A Portable Multi-Sensor Unit and Layered Kalman Filtering for State Estimation on Legged and Wheeled Mobile Robots

Eren Cem Goksuluk, Burak Alp Inan, Ismail Uyanik
Department of Electrical and Electronics Engineering
Hacettepe University, 06800 Ankara, Turkiye
{goksuluk, inan, uyanik}@ee.hacettepe.edu.tr

***Abstract*—We present a portable multi-sensor unit and a layered multi-rate Kalman filtering architecture for real-time state estimation on mobile robots. The hardware integrates a 3D LiDAR, a stereo-inertial camera, an industrial intertial measurement unit (IMU), real-time kinematic GNSS (RTK-GNSS), and on-board computation into a single modular housing that can be mounted, without redesign, on both legged and wheeled platforms. The estimator combines two filters operating at different rates: an error-state Kalman filter (ESKF) that fuses the IMU with leg or wheel kinematics at the IMU rate and an extended Kalman filter (EKF) that anchors the proprioceptive estimate to drift-free exteroceptive sources (LiDAR odometry at $\sim$1 Hz, stereo visual odometry at $\sim$15 Hz). The framework is evaluated on a Unitree Go1 quadruped and a Clearpath Husky A200 using OptiTrack ground truth across standard, static-obstacle, and dynamic-obstacle scenarios. Across six logs, the fused estimate tracks the best single-sensor source within a few centimeters and, critically, avoids the catastrophic failures that any single modality exhibits in at least one scenario. We identify covariance miscalibration as the principal cause of the few cases in which a single sensor marginally outperforms the fusion and outline an adaptive-covariance extension as immediate future work.**

***Index Terms*—Mobile robots, state estimation, sensor fusion, Kalman filter, legged robots, LiDAR odometry, visual-inertial odometry.**

I. Introduction

Reliable state estimation is the backbone of every mobile-robot autonomy stack. Proprioceptive sensors such as IMUs and wheel or leg encoders provide high-rate motion information but accumulate drift; exteroceptive sensors such as LiDARs, cameras, and GNSS receivers are comparatively drift-free but operate at lower rates and degrade in characteristic, modality-specific ways (feature-poor corridors for LiDAR, low-light or dynamic scenes for vision, and canopy or indoors for GNSS). The research community has therefore converged on *fusion* as the default strategy, and a wide range of filter- and optimization-based architectures has been proposed [1], [2], [4], [5], [6], [7].

Despite this maturity, three practical obstacles recur when a multi-sensor estimator is deployed on a real robot. First, *transferability*: most published systems are specialized to one locomotion mode, one sensor set, or one dataset, and retargeting them to a new platform requires substantial re-engineering. Second, *multi-rate integration*: LiDAR, camera, IMU, and kinematic sources operate at frequencies that differ by two or three orders of magnitude, and naïve fusion schemes either waste high-rate information or violate the conditional-independence assumptions of the underlying filter. Third, *graceful degradation*: when one modality fails silently, a fusion estimator must detect the fault and downweight that source rather than integrate corrupted measurements.

This paper reports on a system that addresses the first two obstacles directly and the third partially. Our contributions are:

1) A portable multi-sensor unit that integrates a 3D LiDAR, stereo-inertial camera, industrial IMU, RTK-GNSS, and a Jetson Xavier NX into a single mechanically rigid module, with a standardized mounting interface that has been used without modification on a legged quadruped (Unitree Go1 EDU) and a wheeled platform (Clearpath Husky A200).
2) A layered ESKF–EKF architecture in which the ESKF fuses IMU and leg/wheel kinematics at the IMU rate, and the EKF fuses the ESKF prediction with LiDAR and stereo visual pose measurements at their respective rates. Both stages use the Joseph form and Mahalanobis gating for numerical robustness.
3) An experimental characterization on both platforms across three scenarios (nominal, static obstacles, dynamic obstacles), using OptiTrack ground truth, that quantifies when fusion helps, when it matches the best single sensor, and when covariance miscalibration causes it to underperform a single modality.

Relative to prior work, we make no claim to a new filtering algorithm; the mathematical ingredients—error-state formulations [2], [9], Van Loan discretization [3], Joseph-form updates [1]—are standard. The contribution is a systems integration that transfers cleanly between locomotion modes and an honest empirical characterization of its behavior. The paper is explicit about failure modes and identifies the open directions (adaptive covariance, invariant filtering, tight coupling) as future work.

