Article | Received 1 April 2026; Revised 18 May 2026; Accepted 15 July 2026; Published 10 August 2026
https://doi.org/10.55092/rl20260024

Smooth LiDAR–Inertial–Joint Odometry for perception-driven legged locomotion

Bingquan Li[1,2], Jia Pan[3] and Tianwei Zhang[1,4,*]

[1] The Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, China
[2] School of Computer Science, Guangdong University of Technology, Guangzhou, China
[3] School of Computing and Data Science, The University of Hong Kong, Hong Kong, China
[4] School of Science and Engineering, The Chinese University of Hong Kong-Shenzhen, Shenzhen, China

* Correspondence author; E-mail: zhangtianwei@cuhk.edu.cn.

Highlights:

- LIJO: a tightly coupled LiDAR–IMU–joint state estimator for accurate and robust quadruped odometry in complex outdoor environments.
- A joint-velocity factor with adaptive weighting to reduce trajectory jitter and improve low-speed motion consistency.
- Real-world deployment on a quadruped robot, with experiments showing improved accuracy and smoother trajectories over existing LIO methods.

Abstract: Light Detection and Ranging (LiDAR)–Inertial Odometry (LIO), which tightly fuses complementary data from LiDAR and Inertial Measurement Units (IMUs), is a key technology for high-precision state estimation in legged robot navigation. However, conventional Iterative Closest Point (ICP)-based LIO frameworks provide only pose constraints. Their position estimates often exhibit centimetre-level jitter due to LiDAR measurement noise, especially when the robot is stationary or moving slowly. This temporal inconsistency degrades the performance of downstream perception-driven motion planning and control. In this paper, we propose LiDAR–Inertial–Joint Odometry (LIJO), a novel state estimation framework for quadruped robots that integrates LiDAR, IMU, and joint encoder measurements within a manifold extended Kalman filter (EKF). The torso velocity is first estimated from joint angles and angular velocities via forward kinematics and is then used as a proprioceptive velocity factor in the EKF prediction step. To cope with foot slippage and aggressive motions, we further design a dynamic weighting scheme that adaptively adjusts the confidence of the joint-based velocity factor according to the current motion speed, while treating IMU measurements as filter observations rather than inputs. The resulting system maintains the localization accuracy of state-of-the-art LIO methods and significantly suppresses high-frequency jitter in the estimated trajectory. Extensive experiments on a real quadruped robot in challenging outdoor scenarios, including steep staircases and large-scale loop trajectories, show that LIJO produces smoother and more stable odometry than existing LIO and kinematic-inertial baselines, while preserving real-time performance. The proposed approach thus provides more reliable state inputs for perception and control modules in all-terrain legged locomotion.

Keywords: legged robots; LiDAR–Inertial Odometry; sensor fusion; state estimation; perception-driven locomotion

1. Introduction

Quadruped robots have achieved remarkable progress in recent years and are increasingly deployed in industrial inspection, disaster response, and other missions that require agile mobility in unstructured environments [1,2]. As the core enabler of autonomous navigation on such platforms, real-time state estimation directly determines the reliability of environmental perception and strongly influences the performance of downstream motion planning and control [2–4]. Among existing solutions, Light Detection and Ranging (LiDAR)–Inertial Odometry (LIO) has become a cornerstone for mobile robot navigation thanks to its ability to fuse dense LiDAR point clouds with high-frequency micro-electromechanical systems (MEMS) Inertial Measurement Unit (IMU) data in a temporally and spatially complementary manner [5–9].

However, when LIO is applied to legged robots, several challenges emerge. Existing legged odometry methods mainly focus on improving localisation accuracy through contact-aware kinematic constraints. However, in tightly coupled LIO systems for quadruped robots, trajectory smoothness degradation caused by high-frequency velocity fluctuations remains insufficiently studied, particularly during low-speed or quasi-static locomotion [10,11]. At the same time, the centimetre-level measurement accuracy of typical LiDAR sensors causes noticeable position jitter, particularly when the robot is stationary or moving slowly. Although such jitter may appear small in absolute value, it violates the physical expectation that legged robot motion should be continuous and smooth [12] and can severely degrade the stability of perception-driven locomotion controllers that rely on high-bandwidth state feedback.

In parallel, proprioceptive state estimators that fuse joint encoders and IMUs via Kalman filtering or factor-graph optimisation have demonstrated centimetre-level local accuracy for legged systems [13–16]. Under aggressive legged motions and impact vibrations, IMU errors can propagate through the pre-integration process and introduce significant velocity oscillations into the estimated trajectory. Existing multi-sensor fusion frameworks that combine LiDAR, IMU, and leg kinematics largely focus on improving robustness and accuracy, but they seldom explicitly address the low-jitter requirement of legged robot odometry in the context of perception-driven motion [1,2,17,18].

In this work, as shown in Figure 1, we address the above gap by introducing LiDAR–Inertial–Joint Odometry (LIJO), a tightly coupled state estimation framework that explicitly targets low-jitter odometry for quadruped robots. Through adaptive multi-sensor fusion, the proposed system maintains stable and smooth state estimation even under challenging operating conditions, including impact vibrations, low-speed locomotion, and foot slippage. The key idea is to leverage joint-based body velocity estimates as a smooth proprioceptive constraint within a manifold extended Kalman filter (EKF), while retaining the global consistency and accuracy provided by LiDAR–IMU fusion [14,19].

Our main contributions are summarised as follows:

- We propose LIJO, a tightly coupled LiDAR–IMU–Joint state estimation framework based on a manifold EKF, which delivers accurate, robust, and smooth odometry for quadruped robots operating in complex outdoor environments.
- A joint-velocity factor and an adaptive weighting mechanism are introduced to suppress high-frequency trajectory jitter and improve motion consistency during low-speed locomotion.
- The proposed LIJO system is fully implemented and deployed on a real quadruped robot platform. Extensive real-world experiments in challenging outdoor environments demonstrate that LIJO outperforms existing LIO methods in both estimation accuracy and trajectory smoothness.

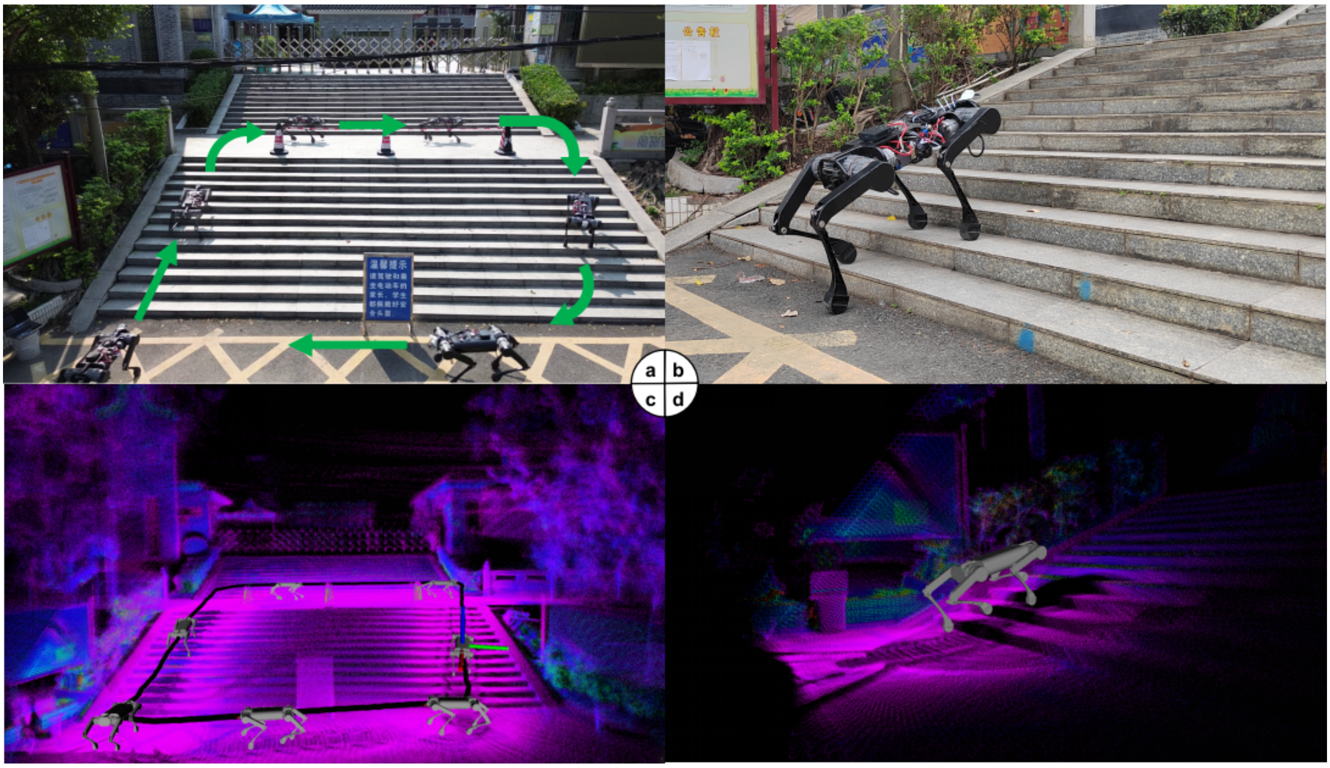

Figure 1. The proposed LIJO method tightly couples proprioceptive information from the robot with LiDAR observations, introducing additional constraints into the state estimation process and thereby improving the stability and continuity of the localisation results. This method provides relatively smooth pose estimates for the perception and control modules, enabling the quadruped robot to exhibit improved adaptability when traversing non-planar terrain.

