

GCLO: Ground Constrained LiDAR Odometry with Low-drifts for GPS-denied Indoor Environments

Xin Wei¹, Jixin Lv¹, Jie Sun, Erbao Dong, Shiliang Pu*

Abstract—LiDAR is widely adopted in Simultaneous Localization And Mapping (SLAM) and High Definition (HD) map production. The accuracy of LiDAR Odometry (LO) is of great importance, especially in GPS-denied environments. However, we found typical LO results are prone to drift upwards along the vertical direction in underground parking lots, leading to poor mapping results. This paper proposes a Ground Constrained LO method named GCLO, which exploits planar grounds in these specific environments to compress the vertical pose drifts. GCLO is divided into three parts. First, a sensor-centric sliding map is maintained, and the point-to-plane ICP method is implemented to perform the scan-to-map registration. Then, at each key-frame, the sliding map is recorded as a local map. Ground points nearby are segmented and modeled as a planar landmark in the form of Closest Point (CP) parameterization. Finally, planar ground landmarks observed at different key-frames are associated. The ground landmark observation constraints are fused into the pose graph optimization framework to improve the LO performance. Experimental results in HIK and KITTI datasets demonstrate GCLO's superior performances in terms of accuracy in indoor multi-floor parking lots and flat outdoor sites. The limitation of GCLO in adaptability for other environments is also discussed.

I. INTRODUCTION

Automated Valet Parking (AVP) is a promising application of autonomous driving technologies, in which vehicles navigate themselves in the parking lots and automatically park into the target spot. Generally, AVP can be divided into two classes, one in private areas [1] and one in public areas. For AVP tasks in public places such as megamalls and hospitals, a pre-defined High Definition (HD) map plays a critical role since it can provide prior information of target parking lots with centimeter-level accuracy. It helps autonomous vehicles locate, plan, perceive, and navigate automatically and safely.

Simultaneous Localization And Mapping (SLAM) is usually the first-step of mass HD map production. Thus the accuracy of SLAM results is of great importance. Visual SLAM (VSLAM) [2] and Light Detection And Ranging (LiDAR) SLAM (LSLAM) [3] are two main research fields. LiDAR provides accurate measurement of the surroundings in the form of point-cloud. The measurement is robust to variations of lighting conditions [4], and the intensity information can also facilitate the HD map labeling procedure. These advantages make LiDAR popular in the field of SLAM and HD map production. Typical LSLAM methods are developed based on the pose graph optimization framework. In the front-end, successive LiDAR scans are aligned [5]-[6],



Figure 1. Top view of distances between ground points of mapping results and fitted ground plane. (a) LO result drifts upward when the vehicle is moving on the ground plane in underground parking lots. As a result, the assembled ground is distorted as an upward bowl. (b) The assembled ground becomes flat after ground observation constraints are fused.

and the sensor trajectory is estimated. This process is commonly referred to as LiDAR Odometry (LO) [3]. In the back-end, loop closure detection is conducted to recognize the re-visited place. A pose graph is then constructed and the residual error is minimized using nonlinear least square algorithms [7]-[8].

LO is a crucial component of LSLAM algorithms, and its accuracy has a significant influence on LSLAM results, especially in environments where the global pose observation is absent, such as GPS-denied underground parking lots. Thus, extensive research has been made to improve the accuracy and robustness of LO results [9]-[10]. However, typical LO results are prone to drift upward along the vertical direction of the road in underground parking lots [11]. As a result, the assembled ground points are distorted as an upward bowl, as shown in Fig. 1(a). A reasonable explanation of this phenomenon is provided in [11]. The depth measured by LiDAR might be biased when the incidence angle is high. As a consequence, when the vehicle is moving on a ground plane, points observed from the road are slightly bent, and the trajectory estimated by LO will drift along vertical directions. To remove the LiDAR measurement bias and compress the predictable pose drift, an incidence angle based mathematical model is provided in [11]. However, this strategy requires adjusting scale factors for different sensors, and the bias might vary in different temperatures [11].

To solve this problem and further improve the accuracy of LO in specific tasks and environments, we leverage that most multi-floor parking lots are well structured with several planar grounds. Thus the grounds are modeled as infinite planar landmarks, and the ground observation constraints are fused

X. Wei, J. Lv, J. Sun, S. Pu are with Hikvision Research Institute (HRI), Hangzhou Hikvision Digital Technology Co., Ltd., Hangzhou, China. Email: {weixin7, lvjixin, sunjie, pushiliang.hri}@hikvision.com.

E. Dong is with University of Science and Technology of China (USTC), Hefei, China. Email: ebdong@ustc.edu.cn.

¹Both authors contributed equally to this manuscript.

*Corresponding author.

into the LO process to improve the LO performance. Key contributions of this paper are as follows:

- A LO method, which exploits ground observation constraints to compress the pose drift in multi-floor indoor and flat outdoor environments, is proposed.
- A ground correspondence method, which associates ground planar landmarks by detecting sharp changes between consecutive key-frames, is proposed.
- Experimental results in KITTI and HIK datasets, which illustrate the superior performance in accuracy and the limitation in generalization, are presented and discussed.

