**A Closely Coupled Framework for** **LiDAR-based Panoramic Segmentation and Dynamic Detection**

***Abstract*** (Panoramic Segmentation是否用词不当)

Long-term mapping(没有地图更新算不算地图维护)是机器人在非静止真实环境中实现自主导航的关键。这篇文章提出了**一种基于LiDAR的全景分割和动态检测的紧耦合框架**，通过single-session SLAM的post-procession完成城镇场景的静态语义地图的构建。The proposed method将这个problem划分为两个相互关联的subproblems: segmentation base on clustering assisted with sensor intensity, high/low dynamic detection in object-level。The mission of segmentation是在hash table管理的curved-voxel坐标系下完成的，同时使用sensor intensity辅助clustering; the procession of dynamic detection 根据前者提供的实例信息进行物体级别的动态检测，并通过多帧融合纠正前者的分割错误。我们验证该方法在复杂的城镇环境中可以有效地构建静态语义地图，从而达到长期维护地图目的，为机器人在不确定环境中实现自主导航奠定基础。

***Introduction*** (感觉中文说不清楚，担心用词不当)

Long-term 3D map maintenance [1, 2] is a fundamental capability required by a robot to explore complex environments. During single-session SLAM real-time reconstruction using ranging (LiDAR) sensor in urban sites，we encounter occlusions and changes as in Fig.1(体现遮挡和动态残影的submap)，which dilute the interpretability of pure-depth map and reduce the reusability of map. In doing so, precise instance segmentation of residual clouds and low-cost elimination of dynamic shadows should be solved suitably toward lifelong map management.

1) **Panoramic segmentation in locally unmeasurable situation**: As reported in [3], panoptic segmentation is an ensemble of both the semantic segmentation for static stuff and the instance segmentation for countable objects. Similar to the method of clustering for segmentation in [4], we utilizes curved voxels assisted with sensor intensity to assemble sparse clouds into instances. Afterward, they will be divided into several categories through various geometric features [5, 6]. The incomplete clouds predicted wrongly due to occlusion can be tracked and corrected in dynamic detection.

2) **Dynamic detection in object level and reclassification**: Some studies [7, 8, 9] regarded change detection as a post-process of comparing multiple pre-built scans associated with temporally distance and time. Following their philosophy, in this work, we take frames and poses result from Lio-sam [10] as prior. Assuming that scans are perfectly aligned, only objects with changing possibility participate in the interactive search, unlike recently methods [8, 9], which need to detect the whole cloud. Any element tracked that is regarded as stationary but with inconsistent label should be reclassified.

In sum, we proposed a novel closely coupled framework for lidar-based panoramic segmentation and dynamic detection. These two modules correlate with each other, the former provides the semantic basis, then the latter remove moving shadow and fixes the mis-segmentation. The proposed has the following contributions: (创新点需补充说明，第一点不确定)

• We designed a model to calibrate the senor intensity by the curvature, which improve the continuance of sensor intensity and reflective property.

• We proposed a non-learning method of clustering for segmentation utilizing curved voxels assisted with sensor intensity to compensate the deficiency of Euclidean clustering.

• We propose a novel dynamic detection based on object tracking, and rectify mis-segmentation according to classification consistence.

***Related Works*** (此部分需要重新组织语言)

1) Intensity calibration: The sensor intensity reading reveals surrounding surface reflectance structure. The intensity channel is noisy since it is affected by not only target surface characteristics (e.g., roughness, surface reflectance), but also acquisition geometry (e.g., distance) and instrument effects (e.g., transmitted energy) [23]. Hence calibration is necessary in order to reduce the disturbance by other factors [24, 25].

2) Clustering for segmentation: Existing point cloud clustering methods can be broadly summarized as four types [3], the Euclidean-based cluster in 3D space, clustering point cloud into supervoxel or superpoint, modified one-pass connected-component labeling on range image and modified two-pass connected-component labeling on range image. Euclidean cluster constructs the kd-tree on the entire point cloud and groups all neighbor points within a radius threshold as one instance [11]. Supervoxel [12] is designed on RGB-D point cloud as a counterpart of superpixel on the 2D image and [4] utilizes curved voxel to deal with sparse cloud. Scan-line Run (SLR) cluster [13] is a row-wise fast scan algorithm based on organized point cloud or range image. The accuracy of clustering is limited only according to the spatial position of points, [14, 15, 16] leverage reflection intensity of lidar to assist clustering.

3) Dynamic detection: Detect high dynamic object is to remove instant shadow, which is closely related to static map construction. The remote-sensing community requires high-cost, dense terrestrial laser scanning (TLS) point cloud data with accurately aligned poses to apply in time-consuming voxel ray casting-based methods [8, 17, 18]. [19] propose the novel concept called pseudo occupancy to express the occupancy of unit space and then discriminate spaces of varying occupancy. [9] as the change detection module in [2], it proposes a multiresolution range image-based false prediction reverting algorithm. Segmentation-based approaches generate correct segment points with dynamic labels and excluding them [20, 21], However, segmentation-based approaches currently rely heavily on the supervised labels and are vulnerable to human error or unknown classes [22].

