

# Multi-Sensor Fusion for Quadruped Robot State Estimation using Invariant Filtering and Smoothing

Ylenia Nisticò<sup>1,\*</sup>, Hajun Kim<sup>2,\*</sup>, João Carlos Virgolino Soares<sup>1</sup>, Geoff Fink<sup>1,3</sup>, Hae-Won Park<sup>2</sup>, Claudio Semini<sup>1</sup>

**Abstract**—This letter introduces two multi-sensor state estimation frameworks for quadruped robots, built on the Invariant Extended Kalman Filter (InEKF) and Invariant Smoother (IS). The proposed methods, named E-InEKF and E-IS, fuse kinematics, IMU, LiDAR, and GPS data to mitigate position drift, particularly along the z-axis, a common issue in proprioceptive-based approaches. We derived observation models that satisfy group-affine properties to integrate LiDAR odometry and GPS into InEKF and IS. LiDAR odometry is incorporated using Iterative Closest Point (ICP) registration on a parallel thread, preserving the computational efficiency of proprioceptive-based state estimation. We evaluate E-InEKF and E-IS with and without exteroceptive sensors, benchmarking them against LiDAR-based odometry methods in indoor and outdoor experiments using the KAIST HOUND2 robot. Our methods achieve lower Relative Position Errors (RPE) and significantly reduce Absolute Trajectory Error (ATE), with improvements of up to 28% indoors and 40% outdoors compared to LIO-SAM and FAST-LIO2. Additionally, we compare E-InEKF and E-IS in terms of computational efficiency and accuracy.

**Index Terms**—sensor fusion, localization, legged robots.

## I. INTRODUCTION

STATE ESTIMATION is crucial for quadruped robots navigating complex environments, as it determines pose, velocity, and key parameters from sensor inputs. Accurate estimation fuses high-frequency proprioceptive inputs (e.g., IMUs, encoders), which suffer from drift, with lower-frequency exteroceptive inputs (e.g., cameras, LiDARs), which may struggle in challenging conditions (e.g., foggy areas, long corridors, etc.). Sensor fusion techniques address these limitations by combining both sensor types. For instance, LIO-SAM [1] employs a factor graph for LiDAR-inertial odometry (LIO), while FAST-LIO [2] uses an iterated Extended Kalman Filter (EKF) to fuse

Manuscript received: December, 31, 2024; Revised February, 18, 2025; Accepted April, 14, 2025.

This paper was recommended for publication by Editor Abderrahmane Kheddar upon evaluation of the Associate Editor and Reviewers' comments.

This work was supported by the ASI Pegasus Project and the Technology Innovation Program (00427719, Dexterous and Agile Humanoid Robots for Industrial Applications) funded by the Ministry of Trade Industry & Energy (MOTIE, Korea).

<sup>1</sup> Ylenia Nisticò, João Carlos Virgolino Soares, and Claudio Semini are with the Dynamic Legged Systems (DLS), Istituto Italiano di Tecnologia (IIT), Genoa, Italy. [ylenia.nisticò@iit.it](mailto:ylenia.nisticò@iit.it), [joao.virgolino@iit.it](mailto:joao.virgolino@iit.it), [claudio.semini@iit.it](mailto:claudio.semini@iit.it)

<sup>2</sup> Hajun Kim and Hae-Won Park are with the Dynamic Robot Control & Design (DRCD) Laboratory, Korean Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. [hajun0219@kaist.ac.kr](mailto:hajun0219@kaist.ac.kr), [haewonpark@kaist.ac.kr](mailto:haewonpark@kaist.ac.kr)

<sup>3</sup> Geoff Fink is with the Dynamic Legged Systems (DLS), Istituto Italiano di Tecnologia (IIT), Genoa, Italy, and also with the Department of Engineering, Thompson Rivers University, Kamloops, BC, Canada. [gefink@tru.ca](mailto:gefink@tru.ca)

\*Equal contribution. Corresponding author: [ylenia.nisticò@iit.it](mailto:ylenia.nisticò@iit.it)  
Digital Object Identifier (DOI): 10.1109/LRA.2025.3564711

LiDAR and IMU data, with FAST-LIO2 [3] further improving accuracy by eliminating the need for feature extraction.

Although effective, these methods do not explicitly consider contact interactions, which can play a significant role in legged robot locomotion. Specifically for legged robots, state estimation incorporating leg kinematics has been studied in previous works, such as [4], [5], which introduced EKF-based frameworks combining IMU and kinematic measurements under the stable contact assumption. To handle violations of this assumption, a slip rejection method that adjusts covariances during slipping was proposed in [6].

Among more recent research works, Barrau and Bonnabel proposed in [7] the use of the group-affine property for state estimation, which enables a system to be formulated as a linear system by using the Lie algebra error of the state variables defined in a Lie group. They also introduced the invariant EKF (InEKF), which leverages group-affine properties and shows better convergence than the standard EKF when initial errors are present. Hartley et al. [8] introduced a Proprioceptive-only InEKF (P-InEKF) for legged robots, demonstrating that fusing IMU data with the robot's kinematics results in a state estimation process that is approximately group-affine. Further studies in P-InEKF have explored the incorporation of a learning-based contact estimator [9] or the use of robust scale-variant cost functions to handle challenging terrains [10]. Moreover, while the P-InEKF processes only the two most recent states, the study in [11] introduced the Invariant Smoother (IS), which leverages the group-affine property within a smoothing framework that preserves the state history. Yoon et al. [12] introduced a Proprioceptive-only Invariant Smoother (P-IS) for legged robots, demonstrating robust performance during dynamic contact events while accounting for contact loop costs.

Despite these advancements, proprioceptive-only methods remain prone to drift over time due to an unobservable absolute position and yaw angle [4]. Several state estimation frameworks for legged robots integrate proprioceptive and exteroceptive sensors to address this issue. For instance, Pronto [13] uses an EKF to fuse IMU, kinematics, stereo vision, and LiDAR for pose corrections. MUSE [14] fuses a nonlinear attitude observer with a Kalman filter for precise state estimation, while STEP [15] integrates pre-computed foot velocities and stereo camera data, removing the need for contact detection and stable contact assumptions. VILENS [16] and Cerberus [17] leverage factor graphs to robustly fuse multiple sensors, with Cerberus also using visual data for kinematic parameter estimation. Leg-KILO [18] tightly integrates IMU, kinematics, LiDAR odometry, and loop

closure via graph optimization for enhanced accuracy.

However, the studies in [13]–[18] do not consider state estimation with the group-affine property, despite its advantages in improving estimation accuracy and convergence stability [7], [11]. Achieving these benefits requires that both the propagation and observation models for sensor measurements are formulated to satisfy the group-affine conditions [7], [11].

Meanwhile, some studies have explored learning-based methods for state estimation [19], [20]. For legged robots, [21] uses a neural measurement model with proprioception, while other works use transformers for the estimation [22], [23]. However, these techniques require large datasets [24], have high computational costs, and may struggle with real-time performance.

Some studies have incorporated exteroceptive sensors into state estimation using the group-affine property. The study in [25] fuses inertial, kinematics, and linear velocity measurements from a tracking camera, capitalizing on the underlying group-affine structure. Similarly, Gao et al. [26] presented an InEKF for dynamic rigid surfaces, combining RGB-D cameras and contact orientation measurements using ArUco markers.

To the best of our knowledge, neither the InEKF nor the IS frameworks for legged robots have integrated LiDAR or GPS. In this work, we aim to design an observation model for these frameworks that integrates LiDAR and GPS, in addition to leg kinematics, while preserving the group-affine structure. The main contributions are as follows:

- We extended InEKF [8] and IS [12] by fusing kinematics and IMU, with LiDAR odometry and GPS measurements to mitigate the position drift inherent in proprioceptive-only methods, naming our approaches as **E-InEKF** and **E-IS**. To the best of our knowledge, this is the first work to incorporate LiDAR odometry and GPS into InEKF and IS for quadruped robots.
- To integrate LiDAR odometry and GPS into the InEKF and the IS, we derive an observation model that satisfies group-affine properties for both sensor types. Additionally, to manage LiDAR’s low frequency of approximately 10 Hz, we calculate LiDAR odometry in a parallel thread using the ICP registration of [27], allowing the estimator thread to maintain a fast computation time.
- Our algorithms were verified on the KAIST Hound2 quadruped robot [28] in indoor and outdoor environments, benchmarking the results against LIO-SAM [1] and FAST-LIO2 [3], two LIO systems. Additionally, ablation studies were conducted to evaluate the performance without LiDAR or GPS, and to compare the accuracy and computation trade-offs between E-IS and E-InEKF.

