Graph-based LiDAR-Inertial SLAM Enhanced by Loosely-Coupled Visual Odometry

Vsevolod Hulchuk Jan Bayer Jan Faigl

***Abstract*— In this paper, we address robot localization using Simultaneous Localization and Mapping (SLAM) with Light Detection and Ranging (LiDAR) perception enhanced by visual odometry in scenarios where laser scan matching can be ambiguous because of a lack of sufficient features in the scan. We propose a Graph-based SLAM approach that benefits from fusing data from multiple types of sensors to overcome the disadvantages of using only LiDAR data for localization. The proposed method uses a failure detection model based on the quality of the LiDAR scan matching and inertial measurement unit data. The failure model improves LiDAR-based localization by an additional localization source, including low-cost black-box visual odometers like the Intel RealSense T265. The proposed method is compared to the state-of-the-art localization system LIO-SAM in cluttered and open urban areas. Based on the performed experimental deployments, the proposed failure detection model with black-box visual odometry sensor yields improved localization performance measured by the absolute trajectory and relative pose error indicators.**

I. Introduction

The localization is important for many mobile robotics applications, including underground exploration, indoor inspection, and outdoor navigation. In these scenarios, the robot's sensors-based localization is needed if external localization systems, such as satellite navigation, are unavailable or do not work reliably because of signal reflections from tall structures. The widely adopted method for localizing a robot using its sensors is *Simultaneous Localization and Mapping* (SLAM) [1], which becomes the de-facto standard in applications where a prior map of the environments cannot be utilized. SLAM can be based on data from various sensors, including *Light Detection and Ranging* (LiDAR) laser scanners [2], visual cameras [3], *Inertial Measurement Units* (IMU), or wheeled odometry, to name just a few.

Using exteroceptive sensors to build a map of the operational environment within which the robot is localized allows for decreasing the localization drift compared to purely proprioceptive incremental methods such as odometry and dead reckoning. Even matching only consecutive frames using *Visual Odometry* (VO) [4] helps to overcome drifts of IMU measurements or slippage of wheeled odometry. Nevertheless, the map's quality is important and related to the data quality, specifically the depth estimates of the range measurements. Current LiDAR sensors provide relatively precise range measurements and can have resolution over one hundred lines [5]. These properties make them suitable for localization, especially in cluttered environments, where LiDAR scans can be precisely matched with respect to (w.r.t.) each other [6]. However, the scan matching may be ambiguous in long corridors or flat fields, leading to localization failure or high drift, as depicted in Fig. 1.

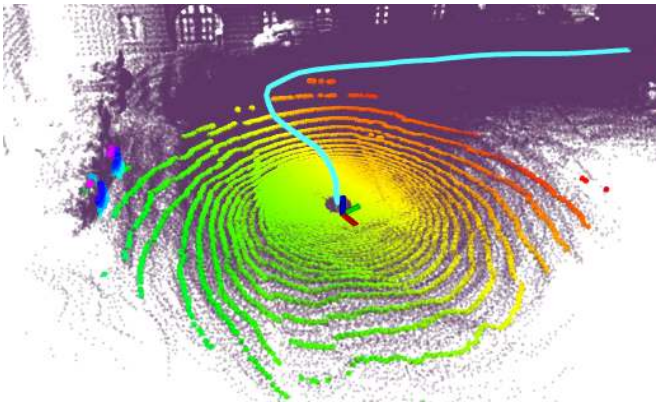


Fig. 1. A situation where LiDAR scan can be aligned with the previous scans (map) in multiple ways since the area covered by the scan is mostly flat, which prompts scan-matching ambiguity. A dense map of the environment is in purple. Distance data of the current scan are denoted in blue to red.

Incremental localization methods, such as IMU and odometry-based methods, including VO, might help to overcome areas where LiDAR scan matching is ambiguous locally, albeit it can lead to higher drift than the LiDAR-based SLAM in the long run. Thus, combining data sources can be advantageous in SLAM, and two main sensor fusion approaches can be found in the literature. The first is tightly-coupled methods that account for sensor raw data, such as in *LiDAR Inertial Odometry via Smoothing and Mapping* [7] (LIO-SAM), where an IMU displacement measurement serves as an initial guess for the scan-matching.

The second class of methods uses a loosely-coupled approach to fuse multiple localization sources, meaning that two displacement outputs from localization systems are fused at the top. Consequently, the resulting estimation tends to be more robust as a failure of one source does not provoke the failure of another one. Also, loose coupling allows the integration of several independent localization systems, making the whole system modular and easily replaceable compared to tightly-coupled systems.

In this paper, we propose an extension of the Pose-

Authors are with the Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague, Czechia.{hulchvse|bayerja1|faiglj}@fel.cvut.cz

The presented work has been supported by the Czech Science Foundation (GAČR) under the research project No. 22-05762S. The support of the Grant Agency of the CTU in Prague under grant No. SGS22/168/OHK3/3T/13 is also gratefully acknowledged.