The rest of this paper is organized as follows: Sec. II presents a review of related work. Notations and preliminaries are explained in Sec. III. Sec. IV presents a detailed explanation of the proposed GCLO method. And the qualitative and quantitative tests and analyses are demonstrated in Sec. V. Sec. VI presents the conclusion along with some future work.

II. RELATED WORK

LiDAR and visual SLAM are popular topics in the robotics community and have been extensively investigated during the past two decades. This paper focuses on 3D LiDAR based methods, together with some algorithms exploiting ground constraints. For a more detailed review of recent SLAM work, please refer to article [12] and the references therein.

A. LSLAM

LO is a fundamental component of LiDAR SLAM. Though the measurements of LiDAR are mostly accurate and robust, some vital shortcomings of LiDAR make the task of building an accurate and robust LO challenging.

First, most mobile LiDAR has a low vertical resolution, and the sparse point cloud it obtains makes robust feature detection and tracking difficult. Therefore, a local map strategy that aggregates past observations is usually utilized in LO [3]. Unfortunately, the dense local map makes point cloud registration time-consuming. From this point, LOAM [3] is a milestone in terms of efficiency and has inspired an amount of other work [13]-[14]. LOAM extracts and aligns sparse corner and plane features in the odometry process and periodically aligns dense features in the mapping process to balance efficiency against accuracy. SuMa [15]-[16] is also impressive, which utilizes a projection method to avoid explicitly finding the closest point. Suma also estimates normal vectors rapidly using the cross product of neighbor vertices. IMLS-SLAM [17] actively samples those points that could provide constraints for pose estimation. With this sampling strategy, only about 10% of the raw LiDAR points would be sampled.

Second, the LiDAR measurement is not a snapshot of environments and thus suffers from motion distortion. This hazardous effect has been gradually realized for the past few years. Typical motion compensation or de-skewing approaches integrate IMU measurements to de-skew raw

LiDAR points [18]-[20]. A constant velocity model is usually utilized [3] when the IMU measurements are not available. In [19], the de-skewing process is fused into graph optimization by separating raw scans into several scan lines, and interpolates the corresponding poses using continuous cubic B-spline in SE3.

Third, the noise of LiDAR measurement has been modeled as zero-mean Gaussian distribution for simplicity during the past two decades [21]. However, as far as our best concern, this assumption can't explain why typical LO results are prone to drift along vertical directions in underground parking lots. Recently, a reasonable explanation is presented that the LiDAR measurement might be biased when the incidence angle is high [11]. An incidence angle based mathematical model is also provided to remove the bias. However, this strategy requires adjusting scale factors for different sensors, and the bias might vary in different temperatures [11], which makes it hard to apply. By contrast, we propose a method to utilize the ground planes in multi-floor buildings and flat outdoor sites to reduce the effect of LiDAR measurement bias. Though introducing ground planes will reduce the generalization capability, the proposed method in this paper is still valuable and promising to improve LO results drastically in specific applications such as AVP, Automated Guided Vehicles (AGV), and service robots.

B. SLAM with Ground Constraints

Ground points are extracted and aligned to estimate 3 DOF of 6 DOF pose in LeGO-LOAM [13] and [22]. However, it only associates ground points in successive observations. Thus, the ground constraint doesn't work for compressing the long-term pose drift, especially the vertical pose drift of LO results.

F. Zheng [23]-[24] proposes a method for ground vehicles that fuses ground constraints into the optimization framework. The vehicle pose is constrained by virtual SE2 ground constraints in [23] or directly parameterized by SE2 parameterization [24]. Koide [25] assumes that one global ground plane exists and uses this ground plane to constrain the vehicle's pose. However, the methods mentioned above are not adaptable for multi-floor environments, which are common in underground parking lots.

III. NOTATIONS AND PRELIMINARIES

A. Notations

In the following section, we denote the homogeneous transformation matrix as $T_a^b \in SE(3)$, which transforms a point $p_i^a \in R^3$ in frame F_a into frame F_b . $R_a^b \in SO(3)$ and $t_a^b \in R^3$ is the rotation matrix and translation vector of T_a^b , respectively. We use L, B, W to represent the coordinate frame of LiDAR, IMU, and global map, respectively. We use Π^a to

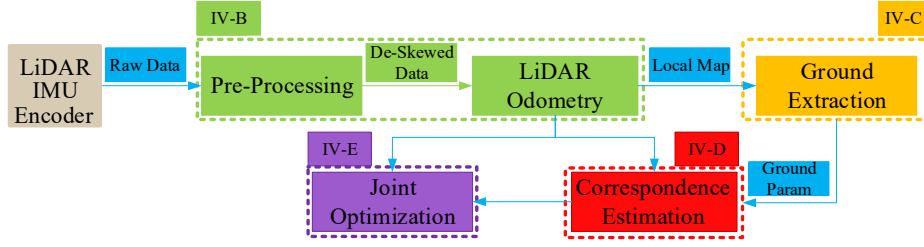


Figure 2. A brief overview of Ground-SLAM framework.

represent the Closest Point (CP) parameter of an infinite plane in frame \mathbf{F}_a .