The paper is organized as follows: Sec. II outlines the preliminaries for group affine property, P-InEKF, and P-IS. Sec. III details the implementation of the observation models in E-InEKF and E-IS. Sec. IV describes the experiments, while Sec. V discusses the results. Sec. VI concludes the paper.

## II. PRELIMINARIES

This section provides the theoretical background based on [7], [11], [29], and an overview of the P-InEKF [8] and P-IS [12].

### A. Matrix Lie Groups and group-affine property

In robots’ state estimation, the state can be represented on a manifold rather than a vector space. For instance, the orientation is represented by the special orthogonal group  $\mathbf{SO}(3) = \{\mathbf{R} \in \mathbb{R}^{3 \times 3} \mid \det(\mathbf{R}) = 1, \mathbf{R}^\top \mathbf{R} = \mathbb{I}_3\}$ , where  $\mathbb{I}_{nd} \in \mathbb{R}^{nd \times nd}$  is the identity matrix. When the state includes additional vector components, such as the position or velocity, the elements of  $\mathbf{SE}_k(3)$  can be adopted, which is represented by the  $(3+k) \times (3+k)$  matrix and defined as follows [8]:

$$\mathbf{X} \triangleq \begin{bmatrix} \mathbf{R} & {}^1p & \dots & {}^kp \\ \mathbf{0}_{k,3} & & & \mathbb{I}_k \end{bmatrix} \in \mathbf{SE}_k(3), \quad (1)$$

where  $\mathbf{R} \in \mathbf{SO}(3)$  is the rotation matrix, and  ${}^ip \in \mathbb{R}^3$ , for  $i = 1, 2, \dots, k$ , is a vector. The element  $\mathbf{X} \in \mathbf{SE}_k(3)$  can be mapped to  $\xi \triangleq [\phi^\wedge, {}^1\xi^\top, \dots, {}^k\xi^\top]^\top \in \mathbb{R}^{3+3k}$ , which lies in an Euclidean vector space, via the logarithmic map represented by  $\mathbf{X} \in \mathbf{SE}_k(3) \rightarrow \text{Log}(\mathbf{X}) \in \mathbb{R}^{3+3k}$  and back via the exponential map described by  $\xi \in \mathbb{R}^{3+3k} \rightarrow \text{Exp}(\xi) \in \mathbf{SE}_k(3)$ . The exponential map in  $\mathbf{SE}_k(3)$  is defined as follows:

$$\text{Exp}(\xi) = \exp(\xi^\wedge) = \begin{bmatrix} \exp(\phi^\wedge) & \mathbf{J}_l(\phi)^1 \xi & \dots & \mathbf{J}_l(\phi)^k \xi \\ \mathbf{0}_{k,3} & & & \mathbb{I}_k \end{bmatrix}, \quad (2)$$

where  $\mathbf{J}_l(\cdot)$  is the left Jacobian of the  $\mathbf{SO}(3)$  manifold, and the *hat* operation  $(\cdot)^\wedge$  on  $\phi$  and  $\xi$ , which is the inverse mapping of the *vee* operation  $(\cdot)^\vee$ , is respectively defined as

$$\phi^\wedge = \begin{bmatrix} 0 & -\phi_z & \phi_y \\ \phi_z & 0 & -\phi_x \\ -\phi_y & \phi_x & 0 \end{bmatrix}, \xi^\wedge = \begin{bmatrix} \phi^\wedge & {}^1\xi & \dots & {}^k\xi \\ \mathbf{0}_{k,3} & & & \mathbf{0}_{k,k} \end{bmatrix} \quad (3)$$

The adjoint matrix, which plays a key role in transforming perturbations in the tangent space, is defined as:

$$\text{Ad}_{\mathbf{X}} = \begin{bmatrix} \mathbf{R} & \mathbf{0}_{3 \times 3} & \dots & \mathbf{0}_{3 \times 3} \\ {}^1p^\wedge \mathbf{R} & \mathbf{R} & \dots & \mathbf{0}_{3 \times 3} \\ \vdots & \vdots & \ddots & \vdots \\ {}^k p^\wedge \mathbf{R} & \mathbf{0}_{3 \times 3} & \dots & \mathbf{R} \end{bmatrix}. \quad (4)$$

Also, we define the  $\odot$  operation, introduced in [12], as follows:

$$\xi^\odot \triangleq \begin{bmatrix} -\phi^\wedge & {}^1\xi \cdot \mathbb{I}_3 & \dots & {}^k\xi \cdot \mathbb{I}_3 \\ \mathbf{0}_{k,3} & \mathbf{0}_{k,3} & \dots & \mathbf{0}_{k,3} \end{bmatrix} \in \mathbb{R}^{(3+k) \times (3+3k)}. \quad (5)$$

The group-affine property provides a significant advantage by ensuring that the observer’s error dynamics do not depend on the current state when the propagation and observation models, defined on the Lie Group, are group-affine [7].

For an arbitrary state  $\mathbf{X}_t$  at time  $t$  and its corresponding estimate  $\bar{\mathbf{X}}_t$  in the Lie Group  $\mathbb{G}$ , the right-invariant error is defined as  $\eta_t^r \triangleq \bar{\mathbf{X}}_t^{-1} \mathbf{X}_t$ , and the corresponding right log-invariant error is  $\xi_t^r \triangleq \text{Log}(\eta_t^r)$ . A system is group-affine if its propagation model  $f(\cdot)$  satisfies the following conditions:  $\frac{d}{dt} \mathbf{X}_t = f(\mathbf{X}_t)$  and  $f(\mathbf{XY}) = \mathbf{X}f(\mathbf{Y}) + f(\mathbf{X})\mathbf{Y} - \mathbf{X}f(\mathbb{I}_{\dim(\mathbf{Y})})\mathbf{Y}$  for  $\forall t \geq 0$  and  $\forall \mathbf{X}, \mathbf{Y} \in \mathbb{G}$ .

When the trajectories  $\mathbf{X}_{0:t}$  and  $\bar{\mathbf{X}}_{0:t}$  are governed by the same group-affine propagation, the right invariant error’s propagation dynamics is  $\frac{d}{dt} \eta_t^r = g(\eta_t^r)$ , where a function  $g(\cdot)$  is represented by  $g(\eta_t^r) = f(\eta_t^r) - \eta_t^r f(\mathbb{I}_{\dim(\mathbf{Y})})$ , as shown in [7]. For small errors, it is also proven in [7] that the

propagation of the log-invariant error  $\xi_t^r$  is a linear differential equation, with a constant matrix  $\mathbf{G}$ :

$$g(\eta_t^r) = (\mathbf{G}\xi_t^r)^\wedge + \mathcal{O}(\|\xi_t^r\|^2), \quad \frac{d}{dt}\xi_t^r = \mathbf{G}\xi_t^r, \quad (6)$$

where  $\mathcal{O}(\cdot)$  indicates higher order terms that are negligible. Due to the constant matrix  $\mathbf{G}$ , the evolution of the error  $\xi_t^r$  is autonomous and independent of the two trajectories  $\mathbf{X}_{0:t}$  and  $\bar{\mathbf{X}}_{0:t}$ , which is referred to as the *log-linear property* of the error [7].

Meanwhile, the group-affine property for the observation model is satisfied under the condition:

$$\mathbf{y}_t = \mathbf{X}_t^{-1}\mathbf{s} + \mathbf{w}_t^{\text{obs}}, \quad (7)$$

where  $\mathbf{y}_t$  is the observation at time  $t$ ,  $\mathbf{s}$  is a constant vector, and  $\mathbf{w}_t^{\text{obs}}$  is Gaussian noise vector of the observation.

### B. State Definition

Inspired by [8], [12], we define the state  $\mathbf{X}_t \in \text{SE}_{N+2}(3)$  for  $N$  contact points, while the IMU bias vector  $\mathbf{x}_t \in \mathbb{R}^6$  is defined in the vector space as follows:

$$\mathbf{X}_t \triangleq \begin{bmatrix} \mathbf{R}_t & \mathbf{v}_t & \mathbf{p}_t & \mathbf{d}_{1,t} & \cdots & \mathbf{d}_{N,t} \\ 0_{N+2,3} & & & & & \mathbb{I}_{(N+2)} \end{bmatrix} \quad \text{and} \quad \mathbf{x}_t \triangleq \begin{bmatrix} \mathbf{b}_t^\omega \\ \mathbf{b}_t^a \end{bmatrix}, \quad (8)$$

where  $\mathbf{R}_t$  is the base orientation,  $\mathbf{v}_t$  is the linear velocity,  $\mathbf{p}_t$  is the robot position, and  $\mathbf{d}_{i,t}$  is  $i$ -th contact position, while  $\mathbf{b}_t^\omega$  and  $\mathbf{b}_t^a$  are the IMU biases for gyroscope and accelerometer.