979-8-3503-0704-7/23/$31.00 ©2023 IEEE

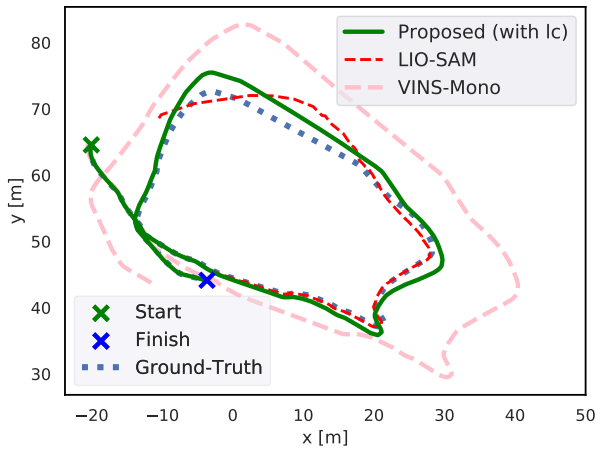


Fig. 2. Result of the proposed Graph-based LiDAR-Inertial SLAM with loosely coupled visual odometry on the rural dataset. Notice that even though the matching of the LiDAR scans was unsuccessful in some areas, the proposed method can use scale and pose drifting visual localization VINS-Mono to overcome such areas and close the loop.

Graph SLAM, combining tightly and loosely coupled ideas. We propose to use tightly coupled sensory fusion between LiDAR and IMU, similar to LIO-SAM. Besides, the developed solution allows utilizing additional sources of pose estimates in a loosely-coupled manner, improving the SLAM performance when LiDAR data matching fails. Various methods of incremental localization can be loose-coupled in the proposed method, such as visual localization, wheel odometry, RADAR-based localization [8], or thermal-inertial odometry [9]. Nevertheless, the properties of the proposed method are demonstrated while using a black box embedded stereo visual localization system, the Intel RealSense T265 (T265) [10], and visual-inertial localization VINS-Mono [11].

We propose a relatively straightforward failure detection model that triggers the incorporation of the additional low-quality pose estimate into the developed Pose-Graph SLAM. The model assesses LiDAR scan matching quality to indicate possible matching failure and IMU-based pose change prediction to confirm the failure for switching the pose estimate. Incorporating the additional localization source is enhanced by an auto-scaling mechanism and improved graph structure.

The triggering threshold has been experimentally established using a real robotic system; the proposed Graph-based SLAM has been deployed in several deployments and compared with the selected state-of-the-art LiDAR-based SLAM. Based on the experimental results, the proposed method demonstrates improvement of the localization performance by the additional source of the incremental localization while not sacrificing LiDAR-based performance in scenarios where LiDAR scan matching performs well, see Fig. 2. We consider the main contributions of the proposed approach as follows.

- Modular enhancement of existing Pose-Graph SLAM by a loosely coupled additional localization system.
- Two-step failure detection model, allowing detection of scan matching failure.

The rest of the paper is organized as follows. Section II overviews the related literature, including a brief description of the selected reference LIO-SAM framework. The proposed method is described in Section III. Experimental results are reported and discussed in Section IV. Finally, the paper is concluded in Section V.

II. Related Work

Many SLAM systems have been proposed [2], [3] and evaluated in the Kitti benchmark [12]. Based on the results reported in [12], most of the top ten performing methods use LiDAR measurements for robot pose estimation. One of the top-performing LiDAR-based methods is LOAM [13], albeit it lacks an explicit loop closure and is limited to only one type of sensor. On the other hand, multiple possible sensors are used in the RTAB-Map [14], which is a general tightly-coupled LiDAR-Visual SLAM framework using multiple graph frameworks. However, failure handling is not resolved in the framework yet, and the authors mention it as a future research direction.

Contrary to the RTAB-Map, the authors of [15] loosely coupled several localization sources. The first step of the coupling is the sanity check, where localization failures are identified for each localization source using the dynamic model of the vehicle. Then, Chamfer distance-based [16] score is used to select the best pose estimate. The advantage of [15] is high robustness, but since the localization sources are completely independent, the visual odometry cannot help the LiDAR-based SLAM to close the loop in the case of temporal LiDAR-based SLAM failure. Furthermore, the Chamfer distance-based score measures the alignment of the LiDAR scans. It does not directly detect when the perfect alignment of LiDAR scans may correspond to a wrong displacement in monotonous corridors or fields.

In [17], the authors review available sensory fusion approaches for LiDAR-Visual SLAM. They mention that the graph-based SLAM [1] is often used for sensor fusion because it abstracts from raw measurements. The approach represents measurements, poses, and observations in a graph structure. Pose-graph SLAM [18] is a specific kind of graph-based SLAM that is the most used nowadays. It restricts the graph's nodes to be poses and positions of robots and landmarks and edges to be measurements-based constraints between them. The authors of [19] demonstrate the computational advantages of the pose-Graph SLAM for large-scale maps, comparing the solution with conventional filtering approaches. The approach is further explored in [20], where the authors review iSAM2 [21], which iteratively re-optimizes only nodes influenced by new observations. Multiple graph optimization frameworks have been proposed, but ORB-SLAM3 [22] uses the g2o library [23] in Loop Closure for Bundle Adjustment [24] to improve the Visual-Inertial Odometry. In VINS-Mono [11], the authors present a Visual-Inertial SLAM solution that fuses a monocular camera and IMU in a tightly-coupled manner for obtaining odometry and optimizing the global trajectory with pose-graph SLAM.