***Overview***

The proposed method including two closely coupled modules. The overall pipeline is composed as in Fig.2 (补充整体方法结构图). The keyframes and pose graph are generated by Lio-sam [10]. Because the method is a post-procession after one single-session SLAM, we are mainly interested in producing a high-quality static and semantic map without concerning the processing speed.

The intensity value affected by multiple factors with noise [23], so it is difficult to completely restore the true value through the mathematical model to reflect the material of the object. In order to use the intensity in the instance segmentation, it is only necessary to ensure the continuity of the values in the adjacent space. We design a model based on curvature to calibrate intensity as in Fig.3(a)(展示intensity calibration结果).

An accurate instance segmentation is the premise of object-level dynamic detection. Firstly, the ground in point cloud should be removed, [26] proposes a method of multi-region ground fitting to deal with rugged scene. Then we utilizes curved voxels (CV) assisted with sensor intensity to cluster base on the baseline of [4]. The calibrated intensity can solve the mis-segmentation caused by the sparsity or occlusion. At the same time, we need to initialize the instance classification into absolute static (AS) and objects with change possibility (D) including high dynamic (HD) and low dynamic (LD), through geometric features [5, 6] as in Fig.3(b)(展示在intensity辅助作用).

After getting keyframes aligned and instance object, only the D and the object that does not satisfy with the classification consistency (NC) participate in CV-level dynamic detection, through interactive searching, the object is determined to be HD or LD, and NC is corrected as in Fig.4(标红高动态物体，并举例被纠正的实例，给出一个较大范围的地图).

***Methodology*** (暂时简要描述，或许在具体实验过程中会修改)

In this section, we give details of the framework proposed.

We define S as

|  |  |
| --- | --- |
|  | (1) |

where the S represents all data from Lio-sam [10], the are a 3D keyframe and a pose, Although using as pose piror for dynamic detection, we allow potential drifts, because the CV-level interactive searching has high robustness.

We define H as

|  |  |
| --- | --- |
|  | (2) |

where the H represents the hash table of one , is a id points to a CV. The H is built at a certain resolution, and includes intensity calibrated (), coordinates () and curvature ().

A. Intensity calibration (此处缺少举例一帧关于intensity的统计图，解释说明为什么)

In the scanning time of a single keyframe, we only consider the influence of distance (), incident angle () and reflective material () on sensor intensity () [23]. To satisfy the intensity continuity of adjacent CV, we can assume that the is constant, so the intensity calibrated is

|  |  |
| --- | --- |
| ( | (3) |

B. Clustering for segmentation and recognition (需要一个完整的流程图，每一步的处理展示)

A large number of ground points are not conducive to instance segmentation, so we need to remove it in advance by patchwork [26].

We divide the keyframe without ground into strict CVs at a certain resolution and deliver them containing points to the hash table for management. Conducting neighborhood search for each effective CV in the hash table, only when the coordinate and intensity condition are satisfied at the same time can they cluster together. After continuous search and iteration, the final segmentation result is obtained. Some point clouds are split or incomplete due to the occlusion of view angle or limited observation range, they always ss incorrectly segmented into multiple instances, then we can conduct neighborhood search at a larger range to cluster them into one instance. (这种情况举例说明图)

When getting several instances, We use geometric features [5, 6] to distinguish them into AS and D, the descriptors are computed resulting in a feature vector . （这边需要一个示意图）

Eigenvalue based: This descriptor of the segment’s point cloud are computed and combined in a feature vector of dimension 1x10. We compute the linearity, planarity, scattering, omnivariance, anisotropy, eigenentropy and change of curvature measures as proposed in [5].

Ensemble of shape histograms: This feature of dimension 1x640 is made of 10 histograms which encode the shape functions D2, D3 and A3 as described in [6]. The D2 shape function is a histogram of the distances between randomly selected point pairs while D3 encodes the area between randomly selected point triplets. The A3 shape function describes the angles between two lines which are obtained from these triplets.

C. Dynamic detection based on object tracked

After obtaining the instance segmentation of two adjacent keyframes , align them by their prior pose and the second keyframe is

|  |  |
| --- | --- |
|  | (4) |

where are the SE(3) pose composition operators [27], represents the inverse of .

After that, only the D and the object that does not satisfy with the NC participate in CV-level dynamic detection, when the NC is tracked by multiple keyframes, its classfication will be judged again. (此处需要一个很长很长的伪代码表和结果展示😊)

*Experiment* (实验部分还没有整理结束)

文章提到的方法整体效果在一些场景还是效果不佳，主要的难点还是视角导致观测不一致，需要继续改进。

目前仅在LT-mapper提供的parkinglot数据集测试过。

倘若测试结束，会把semantickitti视为主要数据集，在其上进行实例分割IOU比对和动态检测测试，不确定是否要与参考文献中剔除的几种动态剔除算法进行客观比较，因为有些算法没有开源或无法完全复现。

除去实验部分和结论部分，以上是整个算法思路简单介绍，其中实例分割与动态检测之间的紧耦合是上次组会至今的一些想法，上次提到“在动态检测方面最好可以在有一个创新点”。

