multi-view merged submaps.

III. SYSTEM OVERVIEW

The overview of the large-scale multi-session mapping framework will be introduced in this section. The framework includes a single-session SLAM workflow, a multi-session map maintenance system and a loop closure detection pipeline for cross-session loop closure detection among submaps, as shown in Fig. 1.

In consideration of real-time performance, while performing the laser-based SLAM algorithm, the whole map is organized by linked submaps, which are accumulated by sequential laser scans following the map updating method in [1] and little high-dynamic points would be remained in the submaps. When a new submap is established, the intra-session loop closure detection module will search the loop closure candidates across the history submaps within this session based on the priori pose estimations, following a overlapping-based loop closure validation step. If a loop closure is detected and validated, a global pose optimization will be applied to improve the accuracy and consistency of the SLAM results.

Besides loop closure detection within a single session, we introduce the LocNet [8] for cross-session loop closure detection. LocNet converts the range information of each scan into a compact feature vector and all of the feature vectors will be maintained in an indexed database which enables quick matching among large numbers of vectors. Therefore when performing a new session of SLAM algorithm, a quick place recognition process can be implemented for each scan based on the existing database constructed from the information of massive history data and the sequential detection information within each submap is utilized for the inter-session loop closure validation.

When a new session of map is constructed, we introduce the inter-session loop closures as edges into the global pose optimization framework to align the new session of map into the history one. After the map alignment, dynamics will be detected within the overlapping areas between each pair of new and history submaps using our proposed sparse point

cloud based dynamic detection method, which remains a more static map that benefits future localization and navigation tasks.

IV. Single Session Laser-based SLAM

3D LiDAR could provide precise geometry information from the environment, which makes the pose estimation more accurate compared with the vision-based method. As for outdoor environment, there exists many high dynamic objects such as the moving cars and pedestrian, which should not be included in the constructed map. In this section we utilize the ICP method implemented from [11] for pose estimation and follow the map updating method introduced in [1] to accumulate the map.

In consideration of real-time performance when the exploration area is large, we only perform the SLAM algorithm on a submap and record the relationship between the neighbor submaps for the global pose graph optimization. When the relative transformation between the current pose and the pose of current submap origin is beyond the designed threshold, the current submap is regarded as finished and a new submap is initialized on the pose of the current frame.

When a submap is finished, the intra-session loop closure detection method will be executed. Some candidates might be selected based on the priori pose estimation information of the existing submap origins. If two submap origins are close in Euclidean space, they are with high probability to claim loop closure and are regarded as a candidate pair. We then apply the ICP algorithm on the two submaps to compute their relative transformation. If the result approximates their priori relative transformation derived from the priori pose estimations and the overlap ratio between the two submaps is reasonable after the iteration result of ICP algorithm, this candidate pair passes the validation process and is regarded as a loop closure. Then a global pose graph optimization is performed

$$\{x_i | x_i \in X\} =$$
$$\arg\min_{x_i} \sum_{x_i} \sum_{x_j \in N_{x_i}} \rho(||(x_i \ominus x_j) \ominus z_{i,j}||^2_{\Omega_{i,j}}) \quad (1)$$

where $X = \{x_0, x_1...x_n\}$ represents the pose estimation set of submap origins within current session, N_{x_i} demonstrates the neighbor set of submap x_i, which includes both the successive submaps while mapping and the submaps introduced by the loop closures. The notation $x_i \ominus x_j$ means the relative transformation between pose x_i and pose x_j as demonstrated in [12]. $z_{i,j}$ represents the observation results computed from the ICP algorithm between the successive or loop closed submaps with the pose of x_i and x_j. $\rho(\cdot)$ is the robust kernel and $\Omega_{i,j}$ represents the information matrix. The whole optimization function is solved using Levenberg-Marquardt algorithm [13].

V. Inter-session Loop closure detection

In our previous work [8], a global localization method is proposed which converts the range information of each laser scan into a feature vector using LocNet then place recognition is achieved by matching the current feature vector with the indexed database that contains the information of the whole map. In this paper we integrate this method into multi-session map maintenance framework for inter-session loop closure detection.