For the sake of readability, since the measurement models for each contact point  $\mathbf{d}_{i,t}$  are identical, we will derive all further equations assuming only a single contact point  $\mathbf{d}_t$ .

We also consider the measurements as  $\mathbf{Z}_t \triangleq [\tilde{\omega}_t^\top, \tilde{\mathbf{a}}_t^\top, \tilde{\mathbf{q}}_t^\top, \dot{\tilde{\mathbf{q}}}_t^\top, \mathbf{p}_{cl,t}^\top, \mathbf{p}_{gps,t}^\top]$ , where  $\tilde{\omega}_t \in \mathbb{R}^3$ ,  $\tilde{\mathbf{a}}_t \in \mathbb{R}^3$ ,  $\tilde{\mathbf{q}}_t \in \mathbb{R}^{3M}$ ,  $\dot{\tilde{\mathbf{q}}}_t \in \mathbb{R}^{3M}$ ,  $\mathbf{p}_{cl,t}$ , and  $\mathbf{p}_{gps,t}$  are the angular velocity, the linear acceleration, joint positions, joint velocities for the  $M$ -legged robot, where  $M$  is the number of legs, LiDAR point clouds, and position data from GPS, respectively.

### C. Invariant Extended Kalman Filter For Legged Robots

In the InEKF [8], both the propagation and observation models are formulated under the group-affine property [8] to ensure that the right-invariant error is independent of the current state. The propagation model integrates IMU measurements to predict the state, whereas the observation model refines the estimates under the stable contact assumption, which implies zero foot velocity during ground contact.

The system dynamics for the propagation that includes the corruption from noise and bias can be described as follows [8]:

$$\frac{d}{dt}\mathbf{R}_t = \mathbf{R}_t(\tilde{\omega}_t - \mathbf{b}_t^\omega - \mathbf{w}_t^\omega)^\wedge, \quad \frac{d}{dt}\mathbf{p}_t = \mathbf{v}_t \quad (9)$$

$$\frac{d}{dt}\mathbf{v}_t = \mathbf{R}_t(\tilde{\mathbf{a}}_t - \mathbf{b}_t^a - \mathbf{w}_t^a) + \mathbf{g}, \quad \frac{d}{dt}\mathbf{d}_t = \mathbf{R}_t\mathbf{w}_t^d \quad (10)$$

$$\frac{d}{dt}\mathbf{b}_t^\omega = \mathbf{w}_t^{b^\omega}, \quad \frac{d}{dt}\mathbf{b}_t^a = \mathbf{w}_t^{b^a} \quad (11)$$

where  $\mathbf{g}$  is the gravity vector,  $\mathbf{w}_t^\omega$ ,  $\mathbf{w}_t^a$ ,  $\mathbf{w}_t^d$ ,  $\mathbf{w}_t^{b^\omega}$ , and  $\mathbf{w}_t^{b^a}$  are the zero-mean Gaussian noise terms of each process. The

propagation model can be described as the noise-corrupted version in the discrete domain [8] of the log-linear property (6):

$$\begin{bmatrix} \xi_{t+1}^r \\ \zeta_{t+1}^r \end{bmatrix} = (\mathbb{I}_{18} + \mathbf{A}_t \Delta t) \begin{bmatrix} \xi_t^r \\ \zeta_t^r \end{bmatrix} + \mathbf{B}_t \bar{\mathbf{w}}_t, \quad \mathbf{B}_t = \Delta t \begin{bmatrix} \text{Ad}_{\bar{\mathbf{X}}_t} & \mathbf{0}_{12,6} \\ \mathbf{0}_{6,12} & \mathbb{I}_6 \end{bmatrix} \quad (12)$$

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{0} & \mathbf{0}_{3,3} & \mathbf{0}_{3,3} & \mathbf{0}_{3,3} & -\bar{\mathbf{R}}_t & \mathbf{0}_{3,3} \\ (\mathbf{g})^\wedge & \mathbf{0}_{3,3} & \mathbf{0}_{3,3} & \mathbf{0}_{3,3} & -(\bar{\mathbf{v}}_t)^\wedge \bar{\mathbf{R}}_t & -\bar{\mathbf{R}}_t \\ \mathbf{0}_{3,3} & \mathbb{I}_3 & \mathbf{0}_{3,3} & \mathbf{0}_{3,3} & -(\bar{\mathbf{p}}_t)^\wedge \bar{\mathbf{R}}_t & \mathbf{0}_{3,3} \\ \mathbf{0}_{3,3} & \mathbf{0}_{3,3} & \mathbf{0}_{3,3} & \mathbf{0}_{3,3} & -(\bar{\mathbf{d}}_t)^\wedge \bar{\mathbf{R}}_t & \mathbf{0}_{3,3} \\ \mathbf{0}_{6,3} & \mathbf{0}_{6,3} & \mathbf{0}_{6,3} & \mathbf{0}_{6,3} & \mathbf{0}_{6,3} & \mathbf{0}_{6,3} \end{bmatrix}, \quad (13)$$

where  $\Delta t$  is the time step of the state estimation,  $\bar{\mathbf{w}}_t = [(\mathbf{w}_t^\omega)^\top, (\mathbf{w}_t^a)^\top, (\mathbf{w}_t^d)^\top, (\mathbf{w}_t^{b^\omega})^\top, (\mathbf{w}_t^{b^a})^\top]^\top$  is the noise vector,  $\mathbf{e}_t \triangleq [(\xi_t^r)^\top, (\zeta_t^r)^\top]^\top$ , and the bias error can be defined as  $\zeta_t^r = \mathbf{x}_t - \bar{\mathbf{x}}_t$ .

For the observation model, the forward kinematic measurement, which captures the relative position of the contact point with respect to the body, is formulated to satisfy the group-affine property of (7). The forward kinematics  $\text{fk}(\tilde{\mathbf{q}}_t)$  can be expressed as:  $\text{fk}(\tilde{\mathbf{q}}_t) = \mathbf{R}_t^\top(\mathbf{d}_t - \mathbf{p}_t) + \mathbf{J}_p(\tilde{\mathbf{q}}_t)\mathbf{w}_t^q$ , where  $\mathbf{J}_p$  denotes the analytical Jacobian of the forward kinematics, and  $\mathbf{w}_t^q$  is the observation noise in forward kinematics. As in (7), the right-invariant form of this measurement can be described as follows, allowing the innovation to remain only dependent on the invariant error [8]:

$$\mathbf{Y}_t^{\text{kin}} = \mathbf{X}_t^{-1}\mathbf{b}_t^{\text{kin}} + \mathbf{V}_t^{\text{kin}}, \quad (14)$$

where  $\mathbf{Y}_t^{\text{kin}} = [\text{fk}(\tilde{\mathbf{q}}_t) \ 0 \ 1 \ -1]^\top$  is the vector of kinematic observation,  $\mathbf{X}_t^{-1}$  is the inverse of the state matrix  $\mathbf{X}_t$ ,  $\mathbf{b}_t^{\text{kin}} = [0_{3,1} \ 0 \ 1 \ -1]^\top$  is a constant vector, while  $\mathbf{V}_t^{\text{kin}} = [\mathbf{J}_p(\tilde{\mathbf{q}}_t)\mathbf{w}_t^q \ 0 \ 0 \ 0]^\top$  is the Gaussian noise vector of the observation model.

Given the observation model of (7), which satisfies the group-affine property, the state and covariance updates of the observation model can be expressed as follows [8]:  $\bar{\mathbf{X}}_t^+ = \text{Exp}(\mathbf{K}_t \Pi \bar{\mathbf{X}}_t^- \mathbf{Y}_t^-)$ ,  $\mathbf{P}_t^+ = (\mathbb{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t (\mathbb{I} - \mathbf{K}_t \mathbf{H}_t)^\top + \mathbf{K}_t \bar{\mathbf{N}}_t \mathbf{K}_t^\top$ , where  $\bar{\mathbf{X}}_t^-$  is the estimated state from system dynamics,  $\mathbf{P}_t$  is the system covariance matrix, and  $\Pi \triangleq [\mathbb{I} \ 0_{3,3}]$  is the auxiliary selection matrix. The Kalman gain matrix  $\mathbf{K}_t$  is computed as  $\mathbf{K}_t = \mathbf{P}_t \mathbf{H}_t^\top (\mathbf{H}_t \mathbf{P}_t \mathbf{H}_t^\top + \bar{\mathbf{N}}_t)^{-1}$ . The matrices  $\mathbf{H}_t$  and  $\bar{\mathbf{N}}_t$  are given by  $\mathbf{H}_t = [0_{3,3} \ 0_{3,3} \ -\mathbb{I} \ \mathbb{I}]$  and  $\bar{\mathbf{N}}_t = \mathbf{R}_t^- \mathbf{J}_p(\tilde{\mathbf{q}}_t) \text{Cov}(\mathbf{w}_t^q) \mathbf{J}_p^\top(\tilde{\mathbf{q}}_t) (\mathbf{R}_t^-)^\top$ .