LiDAR-Inertial odometry is the core of LIO-SAM [7] that uses scan matching based on LOAM [13], where the initial guess of the LiDAR pose is based on integrated

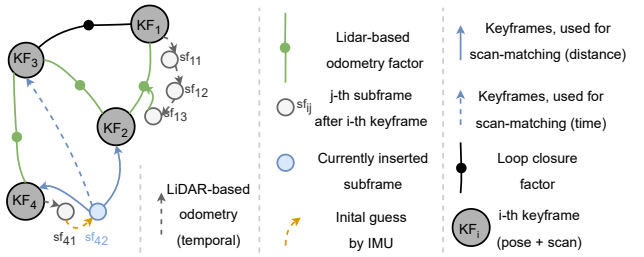


Fig. 3. Map optimization graph in LIO-SAM [7].

IMU measurements. The scans aligned by LiDAR odometry are marked as keyframes if the distance from a pose corresponding to the previous keyframe is above a certain threshold. Otherwise, the pose is treated as a temporal subframe. The relations between the keyframes are represented by constraints that are used to construct a sparse graph within the GTSAM [25] optimization framework. Loop closure is then performed as a parallel process using the *Iterative Closest Point* (ICP) [26], and the loop constraints are added if the ICP converges. For the loop closure detection, the latest keyframe is attempted to be matched against the nearby keyframes, including recent keyframes and keyframes that are close to the current robot pose. If the matching of the keyframes is successful, the transformation between them is inserted into the graph as a constraining factor. The graph structure is illustrated in Fig. 3.

LIO-SAM is further extended by tightly-coupled VO in *LiDAR-Visual-Inertial Odometry via Smoothing and Mapping* (LVI-SAM) [27]. LVI-SAM tightly couples LIO-SAM with Visual SLAM VINS-mono [11] to improve performance in challenging scenarios using sensor-specific failure detectors for LiDAR and VO. However, such an approach does not support flexibility in changing the source of additional localization systems and restricts end-users to specific additional sensors (camera) and algorithms (VINS-mono).

Based on the literature review, we opt for LIO-SAM as a suitable base system for integrating the additional sensor for localization. It provides the advantage of a great performance of LiDAR-based methods [12] while avoiding the disadvantage of the tightly-coupled visual odometry of LVI-SAM, which supports only the specific method of visual odometry. LIO-SAM framework accounts for ambiguities of the scan-matching by checking scan-matching convergence. The convergency is then reflected in uncertainties while optimizing IMU measurements. On the other hand, the system is developed for structure-rich environments. Besides, it does not explicitly handle situations where the scan-matching results are completely unusable. Both the drawbacks are addressed by the proposed loosely-coupled combination of LiDAR-Inertial SLAM and VO.

III. Proposed Method

The proposed loosely coupled VO with the graph-based LiDAR-Inertial SLAM leverages LIO-SAM [7]. It uses the same way of calculating LiDAR-Inertial odometry (referred to as LiDAR-based odometry). However, we modify the factor graph construction to incorporate measurements from an additional localization system, such as VO. The inputs to the proposed method are LiDAR scans, IMU measurements, and pose estimates of the additional localization system(s). Although the proposed approach is general, we consider *Visual-Inertial Odometry* (VIO) as the additional localization system that produces a 6 DoF robot pose estimate to present the proposed concept. The following assumptions are made in the design of the proposed method.

- For simplicity of the description, only a single additional localization system VIO is used, albeit multiple localization sources can be straightforwardly utilized.
- The additional localization system provides pose estimates w.r.t. to the same coordinate frame as the LiDAR-based odometry.
- All sensors' data is synchronized in time.

The proposed method consists of two parts: (i) *failure detection*, which indicates that LiDAR-based odometry failed, and (ii) *visual localization integration*, which integrates the additional localization into the factor graph in the case of the detected failure.

A. Failure Detection

Failure detection starts with the failure indication defined by the *Failure indicator* I_{fail}. If the indication is positive, *Failure resolution* determines if VIO provides a more suitable pose estimate than the LiDAR-based odometry. The overview of the failure detection process is depicted in Fig. 4, and it works as follows.

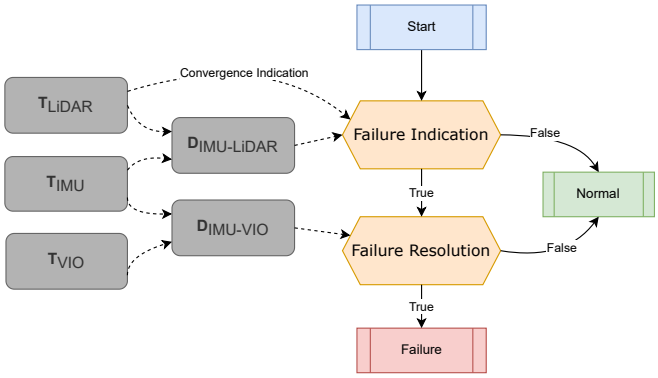


Fig. 4. Failure detection algorithm.

