2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

# STEAM-LIVO: Spatio-Temporally Adaptive Manifold Lidar-Inertial-Visual Odometry for Sensor Degradation in Unstructured Natural Aquatic-Terrestrial Scenes

Yubo Guo[1,2,3], Gang Peng[1,2,3,*], Jialuo Li[1,2,3] and Hai-Tao Zhang[1,2,3,*]

***Abstract*— Sensor degradation in unstructured natural environments—manifesting as LiDAR point cloud sparsity or visual feature dropout—and out-of-sequence measurement challenges critically undermine localization robustness in autonomous systems. To address these limitations, we present STEAM-LIVO, a Spatio-Temporally Adaptive Manifold LiDAR-Inertial-Visual Odometry framework that enables tightly coupled multi-sensor fusion via a spatio-temporal manifold-driven iterative Kalman filter. The proposed method formulates an error-state iterative update mechanism on Lie group manifolds, executes IMU-centric real-time estimation, and ensures resilience under sensor degradation through an incremental observation model integrating LiDAR point-to-plane geometric residuals with visual feature reprojection errors within a shared filtering framework. Comprehensive evaluations in vegetated terrestrial landscapes and dynamic aquatic surfaces demonstrate an average relative pose error of 1.77%, with sustained robustness during partial sensor failures. Rigorous ablation studies further corroborate the efficacy of our spatio-temporal adaptive manifold architecture. Our implementation is publicly available and can be accessed at https://github.com/STEAM-LIVO/STEAM-LIVO.git.**

**Index Terms—unstructured natural scenes; sensor failure; SLAM; out-of-sequence measurement**

## I. INTRODUCTION

Autonomous systems have evolved from theoretical frameworks to multi-domain deployments, extending beyond roadways to aerial and aquatic environments with applications spanning wilderness reconnaissance and hazardous rescue missions.

While modern single-sensor SLAM implementations (visual [1],[2] or LiDAR-based [3],[4],[5],[6]) achieve decimeter-level precision in static settings, inherent limitations persist: visual odometry degrades under lighting changes, rapid motion, and dynamic scenes, whereas LiDAR systems falter in geometrically degenerate or feature-scarce terrains. These challenges motivate multi-sensor fusion research [7]. Current fusion methods achieve centimeter-level precision in structured settings but fail to address the spatiotemporal complexities intrinsic to natural environments.

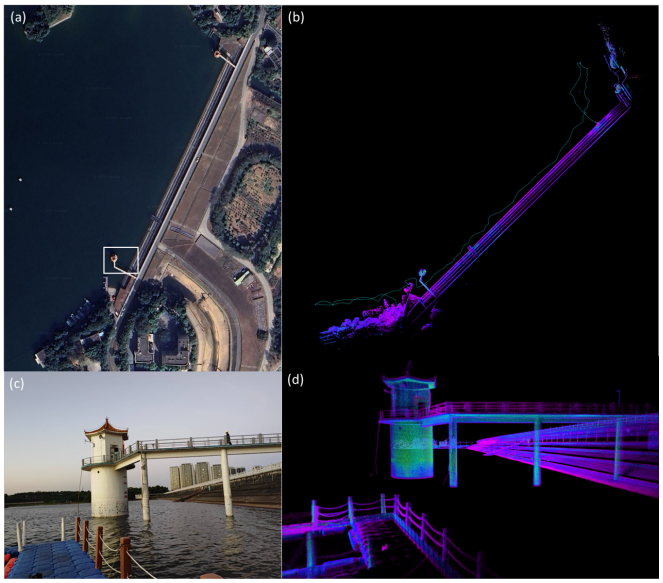


Fig. 1. Illustration of the mapping results of our proposed STEAM-LIVO system. Our method maintains robust pose estimation over 30-minute operations spanning 1km in extreme unstructured environments with invalid sensor features and aggressive motion. (a) Satellite map, (b) Our reconstructed map, (c) real-world view, (d) Map detail comparison

Wilderness environments pose three interrelated challenges for SLAM systems: geometric sparsity from feature-deprived terrains, transient scene dynamics, and extreme photometric distortions. These conditions induce cross-modal sensor degradation—manifesting as GNSS signal deprivation, visual feature loss, and LiDAR point cloud sparsification (Fig. 2)—fundamentally constraining the operational persistence of unmanned platforms. Existing methodologies predominantly target terrestrial wilderness environments [8], [9], [10], [11], yet fail to address sensor degradation in complex natural settings that encompass aquatic domains.

Existing pose estimation methods primarily adopt two frameworks: filter-based propagation and optimization-based integration. Both model IMU kinematics through either recursive state prediction (filtering) or pre-integrated motion constraints (optimization), sharing mathematical equivalence in their inertial formulations. These sequential processing paradigms estimate motion states only upon receiving new measurements, creating inherent delays and measurement

[1] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China
[2] Key Laboratory of Image Information Processing and Intelligent Control, Ministry of Education, Wuhan, China
[3] Engineering Research Center of Autonomous Intelligent Unmanned System, Ministry of Education, Wuhan, China
* Corresponding authors E-mail: penggang@hust.edu.cn, zht@mail.hust.edu.cn

This research is supported by Hubei Province Core Technology Application Research Project for Agricultural Machinery Equipment(No. HBSNYT202213) and Hubei Province Natural Science Foundation of China (No. 2019CFB526).

dependency—sensor failures cause estimation suspension and accumulated drift. While Super Odometry [12] introduces parallel IMU-VIO/LIO fusion to enhance robustness, its independent per-sensor optimization permits conflicting state estimates that degrade overall consistency.

We propose a LiDAR-inertial-visual odometry framework for real-time pose estimation in unstructured harsh environments, with the following key contributions:

- We propose an IMU-centric tightly-coupled LiDAR-camera-IMU framework that integrates LiDAR point-to-plane geometric residuals and visual feature reprojection errors within a shared iterative error-state Kalman filter on Lie group manifolds. Unlike FAST-LIVO2's serial two-stage Kalman update for LiDAR and camera, our unified filtering framework processes all modalities jointly. The visual subsystem directly extracts FAST corners and tracks them via KLT optical flow, rather than relying on LiDAR-projected points, enabling independent visual operation under LiDAR degradation.
- We formulate multi-sensor temporal misalignment as an Out-of-Sequence Measurement (OOSM) problem and develop a manifold-based historical state retrieval and direct correction mechanism: by maintaining a state-covariance buffer and leveraging the implicit temporal encoding in propagated covariances, delayed observations are fused into the current state without re-integrating IMU sequences, preserving real-time capability.
- We validate the framework on the BotanicGarden terrestrial benchmark and a self-collected inland lake aquatic dataset. In water surface environments with dual LiDAR-visual degradation and GNSS denial, all state-of-the-art methods diverge while STEAM-LIVO maintains trajectory continuity. Ablation studies confirm that the OOSM mechanism reduces error by 49.25% compared to ignoring temporal misalignment.

## II. RELATED WORKS

### A. Multi-Sensor SLAM

Current SLAM architectures span vision-inertial (VI), LiDAR-inertial (LI), and LiDAR-inertial-visual (LIV) paradigms. Vision-centric methods such as VINS-Mono [1] and its multi-sensor extension VINS-FUSION [13] achieve tightly-coupled visual-inertial estimation via optical flow tracking and factor graph optimization, yet remain sensitive to illumination variations. LiDAR-based approaches, notably the Fast-LIO series [5], [6], employ direct scan-to-map registration for efficient pose estimation but degrade under geometric sparsity.

