

Article

BIM-Constrained Elastic Fusion Navigation Framework for UAV Bridge Inspection Under Intermittent GNSS Outage

Zeyu Li , Hui Li, Chenhong Xiangli, Yuanyuan Shen, Yi Yu * and Fei Li *

School of Computer Science and Technology, North University of China, Taiyuan 030051, China; 20230101@nuc.edu.cn (Z.L.); 2407024102@st.nuc.edu.cn (H.L.); 2407024120@st.nuc.edu.cn (C.X.); shenyuanyuan@nuc.edu.cn (Y.S.)

* Correspondence: 20030820@nuc.edu.cn (Y.Y.); lifei@nuc.edu.cn (F.L.); Tel.: +86-1383-466-7072 (Y.Y.); +86-1883-416-9656 (F.L.)

Highlights

What are the main findings?

- A novel navigation strategy is proposed to improve the positioning robustness of UAVs in complex sheltered environments, which effectively suppresses cumulative drift and enhances localization stability.
- The proposed method outperforms traditional mainstream algorithms in positioning accuracy and environmental adaptability under GNSS challenging scenarios.

What are the implications of the main finding?

- This research provides a reliable technical reference for practical UAV patrol and inspection missions in similar complex scenes.
- It lays a solid foundation for the further engineering application and popularization of UAV autonomous navigation technology.

Abstract

UAV bridge inspection requires centimeter-level global positioning in the engineering coordinate frame to associate detected defects with BIM component IDs and mileage stakes. Intermittent GNSS outages beneath beams, inside box girders, and in pier-dense regions cause conventional navigation methods to accumulate drift or produce discontinuous pose estimates. This paper presents BCEF-Nav, a BIM-constrained elastic fusion navigation framework for UAV bridge inspection. The framework adaptively adjusts GNSS constraints according to signal availability and replaces degraded GNSS references with semantic-geometric constraints from a high-precision as-built BIM model within a sliding-window optimizer. Semantic-guided feature matching suppresses false correspondences in repetitive bridge structures. BCEF-Nav uses only low-cost commercial sensors and is compatible with mainstream inspection UAVs. Digital twin simulations and field experiments achieved an ATE RMSE of 5.7 cm after 120 s of complete GNSS occlusion and a 100% positioning success rate with positioning errors below 10 cm, outperforming seven recent state-of-the-art baselines.



Academic Editor: Andrey V. Savkin

Received: 16 May 2026

Revised: 16 July 2026

Accepted: 18 July 2026

Published: 21 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: UAV bridge inspection; intermittent GNSS occlusion; BIM-constrained navigation; visual-inertial odometry; multi-sensor fusion; tight coupling optimization

1. Introduction

Bridge inspection is essential for maintaining the safety and serviceability of transportation infrastructure. In China, nearly 40% of more than one million highway bridges

have been in service for over 20 years, placing a large proportion of the bridge inventory in a maintenance-intensive stage [1,2]. Conventional inspection still relies on scaffolding, under-bridge inspection vehicles, and manual visual assessment, making the process costly, time-consuming, and hazardous while providing limited access to beam soffits, box girder interiors, and other confined components. UAV-based visual inspection offers an efficient alternative by enabling close-range data acquisition in these difficult-to-access areas. Its effectiveness, however, depends on centimeter-level positioning in the bridge engineering coordinate system. Without accurate global positioning, detected defects cannot be reliably associated with BIM component IDs or mileage stakes, limiting their value for maintenance planning and lifecycle management of bridge assets [3,4].

Intermittent GNSS occlusion remains one of the main challenges in UAV bridge inspection because GNSS availability varies continuously along the flight path rather than switching between fully available and fully denied conditions. Signals are generally reliable over open bridge decks but degrade beneath beam soffits, inside box girders, and around dense pier groups. Maintaining continuous and globally consistent pose estimates across these localized transitions is therefore essential. Pure visual-inertial odometry provides accurate local motion estimation but inevitably accumulates drift without a global reference. Although LiDAR-inertial odometry improves robustness in low-texture environments, its cost, payload, and power requirements limit deployment on small inspection UAVs. Existing BIM-aided methods are typically loosely coupled, using the structural model mainly for loop-closure detection instead of continuous state estimation, leaving drift largely uncorrected during prolonged GNSS outages. GNSS-VIO fusion systems face a similar limitation. Most rely on hard switching between GNSS-available and GNSS-denied modes, which can introduce pose discontinuities during signal transitions while still lacking a reliable global reference during extended outages. Existing methods therefore leave an important gap: achieving low-cost, tightly coupled navigation that maintains centimeter-level global positioning under intermittent GNSS occlusion while remaining aligned with the bridge engineering coordinate system.

This paper presents BCEF-Nav, a BIM-constrained elastic fusion navigation framework for UAV bridge inspection. It employs a sliding-window nonlinear optimizer that tightly integrates monocular images, MEMS IMU measurements, GNSS observations, and a high-precision as-built bridge BIM model to estimate the UAV pose directly in the bridge engineering coordinate system. An adaptive weighting strategy replaces abrupt switching between GNSS-available and GNSS-denied conditions by continuously adjusting the contribution of GNSS constraints according to signal availability. During GNSS outages, semantic-geometric BIM features provide a global reference that limits odometric drift. Semantic-guided feature matching further suppresses false correspondences in repetitive bridge components, following recent work on semantic feature matching [5]. The framework relies only on low-cost commercial sensors and can be deployed on mainstream small inspection UAVs without specialized high-end payloads.

Although techniques such as VIO, GNSS fusion, and BIM constraints are well established individually, their problem-specific tight coupling for UAV bridge inspection under intermittent GNSS occlusion is novel. This integration enables capabilities unattainable by loose combinations of these methods. On this basis, this work makes three contributions to UAV navigation for infrastructure inspection. It develops an elastic fusion framework with adaptive dynamic weighting, allowing the estimator to pass between valid and invalid GNSS conditions without the pose jumps caused by hard switching. It introduces a BIM semantic-geometric tight-coupling constraint. Global point, line, and plane features, together with component semantics from the bridge BIM model, are embedded into sliding-window optimization as a continuously active global constraint, which to our knowledge

has not been jointly exploited in prior BIM-aided navigation, providing a persistent global reference for odometry drift correction in GNSS-occluded regions. It also proposes a semantic-guided feature matching mechanism. By filtering matches according to bridge component semantics, the method reduces false correspondences from repetitive structural elements and improves robustness in low-texture and low-light inspection scenes. These three methodological contributions are realized in a low-cost, end-to-end implementation that uses only a monocular global-shutter camera, a low-cost MEMS IMU, and a civilian GNSS module, making the framework compatible with mainstream small inspection UAVs without specialized high-end payloads.

We note that the scope of this work is the navigation and global localization stage that provides the centimeter-level pose prerequisite for defect-to-component association; the detection of defects and their registration to BIM components are beyond the present scope and are discussed as future work in Section 5. The rest of the paper proceeds in five sections. Section 2 reviews visual–inertial odometry, model-constrained navigation, and multi-sensor fusion for UAV navigation, and positions the present method within this literature. Section 3 gives the formulation and implementation of BCEF-Nav, covering coordinate definitions, preprocessing, residual construction, and joint optimization. Section 4 reports the validation results, including digital twin simulations, field tests on an in-service expressway bridge, and ablation studies for the main modules. Section 5 analyzes the observed performance, limitations, and possible extensions. Section 6 closes the paper with the main findings and contributions.

2. Related Work

This work relates to three research areas: visual–inertial odometry for GNSS-denied environments, BIM-constrained navigation, and multi-sensor fusion for UAVs. This section briefly reviews representative studies and highlights the limitations of existing methods for bridge inspection under intermittent GNSS outages.

2.1. Visual–Inertial Odometry Under GNSS-Denied Environments

Visual–inertial odometry is widely used for UAV navigation when GNSS is degraded or unavailable. Cameras and IMUs are inexpensive, lightweight, and easy to integrate, making VIO an attractive solution for UAV navigation. Early VIO systems were primarily filter-based, whereas recent methods predominantly adopt tightly coupled optimization frameworks that jointly estimate visual and inertial states rather than processing them separately.

Several representative systems define the baseline for this field. VINS-Mono [6] introduced a tightly coupled monocular VIO framework based on sliding-window nonlinear optimization. ORB-SLAM3 [7], developed from the widely used ORB-SLAM2 framework [8], extended visual SLAM with a multi-map architecture for improved robustness and long-term operation. Point-line VIO formulations further showed that line features can complement point features in structured scenes [9]. These methods have since become common references for evaluating VIO accuracy, robustness, and computational efficiency.

Recent work has adapted VIO to more specific operating conditions. StructVIO exploits structural regularities in man-made environments to improve estimation accuracy in structured scenes [10]. RL-VIO targets dynamic visual–inertial settings with an emphasis on robustness and lightweight execution [11]. Other lightweight and line-feature variants improve the front end, back end, or feature representation to reduce computational cost on embedded platforms. They maintain reliable performance in low-texture and structurally repetitive environments [12–16].

These advances do not remove several limitations of VIO in UAV-based bridge inspection. A VIO system estimates motion relative to its initial frame. By itself, it does not

anchor the trajectory to the bridge engineering coordinate system, which is required for defect localization, component-level indexing, and lifecycle management. Bridge scenes also contain many repeated standardized components, including regularly spaced piers, diaphragms, and similar beam segments. These patterns increase the risk of false visual associations and can corrupt state estimation. The problem becomes more severe beneath concrete beams and inside box girders, where texture is weak and GNSS outages may last for long intervals. Odometry drift then accumulates quickly. Methods such as StructVIO improve estimation in structured environments, but they still lack a fixed global reference and cannot disambiguate repetitive bridge components without additional prior information.

Existing VIO methods provide reliable local motion estimation but lack a persistent global reference in the bridge engineering coordinate system. Moreover, repetitive bridge structures increase the risk of false feature associations. BCEF-Nav addresses this limitation by tightly integrating semantic and geometric BIM constraints within a unified optimization framework. Adaptive BIM constraint weighting further improves the robustness of global positioning under GNSS-degraded conditions.

2.2. Prior Model-Constrained Navigation Methods

Model-constrained navigation uses prior geometric or semantic knowledge of the environment to constrain SLAM and odometry estimates and reduce drift. In civil infrastructure, BIM models are a natural source of such prior information. They contain accurate geometry, component-level semantics, and a predefined engineering coordinate frame. These properties make BIM useful not only for map association, but also for tying the estimated trajectory to project coordinates. Existing BIM-based navigation methods can be broadly divided into loose-coupling and tight-coupling designs, with different levels of interaction between the prior model and the state estimator.

Existing BIM-aided navigation methods generally follow two designs. One group uses loose coupling, as in BIM-SLAM and related BIM-informed visual SLAM systems [17–19]. In these methods, the BIM model is mainly used for map association, localization, or loop-closure detection. Observed features are matched to the BIM model after the system revisits a region, and the match is then used to correct accumulated drift. Other smart-infrastructure visualization frameworks focus more on data organization and inspection visualization than on real-time pose constraints [20]. Loose coupling is easy to deploy, but it does not correct odometry drift continuously during GNSS outages. In long occluded segments, the estimation error can still grow without bound.

The other group uses partial tight coupling. These methods insert BIM-derived geometric features directly into the state estimation problem. Examples include BIM-based geometric constraints for visual-inertial odometry and BIM-LIO/LIO-BIM systems such as LIO-BIM [21], where BIM geometry is incorporated into LiDAR-inertial optimization. Tight coupling improves drift suppression compared with loose coupling. Existing methods, however, mainly use BIM as a geometric prior. The semantic labels in the model are rarely used to disambiguate repeated bridge components, such as similar piers, diaphragms, and beam segments. This limits their robustness in bridge scenes with strong structural repetition. Existing BIM-aided navigation frameworks also do not target the intermittent GNSS occlusion pattern of bridge inspection. They lack an elastic relation between BIM constraint weighting and GNSS validity, which is needed for stable state estimation when GNSS measurements repeatedly disappear and recover.

Although BIM-based navigation provides a global reference for infrastructure inspection, existing methods rarely exploit semantic information and do not explicitly address intermittent GNSS outages. BCEF-Nav addresses these limitations through semantic-geometric tight coupling and adaptive constraint weighting.

2.3. Multi-Sensor Fusion Navigation for UAVs

Multi-sensor fusion of GNSS, vision, and IMU measurements is the standard solution for outdoor UAV navigation. The sensors provide complementary information. GNSS anchors the trajectory to a global frame. Vision estimates short-term relative motion with high local accuracy. The IMU supplies high-rate motion measurements and maintains state propagation when visual observations are sparse or unstable. GNSS–visual–inertial fusion has therefore moved from loosely coupled designs, where each sensor stream is processed largely independently, to tightly coupled estimators that optimize raw measurements within a unified state estimation problem. This shift improves accuracy and robustness, especially when the sensing conditions are incomplete or degraded.