### D. Invariant Smoother For Legged Robots

Unlike the filtering method of [8], which handles the last two states of the system, the smoothing method aims to recover the estimates using Maximum A Posteriori (MAP) estimation, by incorporating the states  $\mathbf{X}_{0:n}$  and measurements  $\mathbf{Z}_{0:n}$  within the specific time window  $n$ . As derived in [12], MAP can be formulated as a nonlinear least-squares optimization problem as follows:

$$\begin{aligned} \mathbf{e}_{0:n}^* = \underset{\mathbf{e}_{0:n}}{\text{argmin}} & (\|\mathbf{r}_{\text{pri}} - \mathbf{J}_{\text{pri}} \mathbf{e}_{0:n}\|_{\Sigma_{\text{pri}}}^2 \\ & + \sum_{t=0}^{n-1} \|\mathbf{r}_t^{\text{pro}} - \mathbf{J}_{\text{pro}}^t \mathbf{e}_{0:n}\|_{\Sigma_{\text{pro}}}^2 + \sum_{t=0}^n \|\mathbf{r}_o^t - \mathbf{J}_o^t \mathbf{e}_{0:n}\|_{\Sigma_o}^2), \end{aligned} \quad (15)$$

where  $n$  is the window size (WS),  $\mathbf{r}_{\text{pri}}$ ,  $\mathbf{r}_{\text{pro}}$  and  $\mathbf{r}_o$  are residuals,  $\Sigma_{\text{pri}}$ ,  $\Sigma_{\text{pro}}$  and  $\Sigma_o$  are covariance matrices, and  $\mathbf{J}_{\text{pri}}$ ,  $\mathbf{J}_{\text{pro}}$  and  $\mathbf{J}_o$  are the Jacobians at the operating points  $\bar{\mathbf{X}}_t$  and  $\bar{\mathbf{x}}_t$  for the distribution of *prior*, *propagation*, and *observation*, respectively.

The prior cost gives an initialization to state estimation, employing marginalization to keep the fixed size of the time window. The detailed derivation is explained in [12] and not repeated here. The propagation cost can be derived from the log-linear property, as shown in [12], and expressed as follows:

$$\mathbf{r}_{\text{pro}}^t = \begin{bmatrix} \text{Log}(f_M^d(\bar{\mathbf{X}}_t)\bar{\mathbf{X}}_{t+1}^{-1}) \\ f_v^d(\bar{\mathbf{x}}_t) - \bar{\mathbf{x}}_{t+1} \end{bmatrix}, \Sigma_{\text{pro}} = \mathbf{B}_t \text{Cov}(\bar{\mathbf{w}}_t) \mathbf{B}_t^\top, \quad (16)$$

$$\mathbf{J}_{\text{pro}}^{t+1} = \mathbb{I}_{18} + \mathbf{A}_t \Delta t, \quad \mathbf{J}_{\text{pro}}^t = \mathbb{I}_{18}, \quad (17)$$

$$f_M^d(\mathbf{X}_t) = \begin{bmatrix} \mathbf{R}_t^d & \mathbf{v}_t^d & \mathbf{p}_t^d & \mathbf{d}_t^d \\ \mathbf{0}_{3,3} & \mathbb{I}_3 \end{bmatrix}, \quad f_v^d(\mathbf{x}_t) = \begin{bmatrix} \mathbf{b}_t^\omega \\ \mathbf{b}_t^a \end{bmatrix}, \quad (18)$$

where  $f_M^d$  is the discretized dynamics of (9) and (10),  $f_v^d$  is the discretized dynamics of (11),  $\mathbf{R}_t^d = \mathbf{R}_t \text{Exp}((\bar{\omega}_t - \mathbf{b}_t^\omega) \Delta t)$ ,  $\mathbf{v}_t^d = \mathbf{v}_t + \mathbf{R}_t(\bar{\mathbf{a}}_t - \mathbf{b}_t^a) \Delta t + \mathbf{g} \Delta t$ , and  $\mathbf{p}_t^d = \mathbf{p}_t + \mathbf{v}_t \Delta t + \frac{1}{2} \mathbf{R}_t(\bar{\mathbf{a}}_t - \mathbf{b}_t^a)(\Delta t)^2 + \frac{1}{2} \mathbf{g}(\Delta t)^2$ .

For the observation cost, the right-invariant kinematics observation of (14) can be employed as in [12]:  $\mathbf{r}_o^t = \bar{\mathbf{X}}_t \mathbf{Y}_t^{\text{kin}} - \mathbf{b}_t^{\text{kin}}, \mathbf{J}_o^t = [(\mathbf{b}_t^{\text{kin}})^\odot \mathbf{0}_{6,6}], \Sigma_o^t = \bar{\mathbf{X}}_t \Sigma_{\text{kin}}(\mathbf{w}^q)_t \bar{\mathbf{X}}_t^\top$ , where the  $\odot$  operator is defined in (5).

When the MAP of (15) is solved, the optimized variables updates the operating states,  $\bar{\mathbf{X}}_i$  and  $\bar{\mathbf{x}}_i$ , at each iteration as:  $\bar{\mathbf{X}}_t^* \leftarrow \text{Exp}(\zeta_t^{r*}) \bar{\mathbf{X}}_t, \mathbf{x}_t^* \leftarrow \bar{\mathbf{x}}_i + \zeta_i^{r*}$ .

### III. PROPOSED OBSERVATION MODEL

The main contribution of this paper is a measurement model that integrates kinematic measurements with exteroceptive sensor data, specifically LiDAR and GPS, while satisfying the group-affine property of (7), thereby enabling the incorporation of sensor measurements into InEKF or IS.

#### A. LiDAR Odometry Factor

For LiDAR odometry, we adopt KISS-ICP [27], an efficient algorithm that estimates the pose by sequentially aligning LiDAR point clouds using a point-to-point ICP method. KISS-ICP applies a constant-velocity motion model to deskew scans, uses voxel-based downsampling to reduce computational load, adapts thresholds for data association based on motion, and refines pose estimates through robust optimization handling point-to-point ICP. By solving the ICP problem, we obtain the position of the LiDAR odometry,  $\mathbf{p}_{\text{lid}}$ , in the world frame.

Unlike many other odometry techniques, KISS-ICP operates independently of additional sensors such as IMUs or leg odometry, enhancing adaptability. Furthermore, we employed a loosely coupled approach, running the LiDAR odometry on a separate thread to preserve the efficiency of proprioceptive state estimation. This loosely coupled setup allows KISS-ICP to deliver reliable LiDAR-based position updates, which integrate seamlessly into our system.

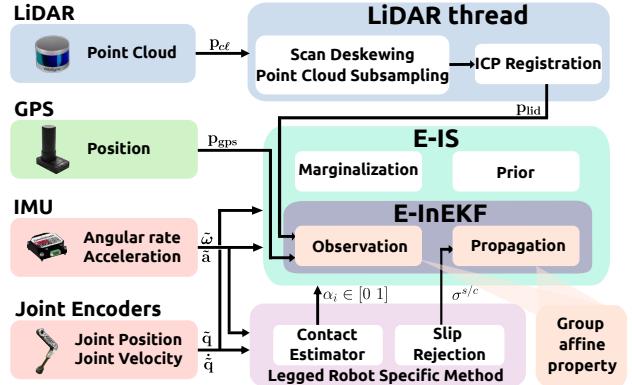


Fig. 1: The structure of the two proposed frameworks. The E-InEKF and E-IS have the propagation and observation modules, but the E-IS also includes the prior and marginalization modules.

