A Modular Loosely-Coupled Multi-Sensor Localization Framework for Low-Cost UGVs in GNSS-Denied Environments

Frederico Rasinhas[1] and Vítor Viegas[1,2] and Bruno Damas[1,3]

[1]CINAV - Centro de Investigação Naval, Almada, Portugal
[2]Instituto de Telecomunicações, Instituto Superior Técnico, Lisbon,Portugal
[3]Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal
{alves.rasinhas, vitor.viegas, bruno.damas}@marinha.pt

Abstract— **This paper presents a modular localization framework for low-cost unmanned ground vehicles operating in GNSS-denied environments. The proposed system adopts a loosely-coupled Error-State Extended Kalman Filter (ESKF) in which inertial measurements are used for state propagation, while wheel odometry, monocular visual odometry, and 2D LiDAR odometry are incorporated as asynchronous correction sources. To improve the practical use of monocular vision in this context, a ground-plane-based scale recovery strategy is integrated into the visual odometry branch, enabling metric motion estimation from a single camera. The framework is evaluated in three representative scenarios, including indoor maneuvering, open outdoor navigation, and outdoor operation in a vegetated environment. Preliminary results show that the contribution of each sensing modality is strongly scenario-dependent. In the predominantly planar trajectories considered in this study, wheel odometry provides the most stable performance, whereas visual and LiDAR-based corrections become more or less useful depending on the surrounding scene structure and texture. The results also show that fusing all available sensors does not necessarily produce the best estimate, motivating future work on more adaptive sensor selection and evaluation in non-planar terrain.**

I. Introduction and Related Work

Reliable self-localization is a key requirement for unmanned ground vehicles (UGVs) operating in contested, cluttered, or infrastructure-poor environments. In many practical deployments, Global Navigation Satellite Systems (GNSS) provide the main source of global position information; however, their performance can degrade severely under signal blockage, multipath, jamming, or spoofing. This limitation motivates the development of navigation systems capable of sustaining acceptable localization performance using only onboard sensing and estimation, especially in missions where low-cost robotic platforms must continue operating when absolute positioning is unavailable or unreliable.

In this context, odometry and multi-sensor fusion have become central building blocks for GNSS-denied navigation. A broad range of solutions has been reported in the literature, combining inertial, visual, laser, and wheel-based sensing according to different architectural choices, from loosely-coupled filtering pipelines to tightly-coupled optimization-based estimators [1]. In general, the main challenge is not only to estimate motion accurately, but also to maintain robustness when the reliability of each sensing modality changes with the environment. This issue is particularly relevant for small UGVs, where sensing and computing resources are constrained, and where a practical solution must remain lightweight, modular, and easy to adapt to different mission profiles.

Among LiDAR-based approaches, LOAM remains one of the most influential references, showing that scan-to-scan geometric registration can provide accurate motion estimation by extracting edge and planar features from point clouds [2]. In the visual-inertial domain, VINS-Mono demonstrated that tightly-coupled fusion between a monocular camera and an IMU can achieve accurate state estimation with sliding-window optimization, while also addressing the scale ambiguity inherent to monocular vision [3]. More recently, multi-modal systems such as LIC-Fusion have combined LiDAR, camera, and IMU data in tightly-coupled frameworks to improve robustness across diverse scenarios [4]. These methods achieve strong performance, but they typically rely on richer sensing suites, heavier optimization back-ends, or tighter calibration dependencies than those desirable for small, low-cost field robots.

For resource-constrained UGVs, a modular loosely-coupled alternative remains attractive. In such a design, each sensing modality can be processed independently and injected into a common estimator as a correction source, which simplifies development, facilitates sensor replacement, and makes the overall system easier to adapt to mission-specific constraints [1], [5]. This flexibility is valuable in mixed indoor/outdoor operation, where the usefulness of a given modality may vary substantially: wheel odometry may be highly informative on structured ground, monocular vision may become more reliable in well-textured outdoor scenes, and a 2D LiDAR may perform well indoors while degrading in open or dynamic environments. An error-state Kalman formulation is particularly suitable for this type of system because it enables high-rate inertial propagation while incorporating asynchronous corrections from heterogeneous sensors in a numerically efficient way [6].

