PVL-Cartographer: Panoramic Vision-aided LiDAR Cartographer-based SLAM for Maverick Mobile Mapping System

Yujia Zhang , Jungwon Kang , Gunho Sohn *

Posted Date: 20 March 2023

doi: 10.20944/preprints202303.0340.v1

Keywords: SLAM, localization, mapping, mobile mapping system, spherical camera, panoramic image, LiDAR, IMU, sensor fusion, pose graph.

Article

PVL-Cartographer: Panoramic Vision-Aided LiDAR Cartographer-Based SLAM for Maverick Mobile Mapping System

Yujia Zhang †, Jungwon Kang and Gunho Sohn *

The Department of Earth and Space Science and Engineering, Lassonde School of Engineering, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

* Correspondence: gsohn@yorku.ca

Abstract: Mobile Mapping System (MMS) plays a crucial role in generating high-precision 3D maps for various applications. However, the traditional MMS that uses tilted LiDAR (light detection and ranging) has limitations in capturing complete information of the environment. To overcome these limitations, we propose a panoramic vision-aided Cartographer simultaneous localization and mapping (SLAM) system for MMS, named "PVL-Cartographer". The proposed system integrates multiple sensors to achieve accurate and robust localization and mapping. It contains two sub-systems, early fusion and middle fusion. In the early fusion, range-maps are created from LiDAR points in a panoramic image space, facilitating the incorporation of visual features. The SLAM system works with both visual features with and without augmented ranges. In the middle fusion, a pose graph combines camera and LiDAR nodes, with IMU (Inertial Measurement Unit) data providing constraints between each node. Extensive experiments in challenging outdoor scenarios demonstrate the effectiveness of the proposed SLAM system in producing accurate results, even in conditions with limited features. Overall, our proposed PVL Cartographer system offers a robust and accurate solution for MMS localization and mapping.

Keywords: SLAM; localization; mapping; mobile mapping system; spherical camera; panoramic image; LiDAR, IMU; sensor fusion; pose graph

1. Introduction

Visual SLAM has received considerable attention in recent years due to its potential for generating dense and accurate 3D maps. The focus has been on developing robust localization solutions using various sensors, including cameras, LiDAR, and IMU, with broad applications in mobile robotics, navigation, and semantic mapping. A typical dense 3D mapping system comprises three major components: (1) sensor pose estimation based on the spatial alignment of consecutive frames; (2) live scene reconstruction based on estimated camera pose and integrated points; and (3) loop closure detection and pose graph optimization [1].

The field of robotics has seen significant advancements in the area of SLAM methods, which are crucial for robots to navigate and perform tasks autonomously. Among these methods, visual SLAM is particularly active, utilizing camera images for motion estimation and mapping. LiDAR sensors, which were initially used for obstacle detection, have also been employed for localization through scan registration algorithms that align point clouds. Feature-based localization in SLAM systems can provide high accuracy when the environment is well-distributed with 2D and 3D features extracted from visual or LiDAR sensors. However, it can be prone to failures in scenes with few visual features and depth variations. On the other hand, inertial sensors such as IMUs do not have such requirements in localization and can estimate motion continuously with high frequency and low latency. However, a consumer-grade IMU can suffer from significant drift over time. This issue can be addressed by fusing different types of sensors, including visual or LiDAR odometry estimates, to achieve an accurate and robust SLAM system. Therefore, the integration of different sensor modalities is highly desirable for the development of next-generation SLAM systems.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

MMS have become an indispensable tool for various applications, such as urban planning, infrastructure inspection, and autonomous driving. The key to the success of an MMS lies in its ability to accurately locate itself and construct a map of its environment in real-time. SLAM is a widely used technique that enables MMS to achieve this goal. By fusing data from different sensors, such as LiDAR and cameras, SLAM can accurately estimate the motion of the MMS and create a map of the surrounding environment. In this context, we introduce a Maverick MMS that is equipped with a tilted multi-beam LiDAR and a panoramic camera, which enables the system to capture a 360^o view of the environment. While the tilted LiDAR provides superior point density, coverage, and accuracy for mapping [2], it poses challenges for SLAM due to its limited horizontal coverage. Nonetheless, by fusing the data from both sensors, our system can overcome this limitation and achieve accurate and robust SLAM performance.

In our Maverick MMS, the LiDAR is tilted to the ground for improved mapping performance, while the panoramic camera offers a wide field of view (FoV) for robust SLAM. The limited FoV of cameras with conventional configurations often leads to feature tracking failures in SLAM, whereas panoramic vision enables long-term feature tracking. Previous works [3–6] have demonstrated the effectiveness of panoramic vision for visual odometry and visual SLAM in various scenes.

While panoramic vision is a promising sensor for SLAM, it has been limited to producing up-to-scale results, which are inadequate for many real-world applications. Previous attempts to produce metric results using GPS [5] or ground control points [6] have proven to be challenging due to their availability and reliability. In this work, we propose an early fusion method that combines the panoramic camera and LiDAR sensor for SLAM, enabling metric scale results without the need for external data [5,6]. Our approach utilizes LiDAR points to produce results with absolute scale, which is achieved by assigning a range value obtained from LiDAR points into visual features. We implement our method on the versatile visual SLAM framework, OpenVSLAM [7], using panoramic vision. Inspired by previous work [8,9] on early-fusion of LiDAR points and visual features, our approach shows promising results in accurately SLAM with the Maverick MMS.

After early-fusion of LiDAR points and visual features, a coupled Visual-LiDAR-IMU SLAM system is developed by our middle fusion via a so-called pose graph formulation [10]. In pose graphs the nodes represent poses and edges between them express spatial information, usually constraints obtained from odometry and loop closures in SLAM systems. The major contributions of our work are as follows:

Our work presents a significant contribution to the field of SLAM by introducing a novel Visual-LiDAR-IMU SLAM system for fusing different sensors. The system consists of two sub-systems: early fusion and middle fusion. In the early fusion, range-maps are created from LiDAR points in a panoramic image space, allowing for straightforward augmentation of ranges to visual features. This enables the SLAM system to work with both visual features with and without augmented ranges. In the middle fusion, a pose graph [10] is used to combine camera and LiDAR nodes, while IMU data provides constraints between each node, including camera-camera, LiDAR-LiDAR, and camera-LiDAR constraints. Our work makes four key contributions:

- The proposed SLAM system combines multiple sensors, including panoramic cameras, LiDAR sensors, and IMUs, to achieve high-accuracy and robust performance.
- The early fusion of LiDAR range-maps and visual features enables our SLAM system to produce results with absolute scale, without relying on external data sources such as GPS or ground control points.
- Our middle fusion using a pose graph formulation allows for the seamless integration of data from different sensors, enabling our SLAM system to provide accurate and consistent localization and mapping results.
- We conducted extensive experiments in challenging outdoor scenarios to demonstrate the effectiveness and robustness of our proposed system, even in conditions where only a few

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

features exist. Overall, our work contributes to the development of more accurate and robust SLAM systems for various real-world applications.

In this paper, we provide a comprehensive review of existing SLAM systems in Section 2, highlighting their limitations and shortcomings. In Section 3, we introduce our novel panoramic vision-aided Cartographer SLAM system. Furthermore, we evaluate the performance of our proposed system using a custom dataset in challenging outdoor scenarios and present our experimental results in Section 4. Finally, in Section 5, we summarize our contributions and highlight the potential implications of our work for future research in this field.

2. Related Work

As described in Section 1, our PVL-Cartographer SLAM system is an extended Cartographer that integrates panoramic camera, tilted LiDAR, and IMU sensors. In this section, we will review some of the state-of-the-art odometry and SLAM methods that are related to our work.