#### B. Observation Model for LiDAR and GPS

1) *LiDAR measurement model*: by computing the LiDAR position estimate  $\mathbf{p}_{\text{lid}}$  from the point-to-point ICP problem [27], we express the LiDAR position estimate  $\mathbf{p}_{\text{lid}}$  with respect to the state  $\mathbf{X}_t$  in the right-invariant form as given in (7):

$$\mathbf{Y}_{\text{lid}} = \bar{\mathbf{X}}_t^{-1} \mathbf{b}_{\text{lid}} + \mathbf{V}_{\text{lid}} \quad (19)$$

where  $\mathbf{b}_{\text{lid}} = [\mathbf{p}_{\text{lid}} \ 0 \ 1 \ 0]^\top$  is the LiDAR observation vector,  $\mathbf{Y}_{\text{lid}} = [\mathbf{0}_{3,1} \ 0 \ 1 \ 0]^\top$  is a constant vector, while  $\mathbf{V}_{\text{lid}} = [\mathbf{J}_{\mathbf{p}_{\text{lid}}} \mathbf{w}_{\text{lid}}^\top \ 0 \ 0 \ 0]^\top$  is the Gaussian noise vector associated with the observation model. This formulation aligns LiDAR measurements with the group-affine property, facilitating consistent integration within the invariant estimation frameworks.

2) *GPS measurement model*: to incorporate GPS position estimates  $\mathbf{p}_{\text{gps}}$  into the invariant state estimation framework, we express the GPS position  $\mathbf{p}_{\text{gps}}$  with respect to the state  $\mathbf{X}_t$  as a right-invariant observation model:

$$\mathbf{Y}_{\text{gps}} = \bar{\mathbf{X}}_t^{-1} \mathbf{b}_{\text{gps}} + \mathbf{V}_{\text{gps}} \quad (20)$$

where  $\mathbf{b}_{\text{gps}} = [\mathbf{p}_{\text{gps}} \ 0 \ 1 \ 0]^\top$  represents the GPS observation vector,  $\mathbf{Y}_{\text{gps}} = [\mathbf{0}_{3,1} \ 0 \ 1 \ 0]^\top$  is a constant vector, while  $\mathbf{V}_{\text{gps}} = [\mathbf{J}_{\mathbf{p}_{\text{gps}}} \mathbf{w}_{\text{gps}}^\top \ 0 \ 0 \ 0]^\top$  is the Gaussian noise vector associated with the observation model.

Based on the measurement model corrupted by zero-mean Gaussian noise, we derived the observation models for LiDAR and GPS measurements. The introduced observation models for LiDAR and GPS of Equations (19) and (20) fulfill the group-affine property required for the invariant state estimation, because our measurement models follow the right-invariant observation form [7], given in Equation (7). Thus, our approach preserves the invariant structure, ensuring that the error dynamics of the estimator remain autonomous and independent of the robot's trajectory.

#### C. E-InEKF: LiDAR-GPS fused Invariant EKF

Building upon the contact-aided InEKF in [8], we extend the measurement model to include LiDAR-odometry and GPS observations. Starting from (14), (19), and (20), we define an aggregated observation vector  $\mathbf{Y} = [\mathbf{Y}_{\text{kin}}^\top, \mathbf{Y}_{\text{lid}}^\top, \mathbf{Y}_{\text{gps}}^\top]^\top$ , and a corresponding constant vector

$\mathbf{b} = [\mathbf{b}_{\text{kin}}^\top, \mathbf{b}_{\text{lid}}^\top, \mathbf{b}_{\text{gps}}^\top]^\top$ . Rewritten in matrix form, the right-invariant observation model becomes:

$$\underbrace{\begin{bmatrix} \mathbf{Y}_{\text{kin}} \\ \mathbf{Y}_{\text{lid}} \\ \mathbf{Y}_{\text{gps}} \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} \mathbf{X}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}^{-1} \end{bmatrix}}_{\mathbf{X}_{\text{aug}}^{-1}} \underbrace{\begin{bmatrix} \mathbf{b}_{\text{kin}} \\ \mathbf{b}_{\text{lid}} \\ \mathbf{b}_{\text{gps}} \end{bmatrix}}_{\mathbf{b}} + \underbrace{\begin{bmatrix} \mathbf{V}_{\text{kin}} \\ \mathbf{V}_{\text{lid}} \\ \mathbf{V}_{\text{gps}} \end{bmatrix}}_{\mathbf{V}} \quad (21)$$

Here  $\mathbf{X}_{\text{aug}}$  is the augmented state matrix, with the state  $\mathbf{X}$  appearing on the diagonal.

The linear update equations follow as described in Sec. II-C, with the Jacobians defined as  $\mathbf{H} = [\mathbf{H}_{\text{kin}}^\top \mathbf{H}_{\text{lid}}^\top \mathbf{H}_{\text{gps}}^\top]^\top$ , and the noise covariance given by the block-diagonal (blkdiag) matrix  $\mathbf{N} = \text{blkdiag}(\mathbf{N}_{\text{kin}}, \mathbf{N}_{\text{lid}}, \mathbf{N}_{\text{gps}})$ , where the individual Jacobians are  $\mathbf{H}_{\text{kin}} = [\mathbf{0}_{3,3} \ \mathbf{0}_{3,3} \ -\mathbb{I}_3 \ \mathbb{I}_3]$ , and  $\mathbf{H}_{\text{lid}} = \mathbf{H}_{\text{gps}} = [\mathbf{0}_{3,3} \ \mathbf{0}_{3,3} \ -\mathbb{I}_3 \ \mathbf{0}_{3,3}]$ , and the individual covariances are  $\mathbf{N}_{\text{kin}} = \mathbf{R}_t \mathbf{J}_p(\tilde{\mathbf{q}}_t) \Sigma(\mathbf{w}_t^q) \mathbf{J}_p^\top(\tilde{\mathbf{q}}_t) \mathbf{R}_t^\top$ ,  $\mathbf{N}_{\text{lid}} = \mathbf{R}_t \mathbf{J}_{\text{lid}} \Sigma(\mathbf{w}_t^{\text{lid}}) \mathbf{J}_{\text{lid}}^\top \mathbf{R}_t^\top$ , and  $\mathbf{N}_{\text{gps}} = \mathbf{R}_t \mathbf{J}_{\text{gps}} \Sigma(\mathbf{w}_t^{\text{gps}}) \mathbf{J}_{\text{gps}}^\top \mathbf{R}_t^\top$ . In these expressions,  $\Sigma(\mathbf{w}_t^q)$ ,  $\Sigma(\mathbf{w}_t^{\text{lid}})$ , and  $\Sigma(\mathbf{w}_t^{\text{gps}})$  are the covariance matrices of the joint position, LiDAR position, and GPS position noise, respectively.

#### D. E-IS: LiDAR-GPS fused Invariant Smoother

The E-IS is formulated such that the *observation* term in the cost function (15) incorporates global position data  $\mathbf{p}_{\text{lid},t}$  and  $\mathbf{p}_{\text{gps},t}$  from LiDAR and GPS, respectively. These global measurements correct the state by comparing the robot's estimated position with observed reference points.

From (15), we continue with the LiDAR observation cost, by expressing the equation for the *LiDAR measurement model* in a right-invariant form, similar to (19), utilizing  $\mathbf{p}_{\text{lid}}$ , which represents the LiDAR position in the robot's body frame. Based on this formulation, we compute the LiDAR residual function, which is subsequently used to update the global robot position. Specifically, the residual  $\mathbf{r}_{\text{lid}}$ , the Jacobian  $\mathbf{J}_{\text{lid}}$ , and the covariance matrix  $\Sigma_{\text{lid}}$  are defined as:  $\mathbf{r}_{\text{lid}} = \mathbf{XY}_{\text{lid}} - \mathbf{b}_{\text{lid}}$ ,  $\mathbf{J}_{\text{lid}} = [\mathbf{b}_{\text{lid}}^\odot \ \mathbf{0}_{6,6}]$ , and  $\Sigma_{\text{lid}} = \mathbf{X} \Sigma_{\text{lid}}(\mathbf{w}_t^{\text{lid}}) \mathbf{X}^\top$  where the definition of  $\mathbf{Y}_{\text{lid}}$  and  $\mathbf{b}_{\text{lid}}$  is the same as in Equation (19).

Similarly, the residual for GPS is obtained from (20), using the same approach as for the kinematic and LiDAR observations. The GPS residual  $\mathbf{r}_{\text{gps}}$ , its Jacobian  $\mathbf{J}_{\text{gps}}$ , and the corresponding covariance  $\Sigma_{\text{gps}}$  are defined as  $\mathbf{r}_{\text{gps}} = \mathbf{XY}_{\text{gps}} - \mathbf{b}_{\text{gps}}$ ,  $\mathbf{J}_{\text{gps}} = [\mathbf{b}_{\text{gps}}^\odot \ \mathbf{0}_{6,6}]$ , and  $\Sigma_{\text{gps}} = \mathbf{X} \Sigma_{\text{gps}}(\mathbf{w}_t^{\text{gps}}) \mathbf{X}^\top$  where  $\mathbf{Y}_{\text{gps}}$  and  $\mathbf{b}_{\text{gps}}$  are defined in Equation (20).