Monocular visual odometry is especially appealing in low-cost platforms because cameras are inexpensive, lightweight, and already available in many robotic systems. However, monocular methods suffer from scale ambiguity, which must be resolved using additional assumptions or external information. For ground vehicles, one practical strategy is to

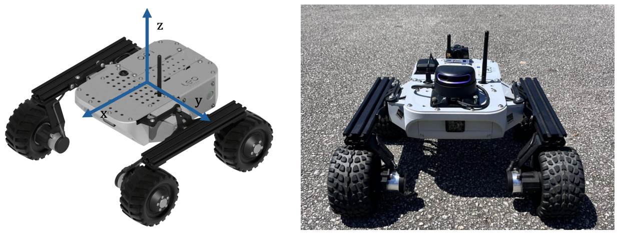

Fig. 1: The Leo Rover platform used in this work, equipped with a monocular camera, IMU, wheel encoders, and a 2D LiDAR.

exploit the ground plane to recover metric scale from the observed scene geometry [7]. In the present work, this idea is combined with feature tracking based on the Lucas-Kanade optical flow method, which provides a computationally efficient mechanism to estimate inter-frame motion from image sequences [8].

This paper presents a working-in-progress modular localization framework for GNSS-denied UGV navigation based on a loosely-coupled Error-State Extended Kalman Filter (ESKF). The proposed system uses the IMU for state propagation and incorporates wheel odometry, monocular visual odometry, and 2D LiDAR odometry as correction sources. Rather than pursuing a tightly-coupled mapping-oriented solution, the goal is to study a lighter architecture that can be deployed on a low-cost rover and reconfigured according to the sensing conditions of the environment. The emphasis is therefore on a transparent and reconfigurable localization pipeline for low-cost UGVs, in which independent odometry branches can be evaluated, combined, or selectively disabled according to the sensing conditions of the environment.

The contribution of this paper is threefold. First, it presents a practical multi-sensor localization architecture tailored to a low-cost UGV operating without continuous GNSS support. Second, it integrates a monocular visual odometry module with ground-plane-based scale recovery into a common ESKF framework alongside wheel and LiDAR observations. Third, it reports preliminary experimental results in representative indoor and outdoor scenarios, highlighting that the best fusion configuration is environment-dependent and that using all available sensors does not necessarily yield the best trajectory estimate.

II. System Overview and Methodology

A. Robotic Platform and Sensing Suite

The proposed localization framework was implemented on a Leo Rover, a small four-wheeled unmanned ground vehicle designed for low-cost robotic experimentation. The platform follows a differential-drive configuration and integrates the onboard sensing suite used throughout this work: a monocular RGB camera, an inertial measurement unit (IMU), wheel encoders, and a 2D LiDAR. These sensors were selected to study a practical localization architecture for UGV operation in GNSS-denied environments while keeping sensing and computational requirements modest.

All measurements are expressed with respect to a body-fixed reference frame attached to the vehicle, with the x axis pointing forward, the y axis to the left, and the z axis upward. Since the rover operates predominantly on approximately planar terrain, the localization problem is largely governed by planar position and heading, while roll and pitch are treated as secondary quantities in the final fusion architecture.

B. Overall Architecture

The system follows a modular loosely-coupled architecture in which each sensing modality first provides an independent motion estimate or kinematic observation, and these measurements are then fused in a common Error-State Extended Kalman Filter (ESKF) [6]. This design choice favours simplicity, modularity, and ease of sensor replacement, which are desirable properties for low-cost field robots operating under changing environmental conditions.

The IMU is used as the main source for state propagation. At each prediction step, angular velocity and linear acceleration measurements are integrated to propagate the nominal state. Wheel odometry, monocular visual odometry, and LiDAR odometry are incorporated asynchronously as correction sources whenever new measurements become available. This organization makes it possible to evaluate different sensing combinations without modifying the estimator core.