The remainder of the paper is organized as follows. Section II describes the multi-sensor unit. Section III develops the estimation architecture. Sections IV and V cover the experimental protocol and results. Section VI discusses limitations,

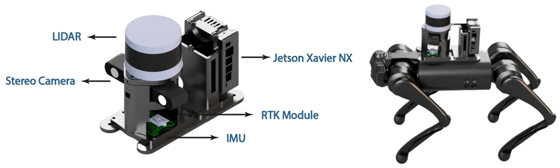

Fig. 1. CAD rendering of the multi-sensor unit. The 3D LiDAR sits at the top; the stereo camera is forward-facing; the IMU is placed close to the centroid; the RTK-GNSS module is above the compute board.

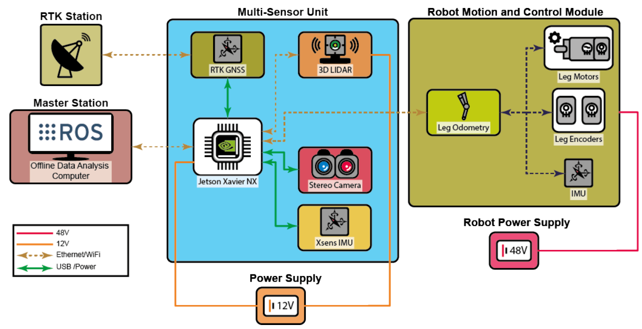

Fig. 2. System architecture. The multi-sensor unit is electrically and computationally independent of the host robot; the robot's odometry enters the unit as an external sensor stream.

and Section VII concludes.

II. Multi-Sensor Unit

A. Mechanical Design

The unit is conceived as a self-contained, portable module bringing together LiDAR, stereo-inertial vision, an industrial IMU, RTK-GNSS, and on-board computation in a single rigid housing (Figure 1). Sensors are placed so that (i) fields of view are complementary and unobstructed, (ii) the IMU lies close to the unit's center of mass to minimize lever-arm effects under rotational motion, and (iii) weight is balanced around the central axis to reduce vibration-induced artifacts.

A key requirement is reusability across host platforms. The mechanical interface is a standardized bolt pattern with thin adapter plates that match the mounting points of each host robot; no internal sensor geometry changes between platforms. Modular cabling and connectorization allow any single sensor or the compute unit to be replaced without disassembling the whole module, which has proved essential during field use.

B. Electronics and Software

All sensors are powered from a dedicated 12 V LiPo battery through regulated DC–DC converters, making the unit electrically independent from the host robot and enabling long experiments without tapping the robot's own power. An NVIDIA Jetson Xavier NX runs a ROS-based acquisition and estimation stack; heterogeneous links (Ethernet, USB 3, UART) are unified through ROS nodes, and all measurements are timestamped on arrival and re-published on a common clock. The host robot exposes its own odometry (leg kinematics on the Go1, wheel encoders on the Husky) over a separate Ethernet link and is treated by the unit as just another sensor. Figure 2 shows the overall data flow between unit, robot, RTK base station, and offline analysis workstation.

III. State Estimation Methodology

We fuse the available streams in two stages. A low-level error-state Kalman filter (ESKF) fuses the IMU with leg or wheel kinematics at the IMU rate and produces a high-rate pose–velocity estimate. A higher-level EKF uses that estimate as its process model and fuses in the exteroceptive LiDAR and stereo pose observations. The design keeps each filter compact and avoids the numerical conditioning issues of a single monolithic filter that would have to update its full state at rates differing by two orders of magnitude. We note that the two-stage design is known to be suboptimal relative to a single centralized filter because the ESKF output is not independent of the IMU-driven prediction used by the upper EKF; we revisit this point in Section VI.