Initially, in the first session we convert each laser scan into the feature vector while performing the SLAM algorithm. In the end besides saving the submaps and their relationships for map maintenance, we also save all of the feature vectors and index them into a loop closure database using the kd-tree algorithm. When a new session of SLAM algorithm is performed, after converting the current laser scan s_i^c into the feature vector l_i^c, we search the closest feature vector l_j^h in the kd-tree based database, which indicates the most resemble laser scan s_j^h in history sessions. We record the submap m_k^h which the laser scan s_j^h belongs to as a label of the current laser scan s_i^c. If the distance between the matched vectors is larger than d_α, we consider there is no matching history information for the current scan.

When this session of data has been processed, we select the inter-session loop closure candidates by statistically checking the place recognition results within each new submap in a voting way based on the following criterion

- Sum up all of the submaps that are matched by the laser scans in current submap m_i^c. Compute the proportions p_j for each matched submap m_j^h.

$$p_j = \frac{number\ of\ scans\ that\ match\ with\ m_j^h}{number\ of\ the\ scans\ in\ m_i^c} \quad (2)$$

- If the largest proportion p_j is larger than γ, the current submap m_i^c and the history submap m_j^h are considered as a pair of loop closure candidate.
- If the sum of top n largest proportions is larger than γ, and the corresponding matched submaps are neighbors, all of the matched submaps are considered as the loop closure candidates to the current submap m_i^c.

After selecting the loop closure candidates, the global ICP algorithm [4] is applied to compute the relative transformation between the two submap within each candidate pair. The overlap-based outlier detection is also utilized to remove the wrong loop closures as mentioned in section IV.

VI. Multi-session laser-based SLAM

This section will introduce the multi-session map maintenance method in our framework. When a new session of map has been constructed and the inter-session loop closures have been detected and validated, we utilize the loop closures as indications to achieve map alignment. Further the dynamics in the overlapping area between the matched new and history submaps will be detected and removed.

A. Multi-session global optimization

We add the loop closure information as edge into the global pose graph. Mathematically, we denote the pose set of the history submap origins as X^h, the pose set of the

Fig. 2: Four places where the inter-session loop closure detection method fails. The pictures in the first line are captured from the sequence Day 3-1. And the pictures in the second line are the corresponding places in sequence Day 1-1.

submap origins in the new session as X^c, the set of inter-session loop closures as $O = \left\{o(x_i^c, x_j^h)\right\}_{x_i^c \in X^c, x_j^h \in X^h}$. The optimization function is

$$
\begin{aligned}
\left\{x_i | x_i \in (X^c \cup X^h)\right\} = \\
\arg\min_{x_i}(\sum_{x_i} \sum_{x_j \in N_{x_i}} \rho(||(x_i \ominus x_j) \ominus z_{i,j}||^2_{\Omega_{i,j}}) + \\
\sum_{o(x_p^c, x_q^h) \in O} \rho(||(x_p^c \ominus x_q^h) \ominus z_{i,j}^o||^2_{\Omega_{i,j}^o}))
\end{aligned}
\tag{3}
$$

where the $z_{i,j}^o$ means the relative transformation computed using the ICP algorithm between the loop closed submaps. We only set the origin of the first submap as the origin of the whole map, therefore the whole trajectory of the new session will be aligned into the same coordination of the history map.

After the alignment of the new session, we once again check loop closure for each submap in the new session as mentioned in IV to expand the set O. And the global pose graph optimization would further be applied for higher accuracy and consistency.

B. Dynamic detection in sparse point clouds

The laser point clouds in the submaps are relatively sparse than the point clouds constructed with RGB-D sensor [10], [14], and the scale is also much larger than the general RGB-D map. So we can not directly apply the existing dynamic detection method in our framework. In this section we introduce a voxel-based dynamic detection method which is designed to process the sparse point clouds.

For each pair of loop closed submaps (m_i^h, m_j^c) indicated in set O, we consider every submap is constructed from two subsets which we denote as $m_i^h = \left\{m_i^{hc}, \overline{m_i^{hc}}\right\}$, $m_j^c = \left\{m_j^{hc}, \overline{m_j^{hc}}\right\}$. Insides m_i^{hc} and m_j^{hc} represent the common parts in both of the submaps; while $\overline{m_i^{hc}}$ and $\overline{m_j^{hc}}$ indicate the different parts of the two submaps. The dynamic detection process includes the following steps

- Transform the submap m_j^h into the local frame of submap m_i^c.