2. Related work

With the increasing deployment of legged robots in complex, unstructured environments, research on state estimation and perception is a key enabler in robotics. High-precision, robust state estimation is essential to ensure stable locomotion and reliable environment perception on challenging terrains.

2.1. Proprioceptive state estimation

In the field of proprioceptive state estimation, the stochastic filtering framework proposed by Bloesch *et al.* [14] is widely regarded as a foundational contribution. By incorporating leg kinematic constraints and modelling slip effects during contact as Gaussian noise, their method provides a principled basis for fusing IMU and joint encoder measurements on legged robots. Building on this line of work, Bledt *et al.* [16] employed a linear Kalman filter (KF) for state estimation and augmented the filter with observations of foot contact height in the global frame. However, because proprioceptive sensors cannot directly measure contact height, this approach requires extensive manual parameter tuning, which limits its practicality in diverse terrain conditions.

Wisth *et al.* [15] proposed a factor-graph-based framework that introduces pre-integrated velocity factors driven by leg odometry and jointly estimates associated biases, significantly improving odometry accuracy on both rigid and deformable surfaces. Yang *et al.* [13] further advanced proprioceptive odometry by exploiting multiple IMUs and leg joint encoders, demonstrating that multi-IMU fusion can enhance the robustness and accuracy of KF-based legged state estimation. Nevertheless, purely kinematic–inertial estimators remain vulnerable to foot slippage and contact model errors, and tend to accumulate drift over long trajectories.

2.2. LiDAR-based state estimation

In LiDAR-based odometry and Simultaneous Localization and Mapping (SLAM), Zhang and Singh introduced LiDAR Odometry and Mapping (LOAM) [5], which jointly optimises LiDAR odometry and mapping and has become a classical baseline. Shan *et al.* later proposed LiDAR Inertial Odometry via Smoothing and Mapping (LIO-SAM) [20], a LiDAR–inertial factor-graph optimisation framework that significantly improves accuracy and robustness compared to pure LiDAR odometry. Xu *et al.* presented FAST-LIO and FAST-LIO2 [7,21], which are renowned for their low-latency performance by tightly coupling LiDAR and IMU in an iterated Kalman filter formulation.

When deployed on legged robots, however, these methods face additional challenges. High-frequency impacts between the feet and the ground induce severe disturbances in accelerometer measurements, often leading to drift along the vertical axis. To mitigate this problem, POINT-LIO [8] treats IMU data as measurements rather than system inputs, thereby reducing the dependence on IMU integration. Experiments have shown that POINT-LIO maintains high robustness even under aggressive motions. Leg-Kinematic-Inertial LiDAR Odometry (KILO) [10] further reduces vertical drift by incorporating contact height detection constraints, while LiDAR–Inertial–Wheel Odometry (LIWO) [12] integrates wheel encoder speed with LiDAR and IMU data in a bundle-adjustment framework to enhance overall state estimation performance.

2.3. Vision and multi-sensor fusion for legged robots

Vision sensors have also been widely adopted for state estimation due to their low cost and rich environmental information. Classical visual and visual–inertial odometry systems such as Oriented FAST and Rotated BRIEF Simultaneous Localization and Mapping (ORB-SLAM) [22] and Vins-Mono [23] have achieved impressive performance in camera-centric SLAM. In recent years, multi-sensor fusion methods have become an active research area. Other works based on factor-graph formulations [2,24] have demonstrated that incorporating visual information can substantially outperform purely kinematic–inertial approaches.

The State Estimator for Legged Robots Using a Preintegrated Foot Velocity Factor (STEP) framework [17] introduces pre-integrated foot velocity factors into Visual–Inertial Navigation System Fusion (VINS-Fusion) [18], achieving high state estimation accuracy without relying on a strict no-slip assumption. Visual-Inertial-Legged Navigation System (VILENS) [1] proposes a comprehensive multi-sensor architecture that tightly couples leg kinematics, IMU, LiDAR, and vision, and achieves robust performance even under single-sensor degradation. However, its high computational complexity limits the online update rate to only a few hertz.

The LIJO method proposed in this paper is inspired by the point-wise matching strategy of POINT-LIO [8] and explicitly incorporates joint kinematic information into a manifold EKF framework to reduce state-estimation noise for legged robots.

3. Methods

In this section, we present the LIJO framework for quadruped locomotion. The system fuses LiDAR, IMU, and joint encoder measurements within a unified manifold EKF to provide low-jitter, drift-resilient state estimates for perception-driven controllers.

As shown in Figure 2, joint encoder and IMU data are processed through a kinematic model to produce smooth body-velocity estimates that drive state propagation, while LiDAR point clouds are registered to an incremental voxel map using a point-to-plane formulation, and the resulting residuals update the state. A speed-aware dynamic weighting scheme modulates the relative influence of proprioceptive and LiDAR cues, allowing the estimator to adapt across motion regimes from quasi-static contact to aggressive trot.

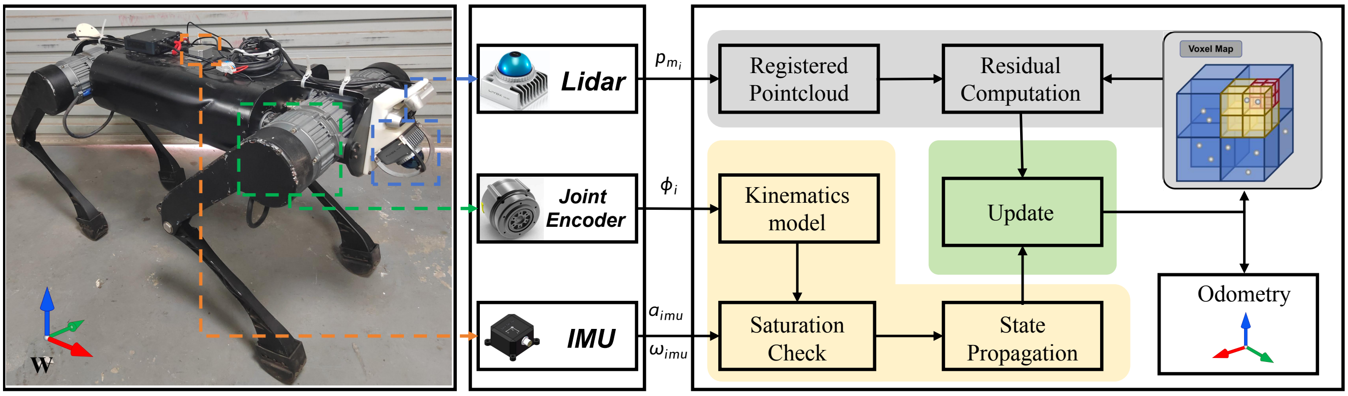

Figure 2. Overview of the state estimation framework, where joint encoders and IMU provide the prediction through forward kinematics, and LiDAR updates the EKF via voxel-based point-to-plane matching.

3.1. Notations and state representation

We define several reference frames commonly used in legged robotics and multi-sensor fusion. The fixed world frame is denoted by W, the robot body (center-of-mass) frame by B, the IMU frame by I, and the LiDAR frame by L. The transformations between B, I, and L are assumed to be time-invariant (see Figure 3) and are calibrated offline.

To avoid singularities associated with Euler angles and to remain consistent with modern visual–inertial and LiDAR–inertial odometry formulations, we adopt a manifold-based state representation.