Fig. 1 provides a schematic overview of the two frameworks, illustrating their shared components and emphasizing the primary differences between them. The diagram highlights the common underlying structure while delineating where the two approaches are different. Fig. 1 also shows the legged robot-specific methods from [6], [12] that we incorporated to enhance the E-InEKF and E-IS. Specifically, a contact estimator based on the momentum observer of [30] determines the contact states  $\alpha_i$  for the  $i$ -th contact, while to handle dynamic events such as slipping, we adopt a slip rejection method that adjusts the covariance of contact-related observations [6].

## IV. EXPERIMENTAL RESULTS

In this section, we present the experimental results of the E-InEKF and E-IS algorithms, using the quadruped robot KAIST



Fig. 2: The indoor experiment involved large impacts and significant changes in the robot's height, as illustrated in the figures.

Hound2. These experiments were conducted in two distinct scenarios: indoor and outdoor environments, as shown in Fig. 2 and Fig. 4. All results were obtained from offline processing and analysis of the collected sensor data. Also, we underline that our state estimators are odometry systems, with no loop closures applied to reduce drift in the trajectory estimation.

For the indoor experiment, the ground truth pose was provided by a Vicon motion capture system. For the outdoor experiment, ground truth was obtained from a Holybro Real-Time Kinematic (RTK) GPS with a helical antenna.

As baselines, we adopted the proprioceptive-only invariant state estimation methods P-InEKF [8] and P-IS [12], as well as the state-of-the-art LiDAR-based methods LIO-SAM [1] and FAST-LIO2 [3]. The performance of the proposed algorithms was quantified using two widely recognized metrics: the mean Absolute Trajectory Error (ATE), which assesses global pose estimation accuracy, and the mean Relative Position Error (RPE), which evaluates local consistency in pose estimation.

#### A. Indoor Experiments

In the indoor experiment, the robot equipped with a Velodyne VLP16 LiDAR sensor performed walking tests in a controlled environment. Since the GPS is not informative in indoor settings, we evaluated the performance of E-IS and E-InEKF, which employ kinematics, IMU, and LiDAR data without GPS data. We refer to E-IS and E-InEKF as **E-IS (w/o GPS)** and **E-InEKF (w/o GPS)**, respectively, to indicate that they exclude the use of GPS in the indoor setting.

The indoor environment simulates real-world challenging terrain, as shown in Fig. 2. The testing area consists of wooden blocks, steps terrains, and slippery surfaces coated with boric acid on acrylic plates. While traversing these terrains, the robot experiences varying heights and high impacts during contact, which affects positional drift [18].

The estimation results of the indoor experiments, as shown in Fig. 3, clearly demonstrate the effectiveness of integrating LiDAR odometry into the invariant estimator through the proposed LiDAR observation model. This integration substantially reduced drift in the z-position, a critical improvement validated by the mean errors summarized in Table I. In contrast, the proprioceptive-only methods exhibit a higher susceptibility to positional drift. These findings underscore that, while the proprioceptive-only approaches reveal certain limitations, incorporating LiDAR data into the estimator significantly mitigates these issues, resulting in more robust and accurate state estimation in indoor environments.

Furthermore, both E-IS (w/o GPS) and E-InEKF (w/o GPS) recorded lower ATE and RPE than the LiDAR-based

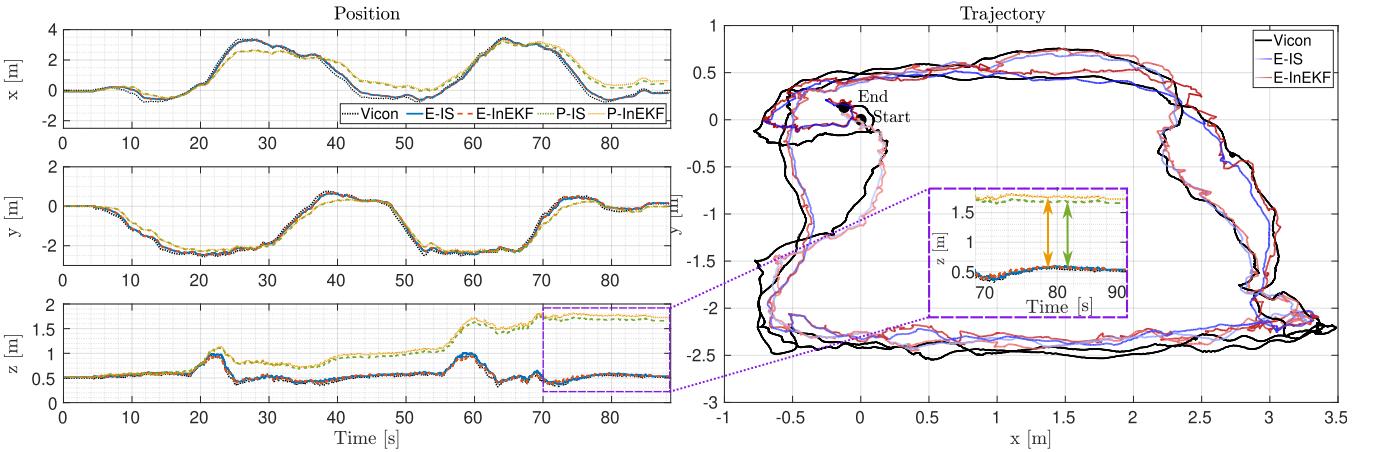


Fig. 3: The results for the indoor experiment, where the robot is prone to experience high contact impacts, show that the proposed frameworks, E-IS and E-InEKF, mitigate the position drift, especially in the z-axis, compared to the proprioceptive-only methods, P-IS and P-InEKF. In the purple dotted box, the position in the z-axis of P-IS and P-InEKF is indicated by the green and yellow arrows. Vicon is used as ground truth. In the right plot, the E-IS and E-InEKF trajectories use a color gradient: the lighter sections indicate the start of the estimates, while the darker sections indicate the end.

baselines. Since these methods fuse IMU and LiDAR data, without taking into account leg kinematics, their accuracy might be compromised by high-impact contacts. High-impact or unexpected motions, such as the slippage, caused larger errors in LIO-SAM and FAST-LIO2, and these errors would likely increase further under even more severe impacts. In contrast, our proposed method maintained consistently robust performance by leveraging the incorporation of leg kinematics as an additional measurement.

The variations in terrain heights and high-impact contacts were critical features of this experiment, emphasizing the importance of incorporating exteroceptive sensors, such as LiDAR, to address positional drift along the z-axis. Positional drift along this axis is inherently unobservable when relying solely on proprioceptive measurements, as previously analyzed in [4]. Our proposed frameworks successfully integrate LiDAR into IS and InEKF, demonstrating how external sensors mitigate drift in invariant estimators [8], [12] and improve pose estimation in challenging conditions.

### B. Outdoor Experiments

To assess the frameworks' ability to reduce long-term drift and maintain robustness in outdoor environments, the robot traversed a long-distance path of approximately 300 m, as shown in Fig. 4. The experiment was designed as a closed route, with the robot returning to its starting point. For the outdoor experiments, we used the KAIST HOUND2 robot equipped with a Velodyne VLP16 LiDAR sensor and a Holybro RTK GPS.

As shown in Fig. 5, both E-IS and E-InEKF achieved lower drift compared to P-InEKF and P-IS, especially in the z-axis, thanks to the incorporation of the exteroceptive measurements.

TABLE I: ATE and RPE statistics over 1 m

Indoor	E-InEKF (w/o GPS)	E-IS (w/o GPS)	LIO-SAM	FAST-LIO2	P-InEKF	P-IS
ATE [m]	0.24	<b>0.23</b>	0.32	0.30	0.90	0.85
RPE [m]	0.09	<b>0.08</b>	0.15	0.16	0.21	0.20

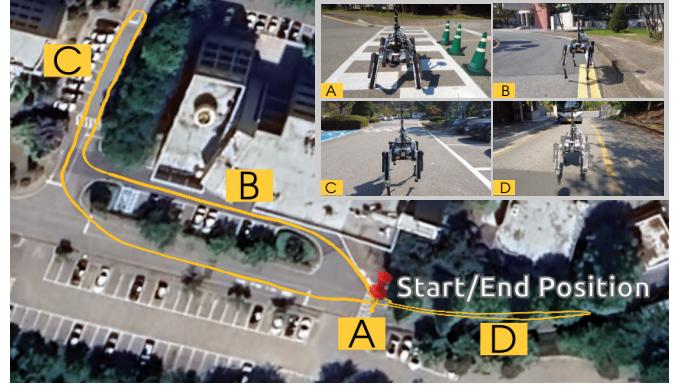


Fig. 4: The illustration of the outdoor experiment: Google Earth view with screenshots of the robot walking along the outdoor path, on the top right corner.

