GENG4412 Engineering Research Project Part 2
Final Report

Visual-Inertial SLAM for Autonomous Shuttle Bus

Anthony Gray
23334769
School of Engineering, University of Western Australia

Supervisor: Thomas Bräunl
School of Engineering, University of Western Australia

Word count: 6930

School of Engineering
University of Western Australia

Submitted: 20 May 2026

Project Summary

LiDAR-based localisation dominates production autonomous shuttle platforms but imposes sensor costs and power demands that constrain fleet-scale deployment. Monocular visual-inertial SLAM offers a significantly lower-cost alternative; the Velodyne VLP-16 used on the nUWAy4 shuttle costs approximately four times the equivalent fisheye camera. This project evaluates whether an ORB-SLAM3 monocular-inertial pipeline, deployed on an embedded NVIDIA Jetson AGX Orin, can meet a 1.0 m Absolute Trajectory Error (ATE) Root Mean Square Error criterion sufficient to substitute the vehicle's existing LiDAR Nav2 baseline.

The pipeline was integrated on the nUWAy4 EasyMile EZ10 shuttle at the University of Western Australia, using a front-facing GMSL2 fisheye camera (30 Hz) and an SBG Ellipse-D IMU (250 Hz). IMU noise parameters were derived from a 3-hour static Allan-variance characterisation; a ×100 white-noise inflation was required to sustain tracking under on-vehicle vibration, exceeding the ×3–10 range reported for standard outdoor platforms. A frame-convention error in the inherited integration pipeline was identified and corrected, resolving an IMU initialisation failure that was blocking deployment. Trajectory accuracy was evaluated against a concurrent EKF-fused GPS/IMU reference using Sim(3)-aligned ATE across twelve trials spanning multiple configurations, a campus loop, and a straight-drive segment. The ablation design separates configuration effects from segment-geometry effects using a segment-matched reference run.

The 1.0 m criterion was not met. Best extended-coverage ATE RMSE was 12.98 m (factor of 13 over criterion); no campus-loop trial achieved complete loop closure (the two longest runs, v3 and v6, each terminated at approximately 170 s of the 180 s route), with Sim(3) scale factors as low as 0.0030 indicating near-total loss of the inertial metric anchor. All five straight-drive trials terminated within 37 s ($\sigma = 2.1$ s), a near-deterministic failure attributed to insufficient feature parallax during forward translation. Reducing IMU noise inflation from ×100 to ×10 approximately doubled ATE on matched segments. Computationally, the Jetson AGX Orin averaged 30.1% CPU, 35.2% GPU, and ≈7 W total power, well within the 15 W TDP budget, confirming that the observed failures are geometric rather than computational.

Two independent root causes are identified: a calibration-deployment gap requiring anomalously high IMU inflation, and a structural degeneracy of monocular-inertial geometry under straight-line forward motion. Addressing either alone is insufficient to meet the criterion. A stereo-inertial configuration, which provides a fixed metric baseline independently of motion profile, is identified as the necessary next step, alongside native Kannala-Brandt fisheye projection and a deterministic rotational startup manoeuvre to initialise metric scale before straight-line service begins.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Thomas Bräunl, and to Zheng Li and Lee Le of the UWA REV team, for their guidance, mentorship, and constant support throughout this project. I would also like to thank my peer Harry Tolcher for his advice and camaraderie during the course of this project.

Nomenclature

Abbrev.	**Definition**	**Abbrev.**	**Definition**
AccRW	Acceleration Random Walk	LiDAR	Light Detection and Ranging
ARW	Angle Random Walk	MAP	Maximum A Posteriori
ATE	Absolute Trajectory Error	MEMS	Micro-Electro-Mechanical Systems (IMU class)
AV	Autonomous Vehicle	ORB	Oriented FAST and Rotated BRIEF (feature)
AVAR	Allan Variance	PCD	Point Cloud Data (file format)
BA	Bundle Adjustment	REP	ROS Enhancement Proposal (eg REP-103)
BI	Bias Instability	ROS 2	Robot Operating System 2
BRIEF	Binary Robust Indep. Elementary Features	RPE	Relative Pose Error
CPU/GPU	Central / Graphics Processing Unit	RRW	Rate Random Walk
DBoW	Distributed Bag-of-Words	RTK	Real-Time Kinematic (GNSS correction)
DDS	Data Distribution Service (ROS 2 mw.)	SE(3)	Special Euclidean Group (rigid transforms)
EKF	Extended Kalman Filter	SO(3)	Special Orthogonal Group (rotations)
FAST	Features from Accelerated Segment Test	SLAM	Simultaneous Localisation and Mapping
FLU	Forward-Left-Up (body frame)	TOF	Time of Flight (LiDAR/depth)
FM	Frequency Modulation (noise process class)	TUM	TUM trajectory format (`t x y z q*`)
FRD	Forward-Right-Down (body frame)	VI-SLAM	Visual-Inertial SLAM
GBA	Global Bundle Adjustment	VIBA	Visual-Inertial Bundle Adjustment
GMSL	Gigabit Multimedia Serial Link	VINS	Visual-Inertial Navigation System
GNSS	Global Navigation Satellite System	VIO	Visual-Inertial Odometry
GPS	Global Positioning System	VO	Visual Odometry
IMU	Inertial Measurement Unit	VRW	Velocity Random Walk
KF	Keyframe	vSLAM	Visual SLAM
LBA	Local Bundle Adjustment	YAML	YAML Ain’t Markup Language