A compact representation of the nominal state is

$$\mathbf{x} = \begin{bmatrix} \mathbf{p} & \mathbf{v} & \mathbf{q} & \mathbf{b}_a & \mathbf{b}_\omega \end{bmatrix}^T,$$

where $\mathbf{p}$ and $\mathbf{v}$ denote position and velocity, $\mathbf{q}$ is the attitude quaternion, and $\mathbf{b}_a$ and $\mathbf{b}_\omega$ are the accelerometer and gyroscope biases, respectively. The corresponding error state is estimated by the ESKF and injected into the nominal trajectory after each correction step [6].

C. Monocular Visual Odometry with Ground-Plane Scale Recovery

Monocular visual odometry is attractive for low-cost UGVs because it requires only a single camera and moderate computational resources. In the proposed system, inter-frame feature motion is estimated using the Lucas–Kanade optical flow method [8]. From the tracked correspondences, the essential matrix is computed and decomposed to recover the relative rotation and translation between consecutive frames. However, as in any monocular formulation, the translation is only recovered up to an unknown scale factor.

To obtain metric motion estimates without resorting to stereo vision or tightly-coupled visual-inertial optimization, the proposed framework adopts a ground-plane-based scale recovery strategy [7]. The method assumes that a subset of the tracked image features belongs to the ground plane visible in front of the rover, which is a reasonable assumption for a UGV operating on mostly planar terrain. In practice, a region of interest is defined in the lower part of the image, where ground features are more likely to appear.

The selected correspondences are triangulated to obtain a local 3D point set, and a plane is then robustly fitted using RANSAC. Let d_{plane} denote the estimated perpendicular

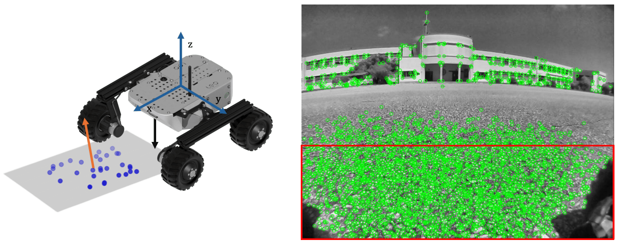

Fig. 2: Ground-plane-based scale recovery: we consider points below the red line to belong to the ground plane.

distance between the camera center and the fitted plane, and let h be the known camera height relative to the ground. The visual scale factor is computed as $s = h/d_{\mathrm{plane}}$ and is applied to the translation recovered from the essential matrix. This converts the relative monocular estimate into a metric motion increment that can be directly fused with the remaining sensor observations.

This strategy is not intended as a fully general solution to monocular scale ambiguity, but rather as a lightweight practical mechanism tailored to ground vehicles. Its main advantage is that it preserves the simplicity of the loosely-coupled architecture while reducing one of the main limitations of monocular visual odometry in field robotics.

D. Additional Motion Sources and ESKF-Based Fusion

In addition to the visual branch, the framework uses wheel odometry and LiDAR odometry as complementary motion sources. Wheel encoder data provide linear and angular velocity observations under the standard non-holonomic assumption for wheeled ground vehicles. This source is computationally inexpensive and particularly useful on structured terrain with good wheel-ground contact.

LiDAR odometry is obtained from the alignment of consecutive scans. Relative planar motion is estimated through scan matching between successive observations, following the general principle of geometric registration methods widely used in LiDAR odometry and mapping [2]. Since a 2D LiDAR is used and the vehicle motion is predominantly planar, only planar translation and yaw increment are retained for fusion.

The fusion module combines high-rate inertial propagation with asynchronous updates from wheel odometry, monocular visual odometry, and LiDAR odometry. During prediction, the IMU drives the temporal evolution of the nominal state. During correction, each available sensing modality contributes an innovation term based on the discrepancy between the predicted state and the corresponding observation.