In Table II, we compared the ATE and RPE of our frameworks with those of the baselines (LIO-SAM and FAST-LIO2). We observed that the proposed frameworks, E-IS and E-InEKF, showed lower ATE and RPE compared to the baseline approaches.

Notably, E-InEKF (w/o GPS) and E-IS (w/o GPS) yielded lower RPE than FAST-LIO2 and LIO-SAM, while FAST-LIO2's ATE was only 2 cm lower than that of our proposed frameworks. Although incorporating leg kinematics can introduce potential noise that may slightly increase the drift over long trajectories, it enhances robustness, especially when LiDAR alone may be unreliable (e.g., environments with dense vegetation or high-impact scenarios), and it is still advantageous to include these measurements whenever possible.

Moreover, our group-affine observation model effectively integrates LiDAR and GPS into the invariant filter and smoother, ensuring robust long-term performance.

## V. DISCUSSION

In this section, we analyze the results presented in Sec. IV, including an ablation study and a discussion on the trade-off

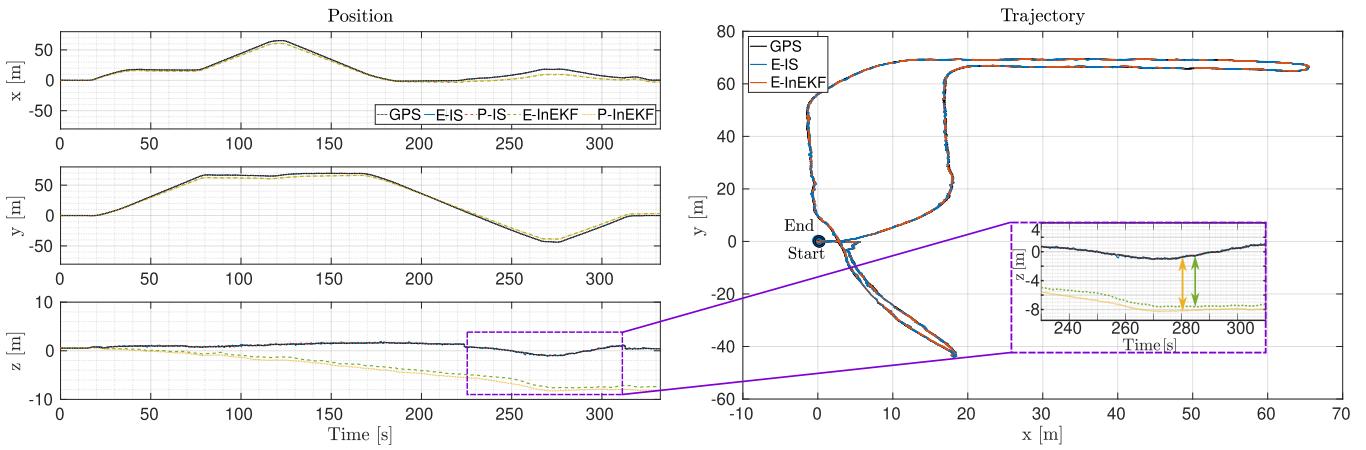


Fig. 5: The results for the outdoor experiment clearly show the improvement in the z-axis drift when using LiDAR and GPS, even in long-distance operations. The drift is indicated by the green and yellow arrows, for the P-InEKF and P-IS, respectively.

between E-IS and E-InEKF. The ablation study in Sec. V-A evaluates the contribution of each module in our approach, while the analysis in Sec. V-B provides valuable insights into the strengths and limitations of both methods.

#### A. Ablation Study

To discuss the effect of exteroceptive sensors, we compare three cases within our frameworks by selectively excluding exteroceptive sensors. Specifically, we discuss E-IS, E-IS (w/o GPS), and P-IS for invariant smoothing and E-InEKF, E-InEKF (w/o GPS), and P-InEKF for invariant filtering.

As shown in Table I and Table II, the experimental results demonstrated that the proposed methods, E-IS and E-InEKF, consistently outperformed LiDAR-based odometry. Both E-IS and E-InEKF effectively corrected positional drift along the z-axis in indoor and outdoor settings, highlighting their robustness in managing vertical displacement errors.

In the outdoor experiments, the reduced accuracy of the E-IS (w/o GPS) and E-InEKF (w/o GPS), compared to E-IS and E-InEKF, is primarily due to the use of GPS as ground truth for evaluation. However, even without GPS, E-IS (w/o GPS) and E-InEKF (w/o GPS) demonstrated better performance than LiDAR-based odometry in the indoor environment, where the robot could experience high-impact contacts, while they showed comparable performance to the baseline, achieving lower RPE in the outdoor experiment. These results highlight the robustness of the proposed frameworks, even in the absence of GPS data, and underscore their ability to outperform state-of-the-art LiDAR-based odometry methods under challenging conditions.

Also, with the exteroceptive sensors of LiDAR or GPS, the positional drift of P-IS and P-InEKF is significantly reduced. Additionally, as shown in Fig. 6, even if we incorporate the exteroceptive sensors, the computation time is not affected by

the low operation rates of LiDAR or GPS, which is attributed to the loosely-coupled manner this work adopted. Notably, since the LiDAR odometry is calculated on a separate thread, we then apply the proposed observation model for LiDAR odometry into E-IS and E-InEKF, while satisfying the group-affine property.

#### B. Comparison between filtering and smoothing

In all tests, the E-IS algorithm achieved better positional accuracy than E-InEKF by incorporating a smoothing window that exploits past states. However, E-InEKF's reduced computational cost makes it better suited for real-time applications. As shown in Table III and Fig. 6, we evaluated both algorithms under different window sizes (WS = 1, 5, 10, and 15) to assess the impact of the window length on accuracy and execution time. Table III shows how the ATE decreases with larger window sizes, while Fig. 6 illustrates the corresponding increase in computation time. In all experiments, E-IS achieved higher positional accuracy than E-InEKF by incorporating state history through smoothing, while E-InEKF showed instead to be more efficient in terms of computation time.

Increasing the window size of E-IS improves accuracy but also increases the computation time. For instance, with WS = 15, E-IS achieves the lowest ATE, but has an average computation time of 4.5 ms per iteration, whereas E-InEKF requires 0.06 ms per iteration. Reducing E-IS's window to WS = 1 lowers the execution time to 0.18 ms, with a marginal decrease in accuracy compared to E-InEKF.

Hence, the choice between the E-IS and E-InEKF depends on the application's requirements. If real-time performance is critical, E-InEKF is preferred. Conversely, tasks demanding higher pose accuracy, such as detailed mapping or high-fidelity localization, benefit from a larger smoothing window in E-IS despite its higher computational cost.

TABLE II: ATE and RPE statistics over 1 m

Outdoor	E-InEKF	E-IS	LIO-SAM	FAST-LIO2	E-InEKF (w/o GPS)	E-IS (w/o GPS)	P-InEKF	P-IS
ATE [m]	0.17	<b>0.15</b>	2.28	1.35	1.68	1.37	<b>6.57</b>	6.32
RPE [m]	0.07	<b>0.06</b>	0.20	0.10	0.09	0.08	0.10	0.09

## VI. CONCLUSION

In this work, we presented two state estimation frameworks for quadruped robots, E-IS and E-InEKF, that combine LiDAR and GPS with the proprioceptive-based InEKF [8] and IS [12]. To incorporate the LiDAR and GPS into the invariant state

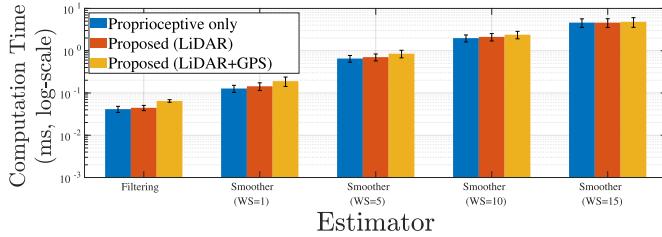


Fig. 6: Comparison of average computation times for E-IS, E-IS (w/o GPS), P-IS across different window sizes (WS) and their InEKF-based counterparts (E-InEKF, E-InEKF (w/o GPS), P-InEKF). The results show that while a larger window size increases the computation times of E-IS and E-IS (w/o GPS), they remain comparable to the proprioceptive-only cases due to the loosely coupled method used in this work.

estimators, we proposed an observation model for LiDAR and GPS that satisfies the group-affine property for the invariant estimators. Moreover, to handle the low frequency of LiDAR measurements, we employ a parallel thread to obtain the LiDAR odometry. To the best of our knowledge, this is the first approach integrating LiDAR and GPS into invariant estimators through a group-affine observation model and a parallel-thread LiDAR odometry module. Indoor and outdoor experiments on Hound2 [28] showed that our methods significantly reduce z-position drift and achieve lower RPE compared to existing proprioceptive and LiDAR-based methods [1], [3], [8], [12].