LiDAR-inertial-visual methods leverage sensor complementarity: LVI-SAM [7] loosely couples LiDAR and visual odometry via factor graph optimization; R3LIVE [14] fuses LiDAR-visual data through photometric error models on colorized maps; FAST-LIVO2 [15] achieves real-time direct LiDAR-visual-inertial odometry via sequential Kalman filter updates with sweep reconstruction. However, all these methods adopt serial measurement-propagation paradigms, making them vulnerable to failures in unstructured environments.

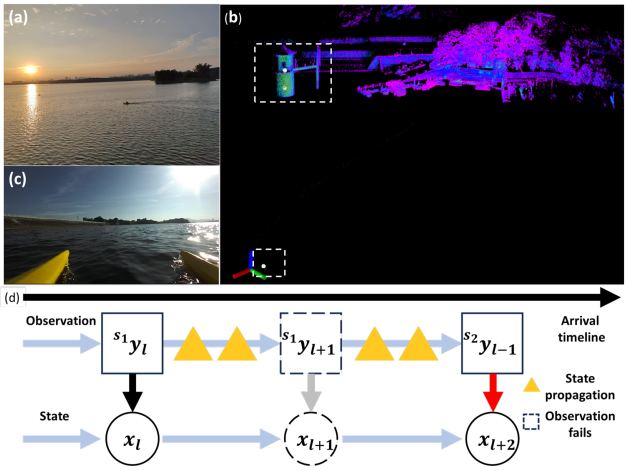


Fig. 2. Challenges in unstructured environments. (a) Vessel operating on water surface with sparse LiDAR returns and insufficient visual features. (b) LiDAR-generated point cloud (white points: sparse latest frame). (c) Camera image with weak features and drastic illumination variations. (d) Key difficulties: 1) State update failure when $s_1$ sensor $y_{l+1}$observations degrade; 2) Out-of-Sequence Measurement problem where $s_2$ sensor's $y_{l-1}$measurements cannot update real-time states directly

### B. SLAM in Unstructured Natural Environments

Current SLAM research for unstructured natural environments remains predominantly terrestrial-centric. ROLO-SLAM [16] mitigates uneven terrain challenges through decoupled pose estimation architecture, isolating translational and rotational components for robust wilderness navigation. Super Odometry [12] implements fault-tolerant modularization with parallelized IMU-centric state prediction and loose-coupling interfaces for LiDAR-inertial and visual-inertial odometry, maintaining continuity during transient sensor outages. LiDAR Graph-SLAM [17] establishes a unified multi-sensor graph optimization framework incorporating adaptive loop closure verification. TRAELS [18] synergizes terrain-referenced navigation, wheel odometry, and IMU measurements via incremental Extended Kalman Filter fusion for kilometer-scale mapping. The orthogonal radar architecture in [19] achieves haze-resilient 3D motion estimation through instantaneous Doppler velocity constraints.

Aquatic SLAM solutions exhibit notable methodological gaps. Roboat II [20] combines 3D LiDAR-IMU-GPS for autonomous surface vessel navigation. The framework in [21] implements GPS-aided omnidirectional visual-inertial odometry via sliding-window bundle adjustment. Recent works such as Coco-LIC [22] address multi-sensor temporal alignment through continuous-time trajectory representations, offering an alternative to discrete OOSM handling; however, they incur higher computational costs that may limit real-time applicability.

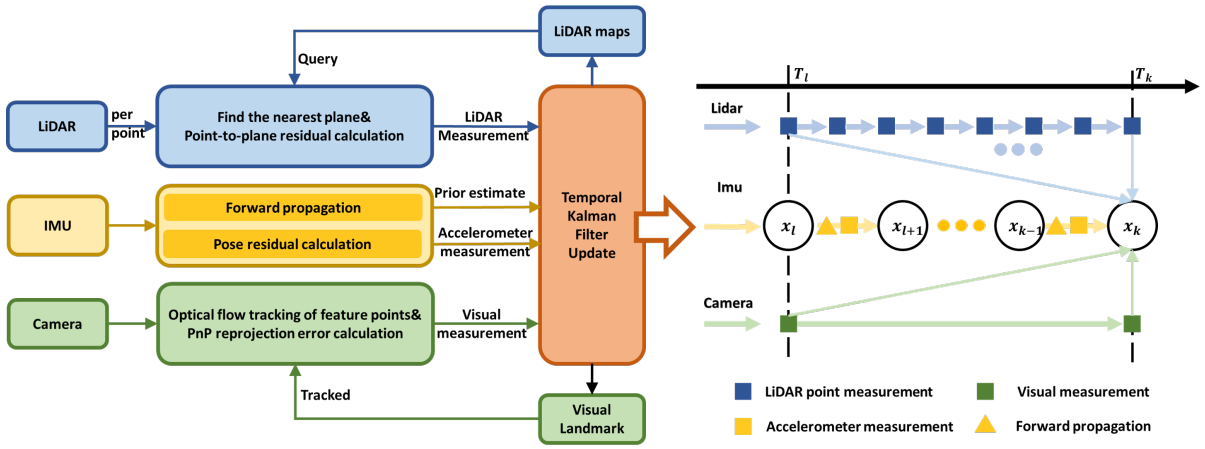


Fig. 3. STEAM-LIVO workflow. The left portion illustrates sensor observation models: LiDAR employs point-to-plane residual computation with local point cloud map accumulated along with the motion of the robot, IMU kinematics enable forward propagation while accelerometer measurements constrain pose estimation, and visual measurements leverage feature reprojection errors. The right section details our temporal fusion mechanism, addressing how data acquired at epoch $l$ arriving at epoch $k$ induces state estimation refinement through asynchronous sensor data fusion.

## III. Problem Formulation

Consider a non-uniform temporal sequence denoted by $\{\mathbb{T}_k\}_{k=1}^{K}$ where temporal intervals need not be equidistant. The system state $x \in \mathcal{M}$ (where $\mathcal{M}$ is a Lie group manifold) from time step $k-1$ to $k$ satisfies:

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}) \tag{1}$$

where $u$ represents sensor-derived inputs and $w$ denotes system noise following $w_{k-1} \sim \mathcal{N}(0, Q_{k-1})$.

When temporal measurement ${}^{s}y_l^k$ from sensor $s$ acquired at epoch $l$ arrives at system time $k$, the observation model becomes:

$${}^{s}y_l^k = {}^{s}h(x_l, v_l) \tag{2}$$

where $v_l$ is observation noise.

For individual sensors, we define observation latency (induced by transmission delay or measurement failure) as $\Delta t = k - l$. The measurement is considered latency-free when $k = l$. In practical scenarios where $k > l$, we categorize the latency as: single-step latency for $\Delta t = 1$, and multi-step latency otherwise.

Spatiotemporal asynchrony emerges from heterogeneous sensor modalities with different sampling rates: Measurements ${}^{s_1}y_{l_1}^{k_1}$ from sensor $s_1$ (acquired at $l_1$, received at $k_1$) and ${}^{s_2}y_{l_2}^{k_2}$ from sensor $s_2$ may exhibit reversed temporal ordering ($l_1 < l_2$ but $k_1 > k_2$), resulting in out-of-sequence measurements (OOSMs). The OOSM problem was first formalized in the target tracking literature [23]; we adapt it to multi-sensor SLAM on Lie group manifolds. Ignoring such temporal disorder may induce state estimation inconsistencies that degrade system performance, as illustrated in Fig.2.