B. Preliminaries

Researchers have proposed several parameterization methods for infinite planes. Hesse Form (HF) consists of a normal vector \mathbf{n} and distance d between the plane and the origin of a given coordinate frame. HF is over-parameterized. Thus, it suffers from the singular information matrix in the least-square optimization, and the identity of the normal vector becomes hard to hold. Spherical coordinate, which represents a normal vector with an azimuth and an elevation angle, is a minimal parameterization. However, it suffers from ambiguities when the elevation angle is equal to $\pm\pi/2$. Unit quaternion [26] is also exploited to represent the infinite plane. However, the physical connection between the quaternion and plane is unclear.

Inspired by [27], CP, which is defined by the closest point on the plane to the origin of a given coordinate frame, is adopted to represent the infinite plane in this paper. CP is a minimal representation, thus also suffers from the singularity problem when the origin of a given coordinate frame lies on the plane. However, the singularity of CP representation could be avoided in this paper since the ground plane has a certain distance from LiDAR equipped on vehicles.

Transforming the ground's parameter between different frames is essential as the ground is modeled as an infinite planar landmark in this paper. HF makes it easy to represent plane equation. Thus HF is used as an intermediate to transform ground parameters. Supposing there is a plane with HF $[\mathbf{n}^a, d^a]$ in frame \mathbf{F}_a . Point \mathbf{p}_i^a lies on this plane, then this point satisfies the plane equation, as shown in (1).

$$(\mathbf{n}^a)^T \mathbf{p}_i^a = d^a \quad (1)$$

Given the transformation matrix \mathbf{T}_b^a between frame \mathbf{F}_a and frame \mathbf{F}_b , $\mathbf{p}_i^a = \mathbf{R}_b^a \mathbf{p}_i^b + \mathbf{t}_b^a$. Substituting this into (1), and we could get (2).

$$((\mathbf{R}_b^a)^T \mathbf{n}^a)^T \mathbf{p}_i^b = d^a - (\mathbf{n}^a)^T \mathbf{t}_b^a \quad (2)$$

Then \mathbf{n}^b and d^b are obtained by using equations in (3). After getting the transformation result that is represented in HF, we could transform it into CP parameterization using (4).

$$\begin{cases} \mathbf{n}^b = (\mathbf{R}_b^a)^T \mathbf{n}^a \\ d^b = d^a - (\mathbf{n}^a)^T \mathbf{t}_b^a \end{cases} \quad (3)$$

$$\Pi^b = d^b \mathbf{n}^b \quad (4)$$

IV. GROUND CONSTRAINED LiDAR ODOMETRY

A. System Review

A brief review of the proposed method is shown in Fig. 2. In the pre-processing module, IMU and wheel encoder measurements are fused using the Extended Kalman Filter (EKF) method to provide high-frequency motion estimation result. The motion distortion of the LiDAR scan is diminished by using the motion estimation result.

The de-skewed LiDAR scan points are used to estimate the relative transformation between consecutive scans using the point-to-plane ICP algorithm. A sensor-centric sliding map is maintained to overcome the sparseness of LiDAR scans and improve the accuracy of LO.

At each key-frame, the sliding map is recorded as a local map. The ground plane is segmented from the recorded local map and the CP parameter is estimated using weighted least-square methods. Afterwards, ground planes observed at different key-frames are associated. Finally, a pose graph is constructed and the residual error is jointly minimized. Details of the critical modules are introduced below.

B. Scan-to-Map with Sensor-Centric Sliding Map

Scan-to-map registration. Assuming at time k , the global pose of LiDAR is $\mathbf{T}_{L_k}^W$ and the sliding map M_{L_k} has been maintained w.r.t. frame L_k . When a new scan S_{k+1} arrives at time $k+1$, the sensor transformation \mathbf{T}_k^{k+1} as well as the corresponding covariance $\Sigma_{\mathbf{T}_k^{k+1}}$, are firstly estimated by using point-to-plane ICP [28] and Censi's method [29], where \mathbf{T}_k^{k+1} is the abbreviation of $\mathbf{T}_{L_k}^{L_{k+1}}$. The coordinate and uncertainty covariance of points in the sliding map is transformed from frame L_k to frame L_{k+1} using (5) and (6), where \mathbf{p}_i^k is the i^{th} point in the sliding map M_{L_k} and $\Sigma_{\mathbf{p}_i^k}$ is the corresponding covariance. $\Sigma_{\mathbf{R}_k^{k+1}}$, $\Sigma_{\mathbf{t}_k^{k+1}}$ are the covariance matrix of estimated rotation and translation component, respectively. $\mathbf{J}_{\mathbf{R}_k^{k+1}}$, $\mathbf{J}_{\mathbf{p}_i^k}$ are the jacobian matrix w.r.t. rotation component and the point, respectively.