- Construct a kd-tree from the submap m_j^h to search the closest point in the submap m_i^c for each point in m_j^h. Keep those matched point pairs that their distances are shorter than d_β as seeds. Further a region-growing method is applied separately in m_i^c and m_j^h based on the two sets of seeds to determine the subsets m_i^{hc} and m_j^{hc}, while during growing both the distance between points as well as the difference of normal vectors are considered. Also $\overline{m_i^{hc}}$ and $\overline{m_j^{hc}}$ are determined as the complementary sets of m_i^{hc} and m_j^{hc}.

- Build a voxel map based on the subset $\overline{m_j^{hc}}$. Use the points in $\overline{m_j^{hc}}$ for raycasting and only keep those occupied voxels which are in the front of each raycasting line. Apply the same process on $\overline{m_i^{hc}}$.

- Use the remained occupied voxels building from $\overline{m_i^{hc}}$ for raycasting and coarse dynamic points could be identified. Then apply a region-growing method to get the final detected dynamic points. Record the points in $\overline{m_i^{hc}}$ that indicate the dynamics as $\widehat{m_i^{hc}}$.

To update the submap m_j^h, remove the detected dynamics and add the points in $\widehat{m_i^{hc}}$ which indicate the dynamics into it for mending the points behind the dynamics. Keep the remaining submaps in the new session that do not share large overlapping with the hisory map and update the whole relative relationship among the submaps.

VII. EXPERIMENTAL RESULTS

In this section we test our multi-session map construction and maintenance method in real-world environment. The data is collected with a four-wheel mobile robot equipped with a VLP-16 Velodyne LiDAR. All of the algorithms are employed in an laptop with Intel i7-6700HQ @ 2.60GHz * 8 CPU and 7.7 GB memory. The environment we aimed for constructing the map is the south part of Yuquan Campus, Zhejiang University, China, which occupies around 250000 square meters.

A. Inter-session loop closure detection

We first test the performance of inter-session loop closure method. For comparison with the single frame loop closure
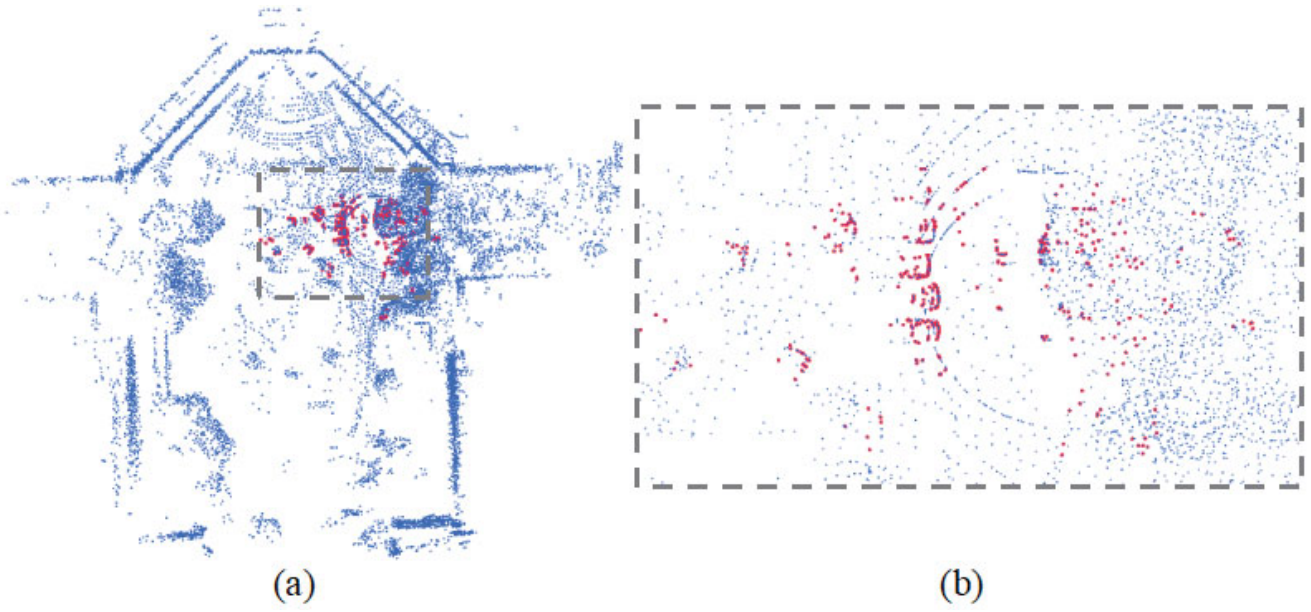
Fig. 3: The dynamic detection result in a submap. The blue points represent the origin laser submap. The red points represent the detected dynamic points. (b) is the enlarged figure of the dotted area in (a).