This formulation offers two practical advantages. First, it enables direct comparison between different sensor combinations without changing the estimator structure. Second, it supports deployment in environments where the reliability of each sensing modality may vary significantly. The resulting system is therefore not intended as a tightly-coupled SLAM solution, but rather as a lightweight and reconfigurable localization framework for low-cost UGVs operating in GNSS-denied environments.

III. Experimental Setup

A. Evaluation Protocol

The proposed localization framework was evaluated through real-world experiments conducted with the Leo Rover platform in indoor and outdoor environments. In all reported experiments, the estimated trajectory was generated exclusively from onboard sensing, namely the IMU, wheel encoders, monocular camera, and 2D LiDAR. External positioning systems were used only to provide reference trajectories for performance assessment and were not included in the nominal fusion process.

For indoor experiments, the reference trajectory was obtained from a beacon-based indoor positioning system and aligned in post-processing with the rover frame by subtracting the initial position and matching the initial heading. For outdoor experiments, the reference trajectory was obtained from GNSS measurements and similarly aligned using the initial motion direction. These external measurements are used as reference trajectories for offline evaluation only; all reported estimates are produced exclusively from onboard sensing. Performance was assessed using three complementary metrics: Absolute Pose Error (APE), Root Mean Square Error (RMSE), and Relative Pose Error (RPE). For the long outdoor experiment, an additional constant bias compensation was applied to the wheel-based angular velocity measurement in order to reduce the systematic drift observed during extended runs.

B. Experimental Scenarios

Three experimental scenarios were selected to assess the proposed modular fusion architecture under distinct operating conditions.

The first scenario corresponds to a short indoor lawnmower-like trajectory, approximately 10 m long, performed in a structured laboratory environment. It includes frequent heading reversals, short straight segments, and repeated turns, making it suitable to stress the estimator under rapid orientation changes in a GNSS-denied setting.

The second scenario corresponds to a long outdoor rectangular trajectory performed on a flat asphalt area, with approximate dimensions of 18 m in length and 10 m in width. This case was selected to evaluate temporal stability over a longer path in an open environment with limited surrounding geometric structure, where accumulated drift becomes more evident.

The third scenario corresponds to an outdoor trajectory of about 12 m in length, performed on a partially vegetated terrain with shadows, irregular ground, and moving foliage. This environment was selected to assess the behavior of the localization framework under conditions that are less favorable to geometric consistency and more variable from the visual point of view.

Together, these three scenarios, depicted in Figure 3, provide a representative evaluation set spanning structured indoor maneuvering, long-duration outdoor navigation in open space, and outdoor operation in a less structured natural environment.

(a) Indoor laboratory. (b) Open area. (c) Vegetated terrain.

Fig. 3: The three experimental environments used for the localization task.

TABLE I: Indoor maneuvering scenario results.

Method	APE [m]	RMSE [m]	RPE [m]
	Isolated methods		
ODO	**0.741**	**0.741**	**0.025**
LIDAR	0.813	0.780	0.026
VO	9.150	10.801	0.555
	ESKF fusion		
IMU+ODO	**0.650**	**0.741**	**0.025**
IMU+LIDAR	0.655	0.746	0.026
IMU+VO	180.241	224.522	0.953
IMU+ODO+LIDAR	0.656	0.742	0.026
IMU+VO+ODO+LIDAR	1.123	1.178	0.028

TABLE II: Open outdoor rectangular scenario results.

Method	APE [m]	RMSE [m]	RPE [m]
	Isolated methods		
ODO	**1.245**	**1.368**	**0.097**
VO	13.034	14.628	0.103
LIDAR	28.721	33.032	0.353
	ESKF fusion		
IMU+ODO	**1.598**	**1.750**	**0.097**
IMU+VO	2.533	2.831	0.113
IMU+LIDAR	2967.917	3788.306	41.389
IMU+ODO+VO	2.973	3.374	0.097
IMU+VO+ODO+LIDAR	3.213	3.663	0.104

C. *Compared Sensor Configurations*

The experimental comparison was carried out at two levels. First, the individual motion estimation modules were evaluated independently, namely wheel odometry (ODO), monocular visual odometry (VO), and LiDAR odometry. Second, multiple fusion configurations were evaluated using the same ESKF core, with the IMU responsible for state propagation and the remaining sensing modalities incorporated as asynchronous correction sources. The reported comparisons focus on representative combinations built from the same modular architecture, namely IMU+ODO, IMU+VO, IMU+LIDAR, IMU+ODO+LIDAR, IMU+ODO+VO, and IMU+VO+ODO+LIDAR.