VINS-Fusion [22] is a representative system in this area. It provides an optimization-based framework for global pose estimation and supports tightly coupled fusion of GNSS, visual, and inertial measurements. Its treatment of GNSS failure, however, is still based on hard switching. Once GNSS is judged invalid, the corresponding constraints are removed from the optimization. This can produce pose discontinuities when the signal drops out or recovers, and it leaves the estimator without a global reference during longer outages. Recent studies have attempted to reduce this problem through more robust GNSS/INS/LiDAR, GNSS–visual–inertial, and federated fusion architectures for UAV and robot navigation under degraded GNSS conditions [23–27]. Elastic fusion methods offer another route. LIC-Fusion [28] adjusts the GNSS measurement weight according to signal quality, allowing smoother transitions between reliable and degraded GNSS states. LIO-SAM [29] further develops adaptive multi-sensor fusion for complex navigation scenes. These methods reduce the pose jumps caused by hard switching. They do not, however, provide an alternative global constraint when GNSS is fully occluded. In that case, the estimator falls back on relative odometry, and drift continues to grow with the outage duration. This limitation is especially restrictive in UAV bridge inspection. Full GNSS occlusion can persist inside box girders and other enclosed bridge regions. Elastic weighting alone cannot maintain centimeter-level global positioning under such conditions, because the missing GNSS constraint is not replaced by another reference tied to the bridge engineering coordinate system.

Existing GNSS–visual–inertial fusion frameworks remain vulnerable to long GNSS outages because they lack an alternative global reference. BCEF-Nav addresses this limitation by adaptively replacing degraded GNSS constraints with BIM semantic–geometric constraints within a unified optimization framework.

3. Materials and Methods

Figure 1 illustrates the overall architecture of BCEF-Nav. The framework adopts a sliding-window nonlinear optimization back end that tightly couples multi-source sensor measurements with prior model constraints. The monocular global-shutter camera (1280×720 , 30 fps) captures undistorted image frames, from which semantic masks, ORB point features, LSD line features, and region-growing plane features are extracted. The MEMS IMU outputs high-rate three-axis angular velocity and linear acceleration measurements, which are used to construct the inter-keyframe pre-integration terms. The GNSS module provides two types of information. The first is the global 3D antenna position in the world coordinate system, which is used to construct the GNSS residual. The second consists of quality indicators, including the number of visible satellites, the PDOP value, and the positioning covariance matrix. These indicators determine GNSS validity and the adaptive constraint weight rather than directly contributing to positioning. This separation enables the proposed elastic fusion scheme by adapting the influence of GNSS constraints according to measurement quality. The high-precision as-built BIM model

serves as an offline-generated global feature library, where point, line, and plane features are represented in the bridge engineering coordinate system and associated with their corresponding component semantic labels. After preprocessing, GNSS validity assessment, residual construction, and joint optimization, the framework outputs real-time global pose estimates in the bridge engineering coordinate system, providing the centimeter-level positioning accuracy required for UAV-based bridge inspection.

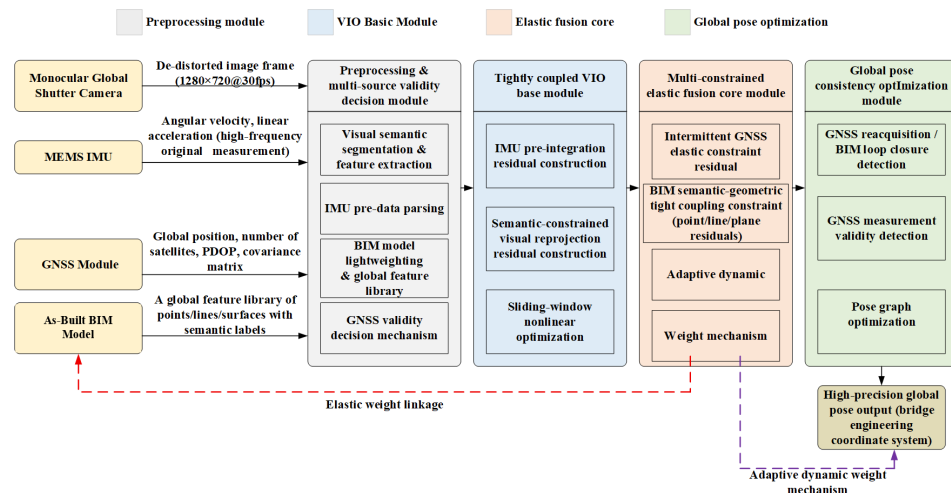


Figure 1. Overall architecture of the proposed BCEF-Nav framework. The framework takes monocular images, MEMS IMU data, GNSS measurements and as-built BIM model as inputs, and outputs high-precision global pose aligned with the bridge engineering coordinate system in real time. Colors distinguish the sensor/model inputs, functional modules, and system output only and carry no quantitative meaning.

3.1. Coordinate System Definition and System State Vector

To place sensor measurements and BIM priors in a single estimation problem, four right-handed coordinate frames are defined. The world frame w is aligned with the bridge BIM model's engineering coordinate system, either CGCS2000 or a local independent coordinate system. It serves as the fixed global reference for state estimation, so the estimated UAV pose can be associated directly with bridge component IDs and mileage stakes. The body frame B is located at the UAV center of mass and coincides with the flight-control frame, which keeps the navigation output compatible with the onboard control system. The camera frame c is located at the camera optical center. Its extrinsic transform relative to the body frame is pre-calibrated and denoted by $T_{BC} \in SE(3)$, describing the fixed rotation and translation between the camera and the UAV body. The BIM frame BIM is fixed to the as-built model. After initial hand-eye alignment, it is aligned with the world frame, and the corresponding transform is denoted by $T_{WM} \in SE(3)$.

With the coordinate systems defined, we construct the full state vector of the system for sliding window optimization. The sliding window contains n keyframes, selected based on parallax and motion magnitude to balance estimation accuracy and computational efficiency, and the full state vector is defined as:

$$\mathcal{X} = [x_0, x_1, \dots, x_n, T_{WM}, \mathcal{P}] \quad (1)$$

where the body state of the i -th keyframe, which captures the full motion state of the UAV at that timestamp, is defined as:

$$x_i = [T_{WB_i}, v_{WB_i}, b_{g_i}, b_{a_i}], i = 0, 1, \dots, n \quad (2)$$

In Equation (2), $T_{WB_i} = (R_{WB_i}, p_{WB_i}) \in SE(3)$ denotes the pose of the UAV body in the world frame at time i , where R_{WB_i} is the rotation matrix and p_{WB_i} is the position vector. The velocity term $v_{WB_i} \in \mathbb{R}^3$ gives the body linear velocity expressed in the world frame. The IMU bias terms $b_{g_i} \in \mathbb{R}^3$ and $b_{a_i} \in \mathbb{R}^3$ correspond to the gyroscope and accelerometer biases, respectively. They are estimated online to compensate for drift in low-cost MEMS IMUs. The BIM-to-world transform T_{WM} is retained as a slow-varying optimization variable after initial alignment, allowing small residual registration errors to be absorbed without breaking global consistency. The set $\mathcal{P}$ contains the 3D world coordinates of visual landmarks tracked within the sliding window.

3.2. Preprocessing and Multi-Source Data Validity Decision Module

Before residual construction, all sensor inputs pass through a preprocessing pipeline. The pipeline extracts usable features, checks measurement validity, and filters unreliable observations before they enter the state estimator.

3.2.1. Multi-Source Data Preprocessing

Visual preprocessing starts from the global-shutter images. Each frame is undistorted using the calibrated camera intrinsics. To reduce false matches from repeated bridge components, YOLOv8-Seg [30] is used for real-time semantic segmentation. YOLOv8-Seg was selected because it provides a favorable balance between segmentation accuracy and computational efficiency, making it suitable for real-time UAV bridge inspection while preserving reliable semantic information for subsequent feature matching. The YOLOv8-Seg segmenter was trained on an in-house bridge-component dataset collected by Shanxi Transportation Holding Group (not publicly available due to commercial confidentiality), containing more than 5300 images randomly split into training, validation, and test sets in an 8:1:1 ratio, with four core semantic classes: beam, pier, bearing, and diaphragm. On the target-bridge test set, the network achieved an overall segmentation accuracy of 92.3% mIoU, with per-class IoU of 95.1% (beam), 93.7% (pier), 88.2% (bearing), and 92.2% (diaphragm). Given that segmentation drives the semantic-guided matching pipeline, mislabeling could introduce erroneous global constraints. To mitigate this, the matching-confidence term s_{match} in Equation (11) down-weights low-confidence associations, preventing occasional segmentation errors from dominating the optimization. The network produces pixel-level masks for major bridge elements, including beams, piers, bearings, and diaphragms [30]. ORB point features [31], LSD line features [32], and region-growing plane features are then extracted within the segmented image regions. Feature matching is allowed only between regions with the same semantic label. This filtering suppresses cross-component associations among repetitive structural elements, which are a common source of estimation error in bridge inspection scenes. Therefore, semantic-guided matching improves the reliability of the constructed BIM constraints before they are incorporated into the optimization framework. In well-lit conditions, the frame-level segmentation failure rate of YOLOv8-Seg is about 3%. Such occasional failures cause only a slight increase in positioning error rather than trajectory jumps or large drift: because the matching-confidence term s_{match} in Equation (11) down-weights low-confidence associations, a mislabeled or missed segment contributes little to the optimization, and the system stays within the engineering accuracy threshold throughout.

IMU preprocessing uses midpoint integration for pre-integration between adjacent keyframes. This avoids propagating the full state at every IMU timestamp and reduces the computational load of the estimator. During pre-integration, the current estimates of IMU bias and gravity are compensated continuously. The formulation follows on-manifold IMU pre-integration to keep the relative motion constraint consistent on $SE(3)$ [17].

GNSS preprocessing parses the receiver output to obtain the global position, the number of visible satellites, the PDOP value, and the position covariance matrix. These quantities are used to build the GNSS residual. They also provide the basis for GNSS validity assessment and for assigning the corresponding constraint weight in the optimization.

BIM preprocessing is completed offline before the inspection mission. The as-built BIM model is first converted into a lightweight triangular mesh to reduce online computation. A global feature library is then extracted from the model, with each feature assigned to its corresponding bridge component label. The library contains plane features represented by normal vectors and global plane equations, line features represented by the global coordinates of their endpoints, and point features represented by their global coordinates in the BIM frame. A KD-tree is built over the feature library to support fast nearest-neighbor queries during online matching, limiting the latency added by BIM-based association.

3.2.2. GNSS Validity Decision Mechanism

To reliably distinguish between GNSS-available and GNSS-denied states in the intermittent occlusion scenario of bridge inspection, we design a multi-dimensional GNSS validity decision rule that accounts for both signal quantity and quality. A GNSS measurement is only classified as valid if all of the following conditions are simultaneously satisfied; if any condition is not met, the measurement is classified as invalid and excluded from the optimization, with its constraint weight set to zero:

$$\left\{ \begin{array}{l} \text{Number of visible satellites} \geq 6 \\ \text{PDOP} \leq 3 \\ \text{Plane positioning covariance} \leq 0.1 \text{ m}^2 \\ \text{Elevation positioning covariance} \leq 0.1 \text{ m}^2 \\ \text{Signal lock time} \geq 10 \text{ s} \end{array} \right. \quad (3)$$

This multi-criterion rule reduces false-valid decisions from degraded GNSS observations, such as measurements with too few visible satellites or large dilution of precision. We adopt a linear-complementary form for two practical reasons. First, it makes the BIM and GNSS weights vary monotonically and continuously over the whole range $\lambda_{\text{GNSS}, \text{norm}} \in [0, 1]$, with no inflection point, matching the intended behavior that the BIM constraint should dominate in proportion to how far the GNSS reference has degraded. Second, unlike an exponential or sigmoid mapping—each of which would add a shape parameter (a decay rate or a steepness term) to be tuned—the linear form needs only the two end values, which keeps the handover easy to interpret and cheap to compute on the embedded platform. A systematic comparison against exponential and sigmoid alternatives is left for future work. As a result, only reliable GNSS measurements are allowed to constrain the global pose estimate. Figure 2 shows the validity decision process and its coupling with the adaptive elastic weighting of BIM constraints.

3.3. Tightly Coupled VIO Base Module

The VIO module provides relative motion estimation through tightly coupled visual and inertial constraints within a sliding-window optimization framework. GNSS and BIM measurements are subsequently introduced as global constraints to suppress long-term drift.