2.1. Visual SLAM

Feature-based visual SLAM approaches have been developed to detect and track corner-like visual features, such as SIFT (Invariant Feature Transform) [11], SURF (Speeded Up Robust Features) [12] and ORB (Oriented FAST and Rotated BRIEF) [13]. ORB-SLAM2 [14] and ORB-SLAM3 [15] have extended these techniques for use with monocular, stereo and RGB-D cameras, as well as the visual-inertial module, and include map reuse, loop closing, and relocalization capabilities. However, feature-based methods struggle to find correspondences in environments with simple or repeated patterns or featureless scenes, leading to motion estimation or tracking failures. LSD (Large-Scale Direct monocular SLAM) [16] and DSO (Direct Sparse Odometry) [17] are state-of-the-art direct visual odometry approaches that have addressed this issue with accurate pose estimation and 3D reconstruction.

However, both monocular feature-based SLAM and direct visual SLAM suffer from scale ambiguity, which can be resolved using depth information. RGB-D SLAM and ToF (Time-of-Fight) SLAM have been developed to provide depth information in addition to images. Previous methods [18,19] have used RGB image and depth information to estimate incremental motion, treating it as a 3D feature matching problem [20] proposes a method that extracts visual features and estimates initial incremental motion with RANSAC-based alignment, then uses the initial motion to initialize the ICP (Iterative Closet Point) estimation. KinectFusion [21] is a seminal RGB-D SLAM system that has been used for real-time tracking and mapping, but it may fail in cases of rapid motion or featureless environments. Visual-inertial fusion [1,22] has been successfully used to overcome such tracking failures.

2.2. Panoramic Visual SLAM

Most visual SLAM systems rely on the common pinhole camera model, which has a limited field of view and can easily fail in tracking due to the lack of features. This can be particularly problematic in cases of rapid motion, changing lighting conditions, or texture-less environments. One promising solution to this problem is to extend the field of view by using fisheye or panoramic cameras.

The OpenVSLAM framework [7] is a versatile visual SLAM solution that can handle a variety of camera models, including pinhole, fisheye, and panoramic cameras. It consists of three main modules for tracking, mapping, and global optimization, which draw inspiration from ORB-SLAM. By leveraging the equirectangular camera model for panoramic vision, OpenVSLAM is able to perform SLAM using panoramic cameras.

RPV-SLAM [23] is a range-augmented panoramic visual SLAM solution that builds upon the OpenVSLAM framework by generating ranges for visual features using LiDAR points. This allows the system to improve the accuracy and robustness of feature tracking in challenging environments. With

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

these advancements, the OpenVSLAM and RPV-SLAM frameworks represent important contributions to the field of visual SLAM and have the potential to enable new applications in robotics, augmented reality, and more.

2.3. LiDAR SLAM

In recent years, LiDAR-based SLAM systems have gained increasing attention due to their ability to provide high-resolution 3D data of the environment. LOAM (Lidar Odometry and Mapping in real-time) [24] is one such LiDAR-based odometry system that has shown promising results without the need for precise range data. This system was proposed in 2014 and is still considered one of the best-performing methods according to the KITTI odometry benchmark dataset [25]. V-LOAM [26], the vision-aided version of LOAM, is its main competitor. To achieve real-time performance, LOAM breaks down the odometry problem into high and low-frequency algorithms that work together. The high-frequency algorithm estimates the velocity, while the low-frequency algorithm performs point cloud registration and mapping for a finer result. This approach allows the system to be fast and computationally low-cost enough to perform in real time, while also guaranteeing low drift and precise mapping. In LOAM, point-to-plane ICP registration is used for point cloud registration, and features are extracted based on their roughness and categorized as point and edge features. LOAM-livox [27] is an extended version of LOAM designed for LiDARs with small FoV and irregular samplings. Another recent development is SA-LOAM [28], a novel semantic-aided LiDAR SLAM based on LOAM that leverages semantics in odometry and loop closure detection.

[29] presented a system that uses transformations computed from ICP to feed a pose graph structure, that in turn is used on loop closings to build an optimization problem that provide updates of keyframes selected along the trajectory. These updates correct the map of environment being built and reduce the accumulated errors from the ICP odometry. However, even if this system improves the estimation of the trajectory through loop closings, it remains prone to local minima in which ICP can converge.

SuMa [30] builds upon previous work in laser-based SLAM, but improves upon it by introducing semantic maps. SuMa++ [31] is an extension of SuMa, which enables laser-based semantic segmentation of the point cloud. This semantic information is used to improve pose estimation accuracy in challenging and ambiguous situations. Specifically, SuMa++ exploits semantic consistencies between scans and the map to filter out dynamic objects and provide higher-level constraints during the ICP process. This allows the system to combine semantic and geometric information based solely on three-dimensional laser range scans to achieve considerably better pose estimation accuracy. Furthermore, unlike other SLAM methods, SuMa++ does not require any data from visual images.

2.4. Sensor-fusion-based SLAM

As discussed earlier, V-LOAM [26] is an innovative extension to the LOAM that integrates vision-based components to enhance its performance. This system has been proposed by the same research group and has shown to outperform LOAM in certain scenarios according to the KITTI benchmark. V-LOAM is particularly effective in detecting sudden and sharp motions, which can be challenging for traditional Lidar-based odometry systems. The high-frequency module of V-LOAM leverages visual features to estimate the velocity of the vehicle while the Lidar ensures precision in smaller movements. Additionally, the point set registration and motion estimation refinement are performed in parallel at a lower rate to achieve accurate and efficient results.

[32] introduced LeGO-LOAM, which is a lightweight and ground-optimized LiDAR odometry and mapping method based on LOAM. Building upon their previous work, the authors extended LeGO-LOAM to include IMU sensors and visual cameras, resulting in tightly-coupled LiDAR-inertial odometry [33] and LiDAR-visual-inertial odometry [34] via smoothing and mapping. This approach leverages the complementary information provided by different sensors to improve the accuracy and robustness of the system, while also reducing drift and enhancing the mapping capability.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

Google's Cartographer [35] has been widely recognized as a real-time solution for indoor mapping. This 2D SLAM system combines scan-to-submap matching with loop closure detection and graph optimization to generate globally consistent maps. To create individual submap trajectories, a local grid-based SLAM approach is employed. In the background, all scans are matched to nearby submaps using pixel-accurate scan matching to create loop closure constraints. The constraint graph of submap and scan poses is periodically optimized to produce a globally consistent map. The final map is generated as a GPU-accelerated combination of finished submaps and the current submap, providing an up-to-date preview for the operator. To extend Cartographer for 3D SLAM, an IMU is required to measure gravity and define the z-direction. Roll and pitch derived from the IMU are used in the scan matcher to reduce the search window in three dimensions. Several works [36,37] have extended Google Cartographer to improve its processing speed and accuracy. Our PVL-Cartographer is an extension of Google's Cartographer SLAM system that adds panoramic visual odometry capabilities.

3. Methodology

In this section, we introduce the proposed system for Cartographer-based Panoramic-Visual-LiDAR-IMU SLAM, which leverages a sensor fusion approach to enhance the accuracy and robustness of the SLAM system.

3.1. Maverick Mobile Mapping System and Notation

MMS have become increasingly popular in recent years due to their ability to provide geospatial data while the platform is in motion. MMSs typically consist of high-resolution cameras and LiDAR as primary sensors for data acquisition, along with other sensor suites such as the global navigation satellite system (GNSS) and IMU for positioning and geo-referencing. While MMSs are primarily used for mapping purposes, they are not typically used for odometry, and post-processing is often required to obtain accurate geo-referencing information.

Recent advancements in MMS technology, such as machine learning, artificial intelligence, object extraction, and autonomous vehicles, have driven the development of increasingly sophisticated MMS systems. However, such systems are often limited to outdoor environments due to the inability to collect accurate GNSS data indoors. To address this issue, the proposed PVL-Cartographer SLAM system for MMS allows for both localization and mapping in GPS-denied environments [38] provides a comprehensive overview of recent MMS technologies, including the types of sensors and platforms utilized in MMS, as well as their capabilities and limitations.