detection results in [8], we also test the submap-based loop closure detection on two sessions of Day 3 in YQ21 dataset. YQ21 dataset is a 21-session dataset recorded in March 3, 7 and 9 in 2017, and each session follows the same routine which is around 1.1km. While performing the single-session SLAM algorithm, we set the translation threshold to create a new submap as 20 meters and rotation threshold as 0.7 rad.

The loop closure detection results of every frame are recorded and the performance is demonstrated in [8]. We set the threshold γ as 0.95 and n as 3. It's important to note that in some frames where there are many dynamic objects moving around the robot, the loop closure detection method might fail to converge since the range information is largely influenced. Also when detect loop closure across sessions to merge a larger map, the distance between the current feature vector collected from a never visited place and the searched feature vector in the database is relatively large. In this case the the loop closure detection method also returns a failure signal. And for those frames the results are excluded from the voting process. If all of the frames in a submap fail the loop closure detection, this submap is regarded as a undefined submap, which could be a new place that has never been visited or a place with many dynamics. We build the map with the data of Day 1-1 and test the inter-session loop closure detection method on session Day 3-1 and Day 3-2. The results are shown in Table I.

TABLE I: The results of the inter-session loop closure detection

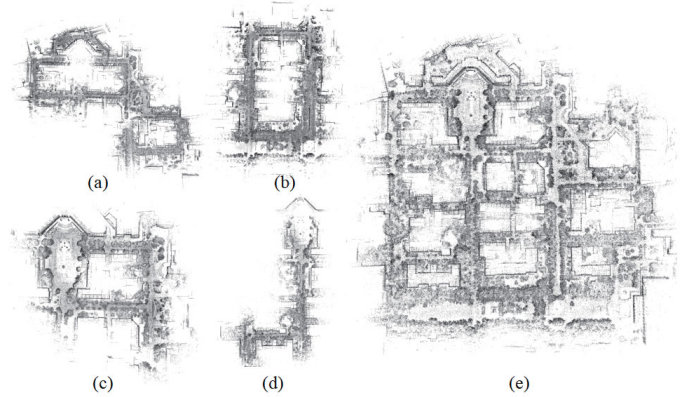| Sequence	total submaps	correctly matched submaps	undefined submaps	wrongly matched submaps
Day 3-1	73	68	5	0
Day 3-2	73	73	0	0

Fig. 4: The merging result from multi-session maps. (a), (b), (c), (d) are four of the single-session of maps used for map maintenance. (e) is the final merged laser map demonstrating the south of Yuquan Campus.

As the results show, in most of the cases our proposed inter-session loop closure detection method could correctly detect the loop closure. And in these two sequences there is no wrong detection result. We show the environment where the undefined submaps collected from in Fig. 2. Since two of the undefined submaps are neighbors, we only show four places in Fig. 2. As we can see, in all of the four places, the environment differs due to the parking cars; and these cars are close to the robot when it collects the data. Those close dynamic objects would cause largely changes in the raw laser scans which leads to the failure of the loop closure detection method.

B. Low dynamics detection in sparse point clouds

Based on the results of inter-session loop closure detection, we can test the low dynamic detection method on any two

388

of the matched submaps. The length of the voxel grid is set as 0.04 m. We show one of the results in Fig. 3. As the result shows, the parking cars could be distinguished out as the dynamic objects. So as some points above the ground which are part of the high dynamic points that are not completely removed by the dynamic detection method in single-session map construction. But many points on the trees would also be detected. We think this is mainly due to the sparsity of the submaps. While constructing the map from one single-session of data, in order to reduce the computation time and storage, we would down-sample the laser points so the submaps could be sparse. So the point clouds of the same tree but constructed from different sessions would be different, which leads to the response of the dynamic detection method. Since the points on the trees take an important part in laser-based localization method, especially for the outdoor environment, they should not be excluded out of the submaps. So when doing dynamic detection, we set a height threshold of 2m to avoid those points above that being deleted from the submap.