The failure indicator I_{Fail} is combined from two components: *convergence indicator* I_{Conv} and *IMU-based indicator* I_{IMU} as

$$I_{\text{Fail}} = I_{\text{IMU}} \texttt{ or } I_{\text{Conv}}. \tag{1}$$

I_{Conv} is triggered when the LiDAR scan matching does not converge, but it might not cover all cases when it is suitable to switch to VIO. Therefore, we also use I_{IMU} to increase the failure detection rate, which is supported by the experimental results reported in Section IV-A. The advantage of I_{IMU} is that it is not directly influenced by a lack of spatial and visual features in the environment. The indicator uses a rough estimation of the robot motion by IMU-based odometry increment $T_{\text{IMU}} \in SE(3)$ to estimate the adequacy

of the LiDAR-based odometry increment $T_{\text{LiDAR}} \in SE(3)$. Using LiDAR scans at 10 Hz to improve the estimate of the robot motion ensures that the IMU-based motion estimates do not suffer from localization drift by integrating IMU measurements for an extended period. The difference of the increments $D_{\text{IMU-LiDAR}}$ is computed as

$$D_{\text{IMU-LiDAR}} = T_{\text{IMU}} \cdot T_{\text{LiDAR}}^{-1}. \tag{2}$$

We analyze the norm of the rotational component and a translational component of the difference defined by

$$\begin{aligned} r_{\text{IMU-LiDAR}} &= ||\operatorname{rot}(D_{\text{IMU-LiDAR}})||^{\text{ANG}} \\ t_{\text{IMU-LiDAR}} &= ||\operatorname{trans}(D_{\text{IMU-LiDAR}})|| \end{aligned} \tag{3}$$

where $\operatorname{rot}(D_{\text{IMU-LiDAR}}) \in SO(3)$ is the rotational component and $\operatorname{trans}(D_{\text{IMU-LiDAR}}) \in R^3$ is the translational component of $D_{\text{IMU-LiDAR}}$. The term $||\cdot||^{\text{ANG}}$ denotes the angular metric of the rotation that is determined as a rotation angle of the angle-axis representation of the rotation.

The IMU indicator I_{IMU} works as an outlier detector [28], and it is defined as logical `or` of two threshold values

$$I_{\text{IMU}} = (r_{\text{IMU-LiDAR}} > c_r) \texttt{ or } (t_{\text{IMU-LiDAR}} > c_t) \tag{4}$$

that triggers when either the rotational or translational component of the difference $D_{\text{IMU-LiDAR}}$ is larger than the corresponding thresholds c_r and c_t, respectively. The thresholds are determined experimentally using outlier detection methodology; see the following section.

The failure resolution begins if the failure indicator I_{Fail} (1) is true. The VIO pose estimate is used if it is significantly closer to the IMU-based odometry than the LiDAR-based odometry. Thus, the resolution is defined by the following condition

$$\begin{gathered} (r_{\text{IMU-VIO}} < \alpha \cdot r_{\text{IMU-LiDAR}}) \\ \texttt{and} \\ (t_{\text{IMU-VIO}} < \alpha \cdot t_{\text{IMU-LiDAR}}) \end{gathered} \tag{5}$$

where $r_{\text{IMU-VIO}}$ and $t_{\text{IMU-VIO}}$ are defined similarly to the IMU-LiDAR difference $D_{\text{IMU-LiDAR}}$ defined in (3).

Note that the LiDAR-based odometry failure might be indicated based on I_{Fail}, but failure resolution (5) would not activate the usage of VIO pose estimate if the latter does not improve the LiDAR-based one. We incorporate an empirically obtained $\alpha = 0.8$ factor in the IMU-LiDAR difference when considering additional odometry over LiDAR. It is done to ensure the significance of any potential improvement by additional odometry and account for IMU noise.

B. Visual Odometry Integration – Scale Self-Adjustment

Let us suppose the LiDAR-based odometry failure is indicated, and VIO provides more precise localization according to the rule (5). In that case, the VIO is incorporated into the factor graph in place of the LiDAR-based odometry, introducing a constraint between the keyframes if the keyframe is inserted. Since the additional odometry (such as visual or wheeled) might suffer from a wrong scale or slow scale drift, the proposed method performs dynamic scale self-adjustment, estimating the scale of the odometry when the LiDAR-based localization is considered sufficiently precise.

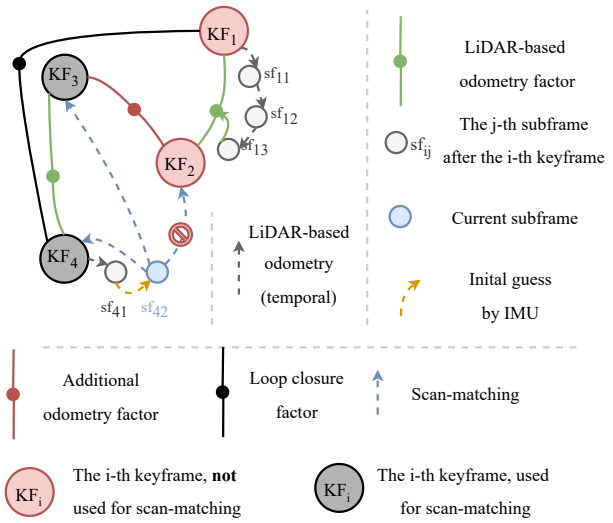


Fig. 5. The proposed method for combining the LiDAR-based odometry with VIO-based pose estimate increments.

We propose to utilize the median value of the moving window to compute the scale. In particular, 500 keyframes-long window includes the past ratios of the absolute values of the translations $t_{\text{VIO}}/t_{\text{LiDAR}}$, where $t_{\langle\text{source}\rangle}$ is the norm of the translational part of the odometry increment. Then, the factor graph structure is created according to the scheme depicted in Fig. 5 as follows.

