2025 IEEE International Conference on Robotics and Automation (ICRA)
May 19-23, 2025. Atlanta, USA

EAR-SLAM: Environment-Aware Robust Localization System for Terrestrial-Aerial Bimodal Vehicles

Wenjun He[1,3], Xingpeng Wang[1,2], Pengfei Wang[1], Tianfu Zhang[1], Chao Xu[1,2], Fei Gao[1,2] and Yanjun Cao[1,2]

***Abstract*— Terrestrial-aerial bimodal vehicles (TABVs) can fly to avoid obstacles and move safely on the ground to save energy, offering enhanced adaptability and flexibility in various challenging environments. However, a robust localization approach becomes a bottleneck to stably applying the TABVs in real-world tasks. Besides the general limitations of visual SLAM methods, large FoV differences between the two modes, abrupt motion strikes in mode transitions, and unstable attitude in ground mode pose great challenges. In this paper, we present an environment-aware robust localization system specifically designed for passive-wheel-based TABVs, which feature two passive wheels alongside a standard quadrotor. The localization system tightly integrates data from multiple sensors, including a stereo camera, Inertial Measurement Units (IMUs), encoders, and single-point laser distance sensors. First, we introduce a terrain-aware odometer model that accurately estimates terrain slope and vehicle's velocity. Then, we propose an anomaly-aware method that senses anomalous sensors and dynamically adjusts the optimization weights accordingly. By explicitly estimating the environmental conditions, such as ground terrain slopes and visual information qualities, the robot can achieve accurate and robust localization results on the ground. To validate our localization approach, we conducted extensive experiments across various challenging scenarios, demonstrating the effectiveness and reliability of our system for real-world applications.**

I. Introduction

Terrestrial-aerial bimodal vehicles (TABVs) have attracted significant attention due to their inherent benefit over single-mode robots [1]–[4]. These vehicles combine the advantages of both aerial and terrestrial modes, providing superior obstacle avoidance capabilities while flying and safe, long-term mobility on the ground. This versatility enhances their adaptability and flexibility in navigating various challenging environments, making them ideal for applications such as search and rescue, environmental monitoring, and infrastructure inspection.

Despite these advantages, a robust localization approach remains a critical challenge that hinders the application of TABVs in real-world tasks. Accurate localization is essential for autonomous navigation and task execution, especially in complex and dynamic environments where GPS signals are weak or unavailable. Simultaneous localization and mapping (SLAM) allows robots to localize themselves and perceive their environment by integrating data from various local sensors—such as cameras, LiDAR, and inertial measurement units (IMUs) [5]–[12]. Visual SLAM is popular for its compact size, light-weight, and low cost. These features make it a good choice for flying robots whose weight and size are critical.

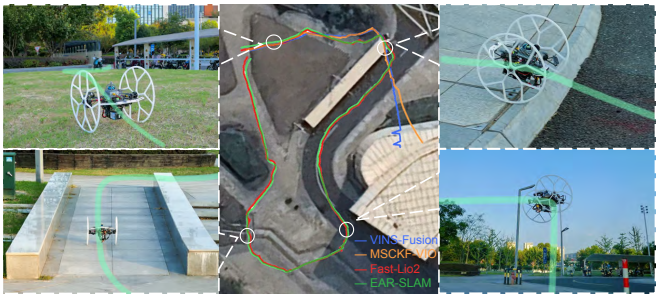


Fig. 1. An outdoor complex scene experiment, including lawn, slope, stepping down, and landing, both VINS-Fusion and MSCKF-VIO fail to localize in middle, while EAR-SLAM continues to maintain localization. Fast-Lio2 trajectories is shown for reference.

Researchers have developed the visual-inertial odometry (VIO) method either by filter-based [13], [14] or optimization-based algorithms [10]–[12]. Although, in recent years, researchers have achieved significant advancements in visual SLAM [15]–[18], the robustness of visual SLAM is still fragile in some conditions, such as illumination changes, textureless surroundings, or fast motion blurs [19], [20]. Localization divergence poses significant risks for autonomous systems, especially for flying robots that may lose control in the air. However, TABVs can leverage their ground mode to mitigate these risks, providing a safe landing and then moving on the ground. Despite the advantages of ground operation, effective localization remains essential for task execution. To enhance adaptability, we aim to develop a robust localization system for passive wheeled TABVs (pwTABVs) to improve position accuracy and stability, enabling the vehicle to navigate effectively in both aerial and terrestrial environments.

Considering the above advantages, we select the visual-inertial SLAM of the camera and ubiquitous IMUs as the base for the localization system for pwTABVs. However, the classical SLAM solutions can not meet the requirement of pwTABVs for the following challenges. **FoV changing** is a special feature of pwTABVs as the motion in the air and on the ground have totally different views for a fixed camera. The views change greatly either the camera is placed in the front or points to the sky or the ground. If it points to the ground, the camera view on the ground could cover a very small area of surroundings, resulting in a few features recognized. **Environmental conditions** are hard to guarantee in

This Work was supported by the National Natural Science Foundation of China under Grants 62103368. * Corresponding author: Yanjun Cao (yanjunhi@zju.edu.cn).
[1]Huzhou Institute of Zhejiang University, Huzhou, 313000, China.
[2]State Key Laboratory of Industrial Control Technology Institute of Cyber-Systems and Control Zhejiang University, Hangzhou, 310027,China
[3]Harbin Engineering University, Harbin, 150001, China.