A. Notation and Nominal State

Let $\mathbf{p}_k \in \mathbb{R}^3$, $\mathbf{v}_k \in \mathbb{R}^3$, and $\mathbf{R}_k \in \mathrm{SO}(3)$ denote position, linear velocity, and orientation of the robot body expressed in a fixed world frame, and let $\mathbf{b}_k^g, \mathbf{b}_k^a \in \mathbb{R}^3$ denote the gyroscope and accelerometer biases. The ESKF nominal state is

$$\mathbf{x}_k = \left[\mathbf{p}_k^\top,\ \mathbf{v}_k^\top,\ \boldsymbol{\theta}_k^\top,\ \mathbf{b}_k^{g\,\top},\ \mathbf{b}_k^{a\,\top}\right]^\top, \tag{1}$$

where $\boldsymbol{\theta}_k$ is a minimal (local) rotation parameterization. Following standard error-state practice [2], we propagate the nominal state with the full non-linear dynamics and maintain a linearized filter on a 15-dimensional error state.

$$\delta\mathbf{x}_k = \left[\delta\mathbf{p}_k^\top,\ \delta\mathbf{v}_k^\top,\ \delta\boldsymbol{\theta}_k^\top,\ \delta\mathbf{b}_k^{g\,\top},\ \delta\mathbf{b}_k^{a\,\top}\right]^\top. \tag{2}$$

The small-angle assumption $\|\delta\boldsymbol{\theta}_k\| \ll 1$ justifies the usual linearization of the rotation error dynamics around the current nominal trajectory.

B. Proprioceptive Filter (ESKF)

Process model: Between IMU samples, the nominal state evolves under the strap-down inertial equations,

$$\begin{aligned} \dot{\mathbf{p}} &= \mathbf{v}, \\ \dot{\mathbf{v}} &= \mathbf{R}(\tilde{\mathbf{a}} - \mathbf{b}^a) + \mathbf{g}, \\ \dot{\mathbf{R}} &= \mathbf{R}\,[\tilde{\boldsymbol{\omega}} - \mathbf{b}^g]_\times, \end{aligned} \tag{3}$$

where $\tilde{\mathbf{a}}, \tilde{\boldsymbol{\omega}}$ are the raw IMU specific-force and angular-rate measurements, $\mathbf{g}$ is the gravity vector, and $[\cdot]_\times$ is the skew-symmetric operator. Bias states evolve as random walks.

Discretization:: Rather than first-order Euler integration of the error dynamics, we obtain the discrete-time transition matrix $\mathbf{F}_d$ and process-noise covariance $\mathbf{Q}_d$ using Van Loan's method [3]. Forming the block matrix,

$$\mathbf{M} = \begin{bmatrix} -\mathbf{F}_c & \mathbf{L}_c\mathbf{Q}_c\mathbf{L}_c^\top \\ \mathbf{0} & \mathbf{F}_c^\top \end{bmatrix} \Delta t, \qquad e^{\mathbf{M}} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}, \tag{4}$$

we recover $\mathbf{F}_d = \mathbf{A}_{22}^\top$ and $\mathbf{Q}_d = \mathbf{F}_d\mathbf{A}_{12}$. Here $\mathbf{F}_c$ and $\mathbf{L}_c$ are the continuous-time Jacobians of the error dynamics with respect to $\delta\mathbf{x}$ and the noise, and $\mathbf{Q}_c$ is the continuous noise spectral density. This closed-form procedure is numerically stable at the IMU rate and removes a tuning knob relative to approximate integrators.

Measurement update: Leg odometry on the Go1 (kinematic body velocity derived from contact-foot Jacobians) and wheel odometry on the Husky are both modeled as noisy observations of the body-frame linear velocity:

$$\mathbf{z}_k^{\text{odom}} = \mathbf{R}_k^\top \mathbf{v}_k + \mathbf{n}_k^{\text{odom}}, \qquad \mathbf{n}_k^{\text{odom}} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^{\text{odom}}). \quad (5)$$

The innovation and its covariance are

$$\mathbf{y}_k = \mathbf{z}_k^{\text{odom}} - h(\hat{\mathbf{x}}_{k|k-1}), \qquad \mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}^{\text{odom}}. \quad (6)$$

We gate updates with the squared Mahalanobis distance,

$$d_k^2 = \mathbf{y}_k^\top \mathbf{S}_k^{-1} \mathbf{y}_k \ \gtrless\ \chi^2_{3,\,0.99} = 11.345, \quad (7)$$

rejecting updates whose d_k^2 exceeds the 99th percentile of the three-DoF chi-squared distribution and inflating $\mathbf{P}$ instead. This protects the state against foot-slip and contact-misdetection events on the Go1 and against wheel slip on the Husky.