- The LiDAR-based odometry creates constraints between the previous and the new keyframes based on scan-matching when the LiDAR-based odometry works successfully. When the new LiDAR scan (frame) is available, it is scan-matched against a reference map combined with the nearby keyframes to create such a constraint. Similarly to LIO-SAM, only if the estimated pose increment exceeds a configurable threshold the frame is inserted into the map as a keyframe. Otherwise, it is treated as a temporal sub-frame to improve the initial guess of the next frame pose and output localization information.
- On the other hand, the VIO constraint is inserted instead of the LiDAR-based one if the failure is detected. However, in contrast to LiDAR-based constraints, the VIO-based ones are not guaranteed to be optimized for the keyframes alignment as they optimize visual features alignment and may suffer from the incorrect and drifting scale. Thus, combining keyframes connected with VIO-based constraints can result in a poorly aligned reference map, and a new LiDAR scan would not be successfully matched against such a reference map. Therefore, only keyframes inserted after the last VIO usage are combined in the reference map when the new LiDAR scan is processed.

Finally, it is necessary to properly handle Loop closure constraints of the graph-based SLAM that aim to match keyframes that are far from each other. These constraints may fix the drift introduced by the VIO constraints. However, false loop closures may appear for the structure-

less keyframes, consequently breaking the graph. Therefore, keyframes corresponding to the LiDAR-based odometry failure are deemed unsuitable for loop closures. Further, the inserted VIO constraints are set to have ten times larger uncertainty than the LiDAR-based ones to ensure that loop closure constraints will fix only VIO constraints without affecting LiDAR-based constraints significantly. The effect of the proposed loop closing system has been experimentally examined, and results are reported in the following section; in particular, the effect is demonstrated in Fig. 10.

IV. Experimental Results

The proposed method has been experimentally validated using a four-wheeled skid-steered robot Husky. The robot was equipped with the Ouster OS0 LiDAR with 128 lines, and the maximum range is approximately $50\,\mathrm{m}$, a 9-axis IMU Xsens MTi-30, and a fisheye stereo tracking camera, the *Intel RealSense T265* (T265). T265 provides out-of-the-box VIO odometry, but its internal loop closures have been disabled to make it compliant with made assumptions on the additional localization systems. The careful extrinsic calibration by measuring the relative pose of T265 w.r.t. the LiDAR was done to comply with the proposed method's assumptions. Thus T265 pose estimations were transformed into the LiDAR frame before using them by the proposed method. The 3 DoF ground truth localization of the robot has been recorded using the Leica TS16 total station, shown in Fig. 6a.

(a) Total station setup (b) Bird's-eye view on urban campus area

Fig. 6. The urban experimental scenario at the Czech Technical University in Prague campus at Charles Square.

The LIO-SAM itself was already evaluated using publicly available datasets in [7]. In this work, we primarily focus on areas that induce the scan matching ambiguity. Thus, two environments have been considered for system performance evaluation. The first environment is the backyard area of the Czech Technical University (CTU) in Prague campus at Charles Square, depicted in Fig. 6b. The second environment is a parking lot at Prague's outskirt visualized in Fig. 7. While the first environment can be considered structure-rich, the parking lot in the rural area contains wide-open locations where LiDAR scans do not provide sufficient features for successful scan matching. The testing environments are denoted as *campus* and *rural* scenarios.

The length of the traveled trajectory is $285\,\mathrm{m}$ and $300\,\mathrm{m}$ for the campus and rural scenarios, respectively. The proposed method is examined with different failure indicators to justify the combined indicator denoted IMU + Convergence. Besides, the performance is compared with the LIO-SAM [7] as the former localization method to show the benefits of the proposed loosely-coupled VIO.

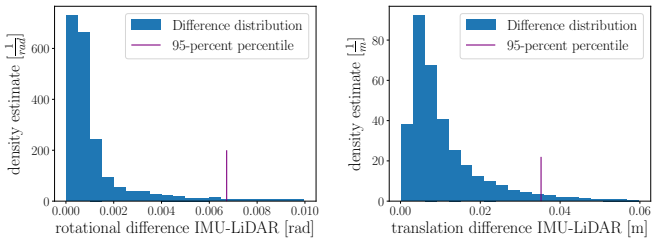
(a) Used wheeled robot (b) Bird's-eye view on a parking lot

Fig. 7. Experimental parking lot scenario in Prague's outskirts.

The evaluation is based on the methodology [29] using medians of the *absolute trajectory error* ATE_t and *relative pose error* RPE_t indicators considering the translational parts of the localization error. In particular, ATE_t evaluates the global accuracy of the trajectory, while the median RPE_t estimates the local consistency of the localization (drift). For RPE_t, the step Δ is set to $1\,\mathrm{m}$, which corresponds to the minimum distance between consecutive poses. Besides, the standard deviation STD_t of the RPE_t is reported to account for outliers. The indication `Fail` is used in cases when the system received corrupted odometry, which led to wrong IMU bias estimation. Such situations prevented the localization system from recovering.

A. Parameterization of the Failure Detection

The failure detection model's parameters have to be estimated, and the following intent describes the estimation. Note that these results are only used to calibrate the proposed method but do not serve to estimate the performance of the proposed method. The proposed IMU-based failure detection model is based on outlier detection [28] for differences between IMU-based and LiDAR-based pose increment $D_{\text{IMU-LiDAR}}$ as of (4) with two established threshold values c_r and c_t, which single out the outliers (failures). The threshold values are set as follows.



(a) Rotational part (b) Translational part

Fig. 8. Histograms of $D_{\text{IMU-LiDAR}}$ differences in the campus scenario. The threshold values c_r and c_t are established as 95 percent quantiles depicted by the vertical line segment.