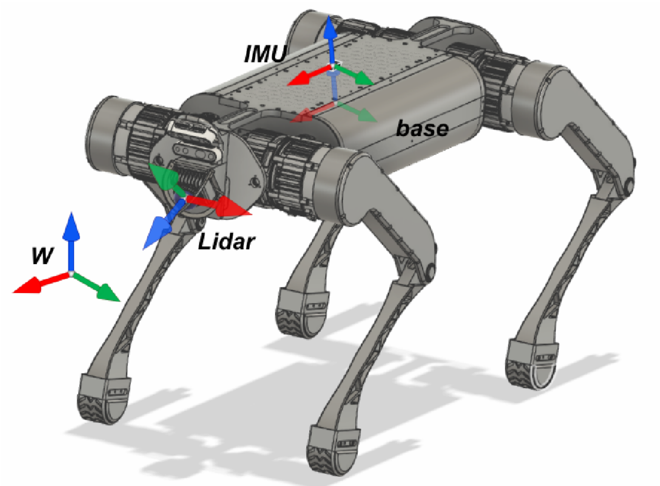

Figure 3. Illustration of the installation layouts and coordinate frame definitions of the IMU, LiDAR, and the robot center of mass on the quadruped robot.

Specifically, orientations lie on the Lie group $\mathrm{SO}(3)$ and positions, velocities, and biases lie in $\mathbb{R}^n$. We use the standard $\oplus$ and $\ominus$ operators to perform perturbation-based updates on the manifold:

$$\oplus : \mathscr{M} \times \mathbb{R}^n \to \mathscr{M}, \quad \ominus : \mathscr{M} \times \mathscr{M} \to \mathbb{R}^n.$$

For $\mathrm{SO}(3)$, the operators are defined via the exponential and logarithmic maps,

$$R \oplus \mathbf{r} = R \cdot \exp(\mathbf{r}), \quad R_1 \ominus R_2 = \log(R_2^\top R_1),$$

$$\exp(\mathbf{r}) = \mathbf{I} + \frac{\sin(\|\mathbf{r}\|)}{\|\mathbf{r}\|}[\mathbf{r}]_\times + \frac{1-\cos(\|\mathbf{r}\|)}{\|\mathbf{r}\|^2}[\mathbf{r}]_\times^2 \tag{1}$$

where $\mathbf{r} \in \mathbb{R}^3$ is a minimal rotation perturbation. For Euclidean components, the operators reduce to simple vector addition and subtraction.

$$\mathbf{a} \oplus \mathbf{b} = \mathbf{a} + \mathbf{b}, \quad \mathbf{a} \ominus \mathbf{b} = \mathbf{a} - \mathbf{b} \tag{2}$$

For a compound manifold $\mathscr{M} = \mathrm{SO}(3) \times \mathbb{R}^n$, the $\oplus$ and $\ominus$ operations are applied component-wise.

$$\begin{bmatrix} \mathbf{a} \end{bmatrix} \oplus \begin{bmatrix} \mathbf{r} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{R} \oplus \mathbf{r} \\ \mathbf{a} + \mathbf{b} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{a} \end{bmatrix} \ominus \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 \ominus \mathbf{R}_2 \\ \mathbf{a} - \mathbf{b} \end{bmatrix} \tag{3}$$

The filter state stacks the robot pose, velocity, and IMU biases in a standard legged–SLAM form

$$x = \left[R_W^I, \mathbf{p}_W^I, \mathbf{v}_W^I, \mathbf{b}_a, \mathbf{b}_\omega\right],$$

where $R_W^I \in \mathrm{SO}(3)$ is the rotation from I to W, $\mathbf{p}_W^I$ and $\mathbf{v}_W^I$ are the IMU position and velocity in W, and $\mathbf{b}_a$, $\mathbf{b}_\omega$ are the accelerometer and gyroscope biases. This representation is compatible with many existing multi-sensor fusion pipelines and facilitates later integration with visual or other exteroceptive cues.

3.2. Joint-based velocity factor for locomotion

Legged locomotion provides rich proprioceptive information through joint encoders. We exploit this signal to construct a joint-based body velocity factor that plays a central role in smoothing the estimated trajectory.

Let $\boldsymbol{\theta}$ and $\dot{\boldsymbol{\theta}}$ denote the joint angles and joint angular velocities obtained from motor encoders. The foot position in the body frame, $\mathbf{p}_B^{\text{foot}}$, is obtained via forward kinematics

$$\mathbf{p}_B^{\text{foot}} = g(\boldsymbol{\theta}, \boldsymbol{\rho}),$$

where $\boldsymbol{\rho}$ denotes the robot's kinematic parameters. Differentiating with respect to $\boldsymbol{\theta}$ yields the Jacobian $\mathbf{J}_g(\boldsymbol{\theta}, \boldsymbol{\rho})$, which maps joint velocities to foot velocities in B:

$$\mathbf{v}_B^{\text{foot}} = \mathbf{J}_g(\boldsymbol{\theta}, \boldsymbol{\rho})\,\dot{\boldsymbol{\theta}}.$$

Assuming the IMU is rigidly attached to the body, with known transformation R_B^I and lever arm, the foot position and velocity in the world frame can be expressed as

$$\mathbf{p}_W^{\text{foot}} = \mathbf{p}_W^B + R_W^B\,\mathbf{p}_B^{\text{foot}},$$
$$\mathbf{v}_W^{\text{foot}} = \mathbf{v}_W^B + R_W^B\left(R_B^I\,\boldsymbol{\omega}_I \times \mathbf{p}_B^{\text{foot}} + \mathbf{v}_B^{\text{foot}}\right),$$

where $\boldsymbol{\omega}_I$ denotes the IMU angular velocity.

Under a no-slip contact assumption for a supporting foot, the foot velocity in the world frame should be zero, $\mathbf{v}_W^{\text{foot}} \approx \mathbf{0}$. Rearranging, we obtain an expression for the center-of-mass (CoM) velocity:

$$\mathbf{v}_W^B = -R_W^B\left(R_B^I\,\boldsymbol{\omega}_I \times \mathbf{p}_B^{\text{foot}} + \mathbf{v}_B^{\text{foot}}\right) + \mathbf{v}_W^{\text{bias}},$$

where $\mathbf{v}_W^{\text{bias}}$ accounts for residual slip and foot rotation effects. In practice, as shown in Figure 4, the rotation of a finite-sized foot introduces additional velocity components with respect to the ground, and we model this as a bias term that scales with the foot diameter and angular velocity.

$$^{\mathscr{W}}\mathbf{v}_{bias} = \tfrac{d}{2} \cdot {}^{\mathscr{W}}\boldsymbol{\omega}_{foot} \tag{4}$$

$$\begin{gathered} ^{\mathscr{W}}\mathbf{v}_{base} = \\ -{}^{\mathscr{W}}\mathbf{R}_{base}({}^{\mathscr{I}}\mathbf{R}_{base}\lfloor {}^{\mathscr{I}}\boldsymbol{\omega} \rfloor {}^{\mathscr{B}}\mathbf{p}_{foot} + {}^{\mathscr{B}}\mathbf{v}_{foot}) + {}^{\mathscr{W}}\mathbf{v}_{bias} \end{gathered} \tag{5}$$

This joint-based CoM velocity $\mathbf{v}_W^B$ is then transformed to an IMU-frame velocity $\mathbf{v}_W^I$ via the known rigid transform between B and I.

$$^{\mathscr{W}}\mathbf{v}_{base} = {}^{\mathscr{W}}\mathbf{v}_{base} + {}^{\mathscr{I}}\mathbf{R}_{base}\lfloor {}^{\mathscr{I}}\boldsymbol{\omega} \rfloor {}^{\mathscr{B}}\mathbf{p}_{imu} \tag{6}$$

The resulting quantity is treated as a proprioceptive velocity "measurement" that is temporally dense and notably smooth compared to LiDAR-derived velocities. Within the EKF, this measurement acts as a prediction factor that drives state propagation between LiDAR updates. However, foot slippage and kinematic inconsistencies are more likely during aggressive motions, which motivates the dynamic weighting strategy described in the next subsection.

Specifically, during low-speed or quasi-static motions, proprioceptive velocity measurements derived from joint kinematics are considered more reliable and are therefore assigned higher confidence. In contrast, during high-speed motions, where foot slippage and modeling errors may become more significant, the confidence of the joint-velocity factor is gradually reduced, and the estimator relies more on LiDAR-inertial measurements. Let $\|\mathbf{v}_W^B\|$ denote the current body speed. We define a speed-dependent scalar α as

$$\alpha = \frac{1}{1+\exp\left(-\beta(\|\mathbf{v}_W^B\|-v_{\mathrm{th}})\right)},$$

where v_{th} is a velocity threshold and β controls the slope of the transition. A larger β results in a faster change of the weighting coefficient around the threshold velocity. This design prevents over-confident proprioceptive constraints during highly dynamic motions while preserving the smoothness benefits of joint-velocity estimation in low-speed scenarios.