Contents

1. Introduction

1.1 Problem Statement

Autonomous shuttle buses have transitioned from research demonstrations to limited public deployment. However, the high cost and complexity involved in integrating their localisation sensor stacks continue to restrict large-scale fleet adoption. The concept is particularly suitable for first- and last-mile mobility on closed routes such as university campuses, hospital precincts, and retirement villages. Yet, the sensors needed to localise the vehicle reliably enough for safe operation on public roads remain the primary capital and integration challenge for each platform.

Autonomous shuttle operation relies on two core functions: localisation (determining the vehicle's position and orientation) and perception (detecting obstacles, lanes, and other road users). This thesis focuses solely on localisation, since it provides the trajectory estimates that inform downstream control. Perception is not directly addressed in this work.

The University of Western Australia's Renewable Energy Vehicle (REV) programme operates the nUWAy autonomous shuttle bus fleet. All results presented in this report were collected using **nUWAy4**, a second-generation EasyMile EZ10 shuttle (Figure 1.1) fitted with a production-shuttle autonomy stack.

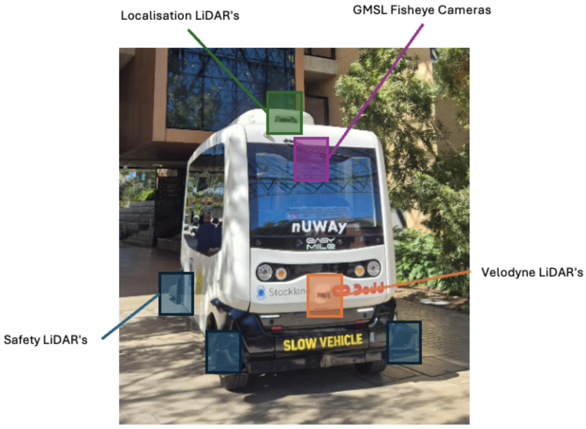


Figure 1.1: The nUWAy4 autonomous shuttle bus, the experimental platform for this study.

1.2 Motivation and Economic Rationale

Current autonomous shuttle systems rely predominantly on LiDAR-based localisation for its robustness and established deployment record, but LiDAR hardware imposes significant cost and power penalties at fleet scale. Visual-Inertial SLAM (VI-SLAM) fuses camera images with IMU measurements to estimate trajectory and build a map in real time, offering a lower-cost alternative as embedded computing has matured. As shown in Table 1.1, the Velodyne VLP-16 used on nUWAy4 costs approximately four times a GMSL fisheye camera; IMU is common to both stacks and excluded.

Table 1.1: Approximate unit cost of the LiDAR and camera sensors used in this study.

Component	Approximate Unit Cost
Velodyne VLP-16 (LiDAR)	\$1,000
GMSL2 1080p fisheye (Camera)	\$250

This project evaluates whether a monocular visual-inertial localisation system can match the existing LiDAR Nav2 baseline within operational safety constraints, and if not, characterises the gap to inform future sensor architecture decisions.

1.3 Literature Review

1.3.1 LiDAR-Based Localisation as the Baseline Standard

Continuous, reliable localisation is a fundamental requirement for autonomous shuttle operation. LiDAR and camera-based sensors represent the two principal modalities used to address this requirement, each with distinct sensing characteristics and performance trade-offs. LiDAR-based systems are widely deployed for vehicle localisation because they directly measure three-dimensional scene geometry via active laser ranging. This produces metric point clouds that enable accurate scan matching and pose estimation in SE(3). The direct observability of geometry makes LiDAR inherently robust to variations in illumination and visually degraded conditions, including night-time operation, shadows, and low-texture environments [6].