We model the baseline distributions of differences $r_{\text{IMU-LIDAR}}$ and $t_{\text{IMU-LIDAR}}$ in the non-failure scenario and set the outliers thresholds as 95 percent quantiles of the

distributions as shown in Fig. 8. LiDAR-based odometry provides satisfactory results that can be treated as "Non-Failure" in the full-range campus dataset; thus, the data is used to model the distribution. Note that the data used does not intersect with data from the campus dataset in reported evaluation tables, ensuring that the model tuning and evaluation are performed using different data.

B. Performance in the Campus Scenario

The robot has been operated in the campus scenario where the total station provides the ground truth data for evaluation. We examine the localization performance of the proposed method based on the scan-matching failure indicator I_{Conv} only and with both indicators I_{Conv} and I_{IMU}. First, we examine the method using only the scan-matching failure indicator and using both indicators. Limiting the LiDAR range to $10\,\text{m}$ has induced the scan-matching ambiguity as illustrated in Fig. 1.

TABLE I

LOCALIZATION PERFORMANCE IN THE CAMPUS SCENARIO WITH AND W/O FAILURE DETECTION AND LIDAR RANGE CROPPED TO $10\,\text{m}$

Method / Failure Indicator	ATE_t [m]	RPE_t [m]	STD_t [m]
LIO-SAM [7] (No indicator)	`Fail`	`Fail`	`Fail`
Proposed IMU	`Fail`	`Fail`	`Fail`
Proposed Convergence	5.35	0.08	0.22
Proposed IMU + Convergence	**4.70**	**0.06**	**0.26**

`Fail` indicates the method has not been able to produce reasonable results.

The results in Table I indicate that a solo IMU-based indicator cannot detect failure by itself but significantly improves the performance when combined with the convergence-based indicator, reflected in more precise localization results.

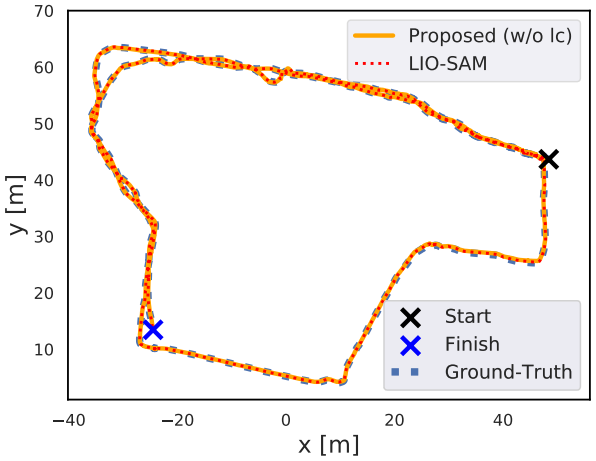
TABLE II

LOCALIZATION PERFORMANCE IN THE CAMPUS SCENARIO WITH FULL LIDAR RANGE AND LIMITED RANGE TO $10\,\text{m}$

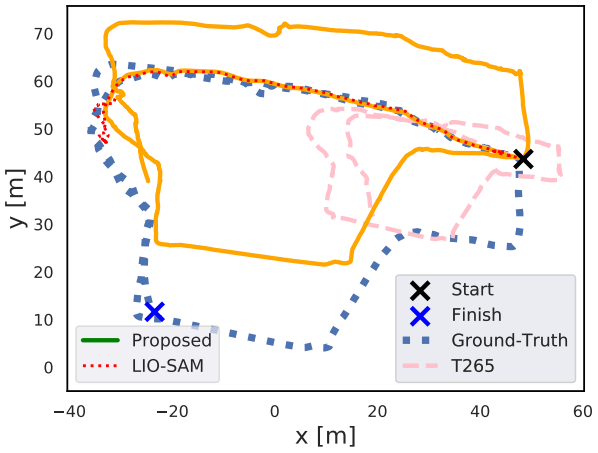
Method	Full range			Limited range		
	ATE_t [m]	RPE_t [m]	STD_t [m]	ATE_t [m]	RPE_t [m]	STD_t [m]
LIO-SAM	0.08	0.04	0.03	`Fail`	`Fail`	`Fail`
T265	16.40	0.83	0.50	16.40	0.83	0.50
T265 scaled*	7.06	0.20	0.30	7.06	0.20	0.30
Proposed method (w/o lc)	0.13	0.04	0.03	4.70	**0.06**	**0.26**
Proposed method (with lc)	0.13	0.04	0.03	**2.8**	0.08	0.3

`Fail` indicates the method has not been able to produce reasonable results.
*Odometry scaled to optimize ATE_t with the constant scale factor after the experiment.

Next, we examine the proposed method and LIO-SAM in two setups: *full range* and *limited range*. Besides, we consider the method in two setups: without and with the *loop closure* (lc). The methods are fed with data directly captured by LiDAR without any range restrictions for the full range. However, for the limited range, LiDAR's range is cropped to $10\,\text{m}$ to examine the localization system performance under conditions where LiDAR scan matching might be ambiguous. In addition to LIO-SAM and the proposed method, we evaluate the localization provided by the T265 with its and with the optimal scale. The optimal scale is the scale minimizing the ATE_t for the T265 trajectory, estimated after the experiment and applied to the entire T265 trajectory. It is considered to estimate the best possible reachable result using T265 with the constant scale. Nevertheless, the proposed method is inputted with the raw T265 data, estimating the scale online using the method introduced in Section III-B. The performance indicators are depicted in Table II and trajectories in Fig. 9.