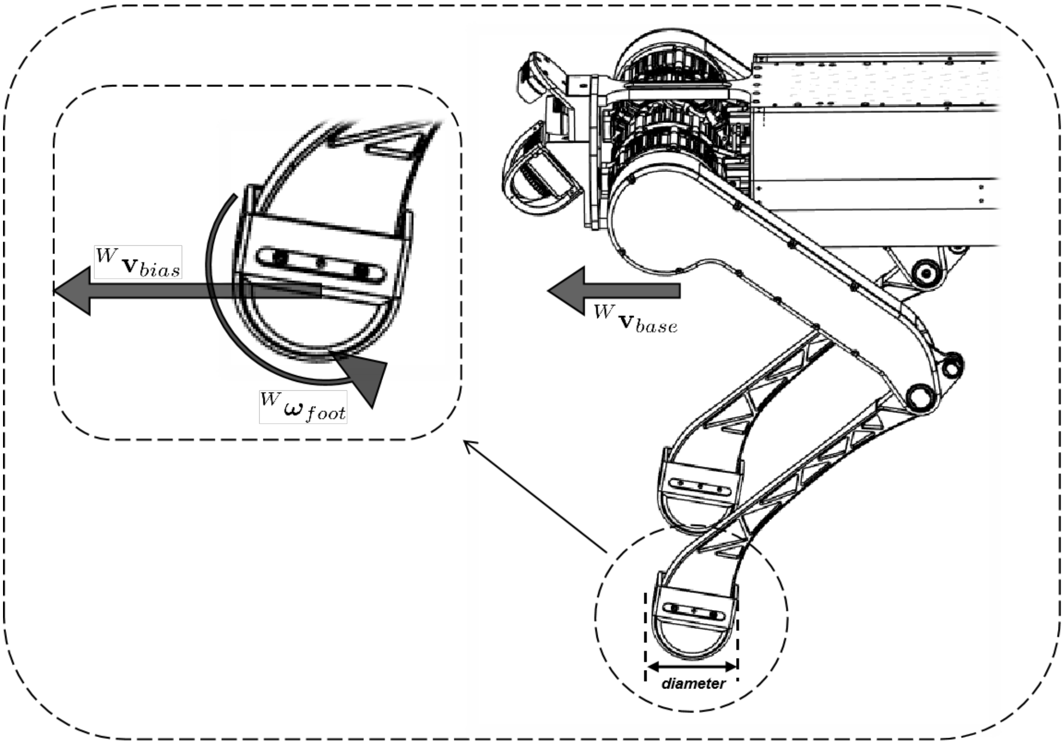

Figure 4. Velocity bias induced by foot rotation.

3.3. State prediction with multi-sensor dynamics

The prediction step of the EKF leverages IMU and joint encoder information to propagate the state between asynchronous LiDAR scans. Given the current state estimate $\bar{x}_k$ and time interval Δt_k, the nominal state at time $k+1$ is propagated as

$$\hat{x}_{k+1} = \bar{x}_k \oplus \left(\Delta t_k\, f(\bar{x}_k, \mathbf{u}_k)\right),$$

where $f(\cdot)$ denotes the continuous-time system dynamics, and $\mathbf{u}_k$ collects the IMU measurements and joint-based velocity factor. Unlike classical LIO formulations, which use only IMU inputs, here the joint-derived velocity provides an additional proprioceptive prediction term tailored to legged locomotion.

The associated covariance is propagated using the standard linearised error-state formulation,

$$\hat{P}_{k+1} = F_{x_k}\bar{P}_k F_{x_k}^{\top} + F_{w_k} Q_k F_{w_k}^{\top},$$

where F_{x_k} and F_{w_k} are the Jacobians of the state error with respect to the previous state and process noise, and Q_k is the process noise covariance. The structure of F_{x_k} and F_{w_k} reflects the coupling between orientation, velocity, and biases typical in multi-sensor legged state estimation. Their explicit forms are given in Equations (7) and (8).

$$\mathbf{F}_{\mathbf{x}_k} = \frac{\partial(\mathbf{x}_{k+1} \ominus \hat{\mathbf{x}}_{k+1})}{\partial \delta \mathbf{x}_k}\bigg|_{\delta \mathbf{x}_k=\mathbf{0}, \mathbf{w}_k=\mathbf{0}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{I}\Delta t_k & \mathbf{0} & \mathbf{0}_{3\times 12} \\ \mathbf{0} & \mathbf{F}_{11} & \mathbf{0} & \mathbf{I}\Delta t_k & \mathbf{0}_{3\times 12} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0}_{3\times 12} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0}_{3\times 12} \\ \mathbf{0}_{12\times 3} & \mathbf{0}_{12\times 3} & \mathbf{0}_{12\times 3} & \mathbf{0}_{12\times 3} & \mathbf{I}_{12\times 12} \end{bmatrix} \tag{7}$$

$$\mathbf{F}_{\mathbf{w}_k} = \frac{\partial(\mathbf{x}_{k+1} \ominus \hat{\mathbf{x}}_{k+1})}{\partial \mathbf{w}_k}\bigg|_{\delta \mathbf{x}_k=\mathbf{0}, \mathbf{w}_k=\mathbf{0}} = \begin{bmatrix} \mathbf{0}_{6\times 9} & \mathbf{0}_{6\times 6} \\ \mathbf{I}_{9\times 9} & \mathbf{0}_{6\times 6} \\ \mathbf{0}_{3\times 9} & \mathbf{0}_{3\times 6} \\ \mathbf{0}_{6\times 9} & \mathbf{I}_{6\times 6} \end{bmatrix} \tag{8}$$

where $\mathbf{F}_{11} = \exp(-\omega_k \Delta t_k)$.

3.4. LiDAR and IMU measurement models

To robustly fuse exteroceptive geometry with inertial and proprioceptive cues, LIJO employs separate measurement models for LiDAR and IMU due to their asynchronous sampling characteristics.

(a) LiDAR point-to-voxel measurements: Following a point-wise registration strategy, LiDAR points are processed at their individual acquisition times to avoid motion distortion. Let $T_I^L = (R_I^L, \mathbf{p}_I^L)$ denote the extrinsic calibration between the IMU and LiDAR frames. A LiDAR measurement point in the IMU frame is modelled as

$$^{\mathscr{I}}\mathbf{p}_{m_k} = {}^{\mathscr{I}}\mathbf{R}_L {}^{\mathscr{L}}\mathbf{p}_{m_k} + {}^{\mathscr{I}}\mathbf{p}_L + \mathbf{n}_{L_k} \tag{9}$$

where $\mathbf{p}_L^m$ is the raw LiDAR point and $\mathbf{n}_L$ is additive Gaussian noise. Given the predicted pose T_W^I at the corresponding time, the point is projected into the world frame

$$^{\mathscr{W}}\hat{\mathbf{p}}_{k+1} = {}^{\mathscr{W}}\hat{\mathbf{R}}_{I_{k+1}} {}^{\mathscr{I}}\mathbf{p}_{m_{k+1}} + {}^{\mathscr{W}}\hat{\mathbf{p}}_{I_{k+1}} \tag{10}$$

and associated with a local planar patch in an incremental voxel map [9]:

$$\begin{aligned} \mathbf{0} &= \mathbf{h}_L(\mathbf{x}_k, {}^{\mathscr{I}}\mathbf{p}_{m_k}, \mathbf{n}_{L_k}) \\ &= {}^{\mathscr{W}}\mathbf{u}_k^T \left({}^{\mathscr{W}}\mathbf{T}_{Ik}\left({}^{\mathscr{I}}\mathbf{p}_{m_k} - \mathbf{n}_{L_k}\right) - {}^{\mathscr{W}}\mathbf{q}_k\right) \end{aligned} \tag{11}$$

The point-to-plane residual takes the form

$$r_L = \mathbf{n}^\top \left(T_W^I \mathbf{p}_I^m - \mathbf{q}\right),$$

where $\mathbf{n}$ is the local plane normal and $\mathbf{q}$ is a point on the plane. Only points with well-conditioned local planar support are used to update the state; others are inserted into the map without contributing residuals.