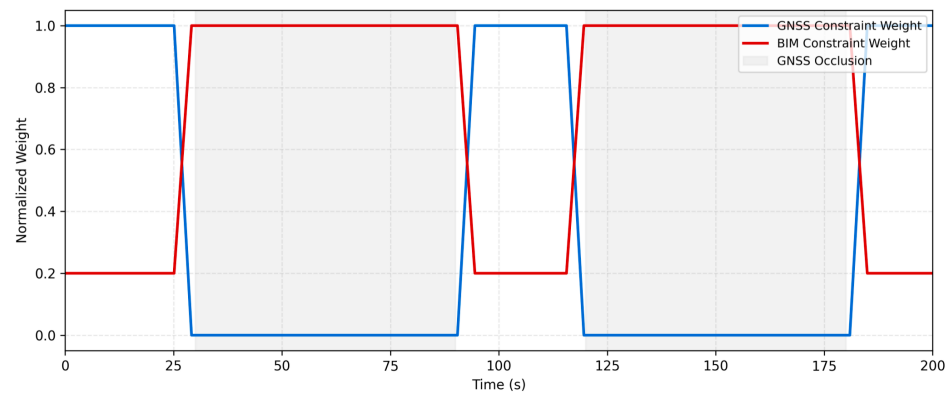


Figure 2. Schematic diagram of GNSS validity decision and adaptive elastic weight mechanism. The weight of BIM constraints automatically increases when GNSS is invalid, replacing GNSS to provide global pose correction, achieving seamless state switching without pose jumps.

3.3.1. IMU Pre-Integration Residual

The IMU pre-integration residual provides a relative motion constraint between adjacent keyframes i and j , based on the pre-integrated IMU measurements collected between the two keyframe timestamps. This residual quantifies the discrepancy between the state transition predicted by the IMU pre-integration and the state transition estimated by the optimization framework, and is defined on the SE(3) manifold as:

$$r_{IMU}(x_i, x_j) = \begin{bmatrix} \delta\alpha_{ij} \\ \delta\beta_{ij} \\ \delta\gamma_{ij} \\ \delta b_{g_i} \\ \delta b_{a_i} \end{bmatrix} = \begin{bmatrix} R_{WB_i}^\top \left(p_{WB_j} - p_{WB_i} - \frac{1}{2} g^w \Delta t_{ij}^2 \right) - \hat{\alpha}_{ij} \\ 2 \text{Log} \left(\hat{\gamma}_{ij}^\top R_{WB_j} \right)^\top \\ R_{WB_i}^\top \left(v_{WB_j} - v_{WB_i} - g^w \Delta t_{ij} \right) - \hat{\beta}_{ij} \\ b_{g_j} - b_{g_i} \\ b_{a_j} - b_{a_i} \end{bmatrix} \quad (4)$$

In this formulation, $\hat{\alpha}_{ij}$, $\hat{\gamma}_{ij}$, and $\hat{\beta}_{ij}$ are the pre-integrated position, rotation, and velocity measurements from the IMU, respectively; Δt_{ij} is the time interval between keyframe i and keyframe j ; and g^w is the gravity vector expressed in the world coordinate system. The residual term includes both the motion discrepancy between the pre-integrated measurements and the optimized state, as well as the change in IMU bias between the two keyframes, which is assumed to follow a random walk model.

3.3.2. Semantic-Constrained Visual Reprojection Residual

The visual reprojection residual quantifies the discrepancy between the observed pixel position of a 3D landmark in an image frame, and the pixel position predicted by projecting the 3D landmark into the camera frame using the optimized camera pose. To eliminate mismatches from repetitive bridge components, we only construct this residual for landmarks that have been filtered by the semantic mask, ensuring that only features from the same bridge component are used for tracking and optimization. For a 3D landmark l with a consistent semantic label, observed in the i -th and j -th keyframes, the reprojection residual for the i -th frame is defined as:

$$r_{VIS}(x_i, x_j, p_l^w) = u_{l,i} - \pi_c \left(T_{CB} T_{WB_i}^{-1} p_l^w \right) \quad (5)$$

where p_l^w is the 3D coordinate of the landmark in the world coordinate system; $u_{l,i}$ is the observed pixel coordinate of the landmark in the i -th keyframe; $\pi_c(\cdot)$ is the pinhole projection model of the camera, which maps a 3D coordinate in the camera frame to a 2D

pixel coordinate; and $T_{CB} = T_{BC}^{-1}$ is the pre-calibrated extrinsic transformation matrix from the body coordinate system to the camera coordinate system.

3.4. Core Module of Multi-Constrained Elastic Fusion

This module provides the main elastic fusion mechanism in BCEF-Nav. It combines GNSS global constraints with BIM semantic–geometric constraints through adaptive dynamic weighting. The estimator does not switch abruptly between GNSS-available and GNSS-denied modes. Instead, the GNSS weight decreases when the signal becomes unreliable, while the BIM constraint weight increases accordingly. When GNSS is invalid, the BIM model serves as the dominant global reference and continues to correct the pose in the bridge engineering coordinate system. This design preserves pose continuity and limits odometry drift during signal outages.

3.4.1. Intermittent GNSS Elastic Constraint

When GNSS signals are classified as valid by the decision mechanism, the GNSS measurement provides the absolute 3D position of the UAV's GNSS antenna in the world coordinate system. The installation offset between the GNSS antenna and the UAV's center of mass (body coordinate system origin) is pre-calibrated and known, enabling us to map the GNSS measurement directly to the body pose. The GNSS residual, which quantifies the discrepancy between the optimized body position and the position measured by GNSS, is defined as:

$$r_{GNSS}(x_i) = p_{WB_i} + R_{WB_i}p_{BC} - p_{GNSS_i} \quad (6)$$

where p_{BC} is the pre-calibrated position offset of the GNSS antenna relative to the body coordinate system origin, expressed in the body frame. It is a fixed extrinsic parameter calibrated once before the mission by combining the mechanical mounting layout with total-station measurement, and remains constant during flight; on the platform used in this work, its components do not exceed ± 0.15 m along the three body axes. p_{GNSS_i} is the 3D position of the GNSS antenna measured by the receiver in the world coordinate system; it is provided by the u-blox ZED-F9P RTK-GNSS module (u-blox AG, Thalwil, Switzerland) directly in the bridge engineering coordinate frame and reaches centimeter-level accuracy under an RTK fixed solution.

To enable elastic fusion of GNSS measurements based on signal quality, we design an adaptive weight for the GNSS constraint that is directly linked to the measurement's covariance matrix. The weight is defined as:

$$\lambda_{GNSS} = \begin{cases} \text{tr}(\Sigma_{GNSS_i}^{-1}), & \text{GNSS valid} \\ 0, & \text{GNSS invalid} \end{cases} \quad (7)$$

where Σ_{GNSS_i} is the positioning covariance matrix output in real time by the receiver together with each position fix (rather than a manually set value), and $\text{tr}(\cdot)$ denotes the trace operation. Under an RTK fixed solution on our platform, the horizontal and vertical variances of Σ_{GNSS_i} typically lie in the range 0.0001–0.01 m², consistent with the ≤ 0.1 m² validity threshold in Equation (3); accordingly, the resulting weight $\lambda_{GNSS} = \text{tr}(\Sigma^{-1})$ varies continuously over several orders of magnitude with signal quality. To make this raw weight comparable with the BIM elastic weight in Equation (11), λ_{GNSS} is normalized to $\lambda_{GNSS, \text{norm}} \in [0, 1]$ by a rolling normalization against the maximum weight observed within a recent history window. We emphasize that the valid/invalid state is determined solely by the decision rule in Equation (3); the weight computation in Equation (7) is decoupled from this decision, avoiding any circular dependency between validity judgement and weighting.

We clarify that setting the GNSS weight to zero upon invalidity is a deliberate design choice, since invalid observations must not enter the optimization. The “elastic” behavior of the framework therefore does not refer to a continuous decay of the GNSS term itself, but to the smooth handover between the GNSS and BIM global constraints: as λ_{GNSS} drops to zero, λ_{BIM} in Equation (11) increases continuously and takes over the global correction, so that the overall pose estimate remains free of discontinuities.

3.4.2. BIM Semantic–Geometric Global Tight Coupling Constraint

The BIM semantic–geometric constraint provides the main global correction term when GNSS is unavailable. It replaces the missing GNSS reference with pose constraints defined in the bridge engineering coordinate system. Based on the pre-extracted feature library from the as-built BIM model, three residuals are constructed from plane, line, and point features and inserted into the sliding-window optimization. Matching is performed only between features with the same semantic label. This restriction reduces false associations among repetitive bridge components, as shown in Figure 3.

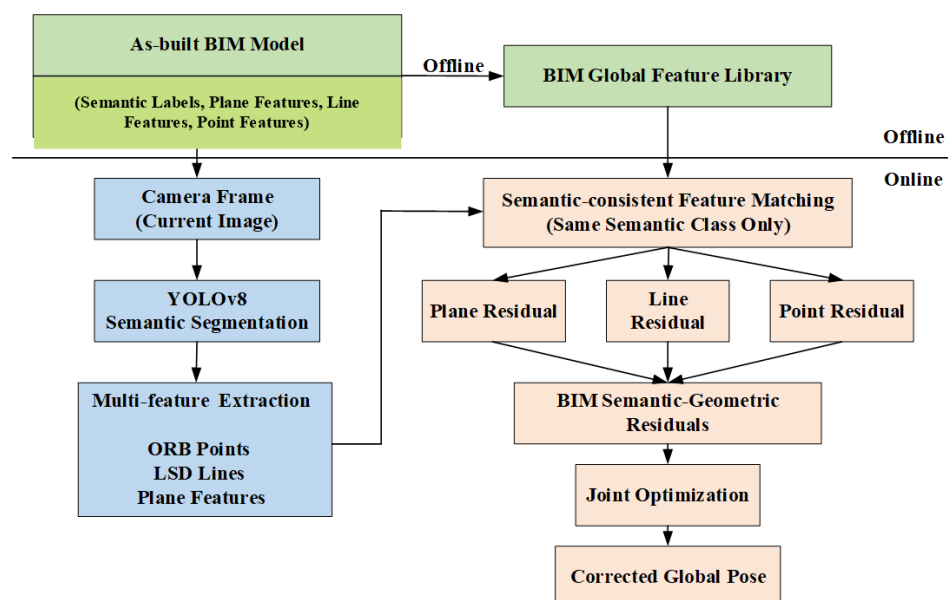


Figure 3. Workflow of the proposed BIM semantic–geometric tight coupling constraint. Offline BIM feature construction and online semantic-guided feature matching jointly generate plane, line, and point residuals for global pose optimization. Green denotes offline BIM data, blue denotes online visual processing, and orange denotes residual construction and optimization; the colors carry no quantitative meaning.

(1) BIM Plane Feature Residual

Plane features are the dominant geometric primitives in bridge scenes. They appear on beam soffits, box girder inner walls, pier surfaces, and deck slabs, and therefore form the main source of BIM constraints. For a real-time plane π_i with semantic label L extracted from the i -th keyframe, its equation in the camera frame is written as $n_{C_i}^\top p_C + d_{C_i} = 0$, where n_{C_i} denotes the plane normal and d_{C_i} denotes the distance offset. The plane is then associated with a BIM plane π_{BIM}^L carrying the same semantic label in the global feature library. The corresponding BIM plane is defined in the world frame as $n_{BIM}^\top p_W + d_{BIM} = 0$, with all parameters fixed by the as-built BIM model.

The plane feature residual is defined as a joint constraint that penalizes both the angle discrepancy between the plane normal vectors and the distance deviation between

the two planes, ensuring that both the orientation and position of the estimated pose are constrained by the BIM plane feature:

$$r_{BIM,plane}(x_i) = \begin{bmatrix} 1 - |n_{BIM}^\top R_{WB_i} R_{BC} n_{C_i}| \\ |n_{BIM}^\top p_{WB_i} + d_{BIM} - n_{C_i}^\top p_{BC} + d_{C_i}| \end{bmatrix} \quad (8)$$

The first residual term measures the angular discrepancy between the real-time plane normal, after transformation into the world frame, and the BIM plane normal. It becomes zero when the two normals are aligned. The second term measures the plane-to-plane distance error in the world frame and vanishes when the two planes coincide.

(2) BIM Line Feature Residual

Line features complement plane constraints, especially along beam edges, pier corners, and bearing seats, where geometric boundaries remain observable even under weak texture. For a real-time line l_i with semantic label L extracted from the i -th keyframe, the corresponding BIM line l_{BIM}^L is retrieved from the feature library using the same semantic label. Its endpoints p_{BIM1} and p_{BIM2} are fixed in the world frame by the BIM model. The line residual is defined as the sum of the Euclidean distances from the two endpoints of the real-time line, after transformation into the world frame, to the matched BIM line. This term measures the geometric consistency between the observed line and its BIM counterpart:

$$r_{BIM,line}(x_i) = d(p_{W1}, l_{BIM}^L) + d(p_{W2}, l_{BIM}^L) \quad (9)$$

where $d(p, l)$ denotes the Euclidean distance from a 3D point p to a 3D straight line l , and p_{W1} and p_{W2} are the two endpoints of the real-time line feature transformed from the camera coordinate system to the world coordinate system using the optimized body pose.

(3) BIM Point Feature Residual

Point features provide sparse but distinctive constraints from salient structural elements, such as bearing corners, embedded parts, and expansion joint edges. These elements have fixed 3D coordinates in the BIM model. For a real-time salient point p_{C_i} with semantic label L extracted from the i -th keyframe, the corresponding BIM point p_{BIM} is retrieved from the feature library using the same semantic label. Its 3D coordinate is fixed in the world frame. The point residual is defined as the Euclidean distance between the transformed real-time point and the matched BIM point:

$$r_{BIM,point}(x_i) = T_{WB_i} T_{BC} p_{C_i} - p_{BIM} \quad (10)$$

This residual directly constrains the absolute global position of the UAV, as the BIM point feature provides a fixed reference in the bridge's engineering coordinate system.