$$\mathbf{p}_i^{k+1} = \mathbf{T}_k^{k+1} \cdot \mathbf{p}_i^k = \mathbf{R}_k^{k+1} \cdot \mathbf{p}_i^k + \mathbf{t}_k^{k+1} \quad (5)$$

$$\Sigma_{\mathbf{p}_i^{k+1}} = \mathbf{J}_{\mathbf{R}_k^{k+1}}^T \Sigma_{\mathbf{R}_k^{k+1}} \mathbf{J}_{\mathbf{R}_k^{k+1}} + \mathbf{J}_{\mathbf{p}_i^k}^T \Sigma_{\mathbf{p}_i^k} \mathbf{J}_{\mathbf{p}_i^k} + \Sigma_{\mathbf{t}_k^{k+1}} \quad (6)$$

Sensor-centric sliding map updating. The observation-based maintenance method filters out the points with considerable uncertainty while the observing features are kept and updated. It contains the following steps:

- 1) First, the registered new scan points are associated with the sliding map points by checking the pre-defined distance metric.
- 2) Then, the associated sliding map points are selected, and their uncertainties are reset to their associated new scan points' covariance matrices, which is commonly referred to as observing error.
- 3) The sliding map points that capture considerable uncertainty will be filtered. In practice, the point is eliminated as long as the corresponding covariance matrix's trace is more significant than the threshold.
- 4) Finally, the registered scan points that are failed to be associated are added to the sliding map as new observations.

C. Ground Extraction and Modelling

Grounds in AVP parking lots are locally flat. Thus grounds are modeled as infinite planes in this paper. At each key-frame, ground points are firstly segmented from the sensor-centric sliding map. Then, an initial CP parameter Π_0 is estimated by applying the RANdom SAMple Consensus (RANSAC) algorithm on the segmented points.

The initial plane parameter estimated by RANSAC might be noisy. Therefore, we formulate a weighted least squares optimization problem as in (7) and (8), and minimize the point-to-plane distances between ground points and the CP parameterized ground plane.

$$\Pi^* = \arg \min_{\Pi} \sum_{i=1}^N \mathbf{r}_i^T \Omega_{r_i} \mathbf{r}_i \quad (7)$$

$$\mathbf{r}_i = \frac{\mathbf{p}_i^T \Pi}{\|\Pi\|} - \|\Pi\| \quad (8)$$

Where \mathbf{p}_i is the i^{th} point of ground points, N is the number of ground points, Ω_{r_i} is the inverse of the uncertainty of the i^{th} measurement \mathbf{r}_i , as in (9), where Σ_{p_i} is the uncertainty of measurement \mathbf{p}_i and Π_l is the linearization point. The cost in (7) is minimized using the Gauss-Newton method. The Jacobian of residual \mathbf{r}_i at Π_l is formulated as in (10). Then the local increment could be formulated in (11). And the plane parameter could be updated in (12). The corresponding uncertainty is approximated as in (13).

$$\Omega_{r_i} = \left(\frac{\Pi_l^T \Sigma_{p_i} \Pi_l}{\|\Pi_l\|_2^2} \right)^{-1} \quad (9)$$

$$\mathbf{J}_i = \frac{\mathbf{p}_i^T}{\|\Pi_l\|} - \frac{\Pi_l^T}{\|\Pi_l\|} - (\mathbf{p}_i^T \Pi_l) \frac{\Pi_l^T}{\|\Pi_l\|^3} \quad (10)$$

$$\Delta \Pi = -(\sum_{i=1}^N \mathbf{J}_i^T \Omega_{r_i} \mathbf{J}_i)^{-1} (\mathbf{r}_i \Omega_{r_i} \mathbf{J}_i) \quad (11)$$

$$\Pi = \Pi_l + \Delta \Pi \quad (12)$$

$$\Sigma_{\Pi} = (\sum_{i=1}^N \mathbf{J}_i^T \Omega_{r_i} \mathbf{J}_i)^{-1} \quad (13)$$

D. Ground Plane Correspondence Estimation

In order to utilize ground constraints, the newly observed ground plane at key-frame \mathbf{F}_{j+1} needs to be associated with historical ground planes. Supposing there are j sequential historical key-frames \mathbf{F}_i , $i = 1, \dots, j$, and each historical key-

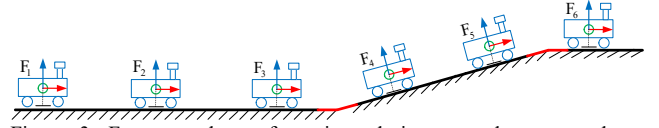


Figure 3. Frames and transformations during ground correspondence estimation, red lines indicate multi-floor junctions in parking lots.

frame possesses an estimated pose \mathbf{T}_i^w provided by LO, and an observed ground plane with the estimated parameter Π^i and its uncertainty Σ_{Π^i} . The remaining problem is to determine which historical ground plane are coplanar with the current ground plane.