For each LiDAR point, search for its 5 nearest neighboring points in the Voxel map. Then, use these nearest neighbors to fit a local planar patch in the measurement model, which has a normal vector ${}^{\mathscr{W}}\mathbf{u}_{k+1}$ and a centroid ${}^{\mathscr{W}}\hat{\mathbf{p}}_{I_{k+1}}$ (see Figure 5). If the 5 nearest points do not lie on the fitted planar patch (*i.e.*, the distance from any point to the plane exceeds a threshold), the LiDAR point ${}^{\mathscr{W}}\hat{\mathbf{p}}_{k+1}$ is merged into the Voxel map without residual computation or state update. Otherwise, Linearising this residual around the current estimate yields

$$
\begin{aligned}
\mathbf{r}_{\mathbf{L_{k+1}}} &= \mathbf{0} - \mathbf{h}_L(\hat{\mathbf{x}}_{k+1}, {}^{\mathscr{I}}\mathbf{p}_{m_{k+1}}, \mathbf{0}) \\
&= \mathbf{h}_L(\mathbf{x}_{k+1}, {}^{\mathscr{I}}\mathbf{p}_{m_{k+1}}, \mathbf{n}_{L_{k+1}}) - \mathbf{h}_L(\hat{\mathbf{x}}_{k+1}, {}^{\mathscr{I}}\mathbf{p}_{m_{k+1}}, \mathbf{0}) \\
&\approx \mathbf{H}_{L_{k+1}} \delta\mathbf{x}_{k+1} + \mathbf{D}_{L_{k+1}} \mathbf{n}_{L_{k+1}}
\end{aligned}
\tag{12}
$$

Here, $\delta\mathbf{x}_{k+1} = \mathbf{x}_{k+1} \ominus \hat{\mathbf{x}}_{k+1}$ represents the state error. $\mathbf{H}_{L_{k+1}}$ and $\mathbf{D}_{L_{k+1}}$ are the Jacobian matrices of the residual with respect to the state error $\delta\mathbf{x}_k$ and the measurement noise $\mathbf{n}_{L_k}$, respectively. The specific forms of $\mathbf{H}_{L_{k+1}}$ and $\mathbf{D}_{L_{k+1}}$ are given below:

$$
\begin{aligned}
\mathbf{H}_{L_{k+1}} &= \left.\frac{\partial \mathbf{h}_L(\hat{\mathbf{x}}_{k+1} \oplus \delta\mathbf{x}, {}^{\mathscr{I}}\mathbf{p}_{m_{k+1}}, \mathbf{0})}{\partial \delta\mathbf{x}_k}\right|_{\delta\mathbf{x}=\mathbf{0}} \\
&= \begin{bmatrix} {}^{\mathscr{W}}\mathbf{u}_{k+1}^T & -{}^{\mathscr{W}}\mathbf{u}_{k+1}^T {}^{\mathscr{W}}\hat{\mathbf{R}}_{I_{k+1}} \lfloor {}^{\mathscr{I}}\mathbf{p}_{m_{k+1}} \rfloor & \mathbf{0}_{1\times 18} \end{bmatrix}
\end{aligned}
\tag{13}
$$

$$
\mathbf{D}_{L_{k+1}} = \left.\frac{\partial \mathbf{h}_L(\hat{\mathbf{x}}_{k+1}, {}^{\mathscr{I}}\mathbf{p}_{m_{k+1}}, \mathbf{n})}{\partial \mathbf{n}}\right|_{\mathbf{n}=\mathbf{0}} = -{}^{\mathscr{W}}\mathbf{u}_{k+1}^T {}^{\mathscr{W}}\hat{\mathbf{R}}_{I_{k+1}}
\tag{14}
$$

$$
r_L \approx H_L \delta x + D_L \mathbf{n}_L,
$$

with Jacobians H_L and D_L computed analytically from the geometry. This formulation is compatible with incremental voxel-based mapping and maintains high update rates suitable for locomotion.

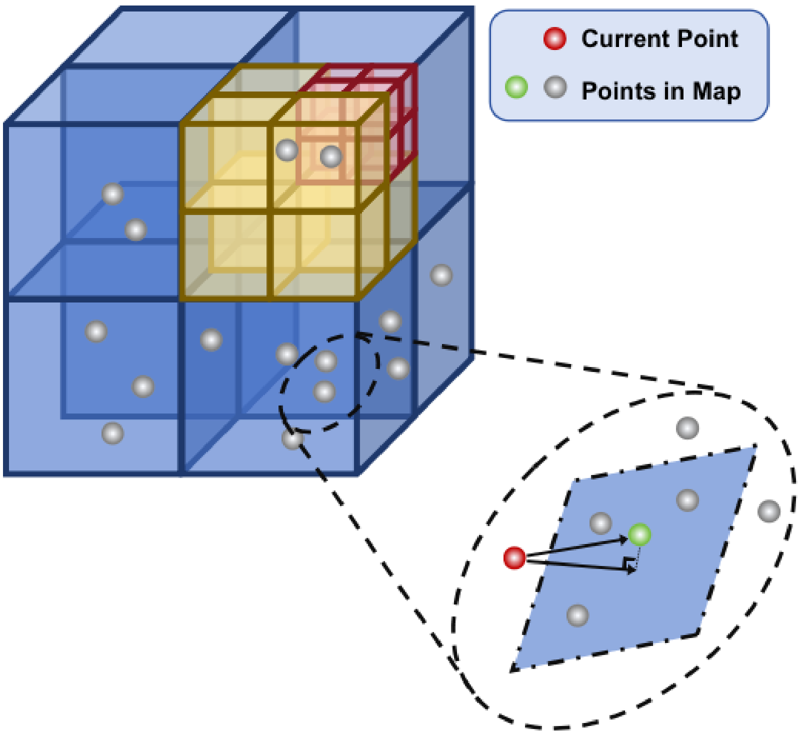

Figure 5. Illustration of direct registration of a LiDAR point to a voxel map.

(b) IMU measurements as observations: Inspired by POINT-LIO-style formulations for robust LiDAR–inertial fusion, we treat IMU measurements as observations rather than direct inputs in the process model. The IMU outputs stacked accelerometer and gyroscope readings,

$$\begin{bmatrix} \mathbf{a}_m \\ \omega_m \end{bmatrix} = \begin{bmatrix} \mathbf{a} + \mathbf{b}_a + \mathbf{n}_a \\ \omega + \mathbf{b}_\omega + \mathbf{n}_\omega \end{bmatrix},$$

where $\mathbf{n}_a$ and $\mathbf{n}_\omega$ denote measurement noise. In the presence of strong mechanical impacts during legged locomotion, accelerometer readings can saturate or exhibit large spikes. We detect such events via a saturation check on $\mathbf{a}_m$ and increase the corresponding measurement covariance to down-weight unreliable IMU data.

The linearised IMU residual is written as

$$\begin{aligned} \mathbf{r}_{\mathbf{I_{k+1}}} &= \mathbf{h}_I(\mathbf{x}_{k+1}, \mathbf{n}_{I_{k+1}}) - \mathbf{h}_I(\hat{\mathbf{x}}_{k+1}, \mathbf{0}) \\ &\approx \mathbf{H}_{I_{k+1}} \delta\mathbf{x}_{k+1} + \mathbf{D}_{I_{k+1}} \mathbf{n}_{I_{k+1}} \end{aligned} \tag{15}$$

where H_I and D_I are the Jacobians with respect to the state error and IMU noise, respectively. The specific forms of $\mathbf{H}_{I_{k+1}}$ and $\mathbf{D}_{I_{k+1}}$ are given below:

$$\begin{aligned} \mathbf{H}_{I_{k+1}} &= \frac{\partial \mathbf{h}_I(\hat{\mathbf{x}}_{k+1} \oplus \delta\mathbf{x}, \mathbf{0})}{\partial \delta\mathbf{x}_k}\bigg|_{\delta\mathbf{x}=\mathbf{0}} \\ &= \begin{bmatrix} \mathbf{0}_{6\times 9} & \mathbf{I}_{6\times 6} & \mathbf{0}_{6\times 3} & \mathbf{I}_{6\times 6} \end{bmatrix} \end{aligned} \tag{16}$$

$$\mathbf{D}_{I_{k+1}} = \frac{\partial \mathbf{h}_I(\hat{\mathbf{x}}_{k+1}, \mathbf{n})}{\partial \mathbf{n}}\bigg|_{\mathbf{n}=\mathbf{0}} = \mathbf{I}_{6\times 6} \tag{17}$$

This measurement model allows LIJO to exploit the high-bandwidth nature of IMU data while explicitly modelling their degradation under impact, which is common in dynamic quadruped locomotion.