C. Multi-session map construction

We finally construct a large laser map of the south part of Yuquan Campus, Zhejiang University which covers around 250000 square meter, as shown in Fig .4 (e). And Fig .4 (a), (b), (c), (d) are four of single-session of maps that used to construct the whole map. For sake of the robustness of inter-session loop closure detection, each new single session would overlap a large part of the environment with the history map.

VIII. CONCLUSIONS

In this paper we propose a multi-session map construction method for large-scale outdoor environment. The different sessions of data are aligned with a laser-based loop closure detection method. And the redundant information is discarded after the dynamic detection method has been applied between overlapped submaps. The experimental results validate the capability of both the inter-session loop closure detection method and the dynamic detection method.

However, there still exists some problems for those two methods. As for inter-session loop closure detection method, the detection results would be largely influenced by the emergency of new objects which do not exist when the previous data was collected, especially if these object are close to the robot. We want to overcome this problem in the future work by merging some priori information to distinguish the dynamics.

Besides, the problem of the dynamic detection method is that there usually exist some false positive detection results. This problem is partially due to the sparsity of the submap, and partially because this method is sensitive to the precision of the submap alignment. If there is some error in the estimation of the relative transformation between the overlapped submaps, especially if the error exists in the pitch or roll direction, wrong detection results would occur. We will try to solve this problem in the future work.

REFERENCES

[1] F. Pomerleau, P. Krüsi, F. Colas, P. Furgale, and R. Siegwart, "Long-term 3d map maintenance in dynamic environments," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3712–3719.

[2] J. McDonald, M. Kaess, C. Cadena, J. Neira, and J. J. Leonard, "6-dof multi-session visual slam using anchor nodes," 2011.

[3] T. Schneider, M. Dymczyk, M. Fehr, K. Egger, S. Lynen, I. Gilitschenski, and R. Siegwart, "maplab: An open framework for research in visual-inertial mapping and localization," *arXiv preprint arXiv:1711.10250*, 2017.

[4] J. Yang, H. Li, and Y. Jia, "Go-icp: Solving 3d registration efficiently and globally optimally," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1457–1464.

[5] P. J. Besl, N. D. McKay, *et al.*, "A method for registration of 3-d shapes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[6] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica: Journal of the Econometric Society*, pp. 497–520, 1960.

[7] R. Dubé, D. Dugas, E. Stumm, J. Nieto, R. Siegwart, and C. Cadena, "Segmatch: Segment based place recognition in 3d point clouds," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 5266–5272.

[8] H. Yin, X. Ding, L. Tang, Y. Wang, and R. Xiong, "Efficient 3d lidar based loop closing using deep neural network," in *Robotics and Biomimetics (ROBIO), 2017 IEEE International Conference on*. IEEE, 2017, pp. 481–486.

[9] A. Walcott-Bryant, M. Kaess, H. Johannsson, and J. J. Leonard, "Dynamic pose graph slam: Long-term mapping in low dynamic environments," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 1871–1878.

[10] M. Fehr, F. Furrer, I. Dryanovski, J. Sturm, I. Gilitschenski, R. Siegwart, and C. Cadena, "Tsdf-based change detection for consistent long-term dense reconstruction and dynamic object discovery," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 5237–5244.

[11] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, "Comparing ICP Variants on Real-World Data Sets," *Autonomous Robots*, vol. 34, no. 3, pp. 133–148, Feb. 2013.

[12] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous robots*, vol. 4, no. 4, pp. 333–349, 1997.

[13] R. Kmmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," in *IEEE International Conference on Robotics and Automation*, 2011, pp. 3607–3613.

[14] O. Kähler, V. A. Prisacariu, and D. W. Murray, "Real-time large-scale dense 3d reconstruction with loop closure," in *ECCV 2016*, 2016, pp. 500–516.

[1]https://github.com/ethz-asl/libpointmatcher

389