A straightforward solution to this problem is associating ground planes by comparing their parameters in one identical coordinate frame under distance metrics (e.g., Euclidean). However, the ground plane might be associated incorrectly if the drift is high. Instead, we leverage that the ground is usually well structured and the ground parameter changes sharply in multi-floor junctions in our target environments, as shown in Fig. 3. We conduct ground correspondence estimation between current key-frame \mathbf{F}_{j+1} and last key-frames \mathbf{F}_j by detecting sharp changes of ground's CP parameters, as in (14) and (15), function f refers to the transformation in (3) and (4), and $\Omega_{\Delta \Pi^i}$ is the inverse of the uncertainty of the relative measurement $\Delta \Pi^i$. In this way, correspondences are free of LO drifts.

$$\Delta \Pi^j = \Pi^j - f(\Pi^{j+1}, \mathbf{T}_j^{j+1}) \quad (14)$$

$$e(\Delta \Pi^j) = (\Delta \Pi^j)^T \Omega_{\Delta \Pi^j} \Delta \Pi^j \quad (15)$$

Uncertainty of variables Π^j , Π^{j+1} and \mathbf{T}_j^{j+1} in (14) will be propagated to $\Delta \Pi^i$. LO remains high-fidelity for a short time. Thus the uncertainty of \mathbf{T}_j^{j+1} is small enough to be neglected. Then we can estimate the uncertainty of relative observation $\Delta \Pi^i$ as in (16) and (17), where \mathbf{R} and \mathbf{t} is the rotation matrix and translation vector of \mathbf{T}_j^{j+1} , respectively. A new planar ground node will be generated when $e(\Delta \Pi^j)$ is large enough.

$$(\Omega_{\Delta \Pi^j})^{-1} = \left(\frac{\partial \Delta \Pi^j}{\partial \Pi^j} \right)^T \Sigma_{\Pi^j} \frac{\partial \Delta \Pi^j}{\partial \Pi^j} + \left(\frac{\partial f}{\partial \Pi^{j+1}} \right)^T \Sigma_{\Pi^{j+1}} \frac{\partial f}{\partial \Pi^{j+1}} \quad (16)$$

$$\frac{\partial f}{\partial \Pi^{j+1}} = \mathbf{R}^T - \frac{\mathbf{R}^T \Pi_l \mathbf{t}^T + \Pi_l^T \mathbf{t} \mathbf{R}^T}{\|\Pi_l\|_2^2} + \frac{2 \Pi_l^T \mathbf{t} \mathbf{R}^T \Pi_l \Pi_l^T}{\|\Pi_l\|_2^4} \quad (17)$$

E. Joint Pose Graph Optimization

After correspondences between observed ground planes have been determined, ground observation constraints are fused into the pose graph optimization framework to compress the pose drift of LO, as shown in Fig. 4(b). Assuming there is a ground plane \mathbf{P} whose CP parameter in

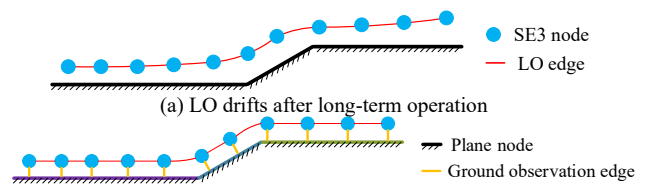


Figure 4. Pose graph of GCLO method.

global coordinate frame W is denoted as Π^w , frame F_j with pose T_j^w observes P with the CP parameter Π^j , which could be transformed to HF $[n^j, d^j]$. With (3) and (4), we transform the ground plane observation in frame F_j into global coordinates. And the difference between the observed and estimated CP parameters is modeled as the residual error, as shown in (19), where R_j^w and t_j^w is the rotation matrix and translation vector of T_j^w , respectively. The Jacobian matrix of residual error r_{ij} with respect to Π^w and T_j^w is shown in (20), (21) and (22). Then the residual errors consist of ground plane observation and LO constraints are jointly minimized using Levenberg-Marquardt (LM) algorithm, as shown in (18), where C means constraints sets, r_{ij} is an abbreviation of $r(x_i, x_j)$, Ω_{ij} is the information matrix of r_{ij} .

$$F(x) = \sum_{(i,j) \in C} r_{ij}^T \Omega_{ij} r_{ij} \quad (18)$$

$$r(\Pi^w, T_j^w) = \Pi^w - d^j R_j^w n^j - (R_j^w n^j)^T t_j^w (R_j^w n^j) \quad (19)$$

$$J_{\Pi^w} = I_{3 \times 3} \quad (20)$$

$$J_{R_j^w} \approx -d^j (R_j^w n^j)^\wedge \quad (21)$$

$$J_{t_j^w} = R_j^w n^j (R_j^w n^j)^T \quad (22)$$

F. Implementation Details

In this paper, we extract one ground plane as a ground observation at each key-frame. However, this strategy doesn't work for all target environments. First, CP parameters might be noisy due to speed bumps. Second, grounds can't always be modeled as planar landmarks since curved and spiral ramps are common in underground parking lots. Third, when the key-frame is located in the intersection area of multi-floor junctions, this key-frame corresponds to two ground planes.