Covariance update: We use the Joseph form

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^\top + \mathbf{K}_k \mathbf{R}^{\text{odom}} \mathbf{K}_k^\top, \quad (8)$$

which preserves symmetry and positive definiteness in fixed-precision arithmetic over long runs on embedded hardware. After each update, $\delta\mathbf{x}$ is injected into the nominal state and reset to zero in the standard ESKF manner.

C. Exteroceptive Odometry Sources

Two drift-free pose sources enter the upper filter.

LiDAR odometry. Edge and planar features are extracted from successive scans and aligned through scan-to-scan and scan-to-map registration, in the spirit of LOAM [8] and its modern variants [5], [6]. We use the map-optimized pose stream (approximately 1 Hz), which maintains global consistency as the map grows. The output is a full $\mathrm{SE}(3)$ pose together with a covariance estimate.

Stereo visual odometry. Stereo visual-inertial pose is obtained from a ZED 2i camera through the manufacturer SDK. The SDK performs real-time stereo matching together with visual-inertial estimation and outputs a 6-DoF camera pose in $\mathrm{SE}(3)$ and a time-varying 6×6 pose covariance at approximately 15 Hz, both of which the EKF uses directly as measurement inputs. The exact internal algorithm is not disclosed by the manufacturer; the output is consistent with a tightly coupled visual-inertial optimization. Visual updates are more frequent than LiDAR but degrade more strongly in scenes with significant dynamic content, as discussed in Section IV.

Both sources are expressed in the body frame via pre-calibrated extrinsics and are timestamped consistently with the IMU through the Jetson acquisition stack.

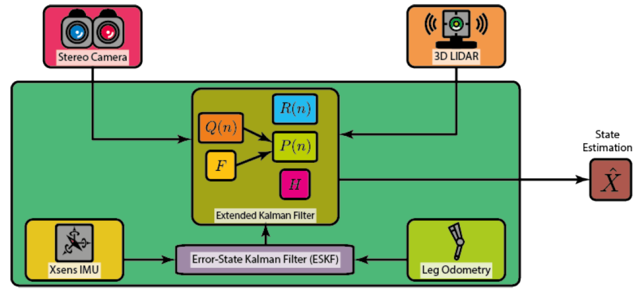

Fig. 3. Block diagram of the layered ESKF–EKF architecture. Proprioceptive (IMU + leg/wheel) streams drive the lower ESKF; exteroceptive LiDAR and camera pose streams update the upper EKF.

D. Layered Multi-Rate Fusion

The upper EKF uses the same state vector as the ESKF (Equation (1)), so the two filters share a consistent representation. The EKF prediction step uses the same high-rate inertial–kinematic motion model that drives the ESKF, propagating the proprioceptive motion estimate forward in time. LiDAR and stereo pose streams enter the EKF as measurement updates when they arrive; the high-rate IMU and leg/wheel data influence the EKF only through the propagation, so that the exteroceptive sources dominate drift correction.

For each exteroceptive source $i \in \{L, C\}$ (LiDAR, camera) we model the measurement as a direct observation of position and orientation:

$$\mathbf{z}_k^{(i)} = \begin{bmatrix} \mathbf{p}_k \\ \boldsymbol{\theta}_k \end{bmatrix} + \mathbf{n}_k^{(i)}, \qquad \mathbf{n}_k^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k^{(i)}), \quad (9)$$

with linearised observation matrix

$$\mathbf{H}^{(i)} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (10)$$

The LiDAR covariance $\mathbf{R}^{(L)}$ is fixed and small, reflecting global consistency of the map-optimised stream. The camera covariance $\mathbf{R}_k^{(C)}$ is time-varying and taken from the SDK output, so that the EKF adaptively weights visual updates based on the visual-inertial optimiser's own uncertainty estimate. Mahalanobis gating and Joseph-form updates are applied as in Equation (7) and Equation (8), adapted to the 6-DoF measurement.

The result is an inherently multi-rate filter: the prediction runs at the IMU rate (hundreds of Hz on our setup), camera updates at $\sim$15 Hz, and LiDAR updates at $\sim$1 Hz. Fast proprioceptive feedback is preserved while slow drift is continuously corrected by whichever exteroceptive source is currently informative.