2025 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3315-4139-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICRA55743.2025.11128310

real-world tasks. Illumination changes, textureless surroundings, and low light conditions easily cause the front-end to be fragile and fail. **Motion constraints** are also the limitation of the VIO approach, which suffers reduced accuracy in specific motion modes, such as stationary states, zero angular velocity motion, or constant local linear acceleration, which introduce unobserved degrees of freedom (DoFs), as discussed in [14]. Robots are particularly susceptible to these degenerate cases during ground compared to handheld mobile devices or aerial robots, which also happens for pwTABVs. Cooperating with encoders for the passive wheels of pwTABVs like ground vehicles is a choice to guarantee robot safety. However, the **unstable orientation** for the current algorithms integrate wheel encoders [21]–[23] is not applicable because of the planar-motion constraints. These studies are based on the assumption of planar motion and therefore only apply to flat terrain. To address this limitation, [24], [25] proposed to approximate the motion manifold for ground robots by a parametric representation and performing pose integration using IMU and wheel odometer measurements, but they are still limited to scenarios with minor slope variations.

To address these challenges, we propose an Environment-Aware Robust SLAM (EAR-SLAM) system for a passive wheeled TABV of our previous work [4]. We tightly integrate data from multiple sensors, including a camera, IMU, encoders, and single-point laser distance sensors. First, we introduce a terrain-aware odometer model that accurately estimates terrain slope and vehicle's velocity by fusing gyroscope, encoder, and single-point laser measurements. Then, we propose an anomaly-aware method that senses anomalous sensors and dynamically adjusts the optimization weights accordingly. By explicitly estimating the environmental conditions, such as ground terrain slope and visual information qualities, the robot can achieve accurate and robust localization results on the ground. To validate our localization approach, we conducted extensive experiments across various challenging scenarios, demonstrating the effectiveness and reliability of our system for real-world applications. The contributions of this paper are as follows:

- We propose an environment-aware robust localization system specifically designed for TABVs, capable of operating effectively in various challenging scenarios.
- We propose a terrain-aware odometer model that estimates terrain slope and TABV's velocity by fusing data from the gyroscope, encoder, and single-point laser measurements, enabling adaptation to various terrains, including flat surfaces and steep slopes.
- We propose an anomaly-aware method that detects anomalous data and dynamically adjusts the optimization weights accordingly, ensuring reliable localization performance.

II. PRELIMINARIES

A. TABV Platform

Building on our previous work on a Quadrotor-based TABV [4], which features a general quadrotor design augmented with two freely rotating passive wheels, we incorporate several additional sensors alongside the Intel Realsense D435i camera. Each wheel of the TABV is equipped with an 18-bit resolution encoder to measure the rotation degrees relative to the body frame. Additionally, three single-point (SP) laser distance sensors are integrated into the robot frame pointing to the ground. The robot's structure and the installation locations for all these sensors are illustrated in Fig. 2.

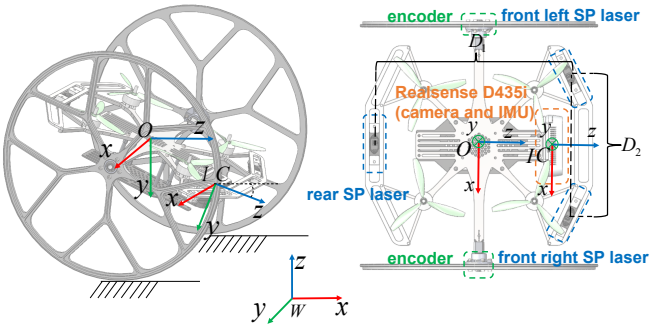


Fig. 2. Camera, IMU and odometer coordinate frames.

B. Notations

In this work, we assume that a TABV navigates in a reference world coordinate system $\{W\}$. We use $\{C\}$, $\{I\}$ and $\{O\}$ to denote the coordinate frames of the stereo camera, IMU and wheel odometer, respectively. The origin of $\{O\}$ is located at the center of the two wheels, and the vehicle moves along the positive z-axis. The $\{O\}$ coordinate system remains consistently parallel to the ground where the TABV is currently located.

III. METHODOLOGY

A. System overview

A terrain-aware odometer model (Section III-B) estimates the terrain slope and the TABV's velocity to enhance localization accuracy. Building on this, we perform terrain-aware odometer pre-integration (Section III-C). Finally, as described in Section III-D, the system incorporates an anomaly-aware method to identify faulty sensor readings and

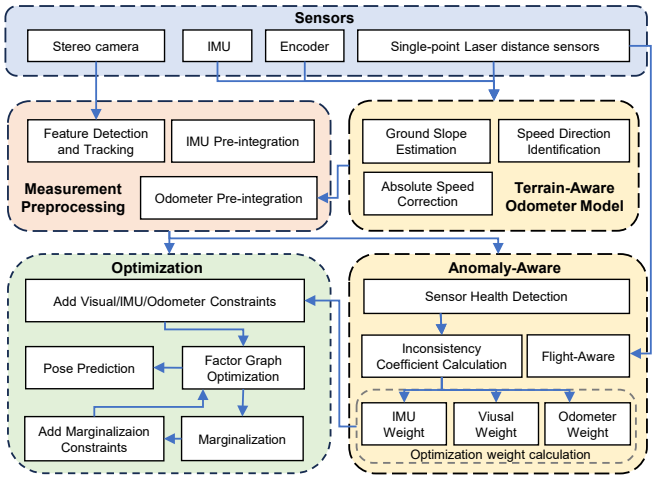


Fig. 3. System Overview.

dynamically adjust the optimization weights, ensuring robust performance in challenging environments.