(4) Elastic Weight Design for BIM Constraints

To enable seamless transition between GNSS and BIM constraints, we design an adaptive dynamic weight for the BIM constraints that is linked to both the feature matching confidence and the validity of the GNSS measurements. The weight is defined as:

$$\lambda_{BIM} = s_{match} \cdot (\lambda_{BIM,base} + (1 - \lambda_{GNSS,norm}) \cdot (\lambda_{BIM,max} - \lambda_{BIM,base})) \quad (11)$$

where $s_{match} \in [0, 1]$ is the feature matching confidence, calculated based on the inlier rate of the matched BIM features and the semantic matching accuracy; $\lambda_{GNSS,norm} \in [0, 1]$ is the normalized GNSS weight, scaled to the range $[0, 1]$ from the raw GNSS weight in

Equation (7); and $\lambda_{BIM,max}$ and $\lambda_{BIM,base}$ are the maximum and base weights of the BIM constraints, respectively, with $\lambda_{BIM,max} > \lambda_{BIM,base}$.

This weighting scheme serves two purposes. As GNSS quality decreases, the BIM constraint weight increases. When GNSS becomes invalid, the BIM term reaches its maximum weight and acts as the main global constraint. The weight also depends on matching confidence. Low-confidence BIM associations are down-weighted, reducing the risk that incorrect correspondences will distort the state estimate. Unlike elastic GNSS-only schemes such as LIC-Fusion [28], which modulate constraint weights based on signal quality but lack an alternative global reference, our method prevents the estimator from falling back solely on relative odometry. Consequently, it avoids the unbounded drift associated with prolonged outages. The coupling in Equation (11) differs in that it ties the BIM weight directly to GNSS validity, so that as $\lambda_{GNSS,norm}$ falls to zero the BIM constraint grows to its maximum and takes over as the dominant global reference. This is exactly the failure mode—sustained full occlusion—that GNSS-only elastic weighting cannot handle, and it is borne out quantitatively in the long-occlusion results presented in Section 4, where LIC-Fusion degrades to 42.3 cm at 120 s while the proposed method remains at 5.7 cm. We adopt a linear-complementary form for two practical reasons. First, it makes the BIM and GNSS weights vary monotonically and continuously over the whole range $\lambda_{GNSS,norm} \in [0,1]$, with no inflection point, matching the intended behavior that the BIM constraint should dominate in proportion to how far the GNSS reference has degraded. Second, unlike an exponential or sigmoid mapping—each of which would add a shape parameter (a decay rate or a steepness term) to be tuned—the linear form needs only the two end values, which keeps the handover easy to interpret and cheap to compute on the embedded platform. Asystematic comparison against exponential and sigmoid alternatives is left for future work. Overall, this adaptive weighting strategy prevents unreliable BIM correspondences from dominating the optimization while allowing BIM constraints to gradually replace GNSS constraints during signal degradation. As a result, the proposed framework achieves a smooth transition between different positioning cues without introducing abrupt changes in the optimization process.

3.4.3. Joint Optimization Objective Function

After defining the residuals, the system objective is written as the weighted sum of their Mahalanobis distances. The resulting nonlinear least-squares problem is solved with the Levenberg–Marquardt algorithm, which is commonly used in sliding-window VIO optimization and provides stable convergence for this form of estimator:

$$\min_{\mathcal{X}} \left(\|r_{IMU}\|_{\Sigma_{IMU}}^2 + \|r_{VIS}\|_{\Sigma_{VIS}}^2 + \lambda_{GNSS} \|r_{GNSS}\|_{\Sigma_{GNSS}}^2 + \lambda_{BIM} \|r_{BIM}\|_{\Sigma_{BIM}}^2 \right) \quad (12)$$

Here, $|r| * \Sigma^2 = r^\top \Sigma^{-1} r$ denotes the squared Mahalanobis norm of residual r , with Σ as the corresponding covariance matrix. The matrices Σ_{IMU} , Σ_{VIS} , Σ_{GNSS} , and Σ_{BIM} define the measurement covariances for the IMU, visual, GNSS, and BIM feature terms, respectively. The BIM residual r_{BIM} stacks the plane, line, and point residuals derived from the BIM feature library. The adaptive weights λ_{GNSS} and λ_{BIM} regulate the two global constraints, allowing the estimator to remain stable across GNSS-available and GNSS-denied intervals.

3.5. Global Pose Consistency Optimization Module

Sliding-window optimization gives accurate local state estimates, but small global errors can still remain over a long inspection trajectory. To correct this residual drift, BCEF-Nav adds a global pose consistency module. When the UAV returns to a GNSS-available

region after an occlusion, or when a loop match is found between online features and the BIM global feature library, a pose graph is built from all keyframe poses in the trajectory. Pose graph optimization is then used to jointly minimize odometry relative-pose residuals and global constraints from GNSS and BIM loop closures. This step removes accumulated full-trajectory drift. The module follows the graph-based smoothing paradigm used in iSAM and g2o-style back ends [33,34]. After optimization, the inspection trajectory remains consistent with the bridge engineering coordinate system. Detected defects can therefore be associated with the corresponding BIM component IDs and mileage stakes throughout the mission.

4. Experimental Validation and Results

To evaluate BCEF-Nav, we conducted three sets of experiments: controlled digital-twin simulations, field tests on an in-service expressway bridge, and ablation studies for the main modules. The evaluation included representative SOTA baselines and used both quantitative trajectory metrics and qualitative trajectory comparisons. Following dataset-style protocols commonly used in long-term robotic navigation benchmarks [35], the experiments assess accuracy, robustness, and global consistency under different GNSS occlusion conditions.

4.1. Experimental Setup

4.1.1. Experimental Platform

All experiments used a mainstream small inspection quadrotor UAV. The hardware was selected to match the typical configuration of commercial platforms used in bridge inspection. Table 1 lists the detailed setup. For the two LiDAR-based baseline methods (LIO-BIM [21] and LIC-Fusion [28]), an auxiliary Livox Mid-70 solid-state LiDAR (Livox Technology Co., Ltd., Shenzhen, China; ranging 0.05–260 m, $70.4^\circ \times 77.2^\circ$ FOV, 144,000 points/s, ± 2 cm ranging accuracy, 720 g, 12 W, time-synchronized with the Pixhawk 6C (Holybro, Tai Po, Hong Kong, China) and Jetson Xavier NX (NVIDIA Corporation, Santa Clara, CA, USA)) was mounted on the same platform; this sensor was used exclusively for these two baselines and is not part of the BCEF-Nav sensor suite. The prior model was a high-precision as-built BIM model of a 1200 m in-service expressway bridge. The model has 5 mm geometric accuracy and contains 32 precast box girders with their corresponding piers. Its coordinate frame is consistent with the bridge as-built engineering coordinate system.

The two BIM constraint weight bounds in Equation (11) were set to $\lambda_{BIM,base} = 1.0$ and $\lambda_{BIM,max} = 6.0$ in all experiments.

Table 1. Hardware configuration of the experimental UAV platform.

Module	Model and Parameters
UAV Platform	Mao 3T-C quadrotor (Mao, Xi'an, China), diagonal wheelbase 380 mm, max takeoff weight ≈ 1930 g; payload: visual sensor and onboard computing unit
Flight Performance	Max horizontal speed 20 m/s; cruise speed in this work ≈ 9 m/s (32.4 km/h). Endurance (windless, sea level): max flight time 43 min (6000 mAh)/35 min (5000 mAh), max hover time 39 min (6000 mAh)/31 min (5000 mAh)
Flight Controller	Pixhawk 6C (Holybro, Tai Po, Hong Kong, China), built-in BMI088 MEMS IMU (Bosch Sensortec GmbH, Reutlingen, Germany; gyro bias stability $10^\circ/\text{h}$, accelerometer bias stability 0.1 mg)

Table 1. *Cont.*

Module	Model and Parameters
Camera	Basler dart daA1280-54um (Basler, Ahrensburg, Germany; AR0134 global-shutter sensor; configured at 1280×720 and 30 fps with an 8 mm lens)
GNSS Module	u-blox ZED-F9P RTK-GNSS (u-blox AG, Thalwil, Switzerland), cm-level positioning (also used as ground-truth source)
On-board Computing	Jetson Xavier NX (NVIDIA Corporation, Santa Clara, CA, USA; 8 GB), ROS Melodic
LiDAR (for baselines only)	Livox Mid-70 solid-state LiDAR (range 0.05–260 m @10% reflectivity, FOV $70.4^\circ \times 77.2^\circ$, 144,000 pts/s in dual-return mode, range accuracy ± 2 cm within 100 m, ≈ 720 g, 12 W, Ethernet output with hardware time synchronization to the Pixhawk and Jetson Xavier NX). Used only to run the LiDAR-based baselines (LIO-BIM, LIC-Fusion); not used by BCEF-Nav.
Prior Data	As-built BIM model (5 mm accuracy, IFC format, 1200 m, 32 spans, precast box girders)

4.1.2. Control Group SOTA Methods

To benchmark BCEF-Nav, we selected seven SOTA methods as baselines. The set includes four widely used benchmark systems and three recent methods published in top conferences or journals between 2024 and 2026. The classic baselines provide reference performance for established UAV navigation pipelines, while the recent methods reflect the latest progress in structured-scene VIO and BIM-aided navigation. Each baseline was chosen for its relevance to bridge inspection. All baselines were run from their official open-source code, with parameters set according to the recommendations in the original papers. The two LiDAR-based baselines (LIO-BIM and LIC-Fusion) were executed on data from an additional Livox Mid-70 solid-state LiDAR mounted on the same platform during acquisition (Table 1); this LiDAR was used solely to provide input for these two baselines and was not used by BCEF-Nav, which relies only on the monocular camera, MEMS IMU, and GNSS module. This configuration is disclosed to ensure a fair and transparent comparison. The control methods are summarized below.

1. VINS-Mono (IEEE TRO 2018) [6]: A representative tightly coupled monocular VIO system and a standard baseline in VIO evaluation. It estimates relative pose by jointly optimizing visual and inertial measurements in a sliding window, but does not include GNSS fusion or any global positioning constraint. We include it to measure the drift behavior of pure VIO during GNSS outages and to provide the lower-bound reference for vision-based navigation in the bridge inspection scenario.

2. ORB-SLAM3 (IEEE TRO 2021) [7]: A visual-inertial SLAM system with a multi-map architecture and loop-closure capability. It is included to evaluate whether advanced VIO with loop closure can remain stable in bridge inspection scenes. Although ORB-SLAM3 improves robustness over conventional VIO, it still lacks a fixed global reference tied to the bridge engineering coordinate system. During extended GNSS outages, it cannot maintain centimeter-level global alignment required for defect localization.

3. VINS-Fusion (arXiv 2019) [22]: A widely used GNSS-visual-inertial fusion framework for UAV navigation. It performs tightly coupled estimation with visual, inertial, and GNSS measurements, but treats GNSS failure through hard switching. When GNSS is invalid, the GNSS factor is removed from the optimization. We include this baseline to

compare hard switching with the proposed elastic fusion strategy and to evaluate the pose discontinuity and drift that occur under intermittent GNSS occlusion.

4. BIM-SLAM (ISARC 2023) [18]: A representative BIM-aided SLAM method based on loose coupling. In this system, the BIM model is used mainly for loop-closure detection rather than as a continuous constraint in the estimator. We include it to examine the behavior of loose-coupling BIM navigation during GNSS outages and to compare it with the proposed tightly coupled BIM constraint formulation.

5. StructVIO (IEEE TRO 2019) [10]: A structure-aware VIO system designed for man-made environments. It exploits regular geometric patterns to improve pose estimation in structured scenes. We include this baseline to assess how structured-scene VIO performs in bridge inspection. Although such priors improve local estimation, StructVIO still operates without a global BIM reference tied to the engineering coordinate system. It therefore cannot eliminate cumulative drift over long GNSS-denied trajectories.

6. LIO-BIM (Advanced Engineering Informatics 2025) [21]: A recent BIM-based tight-coupling method that incorporates BIM geometric features into LiDAR–inertial optimization. We use it as the main baseline for evaluating the BIM constraint module. The comparison tests whether the proposed semantic–geometric coupling offers advantages over purely geometric BIM constraints, especially in suppressing false associations caused by repetitive bridge components.

7. LIC-Fusion (IROS 2019) [28]: An elastic multi-sensor fusion framework that adjusts GNSS measurement weights to smooth the transition between GNSS-available and GNSS-denied states. We use it as the main baseline for the elastic fusion module. The comparison isolates the effect of adaptive weighting and tests its limitation under extended GNSS outages. Without an alternative global reference, such as the proposed BIM semantic–geometric prior, elastic GNSS–VIO fusion still cannot maintain centimeter-level positioning throughout long occluded segments.