3.5. Multi-sensor fusion update

Combining the prediction prior with LiDAR and IMU measurement models, we obtain a standard Gaussian update in the error state. The prior error satisfies

$$\delta\mathbf{x}_{k+1} = \mathbf{x}_{k+1} \ominus \hat{\mathbf{x}}_{k+1} \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{P}}_{k+1}) \tag{18}$$

while each measurement contributes a residual of the form

$$D_{k+1}\, \mathbf{n}_{k+1} = r_{k+1} - H_{k+1}\, \delta x_{k+1} \sim \mathcal{N}(\mathbf{0}, R_{k+1}),$$

where R_{k+1} is the measurement noise covariance (modulated by the dynamic weighting scheme for joint and IMU data).

The posterior estimate is obtained by minimising the sum of squared residuals,

$$\delta x^\star_{k+1} = \arg\min_{\delta x} \left(\|r_{k+1} - H_{k+1} \delta x\|^2_{R_{k+1}} + \|\delta x\|^2_{\hat{P}_{k+1}} \right),$$

whose closed-form solution is given by the Kalman update

$$
\begin{aligned}
\delta\mathbf{x}^{o}_{k+1} &= \mathbf{K}_{k+1}\mathbf{r}_{k+1} \\
\mathbf{K}_{k+1} &= \hat{\mathbf{P}}_{k+1}\mathbf{H}^{T}_{k+1}(\mathbf{H}_{k+1}\hat{\mathbf{P}}_{k+1}\mathbf{H}^{T}_{k+1} + \overline{\mathscr{R}}_{k+1})
\end{aligned} \tag{19}
$$

The nominal state and covariance are then updated on the manifold as

$$\bar{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_{k+1} \oplus \mathbf{K}_{k+1}\mathbf{r}_{k+1} \tag{20}$$

$$\bar{\mathbf{P}}_{k+1} = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{k+1})\hat{\mathbf{P}}_{k+1} \tag{21}$$

By tightly coupling LiDAR, IMU, and joint encoder information in this way, LIJO realises a multi-sensor fusion framework that is explicitly tailored to legged locomotion. The joint-based velocity factor suppresses LiDAR-induced jitter at low speeds, the LiDAR geometry constrains long-term drift, and the IMU provides high-frequency motion cues, together yielding a smooth and accurate state estimate suitable for perception-driven control.

4. Experiments

4.1. Datasets and baselines

Since, to the best of our knowledge, there is no public LiDAR SLAM dataset that provides synchronized joint encoder data for legged robots, all experiments are conducted on a real quadruped platform operating in outdoor environments. The robot is equipped with joint encoders on all legs, an onboard Livox Mid360 LiDAR, and an external high-precision IMU. During the experiments, the robot moves at speeds between 0.3 m/s and 0.5 m/s across three representative scenarios, as illustrated in Figure 6: (i) a steep staircase at the entrance of an elementary school, used to evaluate adaptability to highly uneven terrain; (ii) a loop trajectory around a youth hostel, used to assess performance on mostly flat ground with moderate structural features; (iii) a large loop around the Galaxy World Twin Towers, used to test performance in a large-scale environment with long-term drift.

For each run, the following data streams are recorded: (1) joint angles and angular velocities from the quadruped robot; (2) LiDAR point clouds from the Livox Mid360; (3) inertial measurements from the external IMU.

These datasets allow us to comprehensively evaluate the accuracy, robustness, and smoothness of the proposed LIJO system under both flat and steep terrain conditions. We compare LIJO against the following baselines:

- KF: a motion–inertia-based Kalman filter that fuses leg kinematics and IMU data using a linear KF.
- LIO-Livox: the official LiDAR–inertial odometry algorithm provided by Livox.
- FAST-LIO and Faster-LIO: state-of-the-art tightly coupled LiDAR–inertial odometry methods based on iterated Kalman filtering and incremental voxels.
- POINT-LIO: a robust LiDAR–inertial odometry method that treats IMU as measurement rather than input.

All methods are run with their recommended parameters where applicable. For fairness, we use the same LiDAR and IMU data across methods, and do not perform loop closure or global optimization for any of the real-time odometry pipelines.

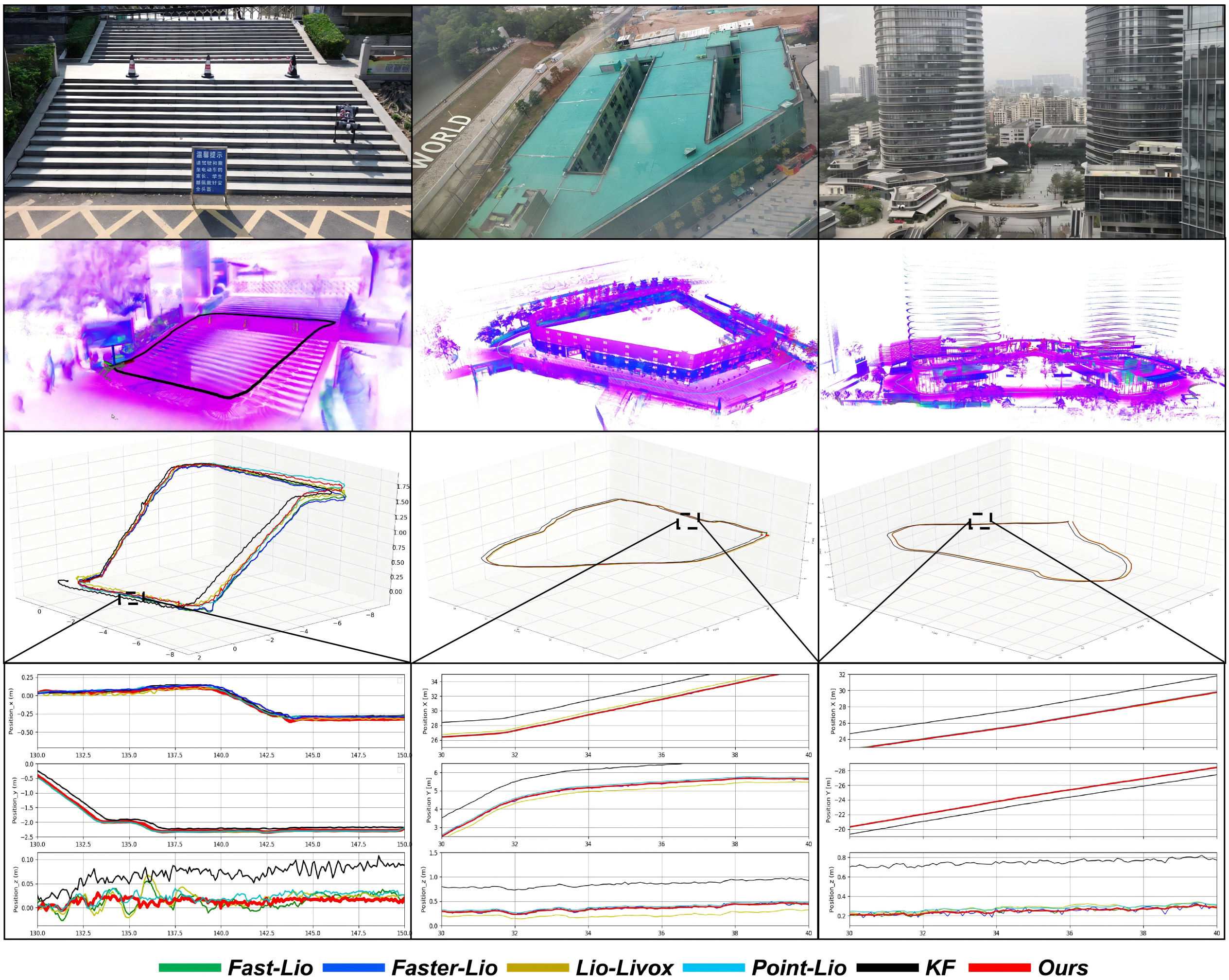

Figure 6. Real scene comparison experiments. The bottom row shows the position-time curves derived from various calculation methods. An enlarged display of the dashed areas is provided.

4.2. Evaluation of smoothness

We first evaluate trajectory smoothness, which is critical for perception-driven locomotion. Figure 6 shows the real-world trajectories and corresponding position–time curves for different methods in the three test scenarios. The bottom row plots the 1D position evolution over time, and zoomed-in views of the highlighted regions reveal clear differences in high-frequency jitter.