## IV. Methodology

### A. Framework Overview and Computational Abstraction

STEAM-LIVO is proposed as an IMU-centric multi-modal odometry system (Fig. 3) that employs manifold-constrained iterative optimization to minimize geometric residuals from LiDAR point clouds and visual feature correspondences. Unlike FAST-LIVO2, which applies sequential two-stage Kalman updates for LiDAR and camera separately, our framework processes both modalities within a shared iterative error-state Kalman filter. Furthermore, the visual subsystem extracts and tracks features directly from camera images via FAST corners and KLT optical flow, rather than relying on LiDAR-projected points as in FAST-LIVO and R3LIVE, enabling independent visual operation under LiDAR degradation. This dual-phase operational paradigm achieves real-time state estimation through IMU-driven prediction-observation cycles that systematically integrate historical sensor data streams.

TABLE I
Notations and Coordinate Conventions

| Symbol | Type | Description |
|---|---|---|
| *Coordinate Frames* | | |
| ${}^{\mathrm{G}}(\cdot)$ | Superscript | Global frame |
| ${}^{\mathrm{I}}(\cdot)$ | Superscript | IMU body frame |
| ${}^{\mathrm{L}}(\cdot)$ | Superscript | LiDAR frame |
| ${}^{\mathrm{C}}(\cdot)$ | Superscript | Camera frame |
| *State Variables* | | |
| ${}^{\mathrm{G}}\mathbf{R}_I$ | $\mathrm{SO}(3)$ | Rotation matrix from IMU to global frame |
| ${}^{\mathrm{G}}\mathbf{p}_I$ | $\mathbb{R}^3$ | IMU position in global frame |
| $\mathbf{x}$ | $\mathcal{M}$ | True state value |
| *State Estimation* | | |
| $\hat{\mathbf{x}}, \bar{\mathbf{x}}$ | $\mathcal{M}$ | Prior/Posterior state estimate |
| $\delta\mathbf{x}$ | $\mathbb{R}^n$ | State error vector |
| $\mathbf{w}_k$ | Noise | Process noise (covariance $\mathbf{Q}_k$) |
| *Operators* | | |
| $\boxplus$ | $\mathcal{M} \times \mathbb{R}^n \to \mathcal{M}$ | Manifold addition (right perturbation model) |
| $\boxminus$ | $\mathcal{M} \times \mathcal{M} \to \mathbb{R}^n$ | Manifold differential operator |
| $\mathrm{Exp}(\cdot)$ | $\mathfrak{so}(3) \to \mathrm{SO}(3)$ | Lie algebra to Lie group mapping |
| $\mathrm{Log}(\cdot)$ | $\mathrm{SO}(3) \to \mathfrak{so}(3)$ | Lie group to Lie algebra mapping |
| $\lfloor\mathbf{v}\rfloor_\times$ | $\mathbb{R}^3 \to \mathfrak{so}(3)$ | Vector to skew-symmetric matrix |

Adopting FAST-LIO's manifold calculus, we formulate state errors through $\boxplus/\boxminus$ operators on $\mathcal{M} = \mathrm{SO}(3) \times \mathbb{R}^n$.

Since rotations in $\mathrm{SO}(3)$ do not form a vector space, we employ Exp/Log maps for local linearization on the tangent space, defining right perturbation as:

$$\begin{bmatrix}\mathbf{R}\\ \mathbf{a}\end{bmatrix}\boxplus\begin{bmatrix}\mathbf{r}\\ \mathbf{b}\end{bmatrix}\triangleq\begin{bmatrix}\mathbf{R}\cdot\mathrm{Exp}(\mathbf{r})\\ \mathbf{a}+\mathbf{b}\end{bmatrix},\begin{bmatrix}\mathbf{R}_1\\ \mathbf{a}\end{bmatrix}\boxminus\begin{bmatrix}\mathbf{R}_2\\ \mathbf{b}\end{bmatrix}\triangleq\begin{bmatrix}\mathrm{Log}\left(\mathbf{R}_2^\top\mathbf{R}_1\right)\\ \mathbf{a}-\mathbf{b}\end{bmatrix}$$

where $\mathbf{r}\in\mathbb{R}^3$, $\mathbf{a},\mathbf{b}\in\mathbb{R}^n$, with $\mathrm{Exp}(\cdot)$ denoting the Rodrigues formula mapping from $\mathfrak{so}(3)$ to $\mathrm{SO}(3)$ and $\mathrm{Log}(\cdot)$ its inverse.

### B. State Propagation

We assume that sensor acquisition timestamps are known, obtainable through offline calibration or hardware synchronization, and that the relative positions among the camera, IMU, and LiDAR remain fixed during operation. We define the system state vector on the manifold $\mathcal{M}\triangleq\mathrm{SO}(3)\times\mathbb{R}^{15}$ as:

$$x=\begin{bmatrix}{}^G\mathbf{R}_I^T & {}^G\mathbf{p}_I^T & {}^G\mathbf{v}_I^T & \mathbf{b}_g^T & \mathbf{b}_a^T & {}^G\boldsymbol{g}^T\end{bmatrix}^T \tag{3}$$

with corresponding error states:

$$\delta\mathbf{x}=\mathbf{x}\boxminus\hat{\mathbf{x}}=\begin{bmatrix}\delta\boldsymbol{\theta}^T & {}^G\delta\mathbf{p}_I^T & \delta{}^G\mathbf{v}_I^T & \delta\mathbf{b}_g^T & \delta\mathbf{b}_a^T & {}^G\delta\boldsymbol{g}^T\end{bmatrix}^T \tag{4}$$

where: ${}^G\mathbf{R}_I$ denotes the rotation matrix from IMU to global frame, ${}^G\mathbf{p}_I$ represents the IMU position in global coordinates, ${}^G\mathbf{v}_I$ is the IMU velocity in global coordinates, $\mathbf{b}_g$ stands for the gyroscope bias, $\mathbf{b}_a$ indicates the accelerometer bias, ${}^G\boldsymbol{g}$ refers to the gravity vector in global coordinates, and $\delta\boldsymbol{\theta}$ corresponds to the right-multiplied rotation error (Lie algebra).

The IMU kinematic equations are formulated as:

$$\begin{aligned}&\dot{\mathbf{R}}=\mathbf{R}(\tilde{\boldsymbol{\omega}}-\mathbf{b}_g-\boldsymbol{\eta}_g)_\times,\quad \dot{\boldsymbol{v}}=\mathbf{R}(\tilde{\boldsymbol{a}}-\mathbf{b}_a-\boldsymbol{\eta}_a)+\boldsymbol{g},\\ &\dot{\boldsymbol{p}}=\boldsymbol{v},\quad \dot{\mathbf{b}}_g=\boldsymbol{\eta}_{bg},\quad \dot{\mathbf{b}}_a=\boldsymbol{\eta}_{ba},\quad \dot{\boldsymbol{g}}=\mathbf{0}.\end{aligned} \tag{5}$$

where $\boldsymbol{\eta}_g\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_g)$ and $\boldsymbol{\eta}_a\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_a)$ denote Gaussian process noise. These continuous-time equations are discretized using the IMU sampling interval $\Delta t$, yielding the nominal and error-state dynamics summarized in Table II.