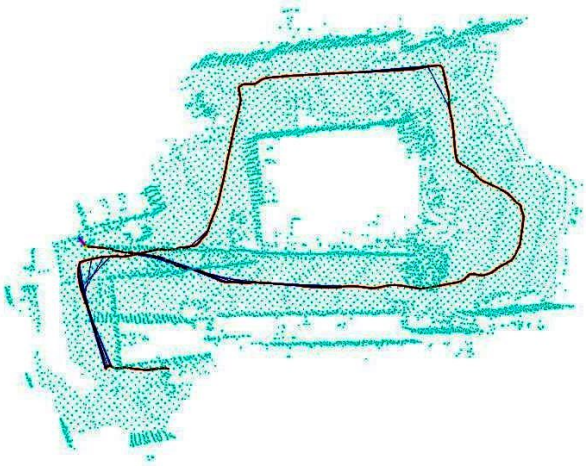
(a) Full range



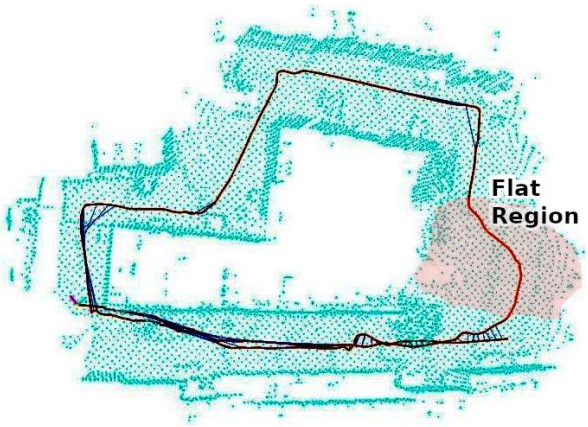
(b) Limited range to $10\,\text{m}$

Fig. 9. Aligned trajectories in the campus dataset.

The presented results support the hypothesis that the environment is structure-rich for the full range and that LIO-SAM and the proposed method provide competitive results. On the other hand, T265 suffers from localization drift and provides worse results, but as rarely used, it only slightly worsens the performance of the proposed method compared to LIO-SAM. However, when the LiDAR range is cropped to $10\,\text{m}$, LIO-SAM fails to output any feasible result once the robot enters the area where it is too far from the buildings. The limited LiDAR scans are ambiguous for the scans-matching algorithm, and the whole localization fails. The proposed method handles these ambiguous LiDAR scans by switching to VIO, as shown in Fig. 9b. Although it introduces



(a) Before loop closure



(b) After loop closure

Fig. 10. Loop closure example conducted by the proposed method in the campus scenario.

a drift caused by the additional odometry, it performs best.

Finally, we examined the loop closure of the proposed method. The obtained maps and trajectories before and after the loop closure are depicted in Fig. 10. It can be observed that the loop closure compensates for the drift introduced by the relatively low-quality VIO. The resulting map is aligned because the loop closure constraint optimized the trajectory where the LiDAR-based odometry was ambiguous, which is the flat region at the right part of the map. At the same time, LiDAR-based constraints that align keyframes with no ambiguity are almost not changed because those have much lower uncertainty in the graph structure. In Table II, it can be seen that for the limited range setup, the loop closure highly improved global consistency reflected by the ATE_t metric while slightly worsening local consistency reflected by the RPE_t metric.

C. Performance in the Rural Scenario

The next deployment took place in the rural scenario with wide open areas. In this case, we use fisheye images from the T265 processed by the VINS-mono [11] odometry to show the flexibility of the proposed method to incorporate measurements from various types of additional localization systems. Thus, we examine the performance of LIO-SAM, VINS-Mono, and two variants of the proposed method, without and with loop closure (lc). The results are summarized in Table III.

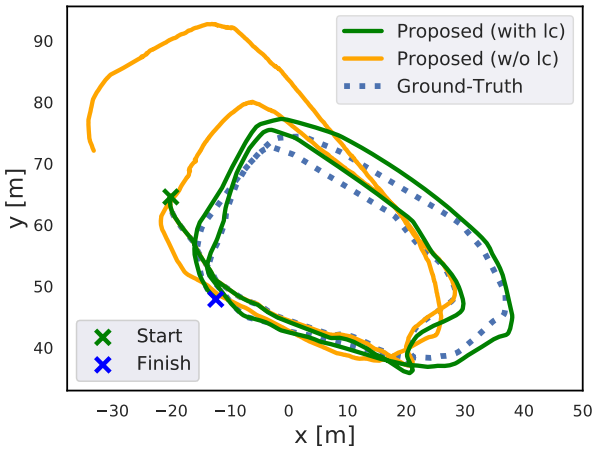
TABLE III

LOCALIZATION PERFORMANCE IN THE RURAL SCENARIO

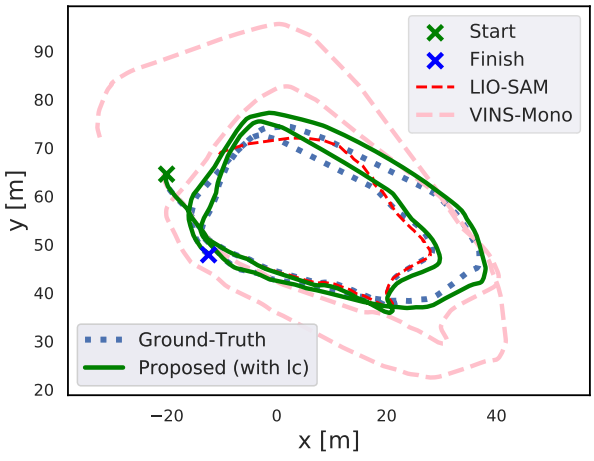
Method	ATE_t [m]	RPE_t [m]	STD_t [m]
LIO-SAM	`Fail`	`Fail`	`Fail`
VINS-Mono	10.9	0.39	0.25
VINS-Mono scaled*	4.97	0.42	0.18
Proposed method (w/o lc)	7.7	0.19	**0.13**
Proposed method (with lc)	**2.4**	**0.15**	1.0