In contrast, camera-based systems estimate structure and motion indirectly from projected image observations. When used alone, this introduces scale ambiguity and increased sensitivity to motion blur, occlusion, and scene texture. To address these limitations, cameras are commonly paired with an Inertial Measurement Unit (IMU) or augmented with learned depth estimation, enabling more complete state observability through visual-inertial fusion.

Rather than being functionally interchangeable, LiDAR and camera-based approaches represent complementary sensing paradigms that differ in how geometric information is acquired and constrained

(Table 1.2). As noted in recent literature, the performance gap between these approaches has narrowed with the development of visual-inertial SLAM and improved feature-learning methods, leading to increased interest in hybrid and cross-modal localisation systems [6–8].

Table 1.2: Qualitative comparison of LiDAR and monocular camera sensing for autonomous vehicle localisation.

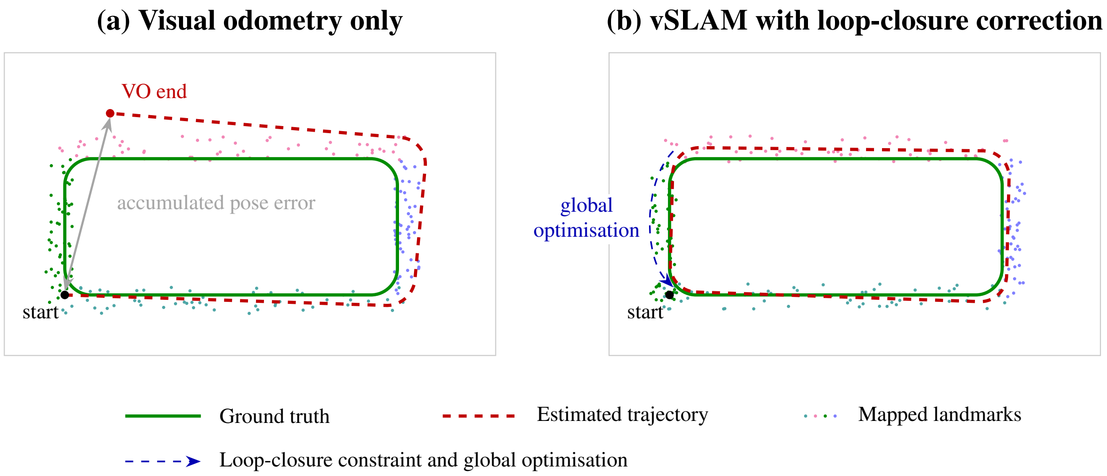
Property	LiDAR	Camera
Sensing principle	Active (emits laser pulses)	Passive (records ambient light)
Output	3D point cloud	2D image
Geometric accuracy	High, direct range measurement	Indirect, recovered via geometry or learned depth
Lighting sensitivity	Insensitive to illumination	Sensitive; degrades in low light, glare, shadow
Weather robustness	Degraded by heavy rain, fog, dust	Similarly degraded; reduced contrast
Semantic information	Limited (no colour, no texture)	Rich (colour, texture, signs, lane markings)
Onboard power draw	High	Low

1.3.2 From Motion Estimation to SLAM-Based Localisation

Visual odometry (VO) estimates ego-motion by tracking feature correspondences between consecutive frames. While effective over short horizons, it suffers from unbounded drift as small frame-to-frame pose errors accumulate. Reported drift rates range from 0.25% to 8% of the travelled distance [9], which for a typical ~ 1 km campus loop corresponds to terminal errors of approximately 2.5 m to 80 m, sufficient to disrupt lane-level navigation. Because VO lacks any global constraint or revisitation mechanism, this error grows monotonically with distance travelled.

Visual SLAM (vSLAM) extends VO by introducing a persistent map representation and enforcing global consistency through loop-closure constraints. When previously visited locations are re-observed, the system adds loop-closure edges to a pose graph and performs global optimisation to redistribute accumulated error across the trajectory. This shifts the error behaviour from unbounded drift to bounded correction, with long-term accuracy depending primarily on the frequency and reliability of loop closures [10].

This contrast is illustrated in Figure 1.2, where visual odometry exhibits continuous drift, whereas vSLAM corrects trajectory error through loop-closure-driven global optimisation.


(a) Visual odometry only

(b) vSLAM with loop-closure correction

Figure 1.2: Unbounded drift in visual odometry (left) versus loop-closure correction in visual SLAM (right).

Monocular vSLAM is well suited to structured shuttle environments due to repeated loop-closure opportunities. However, it remains limited by monocular scale ambiguity and sensitivity to motion blur, occlusion, and low-texture scenes. Paul et al. [11] quantify this limitation in a controlled MATLAB Simulink parking-lot environment, reporting an ORB-SLAM trajectory RMSE of 1.94 m under monocular operation compared to 0.25 m under stereo (approximately an 87% reduction), attributed to the availability of direct metric depth in stereo configurations.