TABLE II
DISCRETIZED NOMINAL AND ERROR STATE DYNAMICS

| State | Nominal State | Error State |
|---|---|---|
| $\mathbf{R}$ | $\mathbf{R}\,\mathrm{Exp}\big((\tilde{\boldsymbol{\omega}}-\mathbf{b}_g)\Delta t\big)$ | $\mathrm{Exp}\big(-(\tilde{\boldsymbol{\omega}}-\mathbf{b}_g)\Delta t\big)\delta\boldsymbol{\theta}$ $-\delta\mathbf{b}_g\Delta t-\boldsymbol{\eta}_\theta$ |
| $\boldsymbol{p}$ | $\boldsymbol{p}+\boldsymbol{v}\Delta t+\frac{1}{2}\mathbf{R}(\tilde{\boldsymbol{a}}-\mathbf{b}_a)\Delta t^2$ $+\frac{1}{2}\boldsymbol{g}\Delta t^2$ | $\delta\boldsymbol{p}+\delta\boldsymbol{v}\Delta t$ |
| $\boldsymbol{v}$ | $\boldsymbol{v}+\mathbf{R}(\tilde{\boldsymbol{a}}-\mathbf{b}_a)\Delta t+\boldsymbol{g}\Delta t$ | $\delta\boldsymbol{v}+\big(-\mathbf{R}(\tilde{\boldsymbol{a}}-\mathbf{b}_a)^\wedge\delta\boldsymbol{\theta}$ $-\mathbf{R}\delta\mathbf{b}_a+\delta\boldsymbol{g}\big)\Delta t-\boldsymbol{\eta}_v$ |
| $\mathbf{b}_g$ | $\mathbf{b}_g$ | $\delta\mathbf{b}_g+\boldsymbol{\eta}_{bg}$ |
| $\mathbf{b}_a$ | $\mathbf{b}_a$ | $\delta\mathbf{b}_a+\boldsymbol{\eta}_{ba}$ |
| $\boldsymbol{g}$ | $\boldsymbol{g}$ | $\delta\boldsymbol{g}$ |

The input to the state transition equations is defined as $\mathbf{u}\triangleq\begin{bmatrix}\tilde{\boldsymbol{\omega}} & \tilde{\boldsymbol{a}}\end{bmatrix}^\top$, which propagates through the nominal state update. The covariance is updated as:

$$\hat{\mathbf{P}}=\mathbf{F}\bar{\mathbf{P}}\mathbf{F}^\top+\mathbf{W} \tag{6}$$

where the state transition matrix $\mathbf{F}$ and noise covariance matrix $\mathbf{W}$ are given by:

$$\mathbf{F}=\begin{bmatrix}\mathrm{Exp}\big(-(\tilde{\boldsymbol{\omega}}-\mathbf{b}_g)\Delta t\big) & \mathbf{0} & \mathbf{0} & -\mathbf{I}\Delta t & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I} & \mathbf{I}\Delta t & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ -\mathbf{R}(\tilde{\boldsymbol{a}}-\mathbf{b}_a)_\times\Delta t & \mathbf{0} & \mathbf{I} & \mathbf{0} & -\mathbf{R}\Delta t & \mathbf{I}\Delta t\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\end{bmatrix} \tag{7}$$

$$\mathbf{W}=\mathrm{diag}\big((\mathbf{J}_r\boldsymbol{\Sigma}_g\mathbf{J}_r^\top)\Delta t,\ \boldsymbol{\Sigma}_a\Delta t^3,\ \boldsymbol{\Sigma}_a\Delta t,\ \boldsymbol{\Sigma}_{bg}\Delta t,\ \boldsymbol{\Sigma}_{ba}\Delta t,\ \mathbf{0}\big) \tag{8}$$

Here, $\boldsymbol{\Sigma}_g$ and $\boldsymbol{\Sigma}_a$ represent the covariance matrices for gyroscope and accelerometer white noise, respectively. The term $\mathbf{J}_r$ denotes the right Jacobian matrix of the Exp mapping.

### C. State Update

To handle delayed or out-of-sequence observations, we maintain a historical state-covariance buffer spanning from timestamp $\tau$ to the current time $k$. When a measurement acquired at a past epoch $l$ arrives at the current time $k>l$, the OOSM mechanism retrieves the corresponding historical state and performs the update without re-integrating the IMU sequence. Specifically, upon receiving delayed measurement $\mathbf{y}_l^k$ at time $k$ (with $l<k$), we retrieve the historical state $\bar{\mathbf{x}}_l$ and covariance $\bar{\mathbf{P}}_l$. The measurement Jacobian is computed as:

$$\mathbf{H}_l=\left.\frac{\partial\mathbf{h}}{\partial\delta\mathbf{x}}\right|_{\bar{\mathbf{x}}_l} \tag{9}$$

The Kalman gain is calculated by:

$$\mathbf{K}_l=\bar{\mathbf{P}}_l\mathbf{H}_l^T(\mathbf{H}_l\bar{\mathbf{P}}_l\mathbf{H}_l^T+\mathbf{R}_l)^{-1} \tag{10}$$

The optimal estimate $\bar{\boldsymbol{x}}_k$ and $\bar{\mathbf{P}}_k$ are obtained through iterative optimization on the tangent plane, solving:

$$\min_{\delta\boldsymbol{x}^j}\left\|\mathbf{y}_l^k-\mathbf{h}(\mathbf{x}^j\boxplus\delta\boldsymbol{x}^j)\right\|_{\mathbf{R}_l}^2+\left\|\mathbf{J}^j(\mathbf{x}^j\boxminus\bar{\mathbf{x}}_l)+\delta\boldsymbol{x}^j\right\|_{(\mathbf{J}^j\bar{\mathbf{P}}_l\mathbf{J}^{jT})}^2 \tag{11}$$

The tangent space projection Jacobian at iteration $j$ is:

$$\mathbf{J}^j=\frac{\partial\left(\mathbf{x}^j\boxplus\delta\mathbf{x}\boxminus\mathbf{x}^j\right)}{\partial\delta\mathbf{x}} \tag{12}$$

$$=\mathrm{diag}\left(\mathbf{J}_\theta,\mathbf{I}_3,\mathbf{I}_3,\mathbf{I}_3,\mathbf{I}_3,\mathbf{I}_3\right) \tag{13}$$

where $\mathbf{J}_{\boldsymbol{\theta}}=\mathbf{I}-\frac{1}{2}\lfloor\delta\boldsymbol{\theta}^j\rfloor_\times$ with $\delta\boldsymbol{\theta}^j=\mathrm{Log}(\mathbf{R}^{(0)T}\mathbf{R}^{(j)})$.