`Fail` indicates the method has not been able to produce reasonable results.
*The odometry scaled to optimize ATE_t with the constant scale factor after the experiment.

From the results, it can be seen that the proposed method performed better than LIO-SAM since it did not fail. VINS-mono provided the robot with smooth but scale and pose drifted odometry. It can be seen in Fig. 11a that due to the loop closure, the proposed method is able to re-estimate the whole trajectory, mainly altering the part where the additional odometry was used. The trajectories the evaluated methods provide are depicted in Fig. 11b.



(a) Loop closure effect



(b) Full trajectories

Fig. 11. Proposed method results in parking dataset.

V. Conclusion

We propose an augmentation of the graph-based SLAM based on LiDAR-Inertial odometry in a modular way for incorporating an additional localization source. Although the additional localization is combined with the LiDAR-Inertial odometry in a loosely-coupled manner, the resulting factor graph can be optimized by identifying loop closures based on LiDAR data even in cases when LiDAR scans matching failed at some part of the trajectory. The proposed improvement is based on failure detection by an IMU model, setting the graph constraints uncertainties according to the nature of localization sources and setting the selection rules for keyframes usage. The proposed method has been tested in urban and rural scenarios demonstrating competitive results compared to LIO-SAM when LiDAR scan matching is not ambiguous. The proposed method outperforms LIO-SAM when the ambiguity of the scan matching induced high localization drift and even a failure of LIO-SAM. The results also indicate that the proposed method can utilize additional localization systems. Moreover, the automatic auto-scale of the data from additional localization supports drifting black-box localization systems like the utilized T265.

For future work, we plan extensive evaluation and comparison of the proposed method with other SLAM methods, including vision-based ones.

References

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, ser. Intelligent Robotics and Autonomous Agents series. MIT Press, 2005.

[2] L. Huang, "Review on lidar-based slam techniques," in *International Conference on Signal Processing and Machine Learning (CONF-SPML)*, 2021, pp. 163–168.

[3] I. Abaspur Kazerouni, L. Fitzgerald, G. Dooly, and D. Toal, "A survey of state-of-the-art on visual slam," *Expert Systems with Applications*, vol. 205, no. 117734, 2022.

[4] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2004.

[5] R. Roriz, J. Cabral, and T. Gomes, "Automotive lidar technology: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6282–6297, 2021.

[6] 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, 2013.

[7] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5135–5142.

[8] E. Ward and J. Folkesson, "Vehicle localization with low cost radar sensors," in *IEEE Intelligent Vehicles Symposium (IV)*, 2016, pp. 864–870.

[9] S. Khattak, C. Papachristos, and K. Alexis, "Keyframe-based direct thermal–inertial odometry," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 3563–3569.

[10] "Intel RealSense Tracking Camera T265," https://www.intelrealsense.com/tracking-camera-t265/, accessed Apr 9, 2023.

[11] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[12] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3354–3361.

[13] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and Systems*, vol. 2, no. 9. Berkeley, CA, 2014, pp. 1–9.

[14] M. Labbé and F. Michaud, "Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019.

[15] A. Reinke, X. Chen, and C. Stachniss, "Simple but effective redundant odometry for autonomous vehicles," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 9631–9637.

[16] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proceedings: Image Understanding Workshop*, 1977, pp. 21–27.

[17] C. Debeunne and D. Vivet, "A review of visual-lidar fusion based simultaneous localization and mapping," *Sensors*, vol. 20, no. 7, p. 2068, 2020.

[18] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based slam," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 31–43, 2010.

[19] K. Konolige, G. Grisetti, R. Kümmerle, W. Burgard, B. Limketkai, and R. Vincent, "Efficient sparse pose adjustment for 2d mapping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 22–29.

[20] Y. J. Xu X, Zhang L, "A review of multi-sensor fusion slam systems based on 3d lidar," *Remote Sensing*, vol. 14, no. 12, p. 2835, 2022.

[21] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 3281–3288.

[22] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[23] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 3607–3613.

[24] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *1999 Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms*, 2000, pp. 298–372.

[25] F. Dellaert, "Factor graphs and gtsam: A hands-on introduction," Georgia Institute of Technology, Tech. Rep., 2012.

[26] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, 1987.

[27] T. Shan, B. Englot, C. Ratti, and D. Rus, "LVI-SAM: tightly-coupled lidar-visual-inertial odometry via smoothing and mapping," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 5692–5698.

[28] P. Filzmoser, R. G. Garrett, and C. Reimann, "Multivariate outlier detection in exploration geochemistry," *Computers & geosciences*, vol. 31, no. 5, pp. 579–587, 2005.

[29] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 573–580.