IV. Experimental Setup

A. Platforms and Ground Truth

The same multi-sensor unit was mounted, in turn, on a Unitree Go1 EDU quadruped and a Clearpath Husky A200 wheeled platform. These platforms span a wide range of motion profiles: the Go1 produces short-duration foot-contact events with non-negligible vertical oscillation, while the Husky

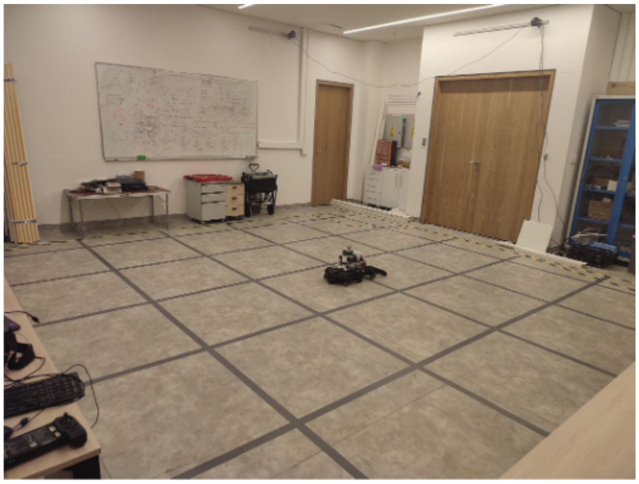

Fig. 4. ROMER indoor arena with the OptiTrack motion-capture system used to obtain ground-truth trajectories.

is close to a planar rigid body on flat ground. On both platforms, the unit communicated with the host robot only through a single Ethernet link carrying odometry.

Experiments took place in the ROMER indoor arena at Middle East Technical University, equipped with an OptiTrack motion-capture system providing pose ground truth at up to 120 Hz with centimetre-level accuracy. Reflective markers were mounted on a rigid bracket attached to the sensor unit.

B. Scenarios

Three scenarios were executed on each platform.

Standard. The robot traversed the arena along varied trajectories, with no obstacles. This served as a baseline for nominal fusion performance and to characterize sensor drift under well-behaved conditions.

Static obstacles. Fixed obstacles of varied size and position were placed in the workspace. The robot navigated around and between them, generating partial LiDAR occlusions and occasional visual feature loss.

Dynamic obstacles. Moving objects while the robot moved. This stressed the visual odometry particularly strongly, since moving pedestrians occupy a large fraction of the camera field of view and violate the static-scene assumption of the underlying visual-inertial estimator.

V. Experimental Results

We evaluate the estimator against OptiTrack ground truth using absolute position error (APE). For each log we report the mean, median, RMSE, and standard deviation in meters.

A. Unitree Go1, Standard Scenario

Figure 6 compares ground truth against the three single-sensor trajectories: leg+IMU odometry (ESKF output using proprioception only), stereo visual odometry, and map-optimized LiDAR odometry. Leg odometry drifts visibly as the experiment proceeds; visual odometry stays closer to the reference but exhibits local excursions; LiDAR odometry is the

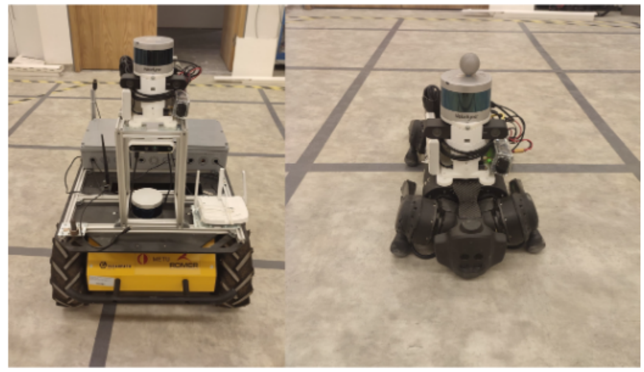

Fig. 5. Go1 EDU quadruped (right) and Husky A200 (left) with the multi-sensor unit installed through a standardized bracket.