Our study presents a Cartographer-based Panoramic-Visual-LiDAR-IMU SLAM system, utilizing the Maverick MMS as depicted in Figure 1. The MMS is outfitted with a Ladybug5 panoramic camera, a Velodyne HDL-32E LiDAR, and high-precision GPS/IMU. Notably, in our experiments, the GPS data is utilized solely as a ground truth to evaluate the performance of the proposed PVL-SLAM system. The camera's optical axis is aligned parallel to the ground when mounted on a car, while the LiDAR's spinning axis is tilted at an angle of 45^o relative to the ground. During data collection, images are captured at an average rate of 7.5 frames per second, while the LiDAR scans are acquired at a spinning rate of 15 revolutions per minute (RPM).

In this study, the GPS/IMU coordinate system is designated as the world coordinate system w, while l and c represent the lidar and camera coordinate systems, respectively. To ensure accurate sensor fusion, all sensors in the Maverick MMS are calibrated using external parameters, with the X-Y-Z axis pointing forward, right, and downward, as illustrated in Figure 2. The sensor's pose is represented by a 4x4 matrix p=[R,t], where R is a 3x3 rotation matrix and t is a translation vector. The variable k denotes a specific time point, where p_k represents the transformation of the local coordinate system at time k relative to its origin. The motion m_k denotes the relative pose between time k-1 and k.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

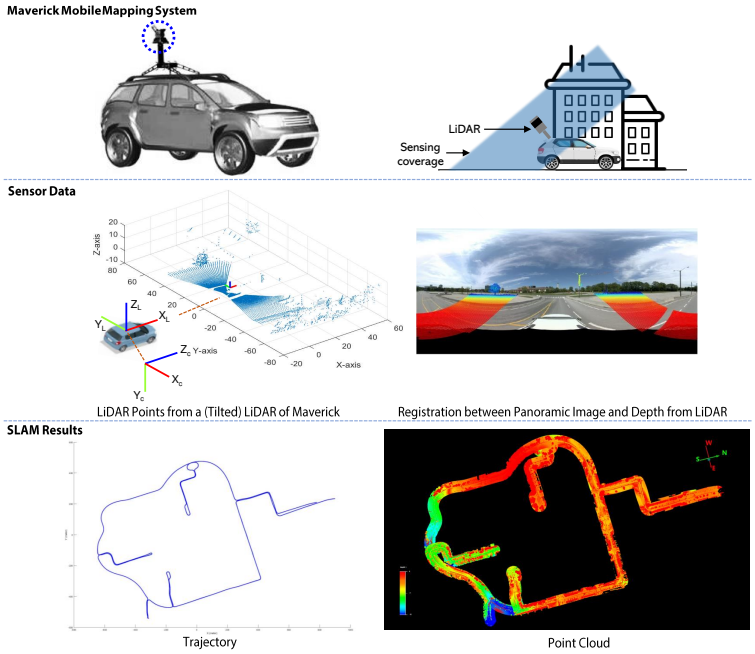


Figure 1. The Maverick MMS with a tilted LiDAR and a panoramic camera.

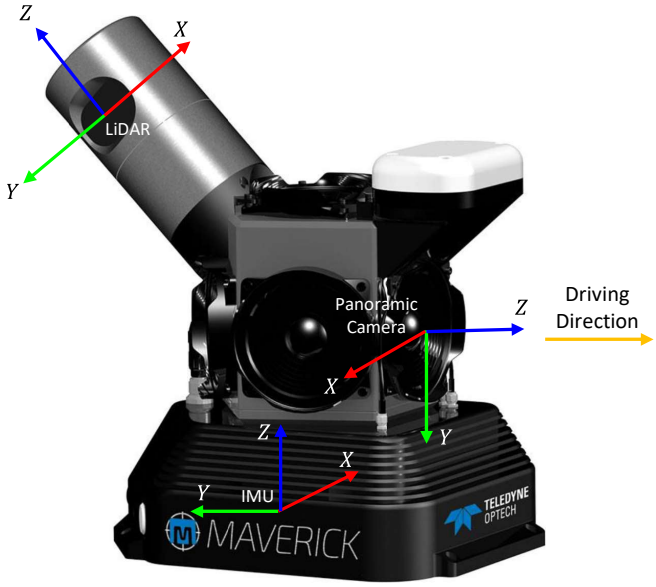


Figure 2. The coordinate system for Maverick MMS with a tilted LiDAR, a panoramic camera and IMU.

3.2. Google Cartographer

This study presents an extension of the Cartographer SLAM technique, a real-time mapping and loop closure system for backpack or vehicle mapping platforms with a 5 cm resolution. The Cartographer approach utilizes a grid map-based representation with flexible resolution and sensor choices, and is divided into four modules, as shown in Figure 3. The data input module requires LiDAR points and IMU data, with optional odometry pose data and fixed frame pose data. The basic processing module includes a voxel filter for processing LiDAR scans. The LiDAR odometry and mapping module utilizes a Ceres scan matcher for feature matching, and a submap to estimate the pose and orientation of the vehicle. Finally, the global adjustment module optimizes the pose by utilizing a larger scan matcher on a global map generated from assembling all submaps. The Cartographer

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

approach is advantageous for its LiDAR-centric SLAM system and the ability to integrate sensor fusion such as IMU and input odometry data.

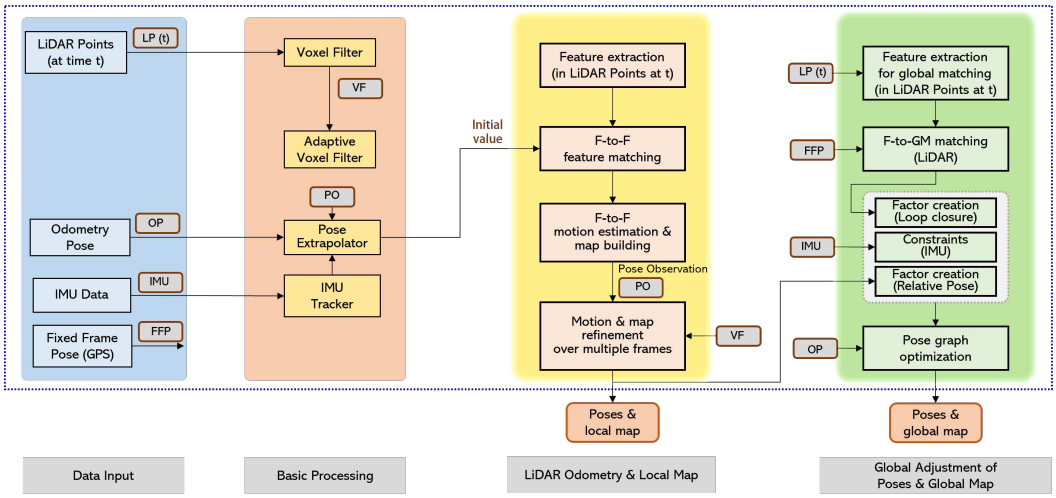


Figure 3. The workflow of the Google Cartographer SLAM.

3.2.1. Local Map Construction

The map generated by Cartographer consists of local submap and a global map that is a combination of accumulated submaps. Then the global map will be corrected or repaired whenever loop closure is identified. The construction of submap is an iterative process that aligns scans from the LiDAR sensor to a submap coordinate frame. The pose transformation of the scan frame ε in the submap is represented as T_ε as defined in Equation (1):

$$T_\varepsilon p = \underbrace{\begin{pmatrix} cos\varepsilon_\theta & -sin\varepsilon_\theta \\ sin\varepsilon_\theta & cos\varepsilon_\theta \end{pmatrix}}_{rotation} p + \underbrace{\begin{pmatrix} \varepsilon_x \\ \varepsilon_y \end{pmatrix}}_{translation} \tag{1}$$

Submap is created with the form of probability grids. The corresponding pixel to consist of all points that are closest to the grid point are defined. A set of grid points is assigned a probability hit or miss if it is in one of these sets. Then, for every hit, the closest grid point is inserted to the hit set. Meanwhile, every miss is inserted to the grid associated with the pixel, excluding grid points which are already in the hit set.