Scale ambiguity is resolved in stereo systems through fixed-baseline depth estimation. Inertial sensing provides complementary motion constraints that improve robustness during visual degradation and short tracking interruptions, but does not directly provide metric depth. As a result, stereo-based, visual-inertial, and combined stereo-inertial configurations are commonly explored in the literature as extensions to monocular vSLAM.

These limitations motivate feature-based visual-inertial SLAM architectures such as ORB-SLAM3, which rely on robust feature extraction and data association to maintain stable tracking under partial observability and degraded visual conditions.

1.3.2.1 Feature Detection and Description

Reliable localisation in monocular vSLAM depends on stable feature correspondence across viewpoint and illumination changes. Corner-like structures with strong two-dimensional intensity gradients are preferred as they remain repeatable under moderate motion and perspective variation. The ORB (Oriented FAST and Rotated BRIEF) descriptor [12] combines the FAST corner detector [13] with a rotation-aware binary descriptor; at 15.3 ms per frame it is approximately 14× faster than SURF [14] while achieving comparable precision and recall on planar scenes [12]. This efficiency makes ORB the standard choice for real-time vSLAM on embedded platforms. Merzlyakov and Macenski [15] note that ORB-based systems fail earlier in low-texture or low-light environments but recover quickly when sufficient features reappear.

☐ Follow Camera
☑ Show Points
☑ Show KeyFrames
☐ Show Graph
☑ Show Inertial Graph
☐ Localization Mode
☐ Step By Step
☐ Show LBA opt

```
SLAM MODE |  Maps: 6, KFs: 22, MPs: 1199, Matches: 752
```

Figure 1.3: ORB keypoints (green squares) detected on a rectified nUWAy4 fisheye frame.

1.3.2.2 Keyframes, Covisibility, and Bundle Adjustment

To balance dense observations against limited compute, vSLAM systems select a sparse subset of frames as keyframes based on significant parallax or appearance change, pruning redundant ones and removing single-observation landmarks [2]. Retained keyframes are linked in a covisibility graph whose edges encode the number of shared landmarks between pairs (Figure 1.4), focusing local optimisation and loop-closure search on the relevant neighbourhood [2, 16].

Bundle adjustment (BA) [17] jointly refines keyframe poses and landmark positions by minimising reprojection errors:

$$\min_{\mathbf{X},\mathbf{C}} \sum_{i=1}^{N} \sum_{j\in \mathrm{obs}(i)} \left\| \mathbf{u}_{ij} - \pi\left(\mathbf{C}_j, \mathbf{X}_i\right) \right\|^2 \tag{1.1}$$

where $\mathbf{X}_i \in \mathbb{R}^3$ are landmarks, $\mathbf{C}_j \in \mathrm{SE}(3)$ are keyframe poses, $\mathbf{u}_{ij}$ is the observation of landmark i in keyframe j, and π is the projection function. ORB-SLAM3 [1] couples a real-time local BA over the recent covisibility neighbourhood with a global BA triggered by loop closures and map merges. Bundle adjustment cannot resolve absolute scale from images alone, due to the projective nature of monocular cameras under SE(3) ambiguity; this is the fundamental limitation that stereo or inertial fusion must address.

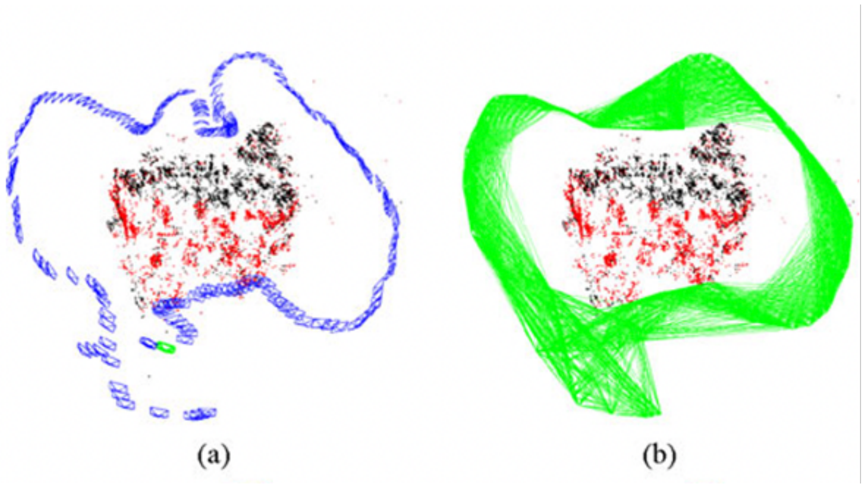


Figure 1.4: Keyframe covisibility graph from an ORB-SLAM reconstruction (adapted from Mur-Artal et al. [2]).