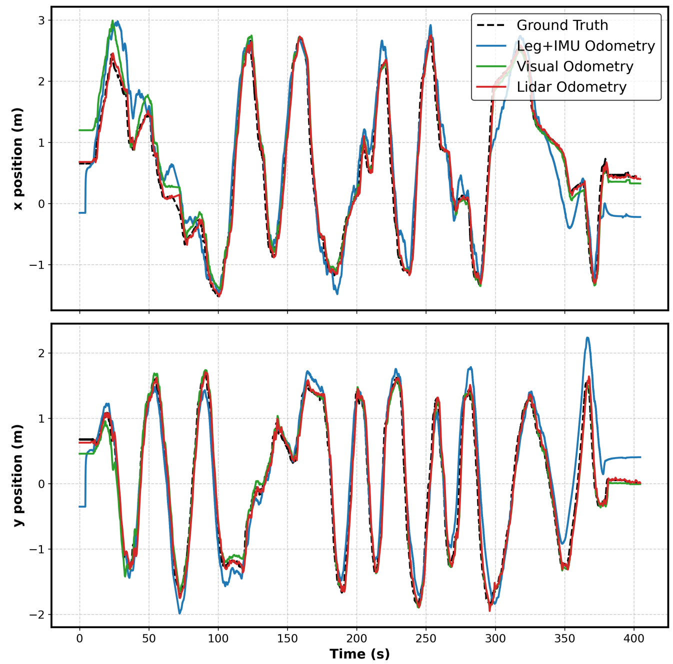

Fig. 6. Standard scenario, Unitree Go1. Ground truth and the three single-sensor odometry estimates (leg+IMU, stereo visual, LiDAR) in the x–y plane.

most consistent single source. Figure 7 and Table I quantify this behavior.

Figures 8 and 9 and Table II report the fused estimates. *Fusion-1* uses the ESKF prediction together with visual odometry, *Fusion-2* uses the ESKF prediction together with LiDAR odometry, and *Fusion-all* includes both exteroceptive sources. The full fusion reduces mean APE to 0.086 m, roughly a factor of two below the best single sensor (LiDAR, 0.200 m) and an order of magnitude below leg odometry. Fusion-1 already recovers most of this gain, which shows that even a single exteroceptive anchor is sufficient to contain the drift of the proprioceptive estimate.

B. Obstacle Scenarios

In the static-obstacle scenario on the Go1, LiDAR odometry remains the strongest single source and the full fusion tracks

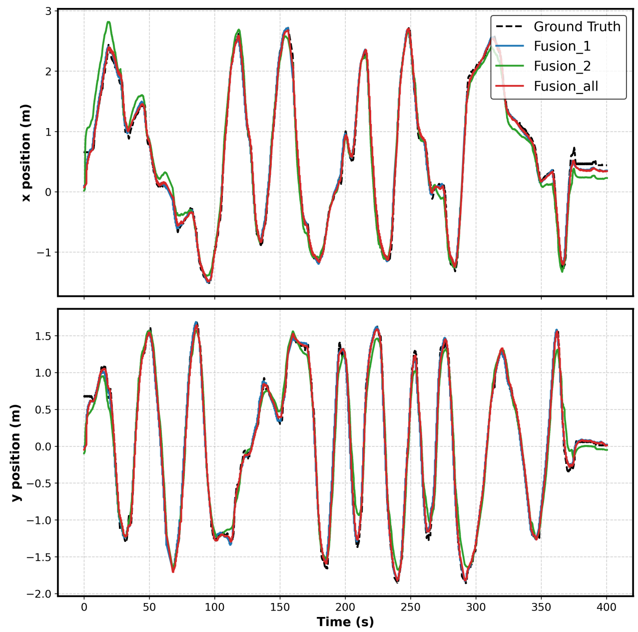

Fig. 7. Standard scenario, Go1. Fused trajectories in the x–y plane: Fusion-1 (ESKF + visual), Fusion-2 (ESKF + LiDAR), Fusion-all (ESKF + visual + LiDAR) against ground truth.

TABLE I
APE (M) FOR SINGLE-SENSOR ESTIMATES, GO1 STANDARD SCENARIO.

	Max	Mean	Median	Min	RMSE	Std
Leg + IMU	1.462	0.720	0.660	0.502	0.738	0.159
Visual (ZED)	0.874	0.312	0.306	0.029	0.344	0.145
LiDAR	0.572	0.200	0.191	0.010	0.230	0.113

it closely while substantially reducing the leg-odometry drift that dominates the ESKF-only output. In the dynamic-obstacle scenario, the balance shifts: the stereo SDK's own covariance estimate grows sharply when pedestrians fill the field of view, and the EKF down-weights the visual updates through $\mathbf{R}_k^{(C)}$. The fused estimate therefore stays close to the LiDAR-anchored trajectory, confirming that the use of the SDK-reported covariance—rather than a fixed $\mathbf{R}^{(C)}$—is essential for graceful degradation under scene-dependent visual failure.