4.1.3. Evaluation Metrics

To comprehensively and objectively evaluate the performance of the proposed framework and control methods, we adopted four widely accepted quantitative evaluation metrics that are standard in the field of robot navigation and UAV inspection. These metrics were selected to assess the global positioning accuracy, local odometry precision, engineering usability, and real-time performance of the system, providing a holistic evaluation of performance in the bridge inspection scenario. The detailed definition, calculation method, and engineering significance of each metric are described below:

1. Absolute Trajectory Error (ATE)

The ATE metric quantifies the discrepancy between the estimated trajectory and the ground truth trajectory in the global coordinate system, providing a direct measure of the system's global positioning accuracy. For a sequence of N estimated poses $T_{est,i} \in SE(3)$ and corresponding ground truth poses $T_{gt,i} \in SE(3)$, the ATE for the i -th frame is defined as the Euclidean distance between the translation components of the aligned estimated and ground truth poses:

$$ATE_i = \|\text{trans}(T_{gt,i}^{-1}T_{est,i})\|_2$$

where $\text{trans}(\cdot)$ returns the translation vector of an $SE(3)$ transformation. For the full trajectory, we report the RMSE, maximum error, and mean error of ATE. The RMSE is used as the primary indicator of global positioning accuracy. ATE is widely used in SLAM and navigation evaluation because it measures the consistency between the estimated trajectory and the ground-truth coordinate frame. This property matches the requirement

of bridge inspection, where UAV poses must remain aligned with the bridge engineering coordinate system.

2. Relative Pose Error (RPE)

The RPE metric evaluates the short-term accuracy of the local odometry system, quantifying the system's ability to estimate relative motion between consecutive frames over a fixed time interval. For a time interval Δt , the RPE for the i -th frame is defined as:

$$\text{RPE}_i = \|\text{trans}\left(\left(T_{gt,i}^{-1}T_{gt,i+\Delta t}\right)^{-1}\left(T_{est,i}^{-1}T_{est,i+\Delta t}\right)\right)\|_2$$

RPE evaluates local odometry drift after removing the effect of global alignment error. It reflects the stability and short-term precision of the VIO backbone. This metric is relevant to flight safety during inspection, since confined regions such as box girder interiors require reliable relative motion estimation for stable UAV control.

3. Positioning Success Rate

The positioning success rate is defined as the proportion of frames during GNSS occlusion where the absolute positioning error is less than 10 cm, calculated as:

$$\text{Success Rate} = \frac{\text{Number of frames with ATE} < 10 \text{ cm}}{\text{Total number of frames during GNSS occlusion}} \times 100\%$$

The 10 cm threshold follows the accuracy requirement of UAV bridge inspection. To bind a detected defect to the correct BIM component ID and mileage stake, the positioning error must remain below 10 cm. Larger errors may assign the defect to the wrong structural element, making the inspection record unusable for maintenance planning. This metric therefore measures engineering usability: the proportion of the mission in which the system satisfies the required centimeter-level positioning accuracy. In this sense, the positioning success rate is not only a navigation metric but also a proxy for detection usability: it directly measures the fraction of observations that can be correctly bound to their BIM component, which is the precondition for any downstream defect-to-component association.

4. Real-Time Performance

Real-time performance is evaluated by the average processing frame rate on the onboard Jetson Xavier NX platform, reported in frames per second. For UAV navigation, the estimator should operate close to the camera capture rate, which is 30 fps in this work, so that pose outputs are available in time for flight control and defect localization. This metric tests whether the system can be deployed on the limited computing hardware of small inspection UAVs.

4.2. Digital Twin Simulation Experiment

We first used digital-twin simulations to evaluate BCEF-Nav under controlled GNSS occlusion, illumination, and structural conditions. The simulation scene was built in Unity as a 1:1 replica of the expressway bridge used in the field tests, with the same geometry, material settings, and component layout. This setting provides two advantages. It allows direct control of variables such as GNSS outage duration, illumination intensity, and structural configuration, which are difficult to isolate in field experiments. It also provides millimeter-level ground-truth trajectories over the full inspection path, including occluded regions where external ground-truth measurement is difficult. Based on this environment, two experiments were conducted: an accuracy comparison under different GNSS occlusion durations and a robustness test under low-illumination and low-texture conditions.

4.2.1. Accuracy Comparison Under Different GNSS Occlusion Durations

To test positioning accuracy under GNSS outages of different lengths, we simulated four occlusion durations: 10 s, 30 s, 60 s, and 120 s. These settings cover the outage range commonly encountered in UAV-based bridge inspection. For each duration, 10 inspection flights were repeated. The average ATE RMSE of each method was then computed across the repeated trials. Table 2 reports the quantitative results, and Figure 4 shows the ATE RMSE trend as the occlusion duration increases.

Table 2. ATE RMSE comparison of each method under different GNSS occlusion durations (unit: cm). Values are reported as mean $\pm$ standard deviation over 10 repeated trials. The best result in each row is shown in bold.

Occlusion	VINS-Mono	ORB-SLAM3	VINS-Fusion	BIM-SLAM	StructVIO	LIO-BIM	LIC-Fusion	BCEF-Nav
10 s	5.2 $\pm$ 0.4	4.8 $\pm$ 0.3	4.5 $\pm$ 0.3	3.8 $\pm$ 0.3	3.6 $\pm$ 0.3	3.2 $\pm$ 0.2	3.9 $\pm$ 0.3	2.1 $\pm$ 0.2
30 s	18.6 $\pm$ 1.7	15.3 $\pm$ 1.4	12.7 $\pm$ 1.1	8.2 $\pm$ 0.7	7.8 $\pm$ 0.7	5.9 $\pm$ 0.4	9.5 $\pm$ 0.8	2.8 $\pm$ 0.2
60 s	42.5 $\pm$ 4.8	36.8 $\pm$ 3.6	31.2 $\pm$ 3.5	15.6 $\pm$ 1.3	22.4 $\pm$ 2.1	10.3 $\pm$ 0.8	24.8 $\pm$ 2.3	3.5 $\pm$ 0.3
120 s	98.3 $\pm$ 12.5	85.6 $\pm$ 10.7	78.9 $\pm$ 9.6	32.4 $\pm$ 3.1	45.2 $\pm$ 5.3	18.6 $\pm$ 1.5	42.3 $\pm$ 4.7	5.7 $\pm$ 0.6

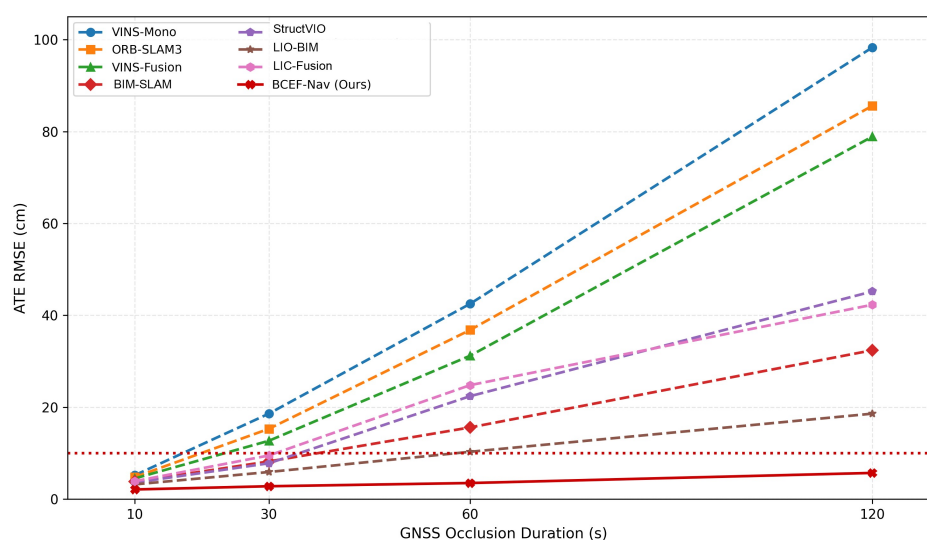


Figure 4. ATE RMSE comparison under different GNSS occlusion durations. The red horizontal dotted line indicates the 10 cm engineering threshold. Baselines shown are VINS-Mono [6], ORB-SLAM3 [7], VINS-Fusion [22], BIM-SLAM [18], StructVIO [10], LIO-BIM [21], and LIC-Fusion [28].

The results show an increasing separation between BCEF-Nav and the baselines as the GNSS outage becomes longer. For pure VIO methods, including VINS-Mono, ORB-SLAM3, and StructVIO, as well as GNSS–VIO fusion methods such as VINS-Fusion and LIC-Fusion, ATE RMSE grows rapidly with occlusion duration. This behavior is expected for relative odometry without an active global constraint. Visual and inertial errors accumulate over time, and the trajectory gradually departs from the engineering coordinate frame. At 120 s of complete GNSS occlusion, the errors of these methods exceed 70 cm, far beyond the accuracy required for bridge inspection.

BIM-SLAM reduces drift compared with pure VIO, but its error still increases with occlusion duration. The reason is its loose-coupling design: the BIM model is used mainly for loop-closure detection and does not provide continuous correction during the occluded segment. LIO-BIM gives the best result among the baselines because BIM geometric features are included in the LiDAR–inertial optimization. Its ATE RMSE nevertheless reaches 18.6 cm under 120 s occlusion, exceeding the 10 cm engineering threshold. This residual error is mainly caused by geometric ambiguity in repetitive bridge components.

Without semantic filtering, incorrect BIM feature associations can enter the optimization and gradually degrade the state estimate.

BCEF-Nav behaves differently. Its ATE RMSE increases only slightly as the occlusion duration grows, reaching 5.7 cm after 120 s of complete GNSS loss. Compared with LIO-BIM, the best baseline, this corresponds to a 69.35% error reduction. The improvement comes from three parts of the design. The tightly coupled semantic–geometric BIM constraint supplies a global reference when GNSS is absent and prevents drift from growing unchecked. Semantic-guided matching removes many false associations caused by repeated bridge components, keeping the BIM constraints reliable in standardized structural regions. The adaptive elastic weighting further allows GNSS and BIM constraints to exchange dominance without hard switching. As a result, the estimator avoids pose discontinuities and remains within centimeter-level global accuracy even under long full-occlusion intervals.

4.2.2. Robustness Test

To test robustness under difficult visual conditions, we evaluated the methods in three scenarios common in bridge inspection: low-texture concrete surfaces, strong backlight, and night-time low illumination. In each scenario, GNSS was fully occluded for 60 s. We measured the ATE RMSE of each method and recorded the average feature-matching inlier rate to assess the stability of the visual front end.

The results show that BCEF-Nav remains stable across the three adverse visual scenarios. Its ATE RMSE differs from the clear-day baseline by less than 2 cm. The baselines degrade much more strongly, with an average 3.5-fold increase in ATE RMSE. Under the most difficult low-light setting, some methods show more than one order of magnitude error growth. This difference mainly comes from two design choices. BCEF-Nav uses line and plane features as the main BIM constraints, rather than depending only on point features. Lines and planes are less sensitive to weak texture and illumination changes, and can still be extracted from concrete surfaces or low-light images where point features become sparse or unstable. The semantic-guided matching strategy further restricts associations to features from the same bridge component. This reduces false matches and improves front-end stability. Across the adverse scenarios, BCEF-Nav keeps an average matching inlier rate of 87%, whereas the baselines drop to 28% under the most challenging conditions. The higher inlier rate keeps the BIM constraints reliable even when the visual front end is degraded.

4.3. Field Experiment

The field experiments were conducted on a large bridge of the Yangqu section of the Erguang Expressway (G55), located in Dayu Town, Yangqu County, Taiyuan, Shanxi Province, China. The bridge is a precast box-girder highway bridge with a total length of about 1200 m, comprising 32 spans with a combined span arrangement of 35 m + 40 m supported by 33 piers, carrying a four-lane dual-carriageway expressway. It was opened to traffic in 2003 and had been in service for about 23 years at the time of the experiments. A high-precision as-built BIM model of this bridge was used as prior data. Four representative inspection scenarios were defined to cover different levels of GNSS occlusion and visual complexity. For the open-sky segments, centimeter-level positions from the Ublox F9P RTK-GNSS (fixed solution) were used as the primary ground truth. In the GNSS-occluded scenarios, where RTK is unavailable, a Leica TS60 total station (Leica Geosystems AG, Heerbrugg, Switzerland) was used to measure the fixed 3D coordinates of key structural points on the bridge, which serve as ground-truth control points. Five control points were measured in each scenario: in the beam-bottom scenario they were spaced about 18 m apart along the bridge (about 51 s apart in time), and in the box-girder-interior scenario about

9 m apart (about 26 s apart). Because a total station yields discrete points rather than a continuous reference, the occluded segments were evaluated by a post-processing alignment based on spatial nearest points, not by interpolation. The locally consistent VIO trajectory of each occluded segment was first registered to the bridge engineering coordinate system using these control points, which removes the global offset inherent to pure visual–inertial odometry. After this alignment, the ATE was computed for every frame using the spatially nearest control point as the global reference, so that the error of the entire flight path—not only that at the measured points—was assessed. Without this total-station calibration, the occluded-segment trajectory would remain detached from the BIM coordinate frame and the resulting per-frame errors would carry no engineering significance.