Future work will focus on integrating RGB-D or thermal cameras and comparing them with tightly-coupled methods to better understand the trade-offs in sensor fusion.

## REFERENCES

- [1] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, “LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping,” in *2020 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2020, pp. 5135–5142.
- [2] W. Xu and F. Zhang, “FAST-LIO: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter,” *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 3317–3324, 2021.
- [3] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, “FAST-LIO2: Fast direct LiDAR-inertial odometry,” *IEEE Trans. Robot.*, vol. 38, no. 4, pp. 2053–2073, 2022.
- [4] M. Bloesch, M. Hutter, M. A. Hoepflinger, S. Leutenegger, C. Gehring, C. D. Remy, and R. Siegwart, “State estimation for legged robots: consistent fusion of leg kinematics and IMU,” *Robotics*, vol. 17, pp. 17–24, Jul. 2013, DOI: 10.15607/RSS.2012.VIII.003.
- [5] M. Bloesch, C. Gehring, P. Fankhauser, M. Hutter, M. A. Hoepflinger, and R. Siegwart, “State estimation for legged robots on unstable and slippery terrain,” in *2013 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2013, pp. 6058–6064.
- [6] J.-H. Kim, S. Hong, G. Ji, S. Jeon, J. Hwangbo, J.-H. Oh, and H.-W. Park, “Legged robot state estimation with dynamic contact event information,” *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 6733–6740, 2021.
- [7] A. Barrau and S. Bonnabel, “The Invariant Extended Kalman Filter as a Stable Observer,” *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1797–1812, 2017.
- [8] R. Hartley, M. Ghaffari, R. M. Eustice, and J. W. Grizzle, “Contact-aided invariant extended Kalman filtering for robot state estimation,” *Int. J. Robot. Res.*, vol. 39, no. 4, pp. 402–430, 2020.
- [9] T.-Y. Lin, R. Zhang, J. Yu, and M. Ghaffari, “Legged robot state estimation using invariant Kalman filtering and learned contact events,” *arXiv preprint arXiv:2106.15713*, 2021.
- [10] H. M. S. Santana, J. C. V. Soares, Y. Nisticò, M. A. Meggiolaro, and C. Semini, “Proprioceptive state estimation for quadruped robots using invariant kalman filtering and scale-variant robust cost functions,” *2024 IEEE-RAS Int. Conf. Humanoid Robots*, pp. 213–220, 2024.
- [11] P. Chauchat, A. Barrau, and S. Bonnabel, “Invariant smoothing on Lie groups,” in *2018 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2018, pp. 1703–1710.
- [12] Z. Yoon, J.-H. Kim, and H.-W. Park, “Invariant smoother for legged robot state estimation with dynamic contact event information,” *IEEE Trans. Robot.*, vol. 40, pp. 193–212, 2024.
- [13] M. Camurri, M. Ramezani, S. Nobili, and M. Fallon, “Pronto: A multi-sensor state estimator for legged robots in real-world scenarios,” *Front. Robot. AI*, vol. 7, 2020, DOI: 10.3389/frobt.2020.00068.
- [14] Y. Nisticò, J. C. V. Soares, L. Amatucci, G. Fink, and C. Semini, “MUSE: A real-time multi-sensor state estimator for quadruped robots,” *IEEE Robot. Autom. Lett.*, pp. 1–8, 2025.
- [15] Y. Kim, B. Yu, E. M. Lee, J. Kim, H. Park, and H. Myung, “STEP: State estimator for legged robots using a preintegrated foot velocity factor,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4456–4463, 2022.
- [16] D. Wisth, M. Camurri, and M. Fallon, “VILENS: Visual, inertial, LiDAR, and leg odometry for all-terrain legged robots,” *IEEE Trans. Robot.*, 2022.
- [17] S. Yang, Z. Zhang, Z. Fu, and Z. Manchester, “Cerberus: Low-drift visual-inertial-leg odometry for agile locomotion,” in *2023 IEEE Int. Conf. Robot. Autom. (ICRA)*, 2023, pp. 4193–4199.
- [18] G. Ou, D. Li, and H. Li, “Leg-KILO: Robust kinematic-inertial-LiDAR odometry for dynamic legged robots,” *IEEE Robot. Autom. Lett.*, vol. 9, no. 10, pp. 8194–8201, 2024.
- [19] M. Zhang, M. Zhang, Y. Chen, and M. Li, “Imu data processing for inertial aided navigation: A recurrent neural network based approach,” in *2021 IEEE Int. Conf. Robot. Autom. (ICRA)*. IEEE, 2021, pp. 3992–3998.
- [20] G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. Van Sloun, and Y. C. Eldar, “Kalmannet: Neural network aided kalman filtering for partially known dynamics,” *IEEE Trans. Signal Process.*, vol. 70, pp. 1532–1547, 2022.
- [21] D. Youm, H. Oh, S. Choi, H. Kim, and J. Hwangbo, “Legged robot state estimation with invariant extended kalman filter using neural measurement network,” *arXiv e-prints*, pp. arXiv–2402, 2024.
- [22] A. Schperberg, Y. Tanaka, S. Mowlavi, F. Xu, B. Balaji, and D. Hong, “Optistate: State estimation of legged robots using gated networks with transformer-based vision and kalman filtering,” in *2024 IEEE Int. Conf. Robot. Autom. (ICRA)*. IEEE, 2024, pp. 6314–6320.
- [23] C. Yu, Y. Yang, T. Liu, Y. You, M. Zhou, and D. Xiang, “State estimation transformers for agile legged locomotion,” in *2024 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2024, pp. 6810–6817.
- [24] J. Kim, H. Kim, S. Jeong, Y. Shin, and Y. Cho, “Diter++: Diverse terrain and multi-modal dataset for multi-robot slam in multi-session environments,” *arXiv preprint arXiv:2412.05839*, 2024.
- [25] S. Teng, M. W. Mueller, and K. Sreenath, “Legged robot state estimation in slippery environments using invariant extended Kalman filter with velocity update,” in *2021 IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 3104–3110.
- [26] Y. Gao, C. Yuan, and Y. Gu, “Invariant filtering for legged humanoid locomotion on a dynamic rigid surface,” *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 4, pp. 1900–1909, 2022.
- [27] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniak, “KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way,” *IEEE Robot. Autom. Lett.*, vol. 8, no. 2, pp. 1029–1036, 2023.
- [28] Y.-H. Shin, S. Hong, S. Woo, J. Choe, H. Son, G. Kim, J.-H. Kim, K. Lee, J. Hwangbo, and H.-W. Park, “Design of KAIST HOUND, a quadruped robot platform for fast and efficient locomotion with mixed-integer nonlinear optimization of a gear train,” in *2022 Int. Conf. Robot. Autom. (ICRA) (ICRA)*, 2022, pp. 6614–6620.
- [29] J. Sola, J. Deray, and D. Atchuthan, “A micro Lie theory for state estimation in robotics,” *arXiv preprint arXiv:1812.01537*, 2018.
- [30] A. De Luca and R. Mattone, “Sensorless robot collision detection and hybrid force/motion control,” in *Proceedings of the 2005 IEEE Int. Conf. Robot. Autom. (ICRA)*. IEEE, 2005, pp. 999–1004.

TABLE III: ATE of the proposed E-InEKF and E-IS with different WS

	Proprioceptive only	Proposed (LiDAR)	Proposed (LiDAR + GPS)
Filter	6.340664 m	1.683859 m	0.171617 m
Smoother (WS:1)	6.331560 m	1.833524 m	0.156991 m
Smoother (WS:5)	6.322465 m	1.638729 m	0.150772 m
Smoother (WS:10)	6.316255 m	1.484747 m	0.150680 m
Smoother (WS:15)	<b>6.313716 m</b>	<b>1.376002 m</b>	<b>0.150467 m</b>