C. Platform Transfer (Husky A200)

The same hardware and estimator, with only the odometry source swapped from leg kinematics to wheel encoders, were deployed on the Husky. Table III summarizes RMSE for all six logs (three on each platform).

Fusion consistently tracks the best single sensor. On every log, the fused RMSE is within a few centimeters of the best individual source. Fusion yields the lowest RMSE on Log 1U and Log 4H; on the remaining logs, it remains close to the best single-sensor estimate while still substantially improving over the other two modalities.

No single modality dominates across scenarios. LiDAR yields the lowest single-sensor error on five of the six logs;

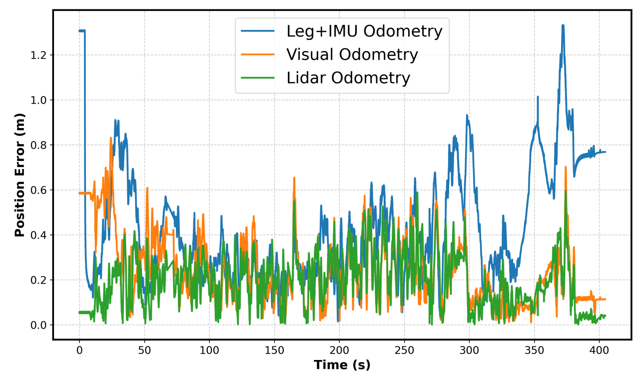

Fig. 8. Absolute position error over time for the three single-sensor estimates on the Go1 standard scenario.

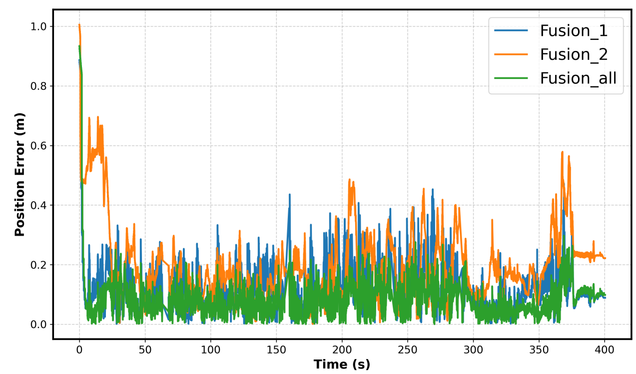

Fig. 9. Absolute position error over time for the fused estimates on the Go1 standard scenario.

visual odometry does so on the remaining one. Because which modality is most informative depends on the environment, a system relying on any fixed single source would incur a substantially larger error on at least one log. The fused estimator maintains an RMSE of $0.201\,\mathrm{m}$ or below across all six logs, giving a uniformly low error envelope across the three scenarios and two platforms.

The present analysis focuses on absolute position error. Heading and full orientation error metrics, per-axis drift rates, and trajectory-level error distributions are natural extensions for a longer-form evaluation.

VI. Discussion and Limitations

Across the six logs, the fused estimator remains within a few centimeters of the best single-sensor source and yields a uniformly low error envelope across platforms and scenarios. A natural direction for further improvement is to replace the static measurement-noise covariances $\mathbf{R}^{(L)}$ and $\mathbf{R}^{(C)}$ with an online adaptation driven by windowed normalized-innovation statistics [13], so that the filter can down-weight a modality whose empirical innovations exceed its expected χ^2 range in a given scenario. This extension is compatible with the layered architecture described here and is the first item of our future work.

TABLE II
APE (M) FOR FUSED ESTIMATES, GO1 STANDARD SCENARIO.

	Max	Mean	Median	Min	RMSE	Std
ESKF only	1.444	0.391	0.341	0.005	0.454	0.231
Fusion-1	0.887	0.127	0.110	0.006	0.152	0.083
Fusion-2	1.006	0.200	0.169	0.006	0.238	0.130
Fusion-all	0.934	0.086	0.072	0.001	0.114	0.074

TABLE III
RMSE (M) FOR ALL LOGS. U: UNITREE GO1; H: HUSKY A200. BOLD MARKS THE BEST VALUE PER ROW.