Initializing with $\mathbf{x}^{(0)}=\hat{\mathbf{x}}_k$, $\mathbf{P}^{(0)}=\bar{\mathbf{P}}_l$, and $\mathbf{K}^{(0)}=\mathbf{K}_l$, the iterative process follows:

$$\delta\mathbf{x}^j=\mathbf{J}^j(\mathbf{x}^j\boxminus\mathbf{x}^0)\sim\mathcal{N}(\mathbf{0},\mathbf{P}^j),\quad \mathbf{P}^j=\mathbf{J}^j\mathbf{P}^{(0)}\mathbf{J}^{jT} \tag{14}$$

The update from iteration $j$ to $j+1$ is:

$$\mathbf{K}^j=\mathbf{P}^j\mathbf{H}^{jT}\left(\mathbf{H}^j\mathbf{P}^j\mathbf{H}^{jT}+\mathbf{R}_l\right)^{-1} \tag{15}$$

$$\delta\mathbf{x}^{j+1}=\mathbf{K}^j\left(\mathbf{y}_l^k-\mathbf{h}(\mathbf{x}^j)+\mathbf{H}^j\delta\mathbf{x}^j\right)-\delta\mathbf{x}^j \tag{16}$$

Nominal state update equations become:

$$\mathbf{x}^{j+1}=\mathbf{x}^j\boxplus\delta\mathbf{x}^{j+1},\quad \mathbf{P}^{j+1}=\left(\mathbf{I}-\mathbf{K}^j\mathbf{H}^j\right)\mathbf{J}^j\mathbf{P}^{(0)}\mathbf{J}^{jT} \tag{17}$$

Upon convergence, the final updates are applied. Since error-state propagation is approximately linear ($\delta\mathbf{x}_k \approx \mathbf{F}_{l\to k}\delta\mathbf{x}_l$), the propagated covariance $\hat{\mathbf{P}}_k$ implicitly encodes the mapping from historical state $l$ to the current time $k$. Consequently, the IEKF correction computed using $\bar{\mathbf{P}}_l$ and $\mathbf{H}_l$ can be directly applied to $\hat{\mathbf{x}}_k$ without re-integrating the IMU sequence—a key efficiency advantage over Recalculation-based approaches:

$$\bar{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k \boxplus \delta\boldsymbol{x}^{j+1}, \quad \bar{\mathbf{P}}_k = \mathbf{P}^{j+1} \tag{18}$$

### *D. Observation Model*

**IMU Observations:** We leverage accelerometer measurements for attitude correction by analyzing specific forces, which combine non-gravitational acceleration with the gravitational component in the body frame. Following the Newtonian equivalence principle, the accelerometer output under dynamic conditions is:

$$\tilde{\mathbf{a}} = {}^G\mathbf{R}_I^T({}^G\mathbf{a} - \boldsymbol{g}) + \mathbf{b}_a + \mathbf{n}_a \tag{19}$$

where ${}^G\mathbf{a} \in \mathbb{R}^3$ represents IMU linear acceleration in global frame and $\mathbf{n}_a \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$ denotes measurement noise.

Under quasi-static conditions (${}^G\mathbf{a} \approx \mathbf{0}$), the observation function $h_a$ simplifies to:

$$h_a(\mathbf{x}) = -{}^G\mathbf{R}_I^T\boldsymbol{g} + \mathbf{b}_a \tag{20}$$

yielding the observation equation:

$$\tilde{\mathbf{a}} = h_a(\mathbf{x}) + \mathbf{n}_a \tag{21}$$

The residual $\mathbf{r}_a$ is computed as:

$$\mathbf{r}_a(\mathbf{x}, \tilde{\mathbf{a}}) = \tilde{\mathbf{a}} - h_a(\mathbf{x}) \tag{22}$$

Given that the accelerometer and gyroscope are sampled synchronously, we assume the measurement time $l$ coincides with the system time $k$. For posterior distribution construction, we perform a first-order Taylor expansion:

$$\mathbf{r}_a(\mathbf{x}, \mathbf{n}_a) \approx \mathbf{r}_a(\hat{\mathbf{x}}_{k+1}, \tilde{\mathbf{a}}) + \mathbf{H}_{k+1}^a\delta\hat{\mathbf{x}}_{k+1} + \mathbf{n}_a = \mathbf{0} \tag{23}$$

where $\tilde{\mathbf{x}}_{k+1}$ denotes the current state estimate and $\delta\tilde{\mathbf{x}}_{k+1}$ the corresponding state error. The Jacobian matrix is:

$$\mathbf{H}^a = \left.\frac{\partial\mathbf{r}_a(\hat{\mathbf{x}}_{k+1} \boxplus \delta\mathbf{x}_{k+1}, \tilde{\mathbf{a}})}{\partial\delta\mathbf{x}}\right|_{\delta\hat{\mathbf{x}}_{k+1}=\mathbf{0}} \tag{24}$$

$$\mathbf{H}^a = \begin{bmatrix} -{}^G\mathbf{R}_I^T\lfloor\boldsymbol{g}\rfloor_\times & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -{}^G\mathbf{R}_I^T \end{bmatrix} \tag{25}$$

The propagated accelerometer noise covariance is:

$$\boldsymbol{\Sigma}_\alpha = {}^G\mathbf{R}_I\boldsymbol{\Sigma}_{n_a}{}^G\mathbf{R}_I^T \tag{26}$$

Total residual covariance becomes:

$$\boldsymbol{\Sigma}_r = \mathbf{H}^a\mathbf{P}(\mathbf{H}^a)^T + \boldsymbol{\Sigma}_\alpha \tag{27}$$

**LiDAR Measurements:** Following the point-to-plane registration framework of [6], each LiDAR point ${}^L\mathbf{p}_m$ is projected to the global frame via ${}^G\mathbf{p}_m = {}^G\mathbf{R}_I({}^I\mathbf{R}_L{}^L\mathbf{p}_m + {}^I\mathbf{t}_L) + {}^G\mathbf{p}_I$ using calibrated LiDAR-IMU extrinsics ${}^I\mathbf{T}_L \in SE(3)$. For each projected point, we compute the signed point-to-plane distance against the local map:

$$h_l(\mathbf{x}, {}^L\mathbf{p}_m) = \boldsymbol{u}^T\left({}^G\mathbf{p}_m - \mathbf{q}\right) \tag{28}$$

where $\boldsymbol{u}$ is the unit normal and $\mathbf{q}$ a reference point on the matched plane. Defining ${}^G\hat{\mathbf{p}}_L \triangleq {}^G\hat{\mathbf{R}}_I({}^I\hat{\mathbf{R}}_L{}^L\mathbf{p}_m + {}^I\hat{\mathbf{t}}_L)$, the measurement Jacobian is:

$$\mathbf{H}^l = \begin{bmatrix} -\boldsymbol{u}^T[{}^G\hat{\mathbf{p}}_L]_\times & \boldsymbol{u}^T & \mathbf{0}_{1\times12} \end{bmatrix} \in \mathbb{R}^{1\times18} \tag{29}$$

with noise propagation $\mathbf{D} = \boldsymbol{u}^T{}^G\hat{\mathbf{R}}_I{}^I\hat{\mathbf{R}}_L$ and total residual covariance $\boldsymbol{\Sigma}_r = \mathbf{H}^l\mathbf{P}(\mathbf{H}^l)^T + \mathbf{D}\mathbf{R}\mathbf{D}^T$.