4.3.1. Accuracy Comparison in Different Scenarios

The field experiments covered four typical inspection scenarios: (1) **Bridge Deck**, with unobstructed GNSS reception; (2) **Pier Surrounding**, with 60 s intermittent GNSS occlusion caused by dense concrete piers; (3) **Beam Bottom**, with 90 s continuous GNSS occlusion beneath the main beams; and (4) **Box Girder Interior**, with 120 s continuous GNSS occlusion inside the closed girder. The box-girder case represents the most demanding navigation condition. Table 3 reports the ATE RMSE of each method in these scenarios.

Table 3. ATE RMSE comparison of each method in typical inspection scenarios (unit: cm).

Inspection Scenario	VINS-Mono	ORB-SLAM3	VINS-Fusion	BIM-SLAM	StructVIO	LIO-BIM	LIC-Fusion	BCEF-Nav (Ours)
Bridge Deck (GNSS fully available)	3.8	3.5	2.2	2.5	2.3	2.1	2.0	1.8
Pier Surrounding (60 s intermittent occlusion)	22.4	18.7	15.2	9.4	8.6	6.5	11.3	2.7
Beam Bottom (90 s full occlusion)	35.6	29.8	26.3	12.5	15.7	8.9	19.6	3.2
Box Girder Interior (120 s full occlusion)	89.2	76.5	71.4	28.7	41.3	16.2	38.5	5.9

The field results are consistent with the simulation results and show how each method behaves under practical inspection conditions. In the Bridge Deck scenario, GNSS reception is unobstructed. All GNSS-aided methods reach centimeter-level accuracy, and their performance differences are small because continuous high-quality GNSS provides a stable global reference. BCEF-Nav obtains the lowest ATE RMSE, 1.8 cm, as the BIM constraints provide additional pose refinement on top of the GNSS measurements.

As GNSS occlusion becomes more severe, the gap between BCEF-Nav and the baselines increases. In the Pier Surrounding scenario, GNSS is intermittently blocked by dense concrete piers. VINS-Fusion suffers from frequent pose discontinuities because its hard-switching strategy repeatedly removes and restores GNSS constraints. Elastic fusion methods avoid this discontinuity, but LIC-Fusion still has no replacement global reference during the short outages, so small drift errors accumulate. BCEF-Nav uses BIM constraints to maintain global correction throughout these intermittent occlusions. It achieves an ATE RMSE of 2.7 cm, reducing the error by 58.5% compared with LIO-BIM, the best-performing baseline.

In the Beam Bottom and Box Girder Interior scenarios, GNSS is fully unavailable for a continuous period. Under this condition, all baselines degrade markedly, whereas BCEF-Nav keeps centimeter-level accuracy. Pure VIO and GNSS–VIO fusion methods lose their global reference once GNSS disappears. Their odometry drift grows throughout the occluded segment, and the ATE RMSE exceeds 70 cm in the Box Girder Interior scenario. Such errors are not acceptable for inspection use.

BIM-SLAM reduces drift to some extent, but its loose-coupling design provides no real-time correction during the occlusion. Its ATE RMSE reaches 28.7 cm after 120 s. LIO-BIM performs best among the baselines, yet its purely geometric matching is vulnerable

to repeated piers and box-girder segments. These false associations gradually corrupt the estimate, giving an ATE RMSE of 16.2 cm in the Box Girder Interior scenario. BCEF-Nav maintains an ATE RMSE of 5.9 cm in the same setting. The semantic–geometric BIM constraint supplies a persistent global reference, while semantic filtering suppresses mismatches among repetitive structural elements.

4.3.2. Positioning Success Rate Comparison

To assess engineering usability, we computed the positioning success rate for the two full-GNSS-occlusion scenarios: Beam Bottom with 90 s occlusion and Box Girder Interior with 120 s occlusion. A pose estimate is counted as successful when its positioning error is below 10 cm. Table 4 reports the results for all methods.

Table 4. Positioning success rate comparison of each method in GNSS-full-occlusion scenarios.

Inspection Scenario	VINS-Mono	ORB-SLAM3	VINS-Fusion	BIM-SLAM	StructVIO	LIO-BIM	LIC-Fusion	BCEF-Nav (Ours)
Beam Bottom (90 s full occlusion)	8.2%	12.5%	17.3%	68.4%	72.1%	85.3%	78.6%	100%
Box Girder Interior (120 s full occlusion)	0%	0%	0%	62.1%	35.7%	79.2%	28.4%	100%

BCEF-Nav is the only method that reaches a 100% positioning success rate in both full-GNSS-occlusion scenarios. In other words, every frame remains within the 10 cm error threshold required for valid defect localization. Pure VIO and GNSS–VIO fusion methods drop to 0% success in the 120 s Box Girder Interior scenario, because their accumulated drift exceeds the threshold early in the sequence and cannot recover without a global reference. LIO-BIM performs best among the baselines, but its success rate in this scenario is 79.2%. More than one fifth of its frames therefore fail to meet the engineering accuracy requirement, which leaves the resulting inspection records incomplete for component-level defect registration.

The 100% success rate of the proposed framework is achieved through the synergistic action of its core innovations. The BIM semantic–geometric tight coupling constraint provides a persistent global reference tied to the bridge’s engineering coordinate system, ensuring that the absolute positioning error never grows unbounded. The semantic-guided feature matching eliminates mismatches from repetitive structural elements, ensuring that the BIM constraints remain reliable throughout the entire occlusion period. The adaptive elastic weight mechanism ensures that the BIM constraint takes over seamlessly as the primary global constraint when GNSS signals are lost, with no drop in positioning accuracy during the state transition. These design elements collectively ensure that the system maintains centimeter-level accuracy for 100% of the inspection mission, even in the most challenging full GNSS occlusion scenarios.

To substantiate the 100% success rate, Figure 5 reports the per-frame absolute positioning error over the 120 s box-girder full-occlusion traversal. The time-series traces (Figure 5a) show that the BCEF-Nav error remains below the 10 cm engineering threshold for the entire sequence, whereas all baselines cross the threshold at various points during the occlusion. The corresponding error CDF (Figure 5b) confirms that 100% of BCEF-Nav frames fall below 10 cm, while 79.2% of LIO-BIM frames remain below 10 cm and 20.8% exceed the threshold. This per-frame evidence supports the aggregate success rate reported in Table 4.

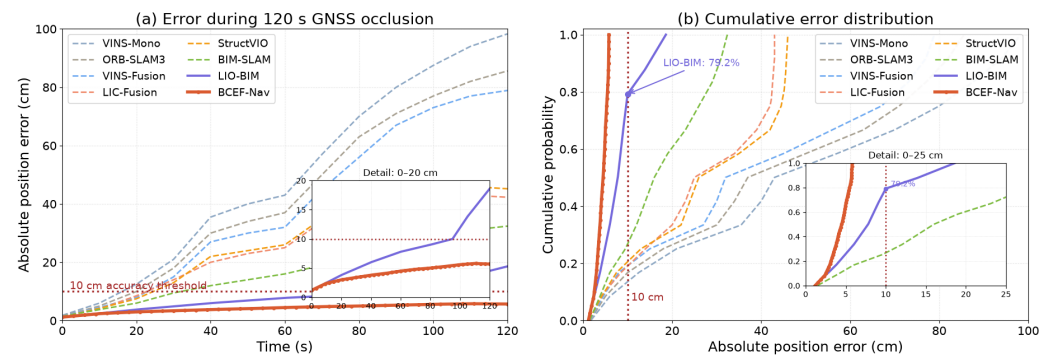


Figure 5. Per-frame absolute positioning error during the 120 s box-girder full GNSS occlusion. (a) Error time series of all methods, with the 10 cm engineering threshold marked; the BCEF-Nav error stays below the threshold for the entire traversal, while all baselines exceed it. (b) Error CDF over the occluded segment; 100% of BCEF-Nav frames are below 10 cm, while 79.2% of LIO-BIM frames are below 10 cm and 20.8% exceed the threshold.

4.3.3. Qualitative Results

Figure 6 shows the estimated inspection trajectories of representative methods in the 120 s box-girder scenario, together with the ground-truth trajectory. The visual comparison agrees with the quantitative results. The BCEF-Nav trajectory remains almost coincident with the ground-truth reference over the full path inside the box girder and stays aligned with the BIM coordinate system despite complete GNSS loss. All shown baselines exhibit visible cumulative drift as the UAV progresses through the occluded segment. VINS-Mono departs strongly from the true trajectory, which would compromise both navigation safety and inspection accuracy. The BIM-based and LiDAR–inertial–camera baselines perform better, but still show clear drift by the end of the 120 s occlusion.

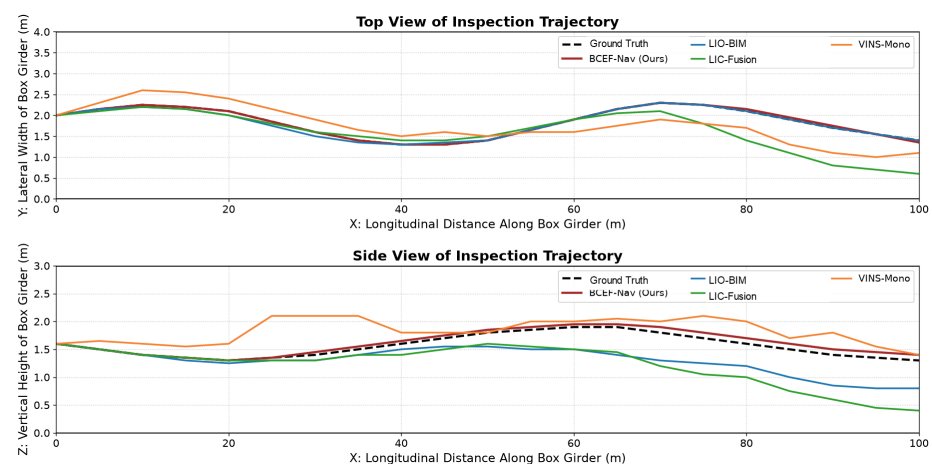


Figure 6. Qualitative comparison of inspection trajectories and ground truth in the 120 s full GNSS occlusion scenario inside the box girder. The proposed BCEF-Nav trajectory is almost coincident with the ground truth, while the representative control methods VINS-Mono [6], LIO-BIM [21], and LIC-Fusion [28] show severe drift.

4.3.4. Real-Time Performance Test

To verify onboard real-time performance, we measured the average frame rate of the full BCEF-Nav pipeline on the Jetson Xavier NX during the complete bridge inspection mission. The system runs at 28 fps, close to the 30 fps camera capture rate and comparable to the lightweight VINS-Mono baseline. It should be noted that this figure is a mission-averaged end-to-end frame rate; per-segment and per-module timing breakdowns were not instrumented in this study, and frame rates during computationally intensive phases—such as

BIM loop-closure triggering and dense feature matching inside the box girder—may differ from this average. This efficiency comes mainly from two implementation choices: offline BIM preprocessing with KD-tree indexing for fast feature association, and a lightweight semantic segmentation network that runs on the embedded GPU. At 28 fps, BCEF-Nav can provide pose estimates in real time for flight control and defect localization. The result indicates that the framework can operate on the computing hardware typically carried by small inspection UAVs, without requiring a heavier or more expensive onboard computer. A finer-grained runtime profile—the per-module timing of the segmentation front end, BIM feature association, and sliding-window optimization, together with the minimum frame rate and the instantaneous rate during compute-intensive phases—is left as future work to further support reproducible real-time evaluation.

Beyond the average frame rate, deploying BCEF-Nav on the edge computing hardware of small inspection UAVs poses several practical challenges. First, the onboard platform (Jetson Xavier NX, 8 GB) must run the YOLOv8-Seg semantic segmentation front end and the sliding-window optimization back end concurrently under tight compute and power budgets; we address this by executing a lightweight segmentation network on the embedded GPU and by moving BIM preprocessing and KD-tree indexing of the global feature library entirely offline, so that online computation is limited to nearest-neighbor queries and residual construction. Second, the BIM global feature library must remain resident in memory during the mission; the lightweight triangular mesh and component-wise partitioned loading keep the memory footprint within the platform’s capacity. Third, synchronized multi-sensor acquisition and parallel scheduling of the visual, inertial, GNSS, and BIM-association threads must be maintained on limited hardware to avoid latency spikes. In terms of minimum requirements, the framework requires an embedded platform with a CUDA-capable GPU and at least 8 GB of memory to sustain real-time operation; lower-end hardware would require further model compression. These considerations indicate that, although BCEF-Nav runs in real time on a mainstream onboard computer, its current compute headroom is limited.