3.2.2. Ceres Scan Matching

Scan matcher is always used in SLAM to process sensor data and estimate pose. This estimated pose with rotation and translation is computed by filtering the residual difference of consecutive scans, as well as a comparison between the current scan states and submap. The Ceres scan matcher used in Cartographer is responsible for finding optimal probability values at the scan points in the submap. This is described as a nonlinear least squares problem as Equation (2).

$$argmin_\varepsilon \sum_{k=1}^{k} (1 - M_{smooth}(T_\varepsilon h_k))^2 \tag{2}$$

where T_ε transforms scan points h_k from the scan frame to the submap frame according to the scan pose. The function M_{smooth} is a bicubic interpolation smooth filter of probability values in the submap.

This mathematical optimization usually gives better precision than the resolution of the grid. The initial estimates from the early fusion is used in the optimization to make the matching process more

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

robust and obtain accurate scan pose estimates. In the naïve Cartographer system, an IMU is used to estimate the rotational component θ of the pose between scan matches. In our PVL Cartographer system, this rotational component is obtained from both IMU and the initial estimates of early fusion.

3.3. RPV-SLAM with Early Fusion

In this section, we will present the basis for the range-augmented panoramic visual SLAM (RPV-SLAM) system [23]. The RPV-SLAM system contains four modules, feature and range module, tracking module, mapping module and loop closing module. In the feature and range module, visual features are extracted from a panoramic image and ranges for the visual features are derived from the LiDAR points. Then the visual features with and without augmented ranges are imputed into the tracking, mapping and loop closing modules. Finally the metrically-scaled results are produced from the pipeline. It is important to note that only results without loop closing are used in the next PVL Cartographer middle fusion, since our PVL Cartographer SLAM has its own loop closing module.

3.3.1. Feature and Range Module

Firstly, given a panoramic image I and the ORB features are extracted from the image. Initialize a range-map R with the same size as the input image I. Then, project the given LiDAR points in the LiDAR frame onto R in the camera frame, where the extrinsic parameters between the panoramic camera and the LiDAR are known through the calibration. After projection, the projected points $P_l = p_i$ are obtained, where p_i is a point i with the range in r_i at image coordinate (u_i, v_i) in R. Secondly, compute ranges for the range-map R over the panoramic image region through range interpolation. Due to the tilted configuration of the LiDAR, the projected points P_l exists only in a limited area having a rainbow shape, as shown in Figure 1. In order to get the dense range-map R having dense ranges over the panoramic image region, a Delaunary-triangulation-based interpolation with P_l is utilized. It is found that the interpolated range at a location (u, v) likely becomes inaccurate if it is far from P_l. Hence, a binary mask M with the same size of range-map R is created. The interpolated ranges inside the mask are kept and ones outside are discarded. Finally, the augmentation of ranges to ORB visual features can be simply done by finding the range in the final generated range-map R_f with respect to the same location of a visual feature. As the result of the augmentation, two kinds of visual features are extracted: (i) visual features augmented with ranges, and (ii) visual features without augmented ranges.

3.3.2. Tracking Module

With the visual features with and without augmented ranges, the tracking module for localization estimates the camera pose for each frame by finding feature matches to the local map. Simultaneously, a motion-only bundle adjustment is performed to minimize the reprojection error. The scaled map points of visual features with augmented ranges can be directly created using the ranges. On the other hand, scaled map points of visual features without augmented ranges are created using triangulation between two frames under an estimated scaled motion. After generating appropriate scaled map points as above, the metrically-scaled SLAM results will be produced.

3.4. PVL Cartographer SLAM with Pose-Graph-based Middle Fusion

The proposed PVL Cartographer SLAM consists of two parts, frontend and backend. The workflow of the proposed system is shown in Figure 4. The frontend includes feature extraction, matching, pose estimation and data association for local map. The backend includes the global map construction, map optimization and loop closure. Comparing with the original google Cartographer (Figure 3), we added early fusion that combines LiDAR points and panoramic image into the module of visual feature tracking. Then the motion estimation is used as initial value in the google Cartographer LiDAR odometry module. Finally, both camera node and LiDAR node are inserted into the global map and will be optimized together.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

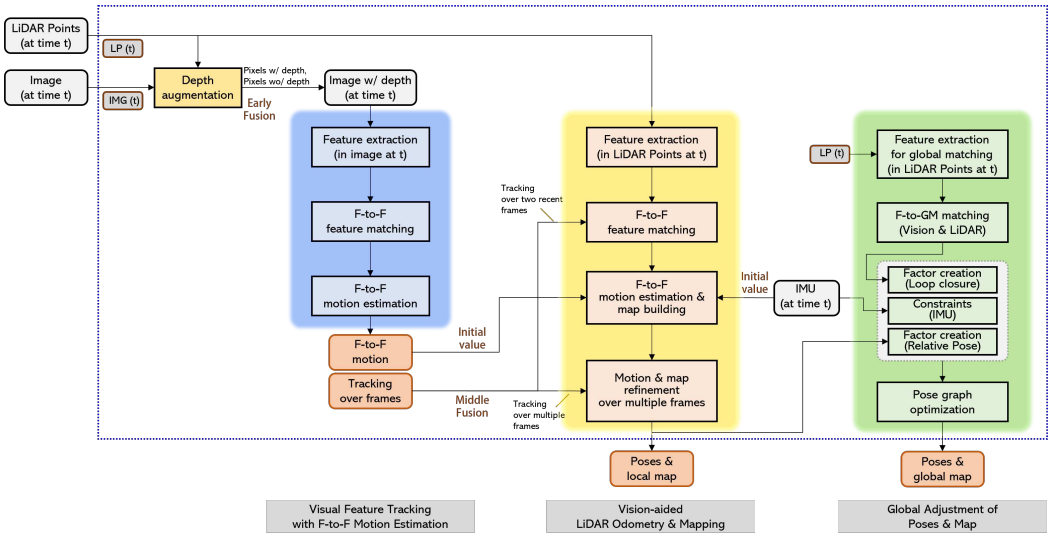


Figure 4. The workflow of the PVL Cartographer SLAM with middle fusion.

First, the early fusion that combines the visual feature and range module generates the pose. Then, the generated pose is inputted into the modified Cartographer system as initial value. Note that any visual odometry algorithm can be applied to the proposed PVL cartographer and the RPV-SLAM-based visual odometry is one of them. In this project, we used RPV as early fusion and it can be changed to other visual odometry. The Ceres scan matcher accepts this initial value and then processes the LiDAR points to obtain a more robust and accurate pose estimates. The estimated camera frame poses from early fusion and estimate LiDAR scan poses from Ceres scan matcher are added as nodes in the global map. The global map structure of the proposed PVL Cartographer SLAM is shown in Figure 5. The global map includes two types of nodes, camera nodes and LiDAR nodes. All nodes are listed based on the timestamp of the frame or the scan. The nodes in the world coordinate system are optimized give some constraints. Since the IMU has the highest frequency, it provides the measurements of angular velocity and acceleration, which can be used as constraints that linking different nodes. In the process, the constraints between camera frames, LiDAR scans and consecutive camera-LiDAR nodes are calculated from the IMU measurements. Then the global map will be optimized as described in the following section.

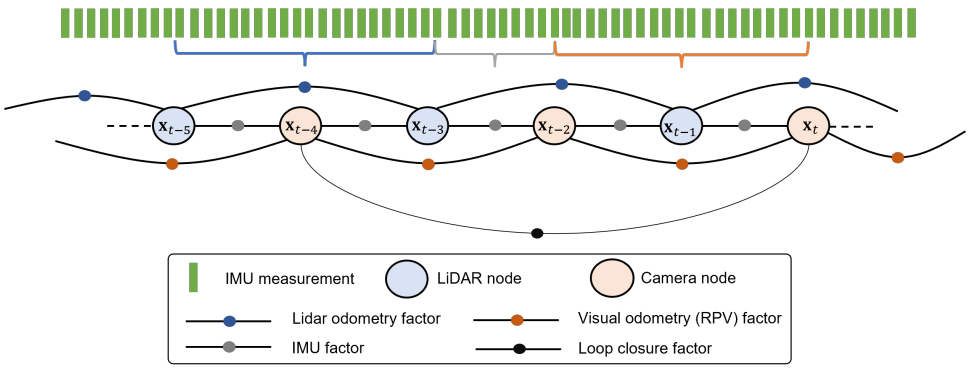


Figure 5. The global map structure of the proposed PVL Cartographer SLAM. The system receives input from a spherical camera, 3D LiDAR and IMU.