**Visual Measurements:** We extract FAST corners from images and track features using KLT optical flow following the pipeline in [1]. Lost or new features undergo triangulation using current poses. For a tracked 3D landmark ${}^G\mathbf{P}_i$ observed as ${}^C\mathbf{p}_i$, the reprojection residual is $\mathbf{r}_c = {}^C\mathbf{p}_i - \pi({}^G\mathbf{P}_i, \mathbf{x})$, where $\pi(\cdot)$ denotes the standard pinhole projection. The landmark is first transformed to the camera frame via ${}^C\mathbf{P}_i = {}^I\mathbf{R}_C^T[{}^G\mathbf{R}_I^T({}^G\mathbf{P}_i - {}^G\mathbf{p}_I) - {}^I\mathbf{p}_C]$ using fixed IMU-camera extrinsics. The Jacobian decomposes as:

$$\mathbf{H}_i = \frac{\partial\pi}{\partial{}^C\mathbf{P}_i}\cdot\frac{\partial{}^C\mathbf{P}_i}{\partial\delta\check{\mathbf{x}}} = \begin{bmatrix} \mathbf{H}_\theta & \mathbf{H}_p & \mathbf{0}_{2\times12} \end{bmatrix} \tag{30}$$

where $\mathbf{H}_\theta = \frac{\partial\pi}{\partial{}^C\mathbf{P}_i}(-{}^I\mathbf{R}_C^T\lfloor{}^G\mathbf{R}_I^T({}^G\mathbf{P}_i - {}^G\mathbf{p}_I)\rfloor_\times)$ and $\mathbf{H}_p = \frac{\partial\pi}{\partial{}^C\mathbf{P}_i}(-{}^I\mathbf{R}_C^T{}^G\mathbf{R}_I^T)$, with the standard pinhole projection Jacobian $\frac{\partial\pi}{\partial{}^C\mathbf{P}_i}$. The observation noise covariance combines pixel measurement and 3D landmark uncertainties: $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_s} = \boldsymbol{\Sigma}_{\mathbf{n}_{\mathbf{p}_s}} + \mathbf{F}_{\mathbf{P}_s}\boldsymbol{\Sigma}_{\mathbf{n}_{\mathbf{P}_s}}\mathbf{F}_{\mathbf{P}_s}^T$ where $\mathbf{F}_{\mathbf{P}_s} = -\frac{\partial\pi}{\partial{}^C\mathbf{P}_i}\cdot{}^I\mathbf{R}_C^T{}^G\mathbf{R}_I^T$.

### *E. Implementation Details*

1. **Map Update Strategy:** For both LiDAR point clouds and visual features, we extract map points and compute residuals against existing map entries. New points are directly added if no correspondence exists. LiDAR points are stored using ikd-tree [6], while visual map maintenance follows VINS [1]. Key hyperparameters include: FAST corner detection threshold of 20, KLT optical flow window size $21 \times 21$, maximum tracked features per frame $C_l = 150$, ikd-tree downsampling resolution 0.3 m, and IEKF maximum iterations 5 with convergence threshold $\|\delta\mathbf{x}\| < 10^{-4}$.

2. **Accelerometer-Assisted Attitude Correction:** We discard accelerometer measurements during high-dynamic motion when residual analysis indicates unreliable data. The residual covariance is:

$$\mathbf{D}_k = \mathbb{E}[\boldsymbol{e}_k\boldsymbol{e}_k^T] = \mathbf{H}_k^a\mathbf{P}_k^-(\mathbf{H}_k^a)^T + \mathbf{R}_k \tag{31}$$

During aggressive maneuvers in unstructured environments, we evaluate the normalized residual:

$$r_k = \boldsymbol{e}_k^T\mathbf{D}_k^{-1}\boldsymbol{e}_k \tag{32}$$

Measurements are rejected when $r_k$ exceeds the $\chi^2$ test threshold $\chi^2_{3,0.95} = 7.815$ (three degrees of freedom corresponding to the accelerometer axes), which constitutes a standard Mahalanobis-distance-based outlier rejection criterion.

3. **Historical State Management:** Given the heterogeneous sensor rates—IMU at 100–200 Hz, camera at 30 Hz,

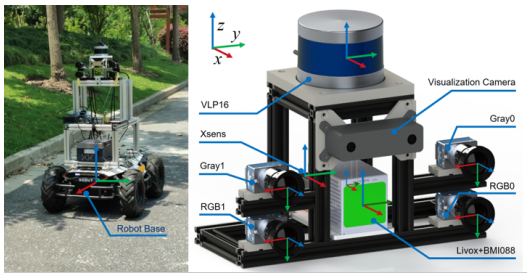


Fig. 4. Garden dataset sensor configuration

and LiDAR at 10 Hz—we retain state estimates only within two consecutive LiDAR cycles to prevent excessive memory consumption. Measurements arriving beyond the buffer horizon are discarded.

4. **Multi-Sensor Temporal Fusion:** Our key insight is that in highly dynamic systems, real-time updates should be conditioned on both the timeliness and reliability of each sensor signal. While the IMU drives forward prediction, accelerometer observations simultaneously provide pose corrections—analogous to attitude estimation on aerial platforms. The complete runtime procedure is summarized in Table III.

TABLE III
STEAM-LIVO MULTI-SENSOR FUSION POSE ESTIMATION ALGORITHM

**Input:** IMU sequence $\{\tilde{\omega}_t, \tilde{a}_t\}$, LiDAR point clouds $\{L_k\}$, camera images $\{I_k\}$
**Output:** Continuous pose sequence $\{\hat{x}_t\}$

**Initialization:**
- The system initializes with a 1-second static phase to estimate initial gravity direction and IMU biases. State $\hat{x}_0 \in \mathcal{M}$ and covariance $P_0$ are set accordingly
- Establish historical buffer $\{\hat{x}_\tau, P_\tau\}$ for delayed observations

**Algorithm Loop:** For each IMU step $t$:
- **IMU State Propagation:**
  - Propagate nominal state using discretized kinematics on manifold $\mathcal{M}$
  - Update error state and covariance via transition matrix $\mathbf{F}$ and noise matrix $\mathbf{W}$
  - Store propagated state $\hat{x}_t^-, P_t^-$ in historical buffer
- **Observation Handling:** If LiDAR or camera measurements arrive:
  - Retrieve historical state $\bar{x}_l, \bar{P}_l$ at measurement time $l$
  - Construct observation residual $\mathbf{r}(\bar{x}_l)$ and Jacobian $\mathbf{H}_l$ on tangent space
  - Formulate iterative optimization problem on manifold: $\min_{\delta x} \|\mathbf{r}(\bar{x}_l \boxplus \delta x)\|^2_{\mathbf{R}_l} + \|\delta x\|^2_{\bar{P}_l}$
  - Apply Gauss–Newton updates with retraction via $\boxplus$ until convergence
  - Update posterior state $\bar{x}_l, \bar{P}_l$ and propagate correction forward to current state $\hat{x}_t$
- **Sensor-specific Updates:**
  - *LiDAR point-wise update:* Each point $p_i$ with timestamp $t_i$ contributes residuals; historical states $\bar{x}_{t_i}$ are iteratively refined to maintain real-time $\hat{x}_t$
  - *Camera feature update:* Visual reprojection residuals are minimized via the same iterative scheme at $\bar{x}_l$
- **Output:** Publish current optimized pose $\hat{x}_t$