This evaluation strategy was designed to assess whether sensor fusion consistently improves over the best isolated modality and whether including all available sensors necessarily leads to the best performance across different environments.

IV. Preliminary Results and Discussion

A. *Indoor Maneuvering Scenario*

Figure 3a shows the indoor laboratory environment used in the lawn-mower-like maneuvering experiment. This scenario is particularly relevant because it combines a structured scene with frequent heading reversals, short straight segments, and repeated turns, making it a useful stress test for loosely-coupled fusion.

The isolated results show that wheel odometry and LiDAR odometry perform similarly well in this indoor structured environment, whereas monocular VO degrades severely. After fusion, the best mean result is obtained with IMU+ODO, with IMU+LIDAR and IMU+ODO+LIDAR remaining very close. In contrast, IMU+VO diverges dramatically, and the IMU+VO+ODO+LIDAR configuration becomes slightly worse than the best partial combinations. This behavior is important because it highlights a key limitation of the present loosely-coupled architecture: when one correction source becomes highly unreliable, adding it to the filter may degrade the global estimate rather than improve it.

B. *Open Outdoor Rectangular Scenario*

Figure 3b presents the open outdoor test area used in the long rectangular trajectory. This scenario is useful to assess temporal stability, since the longer path makes accumulated drift and systematic biases more evident than in shorter indoor runs.

In this open outdoor case, wheel odometry remains the most reliable isolated source after the empirical angular-bias compensation introduced for long runs. VO preserves a relatively low RPE but exhibits large APE and RMSE, indicating that local motion increments remain partially coherent while global drift accumulates over time. LiDAR performs worst, which is consistent with the lack of nearby geometric structure. Among the non-GNSS fused solutions, IMU+ODO is again the best configuration. Adding VO does not improve the estimate, the IMU+VO+ODO+LIDAR solution is worse than IMU+ODO, and IMU+LIDAR fails catastrophically. This scenario therefore reinforces two points: the present 2D LiDAR branch is poorly suited to open outdoor environments, and more sensors do not necessarily yield better localization.

C. *Vegetated Outdoor Scenario*

Figure 3c shows the outdoor vegetated area used in the third scenario. In this case, shadows, moving foliage, and uneven terrain create a more challenging environment for both geometric and visual perception, while also providing richer visual texture than the open rectangular run.

This scenario is arguably the most informative one. As expected, the LiDAR branch again degrades strongly, but monocular VO becomes much more useful than in the indoor maneuvering case. In fact, isolated VO remains reasonably accurate, and the best fused result is obtained with IMU+ODO+VO, which slightly outperforms

TABLE III: Vegetated outdoor scenario results.

Method	APE [m]	RMSE [m]	RPE [m]
Isolated methods			
ODO	**0.215**	**0.211**	**0.021**
VO	0.309	0.417	0.037
LIDAR	12.903	18.396	1.304
ESKF fusion			
IMU+ODO	0.416	0.472	**0.023**
IMU+VO	0.898	1.108	0.034
IMU+LIDAR	15.136	20.385	1.451
IMU+ODO+VO	**0.404**	**0.436**	**0.023**
IMU+ODO+LIDAR	0.477	0.541	0.029
IMU+VO+ODO+LIDAR	0.429	0.462	0.025

both IMU+ODO and the IMU+VO+ODO+LIDAR configuration. This is an important result for the paper because it shows that the contribution of VO is highly environment-dependent: in poor indoor visual conditions it can be harmful, whereas in textured outdoor scenes it can provide useful complementary information. It also shows that the full multi-sensor configuration is not systematically optimal.