没有写中文是因为感觉英文可能在某些概念表达的更准确一点。

***Reference***

|  |  |
| --- | --- |
| [1] | Pomerleau, François, et al. "Long-term 3D map maintenance in dynamic environments." 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2014. |
| [2] | Kim, Giseop, and Ayoung Kim. "LT-mapper: A modular framework for lidar-based lifelong mapping." 2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022. |
| [3] | Zhao, Yiming, Xiao Zhang, and Xinming Huang. "A technical survey and evaluation of traditional point cloud clustering methods for lidar panoptic segmentation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. |
| [4] | Park, Seungcheol, et al. "Curved-voxel clustering for accurate segmentation of 3D LiDAR point clouds with real-time performance." 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019. |
| [5] | Wohlkinger, Walter, and Markus Vincze. "Ensemble of shape functions for 3d object classification." 2011 IEEE international conference on robotics and biomimetics. IEEE, 2011. |
| [6] | Weinmann, Martin, Boris Jutzi, and Clément Mallet. "Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2.3 (2014): 181. |
| [7] | Wellhausen, Lorenz, et al. "Reliable real-time change detection and mapping for 3d lidars." 2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR). IEEE, 2017. |
| [8] | Schauer, Johannes, and Andreas Nüchter. "The peopleremover—removing dynamic objects from 3-d point cloud data by traversing a voxel occupancy grid." IEEE robotics and automation letters 3.3 (2018): 1679-1686. |
| [9] | Kim, Giseop, and Ayoung Kim. "Remove, then revert: Static point cloud map construction using multiresolution range images." 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020. |
| [10] | Shan, Tixiao, et al. "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping." 2020 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2020. |
| [11] | Rusu, Radu Bogdan, and Steve Cousins. "3d is here: Point cloud library (pcl)." 2011 IEEE international conference on robotics and automation. IEEE, 2011. |
| [12] | Papon, Jeremie, et al. "Voxel cloud connectivity segmentation-supervoxels for point clouds." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013. |
| [13] | Zermas, Dimitris, Izzat Izzat, and Nikolaos Papanikolopoulos. "Fast segmentation of 3d point clouds: A paradigm on lidar data for autonomous vehicle applications." 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017. |
| [14] | Tatoglu, Akin, and Kishore Pochiraju. "Point cloud segmentation with LIDAR reflection intensity behavior." 2012 IEEE International Conference on Robotics and Automation. IEEE, 2012. |
| [15] | Reymann, Christophe, and Simon Lacroix. "Improving LiDAR point cloud classification using intensities and multiple echoes." 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015. |
| [16] | Wang, Zhao, et al. "Vehicle detection based on point cloud intensity and distance clustering." Journal of Physics: Conference Series. Vol. 1748. No. 4. IOP Publishing, 2021. |
| [17] | Chen, Chi, and Bisheng Yang. "Dynamic occlusion detection and inpainting of in situ captured terrestrial laser scanning point clouds sequence." ISPRS Journal of Photogrammetry and Remote Sensing 119 (2016): 90-107. |
| [18] | Gehrung, Joachim, et al. " An approach to extract moving objects from mls data using a volumetric background representation." ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences 4 (2017). |
| [19] | Lim, Hyungtae, Sungwon Hwang, and Hyun Myung. "ERASOR: Egocentric ratio of pseudo occupancy-based dynamic object removal for static 3D point cloud map building." IEEE Robotics and Automation Letters 6.2 (2021): 2272-2279. |
| [20] | Ruchti, Philipp, and Wolfram Burgard. "Mapping with dynamic-object probabilities calculated from single 3d range scans." 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018. |
| [21] | Sun, Li, et al. "Recurrent-octomap: Learning state-based map refinement for long-term semantic mapping with 3-d-lidar data." IEEE Robotics and Automation Letters 3.4 (2018): 3749-3756. |
| [22] | Wong, Kelvin, et al. "Identifying unknown instances for autonomous driving." Conference on Robot Learning. PMLR, 2020. |
| [23] | Kashani, Alireza G., et al. "A review of LiDAR radiometric processing: From ad hoc intensity correction to rigorous radiometric calibration." Sensors 15.11 (2015): 28099-28128. |
| [24] | Guo, Jiadong, et al. "Local descriptor for robust place recognition using lidar intensity." IEEE Robotics and Automation Letters 4.2 (2019): 1470-1477. |
| [25] | Levinson, Jesse, and Sebastian Thrun. "Unsupervised calibration for multi-beam lasers." Experimental Robotics. Springer, Berlin, Heidelberg, 2014. |
| [26] | Lim, Hyungtae, Minho Oh, and Hyun Myung. "Patchwork: concentric zone-based region-wise ground segmentation with ground likelihood estimation using a 3D LiDAR sensor." IEEE Robotics and Automation Letters 6.4 (2021): 6458-6465. |
| [27] | Smith, Randall, Matthew Self, and Peter Cheeseman. "Estimating uncertain spatial relationships in robotics." Autonomous robot vehicles. Springer, New York, NY, 1990. 167-193. |