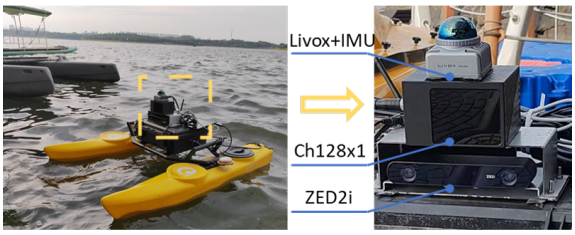


Fig. 5. Boat dataset sensor configuration

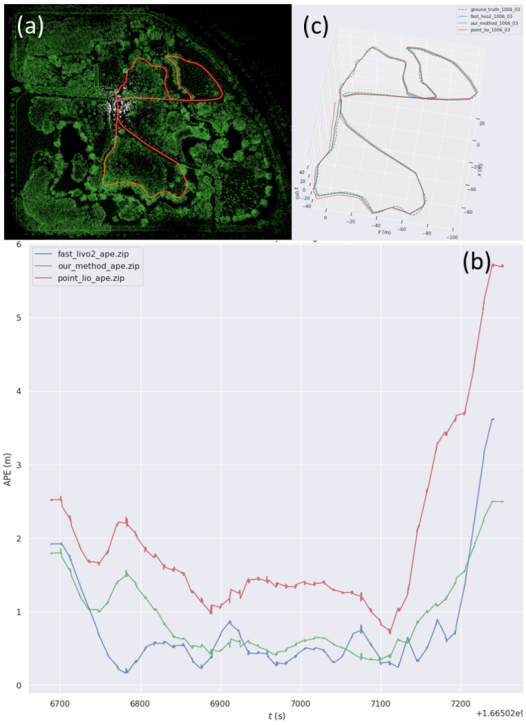


Fig. 6. Trajectory Visualization: (a) Reference trajectory, (b) Absolute Pose Error (APE) progression, (c) Estimated vs. ground truth trajectories

## V. EXPERIMENTS

### A. Experimental Platform and Datasets

Our evaluation framework combines the BotanicGarden benchmark [24] with self-collected inland lake recordings that exhibit dynamic motion profiles and feature-degraded conditions.

As depicted in Fig. 4, the BotanicGarden configuration employs four high-resolution stereo cameras (monochrome/RGB, 1920×1200 @ 40 Hz) supporting simultaneous stereo matching and semantic segmentation. This dataset specifically targets GNSS-denied vegetated environments with reflective water surfaces, providing hardware-synchronized millimeter-accurate ground truth through multi-sensor fusion—filling a critical research gap for unstructured navigation.

Fig. 5 details our aquatic data acquisition platform, which operates under distinctive challenges: persistent GNSS denial, degenerate visual-LiDAR features in open-water domains, and significant wave-induced motion artifacts. All sensor extrinsics were laboratory-calibrated, with IMU–

TABLE IV
DETAILED RELATIVE POSE ERROR (RPE) RESULTS

| Method/Map | 1005_00 | 1005_01 | 1005_07 | 1006_01 | 1008_03 | 1018_00 | 1018_13 | 1006_03 |
|---|---|---|---|---|---|---|---|---|
| Total Trajectory Length | 601.60 | 479.73 | 591.04 | 765.80 | 750.09 | 115.24 | 201.23 | 686.23 |
| FAST-LIVO2 | 2.142 | **2.148** | 1.753 | 2.590 | 2.138 | **0.529** | **0.900** | 1.418 |
| Point-LIO | 2.283 | 2.180 | 2.889 | **2.487** | **1.513** | 0.569 | 0.949 | 1.494 |
| STEAM-LIVO | **1.401** | 2.150 | **1.598** | 2.849 | 2.012 | 1.092 | 1.315 | **1.163** |

TABLE V
ALGORITHMIC PERFORMANCE COMPARISON

| Method | Year | Modality | RPE (%) |
|---|---|---|---|
| ORB-SLAM3 | 2021 | VO | 4.880 |
| ORB-SLAM3 (VI) | 2021 | VIO | 4.272 |
| VINS-Mono | 2018 | VIO | 3.643 |
| Fast-LIO2 | 2022 | LIO | 2.908 |
| LOAM | 2014 | LO | 2.297 |
| R3LIVE | 2022 | LIVO | 2.220 |
| LVI-SAM | 2021 | LIVO | 2.170 |
| Point-LIO | 2023 | LIO | 1.838 |
| FAST-LIVO2 | 2024 | LIVO | 1.742 |
| STEAM-LIVO | 2026 | LIVO | 1.774 |

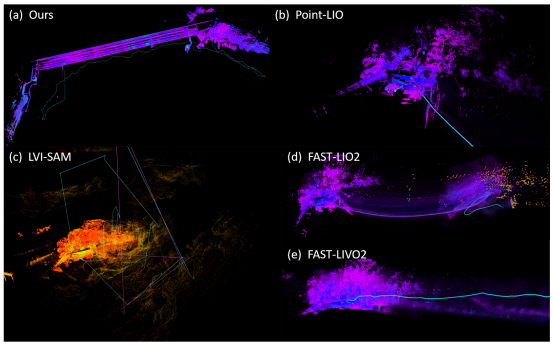


Fig. 7. Aquatic dataset evaluation results. Most pose estimation algorithms exhibit odometry divergence after leaving the harbor area in water surface environments where both LiDAR geometric features and visual texture features are severely degraded.

Livox LiDAR alignment maintained through factory-certified configurations.

### B. Land and Aquatic Unstructured Environment Tests

STEAM-LIVO demonstrates superior adaptability across both datasets. We evaluate relative pose error (RPE) following the BotanicGarden benchmark protocol [24], with quantitative results presented in Table V.

Comparative analysis with reimplemented FAST-LIVO2 and Point-LIO reveals several insights. FAST-LIVO2 shows strong general performance, while Point-LIO exhibits better tolerance to aggressive motion (Table IV). Following official protocols, we compute metrics on sequences 1005_00–1018_13, with 1006_03 representing harsher conditions. Pure LiDAR methods outperform in sequences 1006_01 and 1008_03 due to water surface reflections and lighting variations. FAST-LIVO2's photometric error minimization with exposure time estimation partially mitigates illumination changes, thereby achieving better results in vegetation-rich sequences 1018_00 and 1018_13, as shown in Fig. 6.

The extreme challenges of our aquatic testbed include persistent GNSS denial and progressive sensor degradation when navigating far from shore. As shown in Fig. 7, most SLAM systems fail completely under these conditions, while our method qualitatively maintains trajectory continuity without divergence. We note that quantitative accuracy on this dataset cannot be verified due to the lack of ground truth caused by persistent GNSS denial.

### C. Ablation Study

To evaluate the OOSM mechanism, we compare against two conventional approaches: 1) *Recalculation*, which reorders observations by acquisition timestamps for recomputation (computationally intensive and incompatible with real-time requirements) [25], and 2) *Error Disregard*, which simply ignores temporal misalignment. Quantitative comparisons in Table VI demonstrate our method's superiority, achieving a 49.25% average error reduction (absolute improvement of 1.6488) compared to error disregard, with a maximum single-scene improvement of 69.14% (1018_00).

Notably, our approach exhibits only a marginal 6.76% performance gap (absolute difference of 0.1075) compared to recalculation, while maintaining real-time capability. Although recalculation theoretically eliminates synchronization errors through data reordering, its robustness degrades in extreme scenarios (*e.g.,* 1018_13) due to cumulative numerical instability, ultimately increasing trajectory errors.

TABLE VI
ABLATION STUDY COMPARISONS (RELATIVE POSE ERRORS)

| Sequence/Method | Disregard | Recalculation | Ours |
|---|---|---|---|
| 1005_00 | 3.310 | 1.391 | 1.401 |
| 1005_01 | 3.193 | 2.149 | 2.150 |
| 1005_07 | 2.547 | 1.581 | 1.598 |
| 1006_01 | 4.216 | 2.680 | 2.849 |
| 1008_03 | 3.645 | 1.987 | 2.012 |
| 1018_00 | 3.539 | 0.873 | 1.092 |
| 1018_13 | 4.164 | 1.022 | 1.315 |
| 1006_03 | 2.157 | 1.037 | 1.163 |
| Average | 3.346 | 1.590 | 1.697 |

All methods are deterministic; values are single-run results.