4.4. Ablation Experiment

To assess the contribution of each core module in BCEF-Nav, we performed an ablation study in the most difficult setting: 120 s full GNSS occlusion inside the box girder. Starting from the baseline VIO system, we added the proposed modules one by one and measured the ATE RMSE of each variant. This design isolates the accuracy gain provided by each component. Table 5 reports the ablation results, and Figure 7 shows the corresponding percentage improvement in positioning accuracy.

The ablation results clarify the contribution of each module and support the effectiveness of the proposed design. Variant 1 uses only the baseline VIO module. In the 120 s full-GNSS-occlusion box-girder scenario, its ATE RMSE reaches 89.2 cm. The trajectory shows unconstrained drift and is not usable for bridge inspection. This result confirms that pure VIO, without any global reference, cannot maintain the positioning accuracy required during long GNSS outages.

Table 5. Ablation experiment results (ATE RMSE, unit: cm). Values are reported as mean $\pm$ standard deviation over 10 repeated trials.

No.	VIO Base	Elastic GNSS	BIM Point	BIM Line + Plane	Global Opt.	ATE RMSE
1	✓	×	×	×	×	89.2 $\pm$ 10.3
2	✓	✓	×	×	×	71.4 $\pm$ 8.7
3	✓	×	✓	×	×	22.6 $\pm$ 2.4
4	✓	×	✓	✓	×	12.3 $\pm$ 1.2
5	✓	✓	✓	✓	×	8.7 $\pm$ 0.8
6	✓	✓	✓	✓	✓	5.9 $\pm$ 0.5

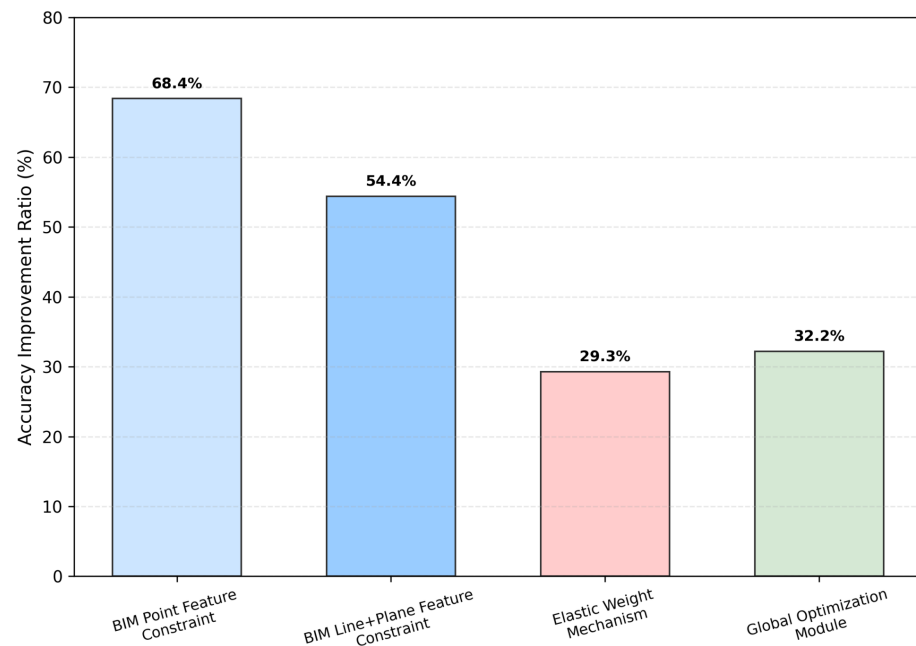


Figure 7. Accuracy improvement ratio of each innovative module in the ablation experiment. The BIM line + plane feature constraint, elastic weight mechanism and global optimization module bring 54.4%, 29.3% and 32.2% accuracy improvement respectively, verifying the effectiveness of the core innovations.

Variant 2 adds the elastic GNSS constraint to the baseline VIO system and reduces the ATE RMSE to 71.4 cm. The improvement is limited, and the error remains well above the 10 cm engineering threshold. The reason is straightforward: elastic GNSS weighting can correct the pose only when GNSS observations are available. During complete occlusion, it provides no global reference and cannot prevent drift accumulation. This result shows that elastic GNSS–VIO fusion alone is insufficient for extended GNSS outages unless another global constraint is introduced.

Variant 3 adds BIM point-feature constraints to the baseline VIO system. The ATE RMSE decreases sharply to 22.6 cm, showing the value of the BIM model as a global reference. BIM point features anchor the estimate to the bridge engineering coordinate system and suppress the unbounded drift observed in pure VIO. The error, however, still remains above the 10 cm threshold. Inside the box girder, salient point features are sparse because of weak texture and repetitive geometry, so point constraints alone are not dense enough to maintain centimeter-level accuracy over the full occlusion period.

Variant 4 adds BIM line and plane constraints on top of Variant 3. The ATE RMSE decreases to 12.3 cm, corresponding to a 54.4% accuracy improvement over point-only BIM constraints. This gain comes from the feature distribution inside the box girder. Lines and planes are far more abundant than salient points and provide denser global constraints over the occluded segment. The result supports the use of point, line, and plane features in combination. In low-texture and highly structured bridge scenes, line and plane constraints are especially useful for maintaining pose accuracy when GNSS is unavailable.

Variant 5 adds the elastic GNSS constraint to Variant 4. The ATE RMSE decreases to 8.7 cm, giving a 29.3% improvement over Variant 4. This gain comes from the adaptive weighting between GNSS and BIM constraints. Before and after the box-girder occlusion, GNSS observations are still available. The elastic fusion mechanism uses these measurements together with BIM constraints to refine the initial and recovered states, reducing the

error carried into the occluded segment. This result confirms that GNSS and BIM constraints are most effective when their weights are adjusted jointly over the full inspection trajectory.

Variant 6 includes all BCEF-Nav modules and adds global pose consistency optimization. The ATE RMSE is further reduced to 5.9 cm, a 32.2% improvement over Variant 5. The gain comes from pose graph optimization, which removes the small residual drift accumulated over the full trajectory. After this step, the complete inspection path remains consistent with the bridge engineering coordinate system.

The ablation study also shows that the modules are not merely additive. The BIM semantic–geometric constraint supplies the global reference needed during GNSS outages. The elastic weighting mechanism regulates the interaction between GNSS and BIM constraints as signal conditions change. The global pose consistency module then removes residual drift over the full trajectory. These components work as a single estimation pipeline, allowing BCEF-Nav to maintain centimeter-level global positioning even under severe GNSS occlusion.

To isolate the contribution of semantic-guided matching, we compared the full BCEF-Nav with a variant in which semantic filtering is disabled (geometry-only BIM matching), holding all other components constant. As shown in Table 6, removing semantic filtering causes the repetitive structural elements of the bridge—such as equally spaced diaphragms and standardized box-girder segments—to trigger numerous cross-component false matches, sharply reducing the reliability of the BIM global constraint: the matching inlier rate drops by about 43%, the ATE RMSE increases by about 132% (from 5.9 cm to 13.7 cm), and the success rate falls below 80%, failing to meet the centimeter-level requirement of bridge inspection. This result verifies that semantic-guided feature matching is a core module for ensuring BIM-constraint validity in repetitive bridge scenes.

Table 6. Ablation of the semantic filtering module under 120 s box-girder full GNSS occlusion.

Configuration	ATE RMSE (cm)	Success Rate (<10 cm)	Avg. Inlier Rate
BCEF-Nav (full, with semantic filtering)	5.9 ± 0.5	100%	87.2% ± 2.1%
Without semantic filtering (geometry-only BIM matching)	13.7 ± 1.4	76.4%	49.6% ± 4.8%

5. Discussion

The experiments show that BCEF-Nav addresses the main navigation challenge in UAV bridge inspection: maintaining centimeter-level global positioning in the bridge engineering coordinate system under intermittent GNSS outages. Compared with existing navigation methods, the framework provides three key improvements.

The first is elastic fusion under intermittent GNSS occlusion. The adaptive weighting mechanism enables smooth transitions between GNSS-available and GNSS-denied conditions without the pose discontinuities caused by hard switching. Even during complete GNSS outages, the as-built BIM model provides a persistent global constraint, keeping odometry drift bounded over extended periods.

The second improvement is the tight coupling of BIM semantic and geometric information. Global point, line, and plane features from the BIM model are incorporated into the sliding-window optimization together with component semantics. Unlike existing BIM-aided methods that primarily use the model for loop closure or geometric association, the proposed framework maintains BIM constraints throughout the optimization process. Semantic-guided feature matching further reduces false correspondences among repetitive bridge components, improving robustness in box girders, beam soffits, and pier-dense regions with weak texture and highly repetitive structures.

The third improvement is engineering deployability. BCEF-Nav uses only a monocular global-shutter camera, a low-cost MEMS IMU, and a civilian GNSS receiver. It does not require LiDAR, fiber-optic IMUs, or high-end onboard computing. The framework therefore provides a practical solution for centimeter-level UAV bridge inspection on mainstream small platforms, reducing the cost barrier to large-scale deployment in GNSS-challenged bridge environments.

Although this work focuses on bridge inspection, the proposed framework can also be extended to other structured infrastructure environments. Tunnels, dams, industrial plants, and port facilities often experience intermittent GNSS availability while providing BIM or CAD priors. In these scenarios, semantic-geometric model constraints can provide the global reference that GNSS alone cannot maintain.

Beyond navigation performance, the proposed framework also supports the broader bridge inspection workflow. Modern bridge inspection increasingly combines multiple sensing modalities, including RGB imagery for surface cracks, thermal imaging for moisture and delamination, LiDAR for geometric deformation, and sensors integrated with structural health monitoring systems. All of these sensing modalities rely on accurate localization so that each observation can be placed in the bridge engineering coordinate system and associated with the correct bridge component. BCEF-Nav provides this capability. Acting as a sensor-agnostic positioning layer, it supports different sensing payloads while registering their measurements to a common global frame, allowing observations from different sensors to be consistently linked to the corresponding component IDs and mileage stakes. Integrating such multi-sensor inspection payloads is beyond the scope of this navigation-focused study and is left for future work. Nevertheless, the centimeter-level localization achieved here provides the foundation for these broader inspection workflows.

BCEF-Nav also has several limitations. It relies on a high-precision as-built BIM model as the global reference for pose correction. Such a model is typically obtained in two ways: for newly constructed bridges, the design-stage BIM is refined using as-built survey data, as in the model adopted in this study, which was provided by the design institute and validated on site; for existing bridges, an as-built model can be reconstructed through laser scanning or photogrammetry. Bridges already incorporated into digital asset management usually satisfy this requirement, allowing the framework to be applied directly. For aging infrastructure without an accurate as-built BIM, however, an initial reverse-modeling process is required before deployment. Model currency is another concern. Settlement, deformation, or structural damage may gradually reduce consistency between the bridge and its BIM model, weakening feature association and degrading positioning accuracy. This issue is more likely for aging bridges whose BIM has not been updated during service, motivating the online structural deformation detection discussed below to maintain a consistent prior throughout the bridge lifecycle.

The monocular visual front end also has inherent limitations. During near-pure rotational motion, image parallax becomes weak, making scale estimation less stable and reducing robustness when the UAV hovers and rotates to inspect local structural details. Reliable BIM geometric constraints also require observable point, line, or plane features. On completely smooth and textureless surfaces, establishing stable feature associations becomes difficult. Although such situations are uncommon in routine bridge inspection, they may occur on newly cast concrete surfaces with few visible edges or surface patterns. From an implementation perspective, the current system also leaves limited computational headroom on the edge platform. Future work will investigate model quantization and operator-level optimization to further reduce hardware requirements and improve deployment on lower-cost onboard computers.

Another limitation concerns generalization. The current validation is based on a single bridge and one high-quality as-built BIM model, so the applicability of the framework to different bridge types and BIM models with varying accuracy has not yet been verified experimentally. Nevertheless, the proposed framework is not specific to the box-girder bridge considered in this study. It relies on point, line, and plane features together with bridge component semantics, which are common to girder, arch, cable-stayed, and other bridge types. In addition, the matching-confidence weight s_{match} in Equation (11) automatically reduces the influence of BIM constraints when match quality decreases, providing a certain degree of tolerance to less accurate BIM models. These observations suggest good potential for generalization, but comprehensive validation across different bridge types and BIM accuracy levels remains future work.

These limitations also define several directions for future research. Online structural deformation detection will be incorporated to update the BIM prior from visual observations, improving positioning consistency during long-term bridge inspections. To reduce the scale uncertainty of the monocular front end, a lightweight binocular depth camera will be introduced to provide direct depth information and improve robustness under weak texture and near-pure rotational motion. As discussed in the Introduction, automated defect detection and defect-to-component registration remain beyond the scope of this navigation-focused study. Future work will integrate the proposed navigation framework with automated defect detection so that identified defects can be directly associated with BIM component IDs and mileage stakes. This integrated workflow will also enable evaluation using detection-oriented metrics, including component localization accuracy, defect registration accuracy, and BIM-element association accuracy, together with comparisons of defect detection performance with and without the proposed BIM-constrained localization. These evaluations are intentionally left for future work because they require a complete defect detection pipeline and real defect datasets, which are outside the scope of the present study. Finally, the upper bound of the BIM constraint weight, $\lambda_{BIM,max} = 6.0$, was selected empirically through repeated field trials conducted with the infrastructure maintenance team of Shanxi Transportation Holding Group. A systematic sensitivity analysis of this parameter under different GNSS occlusion durations and bridge inspection scenarios will be conducted in future work to provide more rigorous guidance for practical parameter selection.