As expected, foot slippage during quadruped motion is unavoidable, which leads to significant cumulative errors for the KF baseline that relies solely on proprioceptive sensors (black curves). In addition, the mechanical impacts generated during locomotion cause substantial disturbances in IMU measurements. This effect is particularly evident in methods such as FAST-LIO and LIO-Livox (green and yellow curves), which use IMU data as system inputs and consequently exhibit noticeable drift, especially along the vertical (z) axis.

Given that the measurement accuracy of most LiDAR systems is currently at the centimeter level (the Livox Mid360 LiDAR used in this study has a measurement accuracy of 2 cm), even the most advanced LIO systems exhibit position estimation jitter of 1–2 cm when stationary, which adversely affects subsequent motion planning for quadruped robots. To address this issue, LIJO innovatively integrates joint-motion information from the robot's body and introduces a velocity-constraint mechanism, significantly improving the smoothness of position estimation. As shown in Figure 7, compared to the traditional LIO velocity curve, the LIJO velocity curve (red) demonstrates.

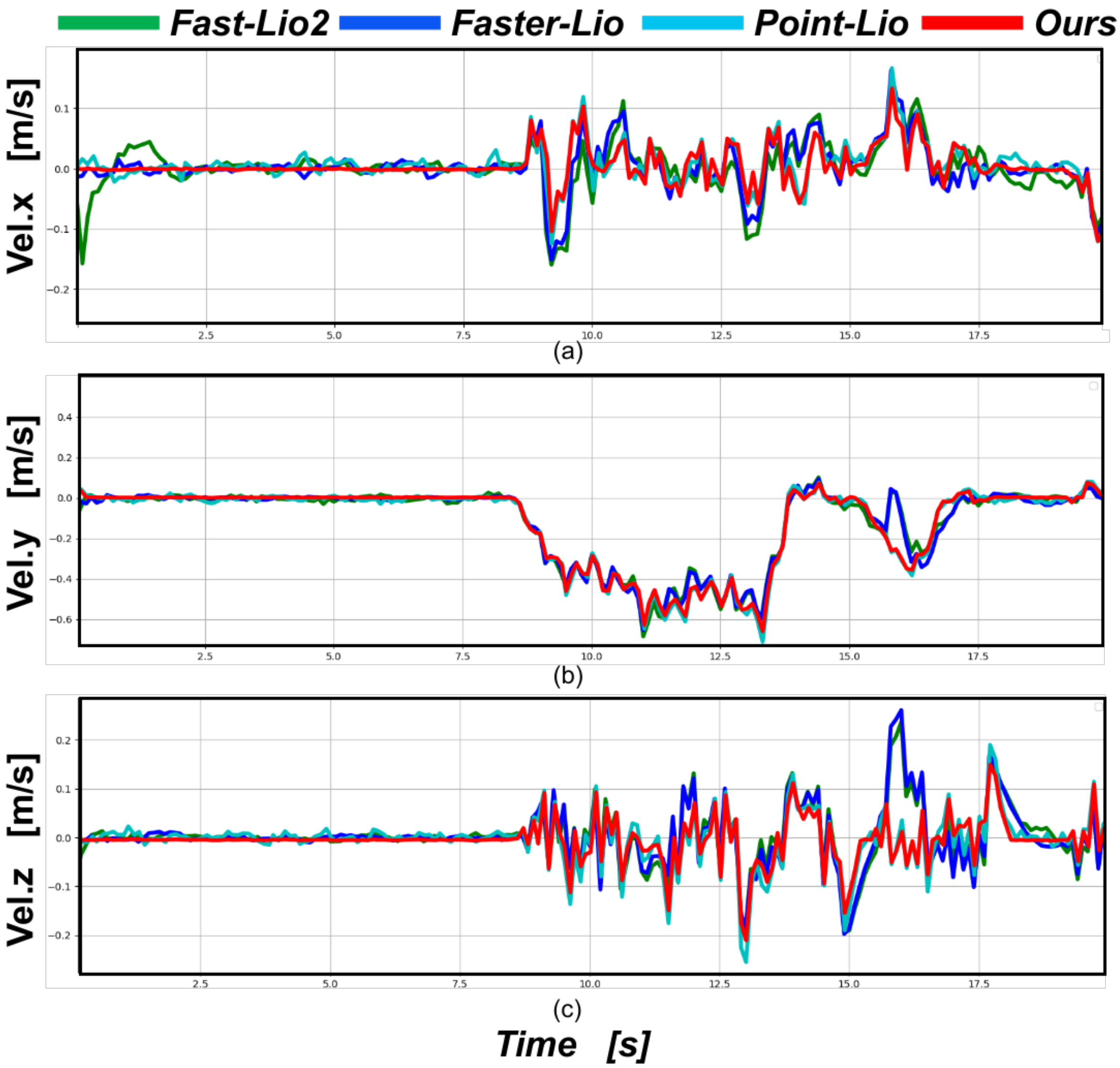

Figure 7. The estimated velocity profiles over time are compared (corresponding to the grey highlighted region in the left panel of Figure 6).

4.3. Loop-closure test and accuracy

To further assess odometry accuracy over longer trajectories, we perform a loop-closure experiment on a standard basketball court. The quadruped robot is commanded to traverse one full lap along the outer boundary of the court at an average speed of 0.5 m/s. Similar to common practice in LiDAR odometry evaluation, we construct a high-quality reference trajectory using an offline optimization pipeline: historical odometry states are stored and used as initial guesses for Iterative Closest Point (ICP)-based refinement, which is iteratively combined with loop-closure detection until convergence. The resulting optimized trajectory is then treated as ground truth for evaluation.

Figure 8 shows the resulting odometry trajectories produced by different methods. In realistic locomotion, foot slippage and modelling errors lead to accumulated drift in KF-based approaches that rely primarily on leg kinematics and IMU data. From a theoretical viewpoint, legged robot motion should be smooth and continuous, rather than exhibiting high-frequency oscillations. The zoomed-in subfigure in the lower-right corner of Figure 8 highlights that LIJO, which incorporates joint information in a tightly coupled manner, produces significantly smoother trajectories than conventional LIO methods and better matches the converged reference.

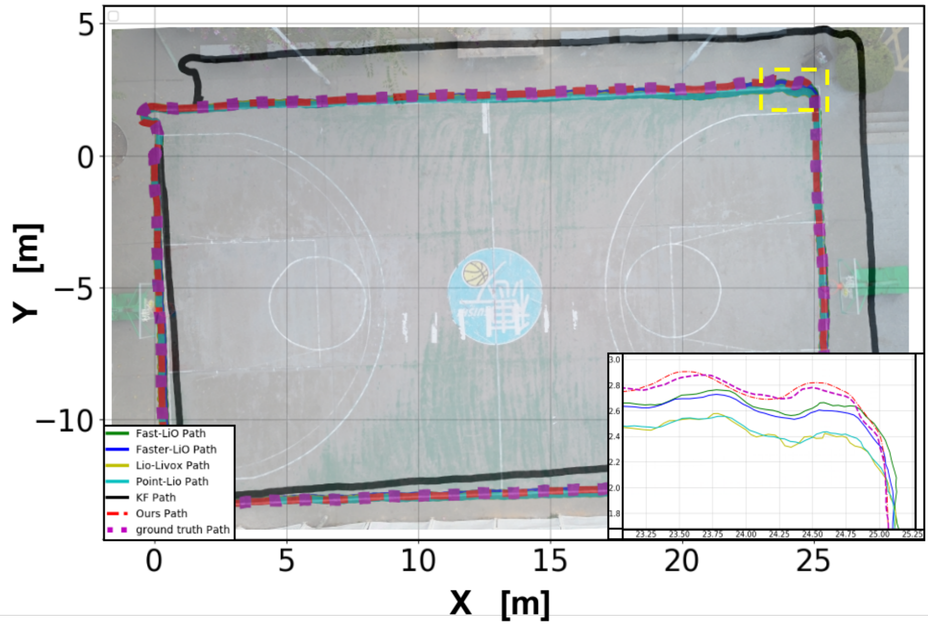

Figure 8. A comparison of trajectories using different methods in the court scene, with the bottom-right corner displaying a magnified view of the region within the yellow dashed box.

Table 1 reports the average Relative Pose Error (RPE) for all methods on the court sequence. LIJO achieves the lowest translation and rotation errors among the tested approaches, improving both metrics over state-of-the-art LIO baselines while maintaining smoothness. These results indicate that fusing joint measurements with LiDAR and IMU not only provides additional constraints to reduce high-frequency jitter, but also yields a more accurate and reliable basis for subsequent perception and motion tasks.