Log	Leg/Wheel	Visual	LiDAR	Fusion
Log 1U	0.738	0.344	0.230	**0.114**
Log 2U	0.899	0.387	**0.083**	0.129
Log 3U	0.817	0.593	**0.186**	0.201
Log 4H	0.176	0.575	0.071	**0.044**
Log 5H	0.460	0.278	**0.074**	0.108
Log 6H	0.195	**0.099**	0.307	0.132

A second limitation is the cascaded ESKF–EKF design itself. Because the upper EKF's prediction uses the IMU (via the ESKF output) and the EKF's measurement updates are derived from an estimate that already incorporates IMU information, the two stages are not strictly independent; a fully centralized filter or a tightly coupled factor-graph formulation [12] would avoid this double-counting. We chose the layered design for engineering reasons. Modularity, independent rate handling, simpler failure-domain analysis, and quantifying the suboptimality it incurs against a centralized baseline are open questions.

A third limitation, noted already in Section III, is the use of a standard EKF on $\mathrm{SO}(3)$-valued states rather than an invariant EKF. Recent work on invariant filtering [10], [11] has established consistency and convergence properties for exactly this class of problems and is now the default in modern legged state estimators. Porting the upper filter to a right-invariant formulation on $\mathrm{SE}_2(3)$ is a natural next step.

Our IMU propagation model omits the Earth's rotation rate. This is acceptable for the short-duration, exteroceptively corrected experiments reported here but would need to be added for long-duration GNSS-denied operation where unaided inertial drift becomes significant.

VII. CONCLUSION AND FUTURE WORK

We presented a portable multi-sensor unit and a layered multi-rate Kalman filtering framework that produces real-time state estimates on both a legged quadruped and a wheeled skid-steer robot without hardware modification. The system achieves sub-decimeter mean position error in the nominal scenario on the Go1, tracks the best single sensor within a few centimeters across six logs spanning three scenarios, and avoids the catastrophic failures that any single modality would suffer in at least one scenario.

Immediate future work targets the three limitations identified in Section VI: (i) online adaptive covariance driven by normalized-innovation statistics; (ii) replacement of the upper EKF by a right-invariant formulation; and (iii) a head-to-head benchmark against tightly coupled systems such as LIO-SAM and FAST-LIO2 on identical logs, including public datasets. RTK-GNSS integration for outdoor operation and loop-closure-aware SLAM coupling complete the near-term roadmap.

ACKNOWLEDGEMENT

The authors thank the ROMER Robotics and Mechatronics Center at Middle East Technical University for access to the OptiTrack arena and for experimental support.

REFERENCES

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA: MIT Press, 2005.

[2] J. Solà, "Quaternion kinematics for the error-state Kalman filter," *arXiv:1711.02508*, 2017.

[3] C. Van Loan, "Computing integrals involving the matrix exponential," *IEEE Trans. Autom. Control*, vol. 23, no. 3, pp. 395–404, 1978.

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

[5] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2020.

[6] 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.

[7] T. Moore and D. Stouch, "A generalized extended Kalman filter implementation for the robot operating system," in *Intelligent Autonomous Systems 13*, 2016, pp. 335–348.

[8] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Robotics: Science and Systems*, 2014.

[9] 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," in *Robotics: Science and Systems*, 2013.

[10] 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.

[11] 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.

[12] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.

[13] R. K. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Autom. Control*, vol. 15, no. 2, pp. 175–184, 1970.

[14] L. Zhang, L. Zhang, R. Gao, L. Pan, C. Xu, and K. Cheng, "Indoor mobile robot localization applying IMU/stereo camera/LiDAR and graph-based optimization," *IEEE Sensors J.*, vol. 24, no. 13, pp. 21466–21478, 2024.

[15] M. Mukherjee, A. Banerjee, A. Papadimitriou, S. S. Mansouri, and G. Nikolakopoulos, "A decentralized sensor fusion scheme for multi sensorial fault resilient pose estimation," *Sensors*, vol. 21, no. 24, p. 8259, 2021.

[16] B. Morrell, K. Otsu, A. Agha, D. D. Fan, S.-K. Kim, M. F. Ginting, X. Lei, *et al.*, "An addendum to NeBula: Toward extending team CoSTAR's solution to larger scale environments," *IEEE Trans. Field Robot.*, vol. 1, pp. 476–526, 2024.