6. Conclusions

This paper presents BCEF-Nav, a BIM-constrained elastic fusion navigation framework for UAV-based bridge inspection. It addresses intermittent GNSS occlusion beneath beams, inside box girders, and around dense piers, where odometry drift can break the centimeter-level alignment required in the bridge engineering coordinate system. Built on a tightly coupled sliding-window optimization back end, the framework introduces semantic-geometric features from a high-precision as-built BIM model into visual-inertial state estimation, allowing the BIM model to serve as a persistent global reference when GNSS is unavailable. An adaptive weighting mechanism regulates the GNSS and BIM constraints: as GNSS degrades, the weight shifts toward the BIM constraint so that global correction continues throughout the occluded interval, while semantic-guided feature matching further suppresses false associations among repetitive bridge components.

Validation in digital-twin simulations and field experiments on an in-service expressway bridge showed that BCEF-Nav maintained an ATE RMSE of 5.7 cm after 120 s of complete GNSS occlusion and achieved a 100% positioning success rate under the 10 cm engineering threshold, while running at 28 fps on the embedded platform of a mainstream small inspection UAV using only low-cost commercial sensors.

The work makes two main contributions. Methodologically, it formulates an elastic multi-constraint fusion scheme for intermittent GNSS occlusion and a BIM semantic–geometric tight-coupling formulation within sliding-window optimization, extending prior model-constrained navigation from loose association or purely geometric constraints to a continuously active global constraint tied to the engineering coordinate system. From an engineering perspective, it provides a low-cost and deployable navigation solution that, given a high-precision as-built BIM model, can also be applied to other structured infrastructure environments such as tunnels, dams, industrial plants, and port facilities.

The proposed framework nevertheless has several limitations. Its performance depends on the accuracy of semantic segmentation and the availability of high-quality BIM models. Positioning accuracy may degrade when semantic predictions are unreliable or when the BIM model differs significantly from the real environment. Future work will investigate more robust semantic perception and dynamic BIM updating strategies.

Author Contributions: Conceptualization, Z.L., Y.Y. and F.L.; methodology, Z.L., H.L. and C.X.; software, C.X. and Y.S.; validation, H.L., Y.S. and Y.Y.; writing—original draft preparation, Z.L., H.L. and C.X.; writing—review and editing, Y.S. and F.L.; supervision, Y.Y. and F.L.; funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Fundamental Research Program of Shanxi Province: 202403021212165, in part by Major Science and Technology Projects of Shanxi Province: 202101010101018.

Data Availability Statement: The data that support the findings of this study are not publicly available due to commercial confidentiality and proprietary restrictions. Reasonable requests for data access will be considered by the corresponding author subject to approval from the relevant institutional and commercial parties.

Acknowledgments: Declaration of AI-Assisted Technologies: During the preparation of this work, the authors used Doubao (May 2026 version) and DeepL (May 2026 version) AI and DeepL solely for language refinement, translation, and grammar checking to improve the readability of the manuscript. The core tasks of research conceptualization, theoretical modeling, algorithm design, experimental implementation, data collection and analysis, result interpretation, and manuscript revision were performed entirely by the authors. The authors have reviewed and edited all AI-generated suggestions and translations, and take full responsibility for the final content of this publication. No AI tools were used to generate research data, results, or conclusions. This use of AI tools is compliant with MDPI's policies on the use of AI in scientific writing.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ATE	Absolute Trajectory Error
BIM	Building Information Modeling
CDF	Cumulative Distribution Function
CGCS2000	China Geodetic Coordinate System 2000
DOP	Dilution of Precision
FOV	Field of View
fps	Frames Per Second
GNSS	Global Navigation Satellite System
IFC	Industry Foundation Classes
IMU	Inertial Measurement Unit
KD-tree	K-Dimensional Tree
LiDAR	Light Detection and Ranging

LSD	Line Segment Detector
MEMS	Micro-Electro-Mechanical System
mIoU	mean Intersection over Union
ORB	Oriented FAST and Rotated BRIEF
PDOP	Position Dilution of Precision
RMSE	Root Mean Square Error
ROS	Robot Operating System
RPE	Relative Pose Error
RTK	Real-Time Kinematic
SE(3)	Special Euclidean Group in 3D
SHM	Structural Health Monitoring
SLAM	Simultaneous Localization and Mapping
UAV	Unmanned Aerial Vehicle
VIO	Visual-Inertial Odometry

References

1. Zhang, C.; Zou, Y.; Wang, F.; del Rey Castillo, E.; Dimyadi, J.; Chen, L. Towards fully automated unmanned aerial vehicle-enabled bridge inspection: Where are we at? *Constr. Build. Mater.* **2022**, *347*, 128543. [\[CrossRef\]](#)
2. Panigati, T.; Zini, M.; Striccoli, D.; Giordano, P.F.; Tonelli, D.; Limongelli, M.P.; Zonta, D. Drone-based bridge inspections: Current practices and future directions. *Autom. Constr.* **2025**, *173*, 106101. [\[CrossRef\]](#)
3. Shu, J.; Xia, Z.; Gao, Y. BIM-Based Trajectory Planning for Unmanned Aerial Vehicle-Enabled Box Girder Bridge Inspection. *Remote Sens.* **2025**, *17*, 682. [\[CrossRef\]](#)
4. Pang, C.; Luo, X.; Wang, J.; Fang, Z. Structural Lines Aided Monocular Visual-Inertial-Wheel Odometry with Online IMU-Wheel Extrinsic Optimization on S^2 Manifold. *IEEE Trans. Intell. Veh.* **2023**, *9*, 4100–4114.
5. Ilter, O.; Armeni, I.; Pollefeys, M.; Barath, D. Semantically guided feature matching for visual slam. In *Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: New York, NY, USA, 2024; pp. 12013–12019.
6. Qin, T.; Li, P.; Shen, S. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020. [\[CrossRef\]](#)
7. Campos, C.; Elvira, R.; Rodríguez, J.J.G.; Montiel, J.M.; Tardós, J.D. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE Trans. Robot.* **2021**, *37*, 1874–1890. [\[CrossRef\]](#)
8. Mur-Artal, R.; Tardós, J.D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [\[CrossRef\]](#)
9. He, Y.; Zhao, J.; Guo, Y.; He, W.; Yuan, K. PL-VIO: Tightly-coupled monocular visual-inertial odometry using point and line features. *Sensors* **2018**, *18*, 1159. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Zou, D.; Wu, Y.; Pei, L.; Ling, H.; Yu, W. StructVIO: Visual-Inertial Odometry With Structural Regularity of Man-Made Environments. *IEEE Trans. Robot.* **2019**, *35*, 999–1013. [\[CrossRef\]](#)
11. Zheng, F.; Lin, W.; Sun, L. RL-VIO: A Robust and Lightweight Visual-inertial Odometry in Dynamic Environments. In *Proceedings of the 2023 42nd Chinese Control Conference (CCC)*, Tianjin, China, 24–26 July 2023; pp. 4243–4248. [\[CrossRef\]](#)
12. Xiaochen, Q.; Zhang, H.; Wenxing, F. Lightweight hybrid visual-inertial odometry with closed-form zero velocity update. *Chin. J. Aeronaut.* **2020**, *33*, 3344–3359. [\[CrossRef\]](#)
13. Ma, Z.; Xiao, J.; Wang, F.; Jiang, P. Fast Visual-Inertial Odometry with Adaptive Feature Coupling. In *Proceedings of the International Conference on Intelligent Robotics and Applications*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 171–186.
14. Chen, Z.; Zhao, W.; Niu, Y.; Deng, T.; Wang, J. OTPL-VIO: Robust Visual-Inertial Odometry with Optimal Transport Line Association and Adaptive Uncertainty. *arXiv* **2026**, arXiv:2603.09653.
15. Yue, K. Multi-sensor data fusion for autonomous flight of unmanned aerial vehicles in complex flight environments. *Drone Syst. Appl.* **2024**, *12*, 1–12. [\[CrossRef\]](#)
16. Tang, H.; Zhang, T.; Niu, X.; Fan, J.; Liu, J. IC-GVINS: A robust, real-time, INS-centric GNSS-visual-inertial navigation system for wheeled robot. *arXiv* **2022**, arXiv:2204.04962.
17. Forster, C.; Carlone, L.; Dellaert, F.; Scaramuzza, D. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Trans. Robot.* **2016**, *33*, 1–21.
18. Vega Torres, M.A.; Braun, A.; Borrmann, A. BIM-SLAM: Integrating BIM Models in Multi-session SLAM for Lifelong Mapping using 3D LiDAR. In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*; International Association for Automation and Robotics in Construction (IAARC): Oulu, Finland, 2023. [\[CrossRef\]](#) [\[PubMed\]](#)

19. Bikandi-Noya, A.; Fernandez-Cortizas, M.; Shaheer, M.; Tourani, A.; Voos, H.; Sanchez-Lopez, J.L. BIM Informed Visual SLAM for Construction Monitoring. *arXiv* **2025**, arXiv:2509.13972.
20. Vipond, N.; Kumar, A.; James, J.; Paige, F.; Sarlo, R.; Xie, Z. Real-time processing and visualization for smart infrastructure data. *Autom. Constr.* **2023**, *154*, 104998. [[CrossRef](#)]
21. Stührenberg, J.; Smarsly, K. LIO-BIM–Coupling lidar inertial odometry with building information modeling for robot localization and mapping. *Adv. Eng. Inform.* **2025**, *66*, 103477. [[CrossRef](#)]
22. Qin, T.; Cao, S.; Pan, J.; Shen, S. A general optimization-based framework for global pose estimation with multiple sensors. *arXiv* **2019**, arXiv:1901.03642.
23. Elamin, A.; Abdelaziz, N.; El-Rabbany, A. A GNSS/INS/LiDAR integration scheme for UAV-based navigation in GNSS-challenging environments. *Sensors* **2022**, *22*, 9908. [[CrossRef](#)] [[PubMed](#)]
24. Hu, J.; Lang, X.; Zhang, F.; Mao, Y.; Huang, G. Square-Root Inverse Filter-based GNSS-Visual-Inertial Navigation. In *Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: New York, NY, USA, 2024; pp. 1646–1652.
25. Negru, S.A.; Geragersian, P.; Petrunin, I.; Guo, W. Resilient multi-sensor uav navigation with a hybrid federated fusion architecture. *Sensors* **2024**, *24*, 981. [[CrossRef](#)] [[PubMed](#)]
26. Zheng, C.; Xu, W.; Zou, Z.; Hua, T.; Yuan, C.; He, D.; Zhou, B.; Liu, Z.; Lin, J.; Zhu, F.; et al. Fast-livo2: Fast, direct lidar–inertial–visual odometry. *IEEE Trans. Robot.* **2024**, *41*, 326–346.
27. Yin, J.; Jiang, H.; Wang, J.; Yan, D.; Yin, H. A robust and efficient ekf-based gnss-visual-inertial odometry. In *Proceedings of the 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO)*; IEEE: New York, NY, USA, 2023; pp. 1–5.
28. Zuo, X.; Geneva, P.; Lee, W.; Liu, Y.; Huang, G. Lic-fusion: Lidar-inertial-camera odometry. In *Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2019; pp. 5848–5854.
29. Shan, T.; Englot, B.; Meyers, D.; Wang, W.; Ratti, C.; Rus, D. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2020; pp. 5135–5142.
30. Sapkota, R.; Karkee, M. Ultralytics YOLO evolution: An overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 object detectors for computer vision and pattern recognition. *arXiv* **2025**, arXiv:2510.09653.
31. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the 2011 International Conference on Computer Vision*; IEEE: New York, NY, USA, 2011; pp. 2564–2571.
32. Von Gioi, R.G.; Jakubowicz, J.; Morel, J.M.; Randall, G. LSD: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *32*, 722–732. [[CrossRef](#)]
33. Kaess, M.; Johannsson, H.; Roberts, R.; Ila, V.; Leonard, J.J.; Dellaert, F. iSAM2: Incremental smoothing and mapping using the Bayes tree. *Int. J. Robot. Res.* **2012**, *31*, 216–235. [[CrossRef](#)]
34. Kümmerle, R.; Grisetti, G.; Strasdat, H.; Konolige, K.; Burgard, W. g 2 o: A general framework for graph optimization. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*; IEEE: New York, NY, USA, 2011; pp. 3607–3613.
35. Maddern, W.; Pascoe, G.; Linegar, C.; Newman, P. 1 year, 1000 km: The oxford robotcar dataset. *Int. J. Robot. Res.* **2017**, *36*, 3–15.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.