Table 1. RPE on the court sequence (transposed). Bold indicates the best performances.

Method	Translation Error [m]	Euler Error [deg]
FAST-LIO	0.1486	0.0691
Faster-LIO	0.1661	0.0755
LIO-Livox	0.4135	0.2026
POINT-LIO	0.1574	0.0587
KF	0.1205	0.0495
LIJO (Ours)	**0.1114**	**0.0357**

4.4. Perceived motion and locomotion quality

Finally, we examine how the different odometry outputs influence perception-driven locomotion. In many legged robot applications, the odometry module is not only responsible for self-localisation but also directly feeds into mapping, terrain perception, and motion planning pipelines. This places stringent requirements on both accuracy and temporal smoothness.

In practice, we observe that traditional LIO methods such as FAST-LIO and POINT-LIO generally provide adequate accuracy for global localization. However, due to LiDAR measurement limitations, their position estimates still exhibit noticeable jitter, as evidenced in Figure 6. Applying simple low-pass filters can reduce this jitter but introduces additional latency, which can severely degrade the responsiveness of whole-body controllers and high-level planners. On the other hand, kinematic–inertial odometry based on leg encoders and IMU (e.g., the KF baseline) typically yields very smooth trajectories but suffers from drift over time and is sensitive to contact slip.

LIJO combines the strengths of these two families of methods. By injecting a joint-based velocity factor into the multi-sensor fusion process and adaptively weighting it against LiDAR updates, LIJO effectively suppresses high-frequency jitter without sacrificing real-time performance or long-term accuracy. Figure 1 illustrates a representative scenario in which the quadruped robot performs perception-driven motion using LIJO: the resulting smooth odometry enables stable map construction and reliable terrain assessment, which in turn supports robust locomotion over irregular terrain. This demonstrates that LIJO is a practical and effective localization solution for legged robots operating in challenging outdoor environments.

5. Conclusion

In this paper, we presented LiDAR–Inertial–Joint Odometry, a manifold EKF–based multi-sensor fusion framework that tightly combines LiDAR, IMU, and joint encoder measurements to provide low-jitter, real-time state estimation for quadruped robots in outdoor environments. By leveraging a joint-based body-velocity factor with dynamic weighting, LIJO effectively suppresses LiDAR-induced pose jitter while preserving the accuracy and robustness of state-of-the-art LiDAR–inertial odometry, thereby offering more reliable state inputs for perception-driven locomotion.

Data availability statement

All data included in this study are available upon request by contacting the corresponding author.

Declaration of generative AI and AI-assisted technologies

The authors did not use generative AI or AI-assisted technologies in the writing of this manuscript.

Acknowledgments

This work was supported by the Shenzhen Science and Technology Program Grant (No. JSGGKQTD20221101115656029, ZDCY20250901094531003 and KJZD20230923113801004.)

Authors' contribution

Conceptualization, T.Z. and J.P.; methodology, B.L. and T.Z.; software, B.L.; validation, B.L. and T.Z.; formal analysis, B.L.; writing—original draft preparation, B.L.; writing—review and editing, J.P. and

T.Z.; supervision, J.P.; funding acquisition, T.Z. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

Jia Pan holds the position of Associate Editor for *Robot Learning* and has not peer reviewed or made any editorial decisions for this paper. The other authors declare no conflicts of interest.

References

[1] Wisth D, Camurri M, Fallon M. VILENS: visual, inertial, LiDAR, and leg odometry for all-terrain legged robots. *IEEE Trans. Rob.* 2022, 39(1):309–326.

[2] Yang S, Zhang Z, Fu Z, Manchester Z. Cerberus: low-drift visual-inertial-leg odometry for agile locomotion. In *Proceedings of 2023 IEEE International Conference on Robotics and Automation (ICRA)*, London, UK, May 29–June 2, 2023, pp. 4193–4199.

[3] Bin T, Yan H, Wang N, Nikolić MN, Yao J, *et al.* A survey on the visual perception of humanoid robot. *Biomimetic Intell. Rob.* 2025 5(1):100197.

[4] Wang S, Nikolić MN, Lam TL, Gao Q, Ding R, *et al.* Robot manipulation based on embodied visual perception: a survey. *CAAI Trans. Intell. Technol.* 2025, 10(4):945–958.

[5] Zhang J, Singh S. LOAM: LiDAR odometry and mapping in real-time. In *Proceedings of Robotics: Science and Systems (RSS 2014)*, Berkeley, USA, July 14–16, 2014, pp. 1–9.

[6] Xu W, Cai Y, He D, Lin J, Zhang F. FAST-LIO2: fast direct LiDAR-inertial odometry. *IEEE Trans. Rob.* 2022, 38(4):2053–2073.

[7] Xu W, Zhang F. FAST-LIO: a fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter. *IEEE Rob. Autom. Lett.* 2021, 6(2):3317–3324.

[8] He D, Xu W, Chen N, Kong F, Yuan C, *et al.* Point-LIO: robust high-bandwidth light detection and ranging inertial odometry. *Adv. Intell. Syst.* 2023, 5(7):2200459.

[9] Bai C, Xiao T, Chen Y, Wang H, Zhang F, *et al.* Faster-LIO: lightweight tightly coupled LiDAR-inertial odometry using parallel sparse incremental voxels. *IEEE Rob. Autom. Lett.* 2022, 7(2):4861–4868.

[10] Ou G, Li D, Li H. Leg-KILO: robust kinematic-inertial-LiDAR odometry for dynamic legged robots. *IEEE Robot. Autom. Lett.* 2024, 9(10):8194–8201.

[11] Li X, Zhang T. DOF-LIO: LiDAR–inertial odometry with lightweight dynamic object filter. *IEEE Trans. Instrum. Meas.* 2026, 75:1–10.

[12] Yuan Z, Lang F, Xu T, Yang X. LIWO: LiDAR-inertial-wheel odometry. In *Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Detroit, USA, October 1–5, 2023, pp. 1481–1488.

[13] Yang S, Zhang Z, Bokser B, Manchester Z. Multi-IMU proprioceptive odometry for legged robots. In *Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Detroit, USA, October 1–5, 2023, pp. 774–779.

[14] Bloesch M, Gehring C, Fankhauser P, Hutter M, Hoepflinger MA, *et al.* State estimation for legged robots on unstable and slippery terrain. In *Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Tokyo, Japan, November 3–7, 2013, pp. 6058–6064.

[15] Wisth D, Camurri M, Fallon M. Preintegrated velocity bias estimation to overcome contact nonlinearities in legged robot odometry. In *Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, May 31–August 31, 2020, pp. 392–398.

[16] Bledt G, Powell MJ, Katz B, Di Carlo J, Wensing PM, *et al.* MIT Cheetah 3: design and control of a robust, dynamic quadruped robot. In *Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, October 1–5, 2018, pp. 2245–2252.

[17] Kim Y, Yu B, Lee EM, Kim JH, Park HW, *et al.* STEP: state estimator for legged robots using a preintegrated foot velocity factor. *IEEE Rob. Autom. Lett.* 2022, 7(2):4456–4463.

[18] Qin T, Cao S, Pan J, Shen S. A general optimization-based framework for global pose estimation with multiple sensors. *arXiv* 2019, arXiv:1901.03642.

[19] Maybeck PS. *Stochastic Models, Estimation, and Control*, 1st ed. New York: Academic Press, 1982.

[20] Shan T, Englot B, Meyers D, Wang W, Ratti C, *et al.* LIO-SAM: tightly-coupled LiDAR inertial odometry via smoothing and mapping. In *Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, USA, 24 October–24 January, 2020–2021, pp. 5135–5142.

[21] Xu W, Cai Y, He D, Lin J, Zhang F. FAST-LIO2: fast direct LiDAR-inertial odometry. *IEEE Trans. Rob.* 2022, 38(4):2053–2073.

[22] Campos C, Elvira R, Rodríguez JJG, Montiel JM, Tardós JD. ORB-SLAM3: an accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Trans. Rob.* 2021, 37(6):1874–1890.

[23] Qin T, Li P, Shen S. VINS-mono: a robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Rob.* 2018, 34(4):1004–1020.

[24] Hartley R, Ghaffari Jadidi M, Gan L, Huang JK, Grizzle JW, *et al.* Hybrid contact preintegration for visual-inertial-contact state estimation using factor graphs. In *Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, October 1–5, 2018, pp. 3783–3790.