3.5. Global Map Optimization and Loop Closure for PVL Cartographer

In the process of estimating pose in local SLAM, errors are created caused by the presence of grid map resolution, sensor noises and different map features fusion. Although the error between frames or scans is small, it accumulates and creates bigger drift error after long distance of travel. The original

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

google Cartographer tries to reduce this error using the sparse pose adjustment (SPA) optimization method. In this optimization method, a similar scan matcher in local SLAM is used for pose correction but in a more extensive global map range. In the mean time, loop closure detection is included in the optimization process. The optimization problem is formulated as nonlinear squares problem. The Ceres is utilized to compute a solution (3) every few seconds.

$$argmin_{\Xi^m \Xi^n} \frac{1}{2} \sum_{ij} \rho(E^2(\varepsilon_i^m, \varepsilon_j^n, \sum_{ij}, \varepsilon_{ij})) \tag{3}$$

where the submap poses $\Xi^m = \{\varepsilon_i^m\}_{i=1,\dots,m}$ and the camera or scan poses $\Xi^n = \{\varepsilon_j^n\}_{j=1,\dots,n}$ in the world are optimized with given constraints. These constraints take the form of relative poses ε_{ij} and associated covariance matrices $\sum_{ij}$.

The transformation between two nodes p_i and p_j can be computed by Equation (4).

$$T(p_i, p_j) = \begin{pmatrix} R_{\varepsilon_i^m}^{-1}(t_{\varepsilon_i^m} - t_{\varepsilon_j^n}) \\ \varepsilon_{i;\theta}^m - \varepsilon_{j;\theta}^n \end{pmatrix} \tag{4}$$

Then the residual E for such a constraint is computed by Equation (5) and (6).

$$E^2(\varepsilon_i^m, \varepsilon_j^n; \sum_{ij}, \varepsilon_{ij}) = e(\varepsilon_i^m, \varepsilon_j^n; \sum_{ij}, \varepsilon_{ij})^2 \sum_{ij} e(\varepsilon_i^m, \varepsilon_j^n; \sum_{ij}, \varepsilon_{ij})^2 \tag{5}$$

$$e(\varepsilon_i^m, \varepsilon_j^n; \sum_{ij}, \varepsilon_{ij}) = \varepsilon_{ij} - \begin{pmatrix} R_{\varepsilon_i^m}^{-1}(t_{\varepsilon_i^m} - t_{\varepsilon_j^n}) \\ \varepsilon_{i;\theta}^m - \varepsilon_{j;\theta}^n \end{pmatrix} \tag{6}$$

The residuals of camera-camera, lidar-lidar, camera-lidar are computed separately and then combines as Equation (7).

$$argmin_{\Xi^m \Xi^n} \frac{1}{2} \sum_{ij} \rho(E_{c-c}^2 + E_{l-l}^2 + E_{c-l}^2) \tag{7}$$

The estimated poses from the early fusion can also be viewed as landmark in the global map. However, the camera frame and lidar scan are captured in different time. This asynchronization problem can be solved by interpolation.

Then, the full weighted landmark cost function can be written as Equation (8).

$$f(p_0^l, p_i^c, p_j^c) = f(p_0^l, p_0^c) = \begin{pmatrix} w_t & w_r \end{pmatrix} (T_{cl}^m - T(p_0^l, p_0^c)) \tag{8}$$

The translation and rotation weights w_t, w_r are part of the landmark observation data and T_{cl}^m is the transformation between camera and lidar, which is fixed value for the Maverick MMS.

4. Experiments

In this section, we present evaluation results of the proposed SLAM system using a MMS in real outdoor environments.

4.1. Dataset

The Maverick sensor was mounted on a vehicle, and our experiments were conducted in diverse outdoor environments, including parking lots, roads, campuses, and residential areas. We selected four sequences of data to evaluate the performance of our PVL-Cartographer SLAM system. Table 1 summarizes the selected data, including the total number of camera and LiDAR frames, image size, distance travelled, running time, scene descriptions, and evaluation methods. The ground-truth trajectories, which are post-processed GPS data with cm-level accuracy, were obtained by bundle adjustment (BA) of images using ground control points (GCPs) in the LMS (LiDAR Mapping Suite)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

software from Teledyne Optech. With high-precision GPS/IMU data and LiDAR points, an offline bundle adjustment was applied to synchronize data from multiple sensors. The ground-truth trajectories overlaid on satellite images are presented in Figure 6, providing a visual representation of the data and the accuracy of our results.

Table 1. Details of our four dataset captured by Maverick MMS to evaluate different methods.

	Sequence A	**Sequence B**	**Sequence C**	**Sequence D**
Sensors	Maverick MMS: Ladybug-5 + Velodyne HDL-32 + IMU			
Region	Parking lot	Campus area	Residential area	Residential area
Camera frames	717	8382	10778	4500
Image size	4096 x 2048	8000 x 4000	8000 x 4000	8000 x 4000
LiDAR frames	1432	17395	22992	9615
Distance travelled	324 meters	7035 meters	7965 meters	3634 meters
Running time	94 seconds	19 minutes	22 minutes	10 minutes
Ground truth	GNSS/IMU	GNSS/IMU	GNSS/IMU	GNSS/IMU
Loop	One small loop	One large loop + a few small loops	Many medium-size loops	A few loops
Dynamic objects	Parking, barrier and person	Car, bus and person	Car, bus and person	Car, bus and person
Compared methods	ORB-SLAM2 (camera-only) VINS-Mono-SLAM (camera + IMU) LOAM (LiDAR) Google-Cartographer-SLAM (LiDAR + IMU) RPV-SLAM (Panoramic camera + LiDAR) Our PVL-SLAM (Panoramic camera + LiDAR + IMU)			

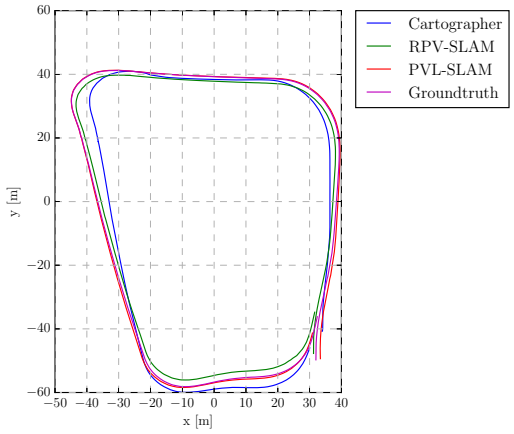
Figure 6. Ground-truth trajectories (marked by green dots) overlaid on satellite images for the sequence A, B, C, D.