We use the bundle adjustment (BA) to minimize the residuals of all observations and obtain the maximum a posteriori estimate of the robot state:

$$\boldsymbol{\chi}^* = \arg\min\{\|\mathbf{r}_p - \mathbf{H}_p\boldsymbol{\chi}\|^2 + \sum_{k\in I}\|w_I\mathbf{r}_I(\hat{\mathbf{z}}_{I_j}^{I_i},\boldsymbol{\chi})\|^2_{\mathbf{p}_{I_j}^{I_i}} + \sum_{k\in O}\|w_o\mathbf{r}_o(\hat{\mathbf{z}}_{o_j}^{o_i},\boldsymbol{\chi})\|^2_{\mathbf{p}_{o_j}^{o_i}} + \sum_{(l,j)\in C}\rho(\|w_c\mathbf{r}_c(\hat{\mathbf{z}}_l^{c_j},\boldsymbol{\chi})\|^2_{\mathbf{p}_l^{c_j}})\} \tag{1}$$

where $\mathbf{r}_o(\hat{\mathbf{z}}_{o_j}^{o_i},\boldsymbol{\chi})$ is the residual of terrain-aware odometer, which will be elaborated in Eq. 12. Other residuals are the same as in VINS-Fusion. w_I, w_o, w_c are the optimization weights of the IMU, terrain-aware odometer and camera, respectively, which can be obtained from Eq. 15.

B. Terrain-Aware Odometer Model

1) ***Ground Slope Estimation****:* With the three single-point laser distance sensors equipped on the TABV, we can obtain distance measurements from three critical positions: the left front, right front, and rear of the vehicle to the ground. Through geometric relationships, we can determine the rotation relationship between the plane of the vehicle body and the current ground surface:

$$\begin{aligned}\mathbf{R}_{oI_k}(\psi_k,\phi_k) &= \mathbf{R}_x(\psi_k)\mathbf{R}_z(\phi_k)\\ &= \begin{bmatrix} cos(\psi_k) & 0 & 0\\ cos(\phi_k)sin(\psi_k) & cos(\phi_k)cos(\psi_k) & -sin(\phi_k)\\ sin(\phi_k)sin(\psi_k) & sin(\phi_k)cos(\psi_k) & cos(\phi_k)\end{bmatrix}\end{aligned} \tag{2}$$

with

$$\begin{aligned}\phi_k &= \arctan(\frac{L_b - \frac{1}{2}(L_l+L_r)}{D_1})\\ \psi_k &= \arcsin(\frac{L_r - L_l}{D_2})\end{aligned} \tag{3}$$

where L_l, L_r and L_b are the distances from the ground to the front left, front right and rear of the vehicle respectively. D_1 is the distance from the rear single-point laser distance sensor to the midpoint between the two front single-point laser distance sensors. D_2 is the distance between the two front laser distance sensors. $\mathbf{R}_{oI_k}(\psi_k,\phi_k)$ is the rotation transformation from the IMU body frame to the odometer frame. ϕ_k, ψ_k are the roll and pitch between odometer frame and ground, respectively.

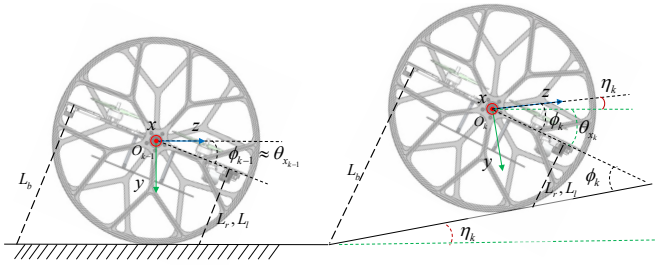


Fig. 4. Illustration of the slope calculation of the terrain on which the TABV is located.

As shown in Fig. 4, we can derive the slope of the ground on which the vehicle is currently standing as:

$$\eta_k = \phi_k - \theta_{x_k} \tag{4}$$

where ϕ_k is the angle between the ground and the TABV, which can be obtained by the single-point laser distance sensors through Eq. 3. θ_{x_k} is the angle between the body and the horizontal plane, which can be obtained by the gyroscope.

2) ***Absolute Speed Correction****:* Due to the passive nature of the wheels on our TABV, any relative motion between the central body and the wheels can be detected with the encoder sensors. Note that the part of wheel speed does not accurately reflect the true overall speed of the vehicle because of the TABV's pitching motion. Therefore, to obtain the accurate speed of the entire vehicle, it is essential to correct the speed obtained directly from the encoder by removing the component generated by the vehicle's pitching motion.

$$v_{o_k} = v_{e_k} - (\omega_{g_p k} - b_{g_p})r \tag{5}$$

where v_e is the speed measured by the encoder, ω_{g_p} is the pitching angular velocity measured by the gyroscope. b_{g_p} is the pitching gyroscope bias. r is the wheel radius. Note that the speed v_o here is a scalar, not a vector.

3) ***Speed Direction Identification****:* We can obtain the current direction vector of the overall vehicle motion from the single-point laser distance sensors and IMU gyroscope, which we use as the direction of the odometer speed.

$$\mathbf{e}_{\mathbf{o_{k-1}o_k}} = \begin{bmatrix}\cos(\delta\phi_k - \delta\theta_{p_k})\sin(-\delta\varphi_{o_k})\\ \sin(\delta\theta_{p_k} - \delta\phi_k)\\ \cos(\delta\phi_k - \delta\theta_{p_k})\cos(\delta\varphi_{o_k})\end{bmatrix} \tag{6}$$

with

$$\begin{aligned}\varphi_{o_k} &= \cos(\psi_k)\sin(\phi_k)\theta_{x_k}\\ &\quad + \cos(\psi_k)\cos(\phi_k)\theta_{y_k} - \sin(\psi_k)\theta_{z_k}\end{aligned} \tag{7}$$

where θ_x, θ_y and θ_z represent the angles measured by the gyroscope around the x, y, and z axes, respectively.