Several methods have been implemented to handle the aforementioned cases. First, the Huber loss function is utilized to decrease the influence of outliers (e. g., speed bumps) for CP parameter estimation. Second, the ground's planarity is evaluated by computing the residual error and uncertainty as in (7) and (13). Grounds with low planarity will not be modeled as planar landmarks, and the corresponding observation constraints will not be fused. Third, only one ground plane is selected under certain metrics (planarity, points count) when ground correspondences are ambiguous. In this way, each key-frame possesses only one ground plane with precisely estimated CP parameter, no matter where the key-frame is located in the parking lots.

V. EXPERIMENTAL EVALUATION

We evaluate GCLO on two datasets. The first one is the HIK dataset which contains four kinds of scenarios. A LiDAR with 40 scan-lines configured at 10 Hz, an IMU and a wheel odometer configured at 100 Hz, together with a RTK-GPS running at 10 Hz are mounted on our vehicle.

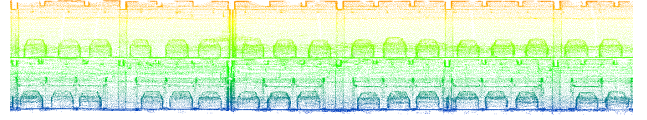


Figure 5. Side view of mapping points of GCLO in underground multi-floor parking lots, color encodes height. Best viewed in color.

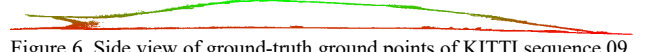


Figure 6. Side view of ground-truth ground points of KITTI sequence 09, color encodes height. Best viewed in color.

The second one is the odometry datasets of the KITTI Vision Benchmark. We include here for comparison GCLO*, and the reported results of LOAM [3], SuMa (Frame-to-Model, w/o loop closure) [15], LiTAMIN2 (ICP+Cov) [30], and PSF-LO [31]. The only difference between GCLO* and GCLO is that the ground observation constraints are cut off in GCLO*.

A. HIK Dataset

Indoor multi-floor parking lots. Experiments in indoor multi-floor parking lots are firstly conducted. The vehicle moves from the B2 floor to the B3 floor, and the assembled points are shown in Fig. 5. Cars and walls are clear, grounds are thin and flat. Besides, the B2 floor and B3 floor are parallel to each other. We also assemble the ground points of the B2 floor. A plane is fitted, and the distances between ground points to this plane are calculated, as shown in Fig. 1. GCLO achieves 0.028 m of the distance on average, compared to 0.236 m of GCLO*. These results qualitatively demonstrate the effectiveness of GCLO.

Outdoor sites. Experiments in the other three sites are conducted to evaluate GCLO quantitatively. The first site is the HIK campus, the second is the outdoor HIK parking lot, and the last is an urban road in Hangzhou. Grounds in the first two sites are almost flat, and the third site is non-ideal since the slope of ground varies gradually. RTK-GPS based inertial navigation system provides ground-truth trajectories in these three sites. Experimental results are listed in Table I. GCLO achieves 1.50% and 1.52 deg/100m of translational and rotational errors, compared to 1.61% and 1.64 deg/100m of GCLO* in the campus dataset. And GCLO achieves 0.88% and 1.14 deg/100m of translational and rotational errors, compared to 0.90% and 1.17 deg/100m of GCLO* in the outdoor parking lots dataset. The comparison of z -translational errors between GCLO and GCLO* further demonstrates the ability to compress vertical drifts. Furthermore, roll and pitch rotational accuracy is improved since a planar ground landmark provides 3 DoF constraints. However, GCLO performs poorly in the third site. The rotational and translational error of GCLO results grow by 7.14% and 21.0% compared to GCLO* results. The main reason is that the slope of grounds varies gradually in this site. As a result, grounds are associated improperly, which lead to higher errors as expected.

B. KITTI Odometry Dataset

As far as our best concern, there are no public datasets

Table. I Results on the HIK dataset

Scenario	x	y	z	trans	roll	pitch	yaw	rot
Campus	0.15/0.15	0.23/0.23	1.25/ 1.14	1.61/ 1.50	1.25/ 1.14	0.81/ 0.74	0.28/0.27	1.64/ 1.52
Parking lots	0.18/0.18	0.36/0.36	0.71/ 0.69	0.90/ 0.88	0.85/ 0.83	0.55/ 0.52	0.32/0.32	1.17/ 1.14
Urban road	0.18/0.18	0.49/0.49	0.85 /0.94	1.12 /1.20	0.41 /0.56	0.29 /0.31	0.20/0.20	0.62 /0.75

Relative errors averaged over trajectories of 100 to 800 m length, relative translational error in %, relative rotational error in degree per 100m, GCLO*/GCLO. Bold numbers indicate top performance.

with 3D LiDAR focusing on multi-floor indoor environments. Since GCLO is also adaptable for flat outdoor sites, it is evaluated in the KITTI odometry dataset. By assessing the ground's flatness based on the ground-truth trajectories, the training set is separated into two subsets, KTI and KTII. Grounds in KTI are almost flat, and KTII is non-ideal for GCLO. Compared with other state-of-the-art methods, the KTI subset is used to demonstrate the superior performance in accuracy of GCLO in our target environments. The KTII subset is used to illustrate the major drawback of adaptability in other environments. Experimental results of KTI and KTII subsets are listed in Table II and Table III, respectively. And the trajectories estimated by our proposed method are shown in Fig. 7.