### D. Real-Time Performance Experiment

To ensure a fair comparison, all methods are evaluated on the same Intel i7-based micro-UAV onboard computer. The

detailed results are reported in Table VII.

TABLE VII
REAL-TIME PERFORMANCE

| Method | Per-frame time (10 Hz) [ms] | Per-frame time (100 Hz) [ms] | Per-point time ($\mu$s) |
|---|---|---|---|
| FAST-LIO2 | 30.7 | 1.2 | 9 |
| Point-LIO | 32.1 | 1.8 | 9 |
| FAST-LIVO2 | 30.03 | –† | 9 |
| STEAM-LIVO | 33 | 2.1 | 9 |

† FAST-LIVO2 does not support per-point 100 Hz update mode.

## VI. CONCLUSION

We present STEAM-LIVO, a spatio-temporally adaptive manifold LiDAR-inertial-visual odometry framework for unstructured environments. Centered on IMU measurements, the proposed method develops an iterative error-state Kalman filter that achieves spatio-temporal alignment of multi-sensor data through tight coupling of LiDAR observations and visual measurements. Evaluated in challenging scenarios with partial sensor failures and aggressive motion, the framework demonstrates superior robustness and accuracy compared to state-of-the-art visual-inertial, LiDAR-inertial, and LiDAR-visual SLAM systems. Ablation studies validate the effectiveness of the proposed method.

Notably, our method exhibits limitations in vegetation-rich environments (sequences 1018_00, 1018_13) where FAST-LIVO2's photometric error model with exposure compensation outperforms our geometric feature-based approach. The current implementation is validated only with Livox solid-state LiDAR; generalization to spinning LiDARs remains to be investigated. Additionally, the OOSM mechanism introduces a marginal computational overhead of approximately 3 ms per frame compared to sequential processing.

## REFERENCES


[1] 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, Aug 2018.

[2] H. Liu, M. Chen, G. Zhang, H. Bao, and Y. Bao, "Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1974–1982.

[3] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time," in *Robotics: Science and Systems*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:18612391

[4] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and R. Daniela, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5135–5142.

[5] W. Xu and F. Zhang, "Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter," 2021. [Online]. Available: https://arxiv.org/abs/2010.08196

[6] 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.

[7] T. Shan, B. Englot, C. Ratti, and R. Daniela, "Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5692–5698.

[8] C. Min, S. Si, X. Wang, H. Xue, W. Jiang, Y. Liu, J. Wang, Q. Zhu, Q. Zhu, L. Luo, F. Kong, J. Miao, X. Cai, S. An, W. Li, J. Mei, T. Sun, H. Zhai, Q. Liu, F. Zhao, L. Chen, S. Wang, E. Shang, L. Shang, K. Zhao, F. Li, H. Fu, L. Jin, J. Zhao, F. Mao, Z. Xiao, C. Li, B. Dai, D. Zhao, L. Xiao, Y. Nie, Y. Hu, and X. Li, "Autonomous driving in unstructured environments: How far have we come?" 2024. [Online]. Available: https://arxiv.org/abs/2410.07701

[9] R. Giubilato, W. Stürzl, A. Wedler, and R. Triebel, "Challenges of slam in extremely unstructured environments: the dlr planetary stereo, solid-state lidar, inertial dataset," *IEEE Robotics and Automation Letters*, pp. 1–8, 2022.

[10] A. Kumar, A. Davatzes, T. F. Shipley, and M. Ani Hsieh, "Lidar slam-based dense reconstructions of natural environments: Field evaluations," in *2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Nov 2022, pp. 116–121.

[11] S. Jeong, H. Kim, and Y. Cho, "Diter: Diverse terrain and multimodal dataset for field robot navigation in outdoor environments," *IEEE Sensors Letters*, vol. 8, no. 3, pp. 1–4, March 2024.

[12] S. Zhao, H. Zhang, P. Wang, L. Nogueira, and S. Scherer, "Super Odometry: IMU-centric LiDAR-Visual-Inertial Estimator for Challenging Environments," Aug. 2021.

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

[14] J. Lin and F. Zhang, "R $^3$ LIVE: A Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package," in *2022 International Conference on Robotics and Automation (ICRA)*. Philadelphia, PA, USA: IEEE, May 2022, pp. 10 672–10 678.

[15] C. Zheng, W. Xu, Z. Zou, T. Hua, C. Yuan, D. He, B. Zhou, Z. Liu, J. Lin, F. Zhu, Y. Ren, R. Wang, F. Meng, and F. Zhang, "FAST-LIVO2: Fast, Direct LiDAR-Inertial-Visual Odometry," Aug. 2024.

[16] Y. Wang, B. Ren, X. Zhang, P. Wang, C. Wang, R. Song, Y. Li, and M. Q.-H. Meng, "Rolo-slam: Rotation-optimized lidar-only slam in uneven terrain with ground vehicle," *Journal of Field Robotics*, vol. n/a, no. n/a. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.22505

[17] L. Montano-Oliván, J. A. Placed, L. Montano, and M. T. Lázaro, "From underground mines to offices: A versatile and robust framework for range-inertial slam," 2024. [Online]. Available: https://arxiv.org/abs/2407.14797

[18] W. J. Wagner, I. Blankenau, M. DeLaTorre, A. Purushottam, and A. Soylemezoglu, "A robust localization solution for an uncrewed ground vehicle in unstructured outdoor gnss-denied environments," 2023. [Online]. Available: https://arxiv.org/abs/2309.02569

[19] Y. S. Park, Y.-S. Shin, J. Kim, and A. Kim, "3d ego-motion estimation using low-cost mmwave radars via radar velocity factor for pose-graph slam," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7691–7698, Oct 2021.

[20] W. Wang, T. Shan, P. Leoni, D. Fernandez-Gutierrez, D. Meyers, C. Ratti, and D. Rus, "Roboat II: A Novel Autonomous Surface Vessel for Urban Environments," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 1740–1747.

[21] Y. Yu, W. Gao, C. Liu, S. Shen, and M. Liu, "A GPS-aided Omnidirectional Visual-Inertial State Estimator in Ubiquitous Environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Macau, China: IEEE, Nov. 2019, pp. 7750–7755.

[22] X. Lang, C. Chen, K. Tang, Y. Ma, J. Lv, Y. Liu, and X. Zuo, "Coco-lic: continuous-time tightly-coupled lidar-inertial-camera odometry using non-uniform b-spline," *IEEE Robotics and Automation Letters*, 2023.

[23] Y. Bar-Shalom, "Update with out-of-sequence measurements in tracking: exact solution," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 769–777, July 2002.

[24] Y. Liu, Y. Fu, M. Qin, Y. Xu, B. Xu, F. Chen, B. Goossens, P. Z. Sun, H. Yu, C. Liu, L. Chen, W. Tao, and H. Zhao, "Botanicgarden: A high-quality dataset for robot navigation in unstructured natural environments," *IEEE Robotics and Automation Letters*, vol. 9, no. 3, pp. 2798–2805, 2024.

[25] E. Allak, R. Jung, and S. Weiss, "Covariance pre-integration for delayed measurements in multi-sensor fusion," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 6642–6649.