Finally, the velocity of the odometer can be represented as:

$$\mathbf{v}_{\mathbf{o_{k-1}o_k}} = \mathbf{e}_{\mathbf{o_{k-1}o_k}} v_{o_k} \tag{8}$$

C. Terrain-Aware Odometer Pre-integration

We use gyroscopes to obtain the rotation transformation from the IMU body frame to the world frame. The discrete odometer pre-integration model between two time instants (corresponding to two image frames) can be constructed as follows:

$$\begin{aligned}\boldsymbol{\alpha}_{I_i o_{k+1}} &= \boldsymbol{\alpha}_{I_i o_k} + \mathbf{R}(\boldsymbol{\gamma}_{I_i I_k})\mathbf{R}_{oI_k}(\psi_k,\phi_k)\mathbf{v}_{o_k o_{k+1}}\delta t\\ \boldsymbol{\gamma}_{I_i o_{k+1}} &= \boldsymbol{\gamma}_{I_i b_{k+1}} \otimes \begin{bmatrix}1\\ \frac{1}{2}(\boldsymbol{\omega}_{g_k} - \mathbf{b}_{g_k})\delta t\end{bmatrix}\end{aligned} \tag{9}$$

where k is discrete moment corresponding to an odometer measurement within $[t_i, t_j]$. δt is the time interval between two odometer measurements k and $k+1$. For ease of writing, we next refer to $\mathbf{R}_{oI_k}(\psi_k,\phi_k)$ by $\mathbf{R}_{oI_k}$, which can be obtained from Eq. 2.

Since the odometer is modeled in discrete time, we directly derive discrete-time dynamics of error terms of (9) as

follows:

$$\begin{bmatrix}\delta\boldsymbol{\alpha}_{k+1}\\ \delta\boldsymbol{\theta}_{k+1}\\ \delta\mathbf{b}_{g_{k+1}}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & f_{01} & f_{02}\\ 0 & \mathbf{I}-[\boldsymbol{\omega}_{g_k}-\mathbf{b}_{g_k}]_\times\delta t & -\mathbf{I}\delta t\\ 0 & 0 & \mathbf{I}\end{bmatrix}\begin{bmatrix}\delta\boldsymbol{\alpha}_k\\ \delta\boldsymbol{\theta}_k\\ \delta\mathbf{b}_{g_k}\end{bmatrix} + \begin{bmatrix}v_{00} & v_{01} & 0 & v_{03} & v_{04} & v_{05}\\ 0 & 0 & \frac{1}{2}\mathbf{I}\delta t & \frac{1}{2}\mathbf{I}\delta t & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \mathbf{I}\delta t\end{bmatrix}\begin{bmatrix}\mathbf{n}_{v_k}\\ \mathbf{n}_{\psi_k}\\ \mathbf{n}_{\omega_k}\\ \mathbf{n}_{\omega_{k+1}}\\ \mathbf{n}_{e_k}\\ \mathbf{n}_{b_{g_k}}\end{bmatrix} \tag{10}$$

with

$$\begin{aligned} f_{01} &= -\mathbf{R}_{I_iI_k}\left\lfloor\mathbf{R}_{I_ko_k}\mathbf{v}_{o_ko_{k+1}}\right\rfloor_\times\delta t\\ f_{02} &= r\mathbf{R}_{I_iI_k}\mathbf{R}_{I_ko_k}\mathbf{e}_{o_ko_{k+1}}\delta t\\ v_{00} &= \mathbf{R}_{I_iI_k}\mathbf{R}_{I_ko_k}\mathbf{e}_{o_ko_{k+1}}\delta t\\ v_{01} &= -\mathbf{R}_{I_iI_k}\mathbf{R}_{I_ko_k}\left\lfloor\mathbf{v}_{o_ko_{k+1}}\right\rfloor_\times\delta t\\ v_{03} &= r\mathbf{R}_{I_iI_k}\mathbf{R}_{I_ko_k}v_{o_{k+1}}\delta t\\ v_{04} &= \mathbf{R}_{I_iI_k}\mathbf{R}_{I_ko_k}v_{o_{k+1}}\delta t\\ v_{05} &= r\mathbf{R}_{I_iI_k}\mathbf{R}_{I_ko_k}\mathbf{e}_{o_ko_{k+1}}\delta t^2 \end{aligned} \tag{11}$$

where $\mathbf{n}_v$, $\mathbf{n}_\psi$, $\mathbf{n}_w$, $\mathbf{n}_e$, $\mathbf{n}_{b_g}$ are the noise of the velocity, the rotation between the odometer and the IMU, the gyroscope, the direction vector of the odometer speed, and bias of the gyroscope.