KTI subset. GCLO generally performs on par with other state-of-the-art methods and often achieves even better results in terms of translational error. GCLO achieves 0.67% of translational error and 0.34 deg/100m of rotational error on average, which ranks 1st and 2nd respectively among listed results. Comparison between GCLO and GCLO* also illustrates the effectiveness of ground constraints proposed in this paper.

KTII subset. Planar ground landmark helps GCLO compress pose drifts significantly in our target environments. However, it restricts the generalization in other non-ideal

environments. Taking sequence 09 as an example, as shown in Fig. 6, though the ground at each key-frame is locally flat, the slope of ground varies gradually. As a result, incorrect association results lead to higher pose drifts instead. Compared to GCLO*, the rotational and translational error of GCLO results grow by 36% and 18%, respectively.

VI. CONCLUSION

In this paper, we exploit ground constraints in multi-floor and flat outdoor environments to compress the pose drifts of LO. Though GCLO is initially designed for AVP applications, it is also valuable for indoor 3D construction, AGV, service robots, et al. Experiments in the HIK and KITTI datasets are conducted to demonstrate the effectiveness of our proposed method. The major drawback is also discussed and illustrated by experimental results.

The modeling and association method of grounds makes GCLO less adaptive for general scenarios. However, we argue that the grounds have not been fully leveraged to facilitate the motion estimation in the SLAM field. In the future, other modeling and association methods of grounds will be further explored to improve the generalization. We also plan to integrate additional semantic information to improve the accuracy and robustness of LO algorithms.

Table. II Results on KTI subset of KITTI odometry training set

Approach	Seq. 00	Seq. 01	Seq. 04	Seq. 05	Seq. 06	Seq. 07	Seq. 08	Average
LOAM	-0.78	-1.43	-0.71	-0.57	-0.65	-0.63	-1.12	-0.84
SuMa	0.3/0.7	0.5/1.7	0.3/0.4	0.2 /0.5	0.2 /0.4	0.3 /0.4	0.4/1.0	0.31 /0.73
LiTAMIN2	0.36/0.78	0.46 /2.10	0.52/1.05	0.31/0.55	0.33/0.55	0.49/0.48	0.35/1.01	0.40/0.93
PSF-LO	-0.64	-1.32	-0.66	-0.45	-0.47	-0.46	-0.94	-0.71
GCLO*	0.47/0.95	0.51/2.21	0.51/1.01	0.39/0.88	0.44/0.90	0.43/0.67	0.40/1.06	0.45/1.11
GCLO	0.23 / 0.49	0.54/1.61	0.20 / 0.40	0.27/ 0.35	0.43/0.42	0.37/ 0.42	0.32 / 0.84	0.34/ 0.67

Table. III Results on KTII subset of KITTI odometry training set

Approach	Seq. 02	Seq. 03	Seq. 09	Seq. 10	Average
LOAM	-0.92	-0.86	-0.77	-0.79	-0.84
SuMa	0.4/1.1	0.5/ 0.7	0.3 / 0.5	0.3 /0.7	0.38 /0.75
LiTAMIN2	0.37 /0.95	0.48 /0.96	0.40/0.69	0.47/0.80	0.43/0.85
PSF-LO	-0.87	-0.75	-0.56	-0.54	-0.68
GCLO*	0.47/1.21	0.53/1.17	0.49/0.99	0.50/1.05	0.50/1.11
GCLO	0.75/1.33	0.56/1.54	0.67/1.01	0.76/1.37	0.68/1.31

Relative errors averaged over trajectories of 100 to 800 m length: relative rotational error in degree per 100m / relative translational error in %. Bold numbers indicate top performance.

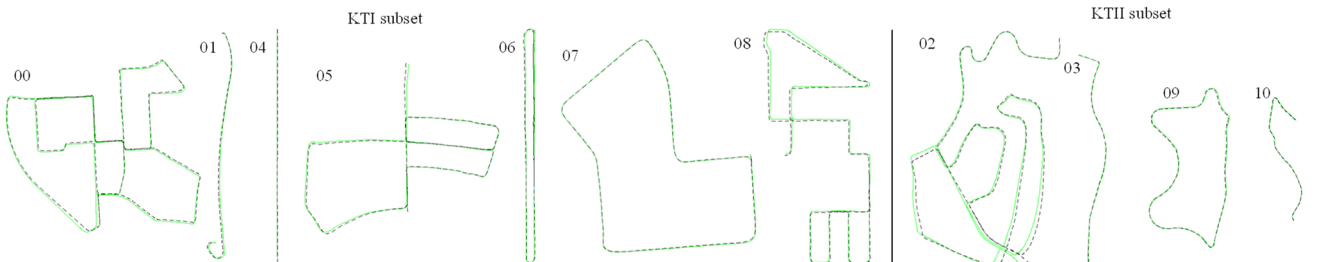


Figure 7. Trajectories in KTI and KTII subset, respectively. Dashed black trajectories correspond to ground-truth trajectories provided by KITTI and green to our approach GCLO. Part of sequences are rotated to save space. Best viewed in color.