4.2. Results

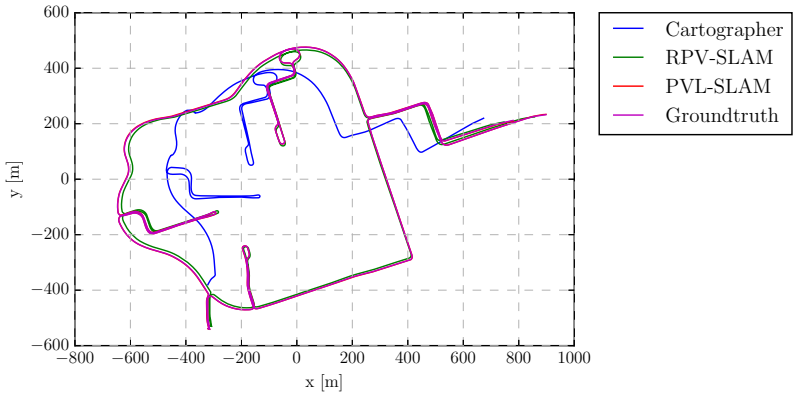
In our evaluation, we compared the performance of the proposed PVL-Cartographer SLAM system with several state-of-the-art SLAM systems as shown in Table 1, including ORB-SLAM2 [14], VINS-Mono [39], LOAM [24], and Google Cartographer [35]. Notably, we tested ORB-SLAM2 with monocular images, which demonstrates the advantage of using panoramic images in visual SLAM. VINS-Mono, a visual-inertial SLAM system, is tested using the monocular images and IMU data from our Maverick MMS. LOAM, a well-known LiDAR-based odometry system, is also included in our evaluation, as well as Google Cartographer, a LiDAR-inertial SLAM system that uses IMU data to define the z-direction.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

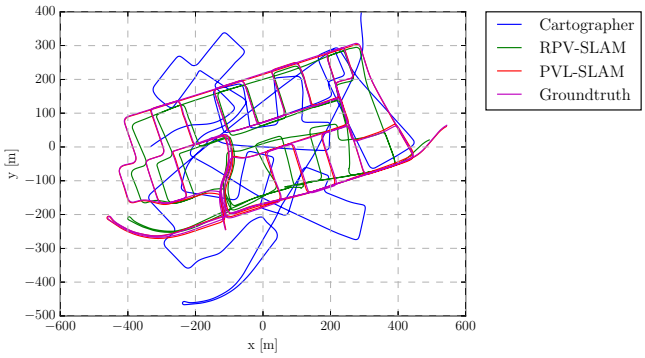
The results demonstrate that both the RPV-SLAM and the proposed PVL-SLAM systems can produce accurate and robust large-scale results. The trajectories and global maps generated by the sensor-fusion-based SLAM systems are presented in Figure 7. Notably, the initial values of the PVL middle fusion module are obtained from the RPV early fusion module, which helps to improve the accuracy of the PVL-SLAM system. By inserting the camera nodes and LiDAR nodes into the pose graph and optimizing them together, the proposed PVL-SLAM system achieves superior performance compared to the state-of-the-art methods.



(a) Sequence A



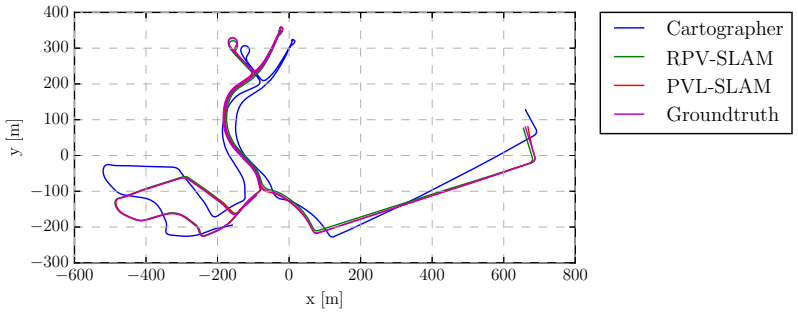
(b) Sequence B



(c) Sequence C

Figure 7. *Cont.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1



(d) Sequence D

Figure 7. Trajectory comparison in each sequence. Note that only a partial trajectory of the Cartographer is shown in (b), as the operation of Cartographer was suspended in the middle of the sequence.

To evaluate the performance of the proposed PVL-Cartographer SLAM system and the compared SLAM methods, we used the rpg trajectory evaluation toolbox [40] and evaluated each method using the absolute trajectory error (ATE), relative translation error (RTE) and relative rotation error (RRE). The ATE is the measured root mean square error (RMSE) using the aligned estimation and the ground truth, and is presented in Table 2. The RTE is the average transnational RMSE in percentage over trajectory segments with length of 10%, 20%, 30%, 40%, 50% of the total length, while the RRE is the average rotation RMSE ($^o/m$) over the same trajectory segments. The results of RTE and RRE are shown in Table 3 and Figure 8, respectively.

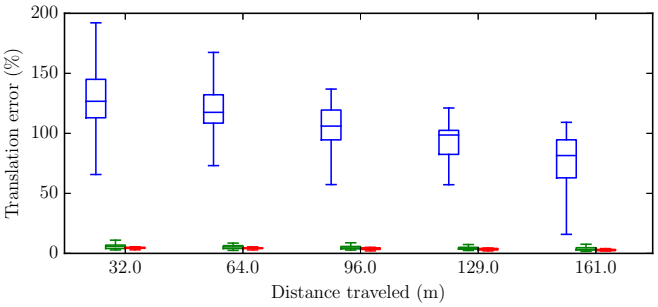
Table 2. Comparison of tranlation accuracy w.r.t. ATE [m].

	ORB SLAM2	VINS-Mono	LOAM	Cartographer	RPV-SLAM	PVL-SLAM
Sequence A	5.894	3.9974	Fail	4.023	1.618	**0.766**
Sequence B	100.870	86.897	Fail	152.230	12.910	**2.599**
Sequence C	155.908	160.765	Fail	183.619	30.661	**3.739**
Sequence D	10.665	12.875	Fail	58.576	5.673	**2.204**
Overall	68.3343	66.1336	Fail	99.612	12.7155	**2.327**

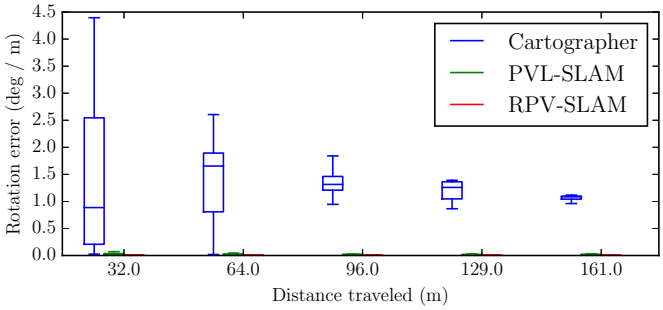
Table 3. Comparison of translation and rotation accuracy w.r.t. relative error, [%] [deg/m] .

	ORB SLAM2	VINS-Mono	LOAM	Cartographer	RPV-SLAM	PVL-SLAM
Sequence A	7.769 0.0677	4.685 0.0410	Fail	6.789 0.0507	3.934 **0.0040**	**3.027** 0.0236
Sequence B	13.770 0.0099	10.779 0.0109	Fail	15.047 0.0090	3.096 **0.0009**	**1.273** 0.0019
Sequence C	4.879 0.0289	3.987 0.0301	Fail	5.764 0.0133	3.752 0.0057	**0.853** **0.0018**
Sequence D	2.878 0.0148	3.085 0.0178	Fail	4.650 0.0137	**1.347** **0.0017**	2.555 0.0035
Overall	7.324 0.030	5.634 0.025	Fail	9.843 0.059	2.393 **0.002**	**1.069** 0.003

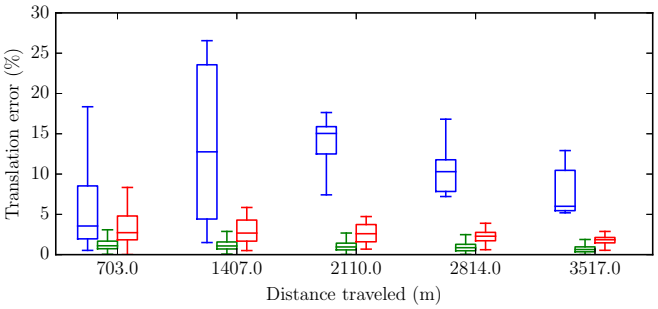
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1