Consider the odometer measurements within two consecutive frames i and j in the sliding window, the residual for pre-integrated odometer measurement can be defined as:

$$\begin{aligned}\mathbf{r}_o(\hat{\mathbf{z}}^{o_i}_{o_j},\boldsymbol{\chi}) &= \begin{bmatrix}\mathbf{r}_p\\ \mathbf{r}_q\\ \mathbf{r}_{b_g}\end{bmatrix} = \begin{bmatrix}\mathbf{q}^*_{wI_i}(\mathbf{p}_{wo_j}-\mathbf{p}_{wo_i})-\boldsymbol{\alpha}_{I_io_j}\\ 2[\boldsymbol{\theta}^*_{I_iI_j}\otimes(\mathbf{q}^*_{wI_i}\otimes\mathbf{q}_{wI_j})]_{xyz}\\ \mathbf{b}_{g_j}-\mathbf{b}_{g_i}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{R}_{I_iw}(\mathbf{R}_{wI_j}\mathbf{p}_{Io}+\mathbf{p}_{wI_j}-\mathbf{R}_{wI_i}\mathbf{p}_{Io}-\mathbf{p}_{wI_i})-\boldsymbol{\alpha}_{I_io_j}\\ 2[\boldsymbol{\theta}^*_{I_iI_j}\otimes(\mathbf{q}^*_{wI_i}\otimes\mathbf{q}_{wI_j})]_{xyz}\\ \mathbf{b}_{g_j}-\mathbf{b}_{g_i}\end{bmatrix}\end{aligned} \tag{12}$$

where $[\cdot]_{xyz}$ represents the operation of extracting the vector part of a quaternion, used in error-state representation.

D. Anomaly-Aware Method

*1) **Sensor Health Detection**:* Inspired by [26], we determine whether the sensor is abnormal by sensing the consistency of the sensor.

We use the IMU and odometer measurements to calculate the respective pre-integrated values $\mathbf{p}^I_{I_iI_j}$, $\mathbf{p}^o_{I_iI_j}$ between the two frames i and j, respectively, while using the PnP algorithm to calculate the position increment obtained from the visual matching $\mathbf{p}^c_{I_iI_j}$. The mean and variance of these three position increments can be calculated as follows:

$$\begin{aligned}\boldsymbol{\mu} &= \frac{\mathbf{p}^I_{I_iI_j}+\mathbf{p}^o_{I_iI_j}+\mathbf{p}^c_{I_iI_j}}{3}\\ \sigma^2 &= \frac{\|\mathbf{p}^I_{I_iI_j}-\boldsymbol{\mu}\|^2+\|\mathbf{p}^o_{I_iI_j}-\boldsymbol{\mu}\|^2+\|\mathbf{p}^c_{I_iI_j}-\boldsymbol{\mu}\|^2}{3}\end{aligned} \tag{13}$$

If the variance σ^2 is over a threshold, we consider the sensors to be in poor consistency. In other words, one sensor might be anomalous and inconsistent with others. The threshold value is determined by experimental approach.

*2) **Inconsistency Coefficient Calculation**:* The inconsistency coefficient for each sensor can be calculated by the following equation:

$$\begin{aligned}\lambda_I &= \|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|+\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^c_{I_iI_j}\|\\ \lambda_o &= \|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|+\|\mathbf{p}^c_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|\\ \lambda_c &= \|\mathbf{p}^c_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|+\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^c_{I_iI_j}\|\end{aligned} \tag{14}$$

where a larger inconsistency coefficient represents a larger difference from other sensor measurements.

*3) **Optimazation Weight Calculation**:* We consider the sensor with the highest inconsistency coefficient among the three sensors to be an anomalous sensor and adjusted its optimization weights according to the weighting function as follows.

$$\begin{aligned}w_I &= \frac{\|\mathbf{p}^c_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|}{2\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|}+\frac{\|\mathbf{p}^c_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|}{2\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^c_{I_iI_j}\|}\\ w_o &= \frac{\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^c_{I_iI_j}\|}{2\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|}+\frac{\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^c_{I_iI_j}\|}{2\|\mathbf{p}^c_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|}\\ w_c &= \frac{\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|}{2\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^c_{I_iI_j}\|}+\frac{\|\mathbf{p}^I_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|}{2\|\mathbf{p}^c_{I_iI_j}-\mathbf{p}^o_{I_iI_j}\|}\end{aligned} \tag{15}$$

Note that when the PnP algorithm fails to estimate the position increment, we fully rely on the IMU and Odometer, and w_I, w_o are set to one.

Simultaneously, the single-point laser distance sensors detect the flight status of the TABV. If the measurements from all three single-point laser distance sensors exceed the wheel radius, the TABV is considered to be in flight, and the odometer weight w_o is set to zero.

IV. EXPERIMENTS

To evaluate our work, we collected fifteen datasets in various environments, including flat terrain, uneven terrain, different modes of motion, camera occlusion, and during take-off and landing. Fig. 5 illustrates some of these environmental conditions. We used the NOKOV motion capture system to collect ground truth data for the TABV in indoor environments. For outdoor environments, we used the results from FAST-LIO2 [8] with high-accuracy LiDAR as the ground truth. All the experiments presented here are performed on an Intel NUC12WSK-i7.

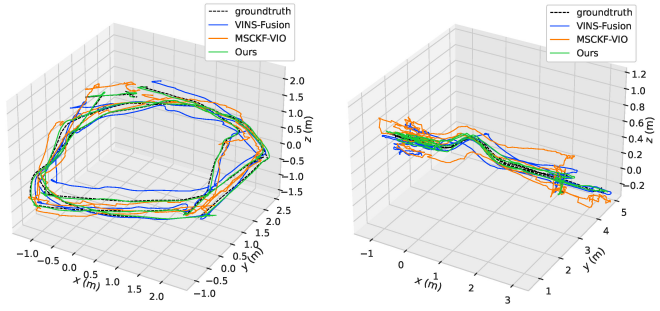
A. Positioning Accuracy

In the first experiment, we evaluate the positioning accuracy of the proposed method by comparing it with VINS-Fusion [27] and MSCKF-VIO [28], which combine the stereo camera and IMU. In the sequences $Flat$, the TABV moves on flat indoor terrain, which presents challenges due to unobservable DoFs. In the $Wave$ sequence, we constructed an indoor wave ramp where the TABV traverses an uneven surface, experiencing occasional wheel slippage. The $Slope$ sequence involves the TABV navigating a steep outdoor incline. In the $Lawn$ sequence, the TABV traverses an outdoor lawn. During the $Wave$ and $Slope$ sequences, FoV changes due to the TABV's aggressive motion control on steep slopes,