REFERENCES

- [1] T. Qin, T. Chen, Y. Chen, and Q. Su, "Avp-slam: Semantic visual mapping and localization for autonomous vehicles in the parking lot," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.
- [2] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004-1020, 2018.
- [3] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real time," in *Robotics: Science and System*, vol. 2, 2014, p. 9.
- [4] A. Carballo, J. Lambert, A. Monrroy-Cano, D. R. Wong, P. Narksri, Y. Kitsukawa, E. Takeuchi, S. Kato, K. Takeda, "LIBRE: The multiple 3D LiDAR dataset," *arXiv preprint arXiv: 2003.06129*, 2020.
- [5] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, 1992.
- [6] P. Biber and W. Strasser, "The normal distribution transform: a new approach to laser scan matching," in *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2003.
- [7] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige and W. Burgard, "g2o: A general framework for graph optimization," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011.
- [8] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," in *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1365-1378, 2008.
- [9] J. Zhang, M. Kaess, and S. Singh, "On degeneracy of optimization-based state estimation problems," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016.
- [10] J. Zhang, and S. Singh, "Visual-lidar odometry and mapping: Low-drift, robust, and fast," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015.
- [11] J. Laconte, S. P. Deschenes, M. Labussiere and F. Pomerleau, "Lidar measurement bias estimation via return waveform modelling in a context of 3d mapping," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.
- [12] C. Cadena et al., "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," in *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309-1332, 2016.
- [13] T. Shan, and B. Englot, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018.
- [14] J. Lin and F. Zhang, "Loam_livox: A fast, robust, high-precision LiDAR odometry and mapping package for LiDARs of small FoV," in *2020 International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.
- [15] J. Behley and C. Stachniss, "Efficient Surlfel-Based SLAM using 3D Laser Range Data in Urban Environments," in *Robotics: Science and System*, 2018.
- [16] X. Chen, A. Milioto, E. Palazzolo, P. Giguere, J. Behley and C. Stachniss, "SuMa++: Efficient LiDAR-based Semantic SLAM," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019.
- [17] J. Deschaud, "IMLS-SLAM: scan-to-model matching based on 3D data," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018.
- [18] H. Ye, Y. Chen and M. Liu, "Tightly coupled 3d lidar inertial odometry and mapping," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.
- [19] D. Droschel, S. Behnke, "Efficient continuous-time SLAM for 3D lidar-based online mapping," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018.
- [20] F. Neuhaus, T. Kob, R. Kohnen, and D. Paulus, "Mc2slam: Real-time inertial lidar odometry using two-scan motion compensation," in *German Conference on Pattern Recognition (GCPR)*. Springer, Cham, 2018.
- [21] S. Thrun, "Probabilistic robotics," *Communications of the ACM* 45, 2002.
- [22] X. Liu, L. Zhang, S. Qin, D. Tian, S. O and C. Chen, "Optimized LOAM Using Ground Plane Constraints and SegMatch-Based Loop Detection," in *Sensors* 2019.
- [23] F. Zheng, H. Tang and Y. Liu, "Odometry-vision-based ground vehicle motion estimation with se (2)-constrained se (3) poses," in *IEEE Transactions on Cybernetics*, vol. 49, no. 7, pp 2652-2663, 2019.
- [24] F. Zheng and Y. Liu, "Visual-Odometric Localization and Mapping for Ground Vehicles Using SE (2)-XYZ Constraints," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.
- [25] K. Koide, J. Miura and E. Menegatti, "A Portable Three-Dimensional LIDAR-Based System for Long-Term and Wide-Area People Behavior Measurement," in *International Journal of Advanced Robotic Systems*, 2019.
- [26] M. Kaess, "Simultaneous localization and mapping with infinite planes," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015.
- [27] P. Geneva, K. Eickenhoff, Y. Yang, G. Huang, "LIPS: Lidar-inertial 3d plane slam," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018.
- [28] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," in *1991 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1991.
- [29] A. Censi, "An ICP variant using a point-to-line metric," in *2008 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2008.
- [30] M. Yokozuka, K. Koide, S. Oishi, A. Banno, "LiTAMIN2: Ultra Light LiDAR-based SLAM using Geometric Approximation applied with KL-Divergence," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021.
- [31] G. Chen, B. Wang, X. Wang, H. Deng, B. Wang, S. Zhang, "PSF-LO: Parameterized Semantic Features based Lidar Odometry," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021.