Overall, the three scenarios support a consistent conclusion. Wheel odometry is the most stable source across all experiments, LiDAR odometry is highly dependent on environmental structure and becomes unreliable in open or dynamic outdoor scenes, and monocular VO is the most variable modality, ranging from strongly detrimental to clearly beneficial depending on feature quality and scene texture. The comparison between isolated and fused configurations is intended to assess not only the best attainable error in each scenario, but also how each sensing modality affects the estimator when environmental conditions change. Consequently, the main outcome of this working-in-progress study is not that all sensors should always be fused, but rather that a modular architecture should enable selective and context-dependent sensor usage.

V. Conclusions and Ongoing Work

This paper presented a working-in-progress modular localization framework for low-cost UGV navigation in GNSS-denied environments. The proposed system adopts a loosely-coupled architecture based on an Error-State Extended Kalman Filter, using the IMU for state propagation and wheel odometry, monocular visual odometry, and LiDAR odometry as asynchronous correction sources. A ground-plane-based scale recovery strategy was also integrated in the monocular visual odometry branch, enabling metric visual motion estimates while preserving the simplicity of the overall architecture.

The reported experiments show that the proposed framework is able to provide promising localization performance across distinct operating scenarios, but also highlight that the usefulness of each sensing modality is strongly environment-dependent. In the tested scenarios, wheel odometry was the most stable source of motion information, while LiDAR odometry was more sensitive to the availability of environmental structure. Monocular visual odometry showed the largest variability, ranging from clearly detrimental in the indoor maneuvering experiment to beneficial in the outdoor vegetated scenario.

An important outcome of this study is that the full multi-sensor configuration was not systematically the best solution. In several cases, more selective sensing combinations produced better results than fusing all available modalities. This suggests that, in a loosely-coupled framework of this type, localization performance depends not only on sensor diversity but also on the consistency of the observations provided by each modality under the current operating conditions.

At the same time, the present conclusions must be interpreted in light of the experimental conditions. The reported scenarios are predominantly planar, which naturally favors wheel odometry, since its motion model is well matched to this type of vehicle motion. In more challenging terrains involving slopes, ramps, steps, uneven ground, or other clearly non-planar motion, the assumptions underlying wheel-based odometry may become less valid. In such situations, visual and LiDAR-based motion estimates may provide more relevant complementary information and play a more important role in the fusion process.

Ongoing work will therefore focus on evaluating the proposed framework in more geometrically complex and non-planar environments, in order to better understand the limitations of the current planar formulation and the conditions under which visual and LiDAR information become more advantageous. Future developments will also address more adaptive use of the available sensing modalities, with the goal of improving robustness when observation quality changes significantly across environments.

References

[1] S. A. Mohamed, M. H. Haghbayan, T. Westerlund, J. Heikkonen, H. Tenhunen, and J. Plosila, "A survey on odometry for autonomous navigation systems," *IEEE Access*, vol. 7, pp. 97 466–97 486, 2019.

[2] J. Zhang, S. Singh *et al.*, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and systems*, vol. 2, no. 9. Berkeley, CA, 2014, pp. 1–9.

[3] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE transactions on robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[4] X. Zuo, P. Geneva, W. Lee, Y. Liu, and G. Huang, "Lic-fusion: Lidar-inertial-camera odometry," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5848–5854.

[5] T. Barata, B. Damas, V. Viegas, and V. Conceição, "A dataset for robot navigation in gnss-disrupted environments using low-cost sensors," in *2025 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2025, pp. 132–137.

[6] J. Sola, "Quaternion kinematics for the error-state kalman filter," *arXiv preprint arXiv:1711.02508*, 2017. [Online]. Available: https://arxiv.org/abs/1711.02508

[7] D. Zhou, Y. Dai, and H. Li, "Ground-plane-based absolute scale estimation for monocular visual odometry," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 791–802, 2019.

[8] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI'81: 7th international joint conference on Artificial intelligence*, vol. 2, 1981, pp. 674–679.