Fig. 5. Top: the actual scenarios corresponding to *wave*, *slope*, *lawn*, and *occlusion*, respectively. Bottom: the camera view for the scene.

leading to a temporary loss of visual features. Similarly, in the *Lawn* sequence, tall grass obstructs the camera's view. When the camera is directed toward the ground, its narrow field of view limits the number of recognized features, impacting the SLAM system's performance.

As shown in Fig. 6 and Table I, compared to the VIO methods, our method presents higher localization accuracy in both indoor and outdoor environments, as well as on flat and uneven terrain. The percentages represent the localization accuracy, calculated as the ratio of the RMSE to the total trajectory length.



(a) Flat Terrain (b) Wave Ramp

Fig. 6. Estimated trajectories and ground truth trajectories of EAR-SLAM, VINS-Fusion and MSCKF-VIO in flat terrain and wave ramp.

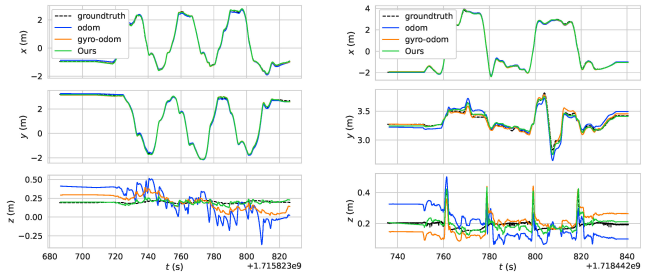
TABLE I

RMSE(M) AND ACCURACY COMPARISON

Seq.	VINS-Fusion		MSCKF-VIO		Ours	
Flat1	0.308	0.782%	0.249	0.633%	**0.089**	**0.227%**
Flat2	0.350	0.738%	0.269	0.567%	**0.062**	**0.130%**
Flat3	0.252	0.568%	0.255	0.573%	**0.098**	**0.221%**
Wave1	0.122	0.318%	0.323	0.814%	**0.097**	**0.246%**
Wave2	0.263	0.711%	0.241	0.653%	**0.055**	**0.148%**
Wave3	0.206	0.728%	0.313	1.107%	**0.071**	**0.252%**
Slope	0.214	0.402%	0.489	0.815%	**0.131**	**0.252%**
Lawn	1.177	1.935%	0.871	1.145%	**0.198**	**0.308%**
Fly1	0.156	0.470%	0.267	0.807%	**0.119**	**0.360%**
Fly2	0.406	0.690%	0.429	0.728%	**0.319**	**0.542%**

To further evaluate the terrain-aware odometer model, we conducted ablation experiments. We implemented a two-wheel differential odometer model (**odom**) as a baseline method and tested it on the same sequences. Results show that coupling VIO with this model improves positioning accuracy on flat terrain but is ineffective on non-flat terrain. Replacing the wheel's angular velocity with the gyroscope's angular velocity (**gyro-odom**) enhanced localization accuracy and extended applicability to non-flat terrain. However, even with **gyro-odom**, z-axis drift persists on flat terrain lacking sufficient z-axis excitation, unless planar-motion constraints are incorporated.

As shown in Fig. 7 and Table II, our method directly senses the terrain information, so it not only has estimated trajectories in the x-y plane but also estimates the z-axis position and suppresses the z-axis drift, especially when the z-axis excitation is not sufficient.



(a) Flat Terrain (b) Wave Ramp

Fig. 7. Estimated and ground truth trajectories of the proposed approach and other odometer model in flat terrain and wave ramp on the x-y plane.

TABLE II

RMSE(M) OF ODOMETER MODEL COMPARISON

Seq.	RMSE(m)	VINS-Fusion	odom	gyro-odom	Ours
Flat1	pos. err.	0.308	0.160	0.113	**0.089**
	z err.	0.109	0.076	0.079	**0.039**
Flat2	pos. err.	0.350	0.210	0.119	**0.062**
	z err.	0.206	0.078	0.097	**0.022**
Flat3	pos. err.	0.252	0.148	0.143	**0.098**
	z err.	0.199	0.056	0.059	**0.028**
Wave1	pos. err.	0.122	0.140	0.128	**0.097**
	z err.	0.047	0.072	0.041	**0.031**
Wave2	pos. err.	0.263	0.120	0.083	**0.055**
	z err.	0.070	0.088	0.054	**0.024**
Wave3	pos. err.	0.218	0.161	0.143	**0.071**
	z err.	**0.033**	0.060	0.038	0.039
Slope	pos. err.	0.214	0.235	0.150	**0.131**
	z err.	0.097	0.180	0.096	**0.061**
Lawn	pos. err.	1.177	0.643	0.524	**0.198**
	z err.	**0.056**	0.367	0.089	0.060

In addition, to verify the effect of VIO degradation when the TABV moves on the ground on the overall positioning accuracy, we conducted experiments on sequences *Fly* containing both flying and rolling modes. As shown in Table I and Fig. 8, our method enhances overall localization accuracy and improved z-axis estimation.