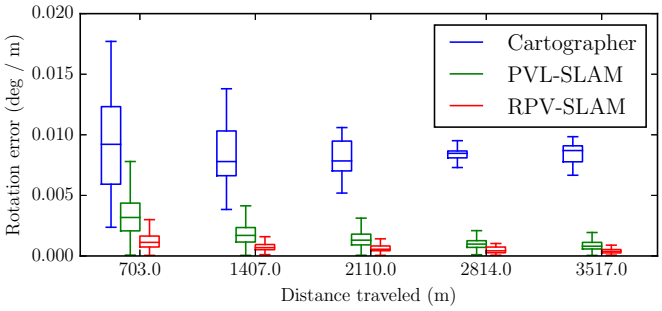
(a) Relative translation error for sequence A



(b) Relative rotation error for Sequence A



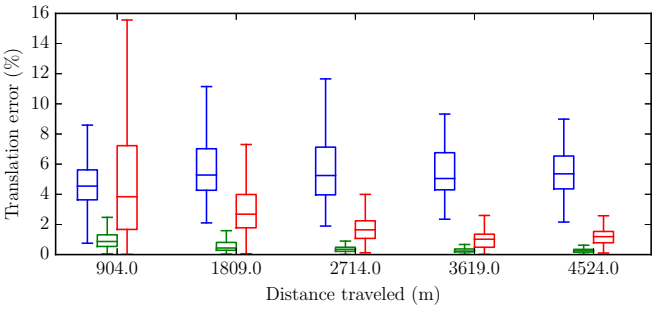
(c) Relative translation error for sequence B



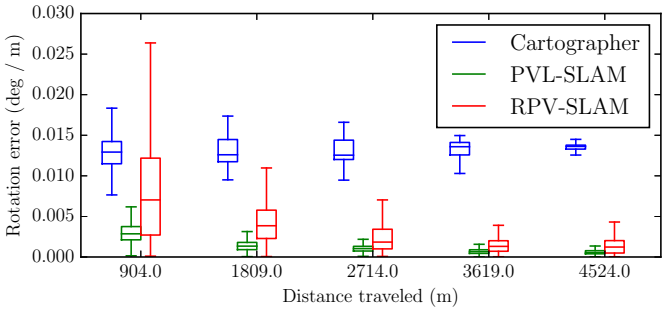
(d) Relative rotation error for Sequence B

Figure 8. *Cont.*

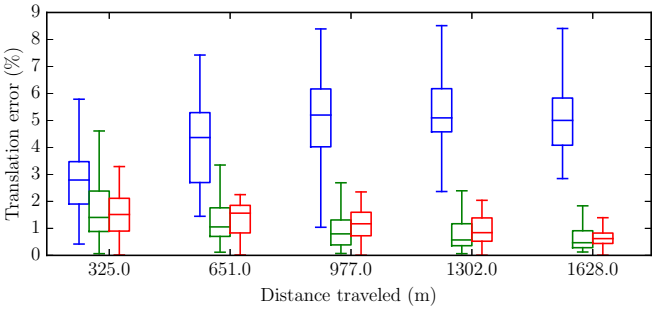
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1



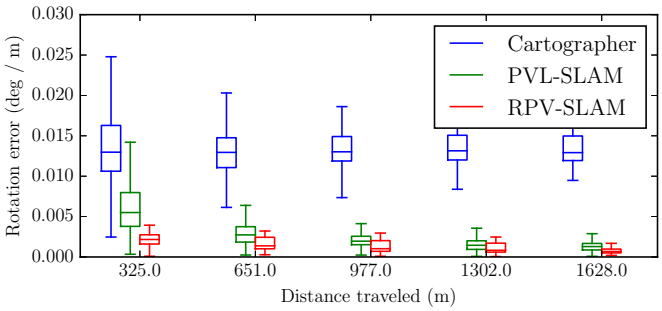
(e) Relative translation error for sequence C



(f) Relative rotation error for Sequence C



(g) Relative translation error for sequence D



(h) Relative rotation error for Sequence D

Figure 8. Relative translation and rotation errors for different sub-trajectory lengths shown as a series of boxplots.

4.3. Discussion

Our proposed PVL-Cartographer SLAM system demonstrated superior performance compared to google Cartographer in all four tested sequences, achieving improvements in ATE by 80.96%, 98.29%, 97.96%, 96.24%, respectively. Moreover, our system outperformed google Cartographer in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0340.v1

terms of RTE by 55.41%, 91.54%, 85.20%, 45.05%, and in terms of RRE by 53.45%, 78.89%, 86.47%, 74.45% for sequences A to D. It should be noted that for RRE, RPV-SLAM performed better than our PVL-Cartographer. Overall, our results suggest that camera-centric SLAM systems, such as RPV-SLAM, perform better in orientation estimation, while LiDAR-centric systems perform better in translation estimation. Among the collected data with different scenarios, Sequence C posed the most challenges due to its long traveled distance and complex loop structures. Nevertheless, our PVL-Cartographer system delivered the best performance in terms of ATE, RTE, and RRE. This can be attributed to the highly effective loop closure module, which was able to significantly reduce errors through map optimization when multiple loops were detected. Despite the challenges posed by this scenario, our PVL-Cartographer system demonstrated robustness and superior performance.

The results showed that both the RPV and PVL-Cartographer SLAM systems outperformed the state-of-the-art methods in terms of accuracy and robustness, which demonstrated the superiority of sensor-fusion-based SLAM. Figure 8 further demonstrated that the translation and rotation errors were efficiently reduced as the distance traveled increased, indicating that the loop-closure model worked effectively over long distances. In addition, Figure 9 showed that the proposed SLAM system could generate accurate trajectories by successfully closing the loops, even in challenging scenarios with complex loops and long traveled distances. However, ORB-SLAM2, VINS-Mono, and Cartographer only produced part of the trajectory due to the insufficient features for matching, while LOAM generated incorrect trajectories, especially the orientation angles, due to the incorrect orientation estimation if points was only scanned by the tilted LiDAR and no IMU is used to point out the z-direction.

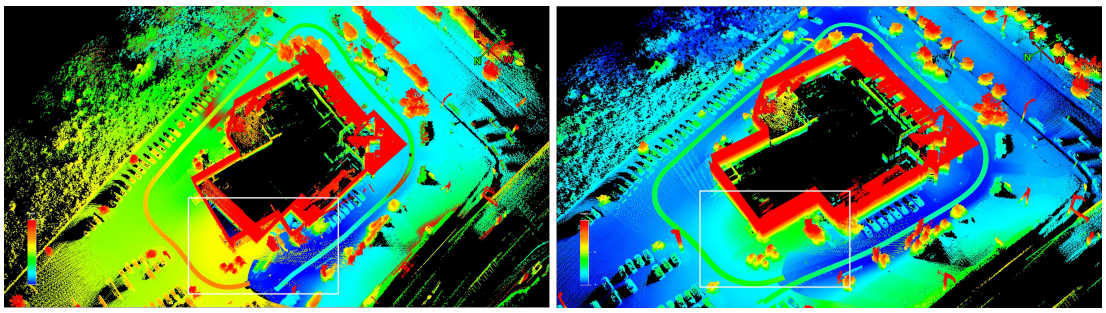


Figure 9. For Sequence A, the left image shows the misalignment without loop closure and the right image shows the loop closing result.

Overall, our results showed that the proposed sensor-fusion-based SLAM systems were more robust than the camera-only and LiDAR-only approaches, especially for challenging data collected by a MMS with a panoramic camera and tilted LiDAR. The results also indicated that the PVL-SLAM system, which integrated panoramic visual odometry, tilted LiDAR, and IMU sensors, was effective in producing accurate and robust large-scale results.

5. Conclusion

We have developed a sensor-fusion-based SLAM system, PVL-Cartographer-SLAM, for a Maverick MMS equipped with a panoramic camera and a tilted LiDAR. Our approach combines multiple sensors, including a panoramic camera, LiDAR, and IMU, tightly through pose graph to enable robust and accurate SLAM. The system includes an early fusion range augmented panoramic visual odometry system, RPV, which produces metrically-scaled trajectories using visual features augmented with ranges derived from LiDAR points. Our experiments demonstrate that our proposed system outperforms existing state-of-the-art SLAM systems, including camera-centric, camera-inertial, LiDAR-centric, and LiDAR-inertial SLAM systems, even when only a limited number of visual features are augmented with ranges due to limited overlap between an image and points from the tilted LiDAR.