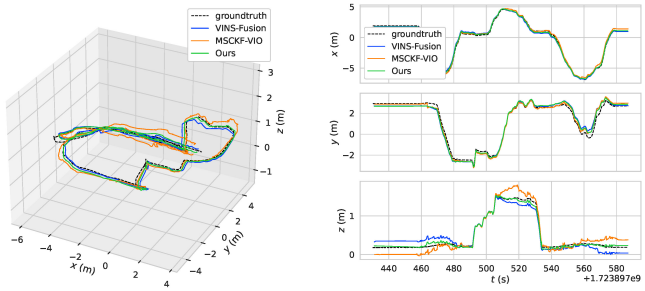


Fig. 8. Estimated and ground truth trajectories of EAR-SLAM, VINS-FUsion and MSCKF-VIO in *Fly*2 containing both flying and rolling modes.

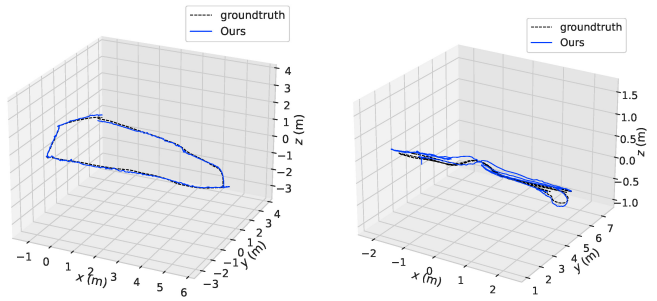
B. Localization Robustness

The following experiments evaluate the robustness of the proposed method in various corner cases.

*1) **Visual Occlusion**:* Localization drift or failure can occur in visual SLAM due to the absence of valid feature points. In the sequences *Occlusion*1 and *Occlusion*2, the TABV moves on flat terrain and wave ramp, respectively, with the camera fully occluded throughout.

As shown in Table III and Fig. 9, in extreme visual occlusion scenarios, both VINS-Fusion and MSCKF-VIO

fail to localize, while our method maintains localization capability.



(a) Flat Terrain (b) Wave Ramp

Fig. 9. Estimated trajectories and ground truth trajectories with full camera occlusion, both VINS-Fusion and MSCKF-VIO fail to localize.

TABLE III

RMSE(M) COMPARISON FOR VISUAL CHALLENGES

		VINS-Fusion	MSCKF-VIO	Ours
Occlusion1		—	—	**0.245**
Occlusion2		—	—	**0.337**
Downsampled Flat3				
15Hz	no LC	0.215	0.291	**0.186**
	LC	0.182	—	**0.097**
10Hz	no LC	1.223	—	**0.239**
	LC	0.843	—	**0.143**
5Hz	no LC	—	—	**0.499**
	LC	—	—	**0.373**

—: Fails due to severe drift LC: Loop closure is enabled

*2) **Low Camera Frame Rates**:* To evaluate performance under computationally constrained scenarios where camera frame rates might be reduced, we downsampled the Flat3 sequence. Note that the image frequency in the algorithm is much lower than the sampling frequency to ensure the synchronization of stereo camera images.

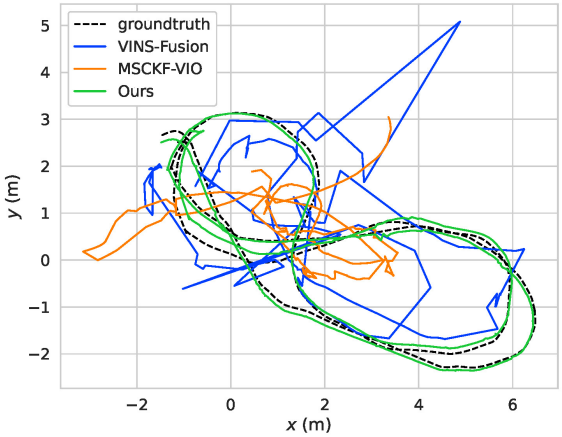


Fig. 10. Estimated and ground truth trajectories in the $Flat3$ of a camera frequency of 10 Hz on the x-y plane.

As shown in Fig. 10 and Table. III, when the camera frame rate is reduced to 10Hz, both VINS-Fusion and MSCKF-VIO fail to localize. However, our method not only maintains localization but also achieves high accuracy after loop closure.

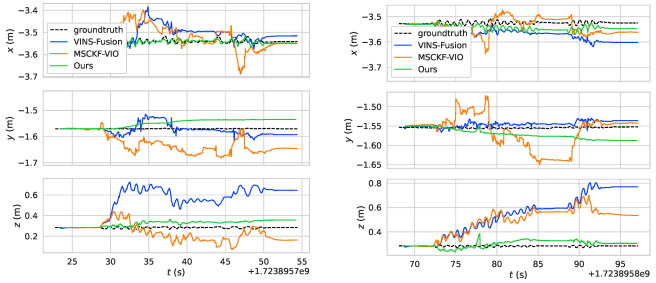
*3) **Visual Matching Failure**:* We fixed the wheels and applied an external force to make its body sway around the wheel axis for about the 30s. During this time, the overall TABV did not have a translation move, and the camera only rotated without translation, which is a challenge for the triangulation process in the visual front-end. Additionally, in the $Sway3$, we allowed the wheels to spin idle for a period of time.

TABLE IV

START-TO-END ERROR(M) FOR VISUAL MATCHING SEQUENCE

Seq.	groundtruth	VINS-Fusion	MSCKF-VIO	Ours
Sway1	0.001	0.371	0.144	**0.081**
Sway2	0.003	0.500	0.252	**0.043**
Sway3	0.023	0.528	0.508	**0.051**

Table IV shows the distance between the start and end points of the estimated trajectory for each method. Our method demonstrates the low trajectory drift compared to the other methods, as detailed in Fig. 11. This result further demonstrates the feasibility of using the IMU's pitch angle to compensate the encoder for accurate velocity estimation.



(a) Sway1 (b) Sway2

Fig. 11. Estimated trajectories and ground truth trajectories in sequences $sway$ with only rotation and no translation.

*4) **Anomaly-Aware**:* To validate the effect of anomaly-aware, we conducted ablation experiments. We enabled and disabled anomaly-aware on the downsampled $Flat3$ and $Sway$, respectively.

TABLE V

RMSE(M) IN FLAT3 AND START-TO-END ERROR(M) IN SWAY.

Seq.	no Anomaly-Aware	Anomaly-Aware
Flat3 15Hz	0.231	**0.186**
Flat3 10Hz	0.321	**0.239**
Flat3 5Hz	0.754	**0.499**
Sway1	0.121	**0.081**
Sway2	0.055	**0.043**
Sway3	0.499	**0.051**

The results show that the anomalous data can be detected and separated from the optimization process after enabling anomaly-aware, thereby improving positioning accuracy and system stability, as shown in the Table V.

V. CONCLUSION

We propose EAR-SLAM dedicated to TABVs in complex scenarios, which senses information in environment to estimate the state. Specifically, our approach includes a terrain-aware odometer model that senses terrain slope and vehicle's velocity by fusing gyroscope, encoder, and single-point laser measurements to estimate position accurately. In addition, we propose an anomaly-aware method based on sensor consistency that senses anomalous sensor data and adjusts the optimization weights accordingly. Finally, we propose conduct extensive experiments in the real world. The results show that EAR-SLAM performs better than alternative state-of-the-art methods in terms of both accuracy and robustness.

REFERENCES

[1] J. Yang, Y. Zhu, L. Zhang, Y. Dong, and Y. Ding, "Sytab: A class of smooth-transition hybrid terrestrial/aerial bicopters," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9199–9206, 2022.

[2] D. D. Fan, R. Thakker, T. Bartlett, M. B. Miled, L. Kim, E. Theodorou, and A.-a. Agha-mohammadi, "Autonomous hybrid ground/aerial mobility in unknown environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 3070–3077.

[3] J. Lin, R. Zhang, N. Pan, C. Xu, and F. Gao, "Skater: A novel bi-modal bi-copter robot for adaptive locomotion in air and diverse terrain," *IEEE Robotics and Automation Letters*, vol. 9, no. 7, pp. 6392–6399, 2024.

[4] R. Zhang, Y. Wu, L. Zhang, C. Xu, and F. Gao, "Autonomous and Adaptive Navigation for Terrestrial-Aerial Bimodal Vehicles," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3008–3015, 2022.

[5] J. Zhang and S. Singh, "Low-drift and real-time lidar odometry and mapping," *Autonomous Robots*, vol. 41, no. 2, pp. 401–416, 2017.

[6] T. Shan and B. Englot, "LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4758–4765.

[7] Weikun Zhen and Sebastian Scherer, "Estimating the Localizability in Tunnel-like Environments using LiDAR and UWB," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 4903–4908.

[8] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "FAST-LIO2: Fast Direct LiDAR-Inertial Odometry," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2053–2073, 2022.

[9] C. Bai, T. Xiao, Y. Chen, H. Wang, F. Zhang, and X. Gao, "Faster-LIO: Lightweight Tightly Coupled Lidar-Inertial Odometry Using Parallel Sparse Incremental Voxels," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4861–4868, 2022.

[10] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[11] 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.

[12] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[13] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3565–3572.

[14] A. Martinelli, "Closed-Form Solution of Visual-Inertial Structure from Motion," *International Journal of Computer Vision*, vol. 106, no. 2, pp. 138–152, 2014.

[15] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 15–22.

[16] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[17] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[18] J. Engel, V. Koltun, and D. Cremers, "Direct Sparse Odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, 2018.

[19] H. A. G. C. Premachandra, R. Liu, C. Yuen, and U.-X. Tan, "UWB Radar SLAM: An Anchorless Approach in Vision Denied Indoor Environments," *IEEE Robotics and Automation Letters*, vol. 8, no. 9, pp. 5299–5306, 2023.

[20] S. Park, T. Schöps, and M. Pollefeys, "Illumination change robustness in direct visual slam," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 4523–4530.

[21] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, "VINS on wheels," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 5155–5162.

[22] F. Zheng and Y.-H. Liu, "Visual-Odometric Localization and Mapping for Ground Vehicles Using SE(2)-XYZ Constraints," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 3556–3562.

[23] J. Yin, A. Li, W. Xi, W. Yu, and D. Zou, "Ground-fusion: A low-cost ground slam system robust to corner cases," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 8603–8609.

[24] M. Zhang, Y. Chen, and M. Li, "Vision-aided localization for ground robots," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 2455–2461.

[25] Y. Su, T. Wang, C. Yao, S. Shao, and Z. Wang, "GR-SLAM: Vision-Based Sensor Fusion SLAM for Ground Robots on Complex Terrain," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5096–5103.

[26] A. W. Palmer and N. Nourani-Vatani, "Robust odometry using sensor consensus analysis," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3167–3173.

[27] T. Qin, J. Pan, S. Cao, and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," 2019, arXiv:1901.03638 [cs].

[28] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust stereo visual inertial odometry for fast autonomous flight," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 965–972, 2018.