Our findings suggest that our proposed sensor-fusion-based SLAM system is a promising approach for challenging outdoor localization and mapping scenarios.

The PVL-Cartographer SLAM system presented in this study can be extended through a number of avenues for future research:

- Firstly, by adopting advanced depth estimation or completion methods, denser range-maps can be created which would enable more visual features to be augmented with the ranges;
- Secondly, incorporating range measurements in both the local and global bundle adjustment would enhance the accuracy of the system;
- Thirdly, efforts are underway to improve the current PVL-Cartographer SLAM to a more tightly coupled visual-LiDAR-IMU SLAM system through pose graph or factor graph;
- Fourthly, the system can be further extended by developing a SLAM pipeline that combines the visual features and LiDAR features;
- Finally, applying deep neural network techniques for feature classification and pose correction would likely improve the system's overall performance.

References

1. Huai, J.; Zhang, Y.; Yilmaz, A. REAL-TIME LARGE SCALE 3D RECONSTRUCTION BY FUSING KINECT AND IMU DATA. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* **2015**, *2*.
2. Alsadik, B. Ideal angular orientation of selected 64-channel multi beam lidars for mobile mapping systems. *Remote sensing* **2020**, *12*, 510.
3. Lin, M.; Cao, Q.; Zhang, H. PVO: Panoramic visual odometry. 2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM). IEEE, 2018, pp. 491–496.
4. Tardif, J.P.; Pavlidis, Y.; Daniilidis, K. Monocular visual odometry in urban environments using an omnidirectional camera. 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2008, pp. 2531–2538.
5. Shi, Y.; Ji, S.; Shi, Z.; Duan, Y.; Shibasaki, R. GPS-supported visual SLAM with a rigorous sensor model for a panoramic camera in outdoor environments. *Sensors* **2012**, *13*, 119–136.
6. Ji, S.; Qin, Z.; Shan, J.; Lu, M. Panoramic SLAM from a multiple fisheye camera rig. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *159*, 169–183.
7. Sumikura, S.; Shibuya, M.; Sakurada, K. OpenVSLAM: A versatile visual SLAM framework. Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 2292–2295.
8. Zhang, J.; Kaess, M.; Singh, S. Real-time depth enhanced monocular odometry. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2014, pp. 4973–4980.
9. Graeter, J.; Wilczynski, A.; Lauer, M. Limo: Lidar-monocular visual odometry. 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2018, pp. 7872–7879.
10. Grisetti, G.; Kümmerle, R.; Stachniss, C.; Burgard, W. A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine* **2010**, *2*, 31–43.
11. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* **2004**, *60*, 91–110.
12. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Computer vision and image understanding* **2008**, *110*, 346–359.
13. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics* **2015**, *31*, 1147–1163.
14. Mur-Artal, R.; Tardós, J.D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE transactions on robotics* **2017**, *33*, 1255–1262.
15. Campos, C.; Elvira, R.; Rodríguez, J.J.G.; Montiel, J.M.; Tardós, J.D. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics* **2021**, *37*, 1874–1890.
16. Engel, J.; Schöps, T.; Cremers, D. LSD-SLAM: Large-scale direct monocular SLAM. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13. Springer, 2014, pp. 834–849.
17. Engel, J.; Koltun, V.; Cremers, D. Direct sparse odometry. *IEEE transactions on pattern analysis and machine intelligence* **2017**, *40*, 611–625.

18. Kerl, C.; Sturm, J.; Cremers, D. Robust odometry estimation for RGB-D cameras. 2013 IEEE international conference on robotics and automation. IEEE, 2013, pp. 3748–3754.
19. Endres, F.; Hess, J.; Sturm, J.; Cremers, D.; Burgard, W. 3-D mapping with an RGB-D camera. *IEEE transactions on robotics* **2013**, *30*, 177–187.
20. Henry, P.; Krainin, M.; Herbst, E.; Ren, X.; Fox, D. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *The international journal of Robotics Research* **2012**, *31*, 647–663.
21. Newcombe, R.A.; Izadi, S.; Hilliges, O.; Molyneaux, D.; Kim, D.; Davison, A.J.; Kohi, P.; Shotton, J.; Hodges, S.; Fitzgibbon, A. Kinectfusion: Real-time dense surface mapping and tracking. 2011 10th IEEE international symposium on mixed and augmented reality. Ieee, 2011, pp. 127–136.
22. Nießner, M.; Dai, A.; Fisher, M. Combining Inertial Navigation and ICP for Real-time 3D Surface Reconstruction. Eurographics (Short Papers), 2014, pp. 13–16.
23. Kang, J.; Zhang, Y.; Liu, Z.; Sit, A.; Sohn, G. RPV-SLAM: Range-augmented panoramic visual SLAM for mobile mapping system with panoramic camera and tilted LiDAR. 2021 20th International Conference on Advanced Robotics (ICAR). IEEE, 2021, pp. 1066–1072.
24. Zhang, J.; Singh, S. LOAM: Lidar odometry and mapping in real-time. Robotics: Science and Systems. Berkeley, CA, 2014, Vol. 2, pp. 1–9.
25. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research* **2013**, *32*, 1231–1237.
26. Zhang, J.; Singh, S. Visual-lidar odometry and mapping: Low-drift, robust, and fast. 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015, pp. 2174–2181.
27. Lin, J.; Zhang, F. Loam livox: A fast, robust, high-precision LiDAR odometry and mapping package for LiDARs of small FoV. 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020, pp. 3126–3131.
28. Li, L.; Kong, X.; Zhao, X.; Li, W.; Wen, F.; Zhang, H.; Liu, Y. SA-LOAM: Semantic-aided LiDAR SLAM with loop closure. 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021, pp. 7627–7634.
29. Mendes, E.; Koch, P.; Lacroix, S. ICP-based pose-graph SLAM. 2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). IEEE, 2016, pp. 195–200.
30. Behley, J.; Stachniss, C. Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments. Robotics: Science and Systems, 2018, Vol. 2018, p. 59.
31. Chen, X.; Milioto, A.; Palazzolo, E.; Giguere, P.; Behley, J.; Stachniss, C. Suma++: Efficient lidar-based semantic slam. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019, pp. 4530–4537.
32. Shan, T.; Englot, B. Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018, pp. 4758–4765.
33. Shan, T.; Englot, B.; Meyers, D.; Wang, W.; Ratti, C.; Rus, D. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. 2020 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2020, pp. 5135–5142.
34. Shan, T.; Englot, B.; Ratti, C.; Daniela, R. LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping. IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021, pp. 5692–5698.
35. Hess, W.; Kohler, D.; Rapp, H.; Andor, D. Real-time loop closure in 2D LIDAR SLAM. 2016 IEEE international conference on robotics and automation (ICRA). IEEE, 2016, pp. 1271–1278.
36. Dwijotomo, A.; Abdul Rahman, M.A.; Mohammed Ariff, M.H.; Zamzuri, H.; Wan Azree, W.M.H. Cartographer slam method for optimization with an adaptive multi-distance scan scheduler. *Applied Sciences* **2020**, *10*, 347.
37. Nüchter, A.; Bleier, M.; Schauer, J.; Janotta, P. Continuous-time slam—improving google's cartographer 3d mapping. *Latest Developments in Reality-Based 3D Surveying and Modelling* **2018**, pp. 53–73.
38. Elhashash, M.; Albanwan, H.; Qin, R. A Review of Mobile Mapping Systems: From Sensors to Applications. *Sensors* **2022**, *22*, 4262.

39. Qin, T.; Li, P.; Shen, S. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics* **2018**, *34*, 1004–1020.
40. Zhang, Z.; Scaramuzza, D. A Tutorial on Quantitative Trajectory Evaluation for Visual(-Inertial) Odometry. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS), 2018.