1.3.2.3 Visual-Inertial Fusion

Visual-inertial fusion adds inertial constraints to monocular vSLAM, resolving scale ambiguity through translational excitation and improving robustness during visual degradation. The accelerometer provides a gravity-direction prior; gyroscope integration gives orientation estimates between frames. IMU preintegration [18] compresses all inertial samples between two keyframes into a single relative-motion factor on the SO(3) manifold, avoiding repeated reintegration during nonlinear optimisation and underpinning both ORB-SLAM3 and VINS-Mono [1, 19]. Performance is sensitive to accurate IMU noise characterisation: under-weighting inertial constraints degrades scale observability, while over-weighting can cause initialisation failure [1].

1.3.2.4 IMU Noise Characterisation via Allan Variance

Allan Variance (AVAR) [3] decomposes inertial noise into stochastic regimes by analysing the variance of consecutive averages over varying integration intervals τ. On a log-log Allan deviation plot (Figure 1.5), each regime appears as a characteristic slope: Velocity/Angle Random Walk (VRW/ARW) at $k = -1/2$; Bias Instability (BI) as a near-zero plateau; and Acceleration/Rate Random Walk (AccRW/RRW) at $k = +1/2$. These three regimes directly populate the noise-density and bias-random-walk parameters used by ORB-SLAM3. However, AVAR assumes a stationary environment; real automotive conditions introduce unmodelled vibrations and vehicle dynamics, so inflation factors of $\times 3$ to $\times 10$ are commonly applied when transferring bench calibrations to outdoor platforms [1].

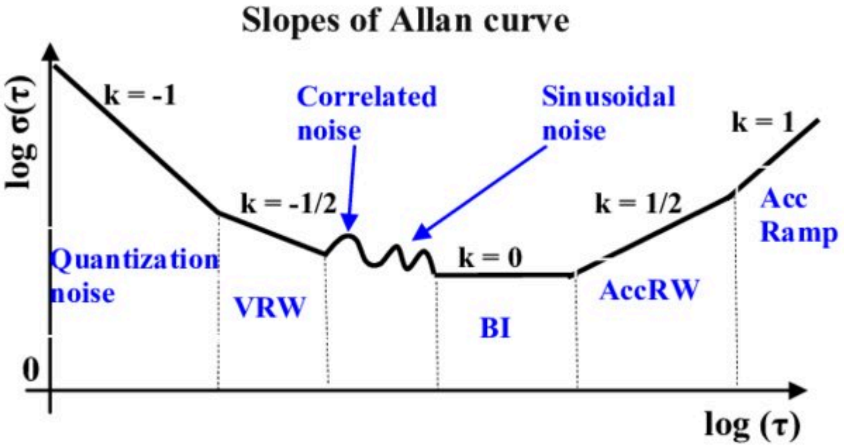


Figure 1.5: Schematic Allan deviation $\sigma(\tau)$ showing characteristic IMU noise regimes (reproduced from Matejček and Šostronek [3]).

1.3.3 ORB-SLAM3: State-of-the-Art Architecture and Automotive Deployment

Visual SLAM systems are typically classified along two axes: feature-based versus direct front-ends, and filter-based versus graph-optimisation back-ends. Filter-based methods (e.g. MSCKF) are computationally efficient but less accurate due to the absence of state relinearisation. Direct methods (e.g. DSO, VI-DSO) optimise photometric error and can perform well in low-texture scenes but are sensitive to illumination changes and outdoor visual variability.

Feature-based graph-optimisation systems such as ORB-SLAM3 [1] and VINS-Mono/Fusion [19] therefore dominate outdoor robotic applications. ORB-SLAM3 is adopted in this work as it integrates ORB feature tracking, covisibility-based optimisation, IMU preintegration, and fisheye camera support into a unified real-time framework [1, 20], making it suitable for long-duration autonomous vehicle operation.

ORB-SLAM3 operates using three parallel threads over a multi-map Atlas [1]. The tracking thread performs real-time pose estimation using ORB features and IMU propagation. The local mapping thread maintains the covisibility graph and performs local bundle adjustment. The loop-and-map-merging thread detects loop closures and performs global optimisation using DBoW2 place recognition. The full architecture is shown in Figure 1.6.

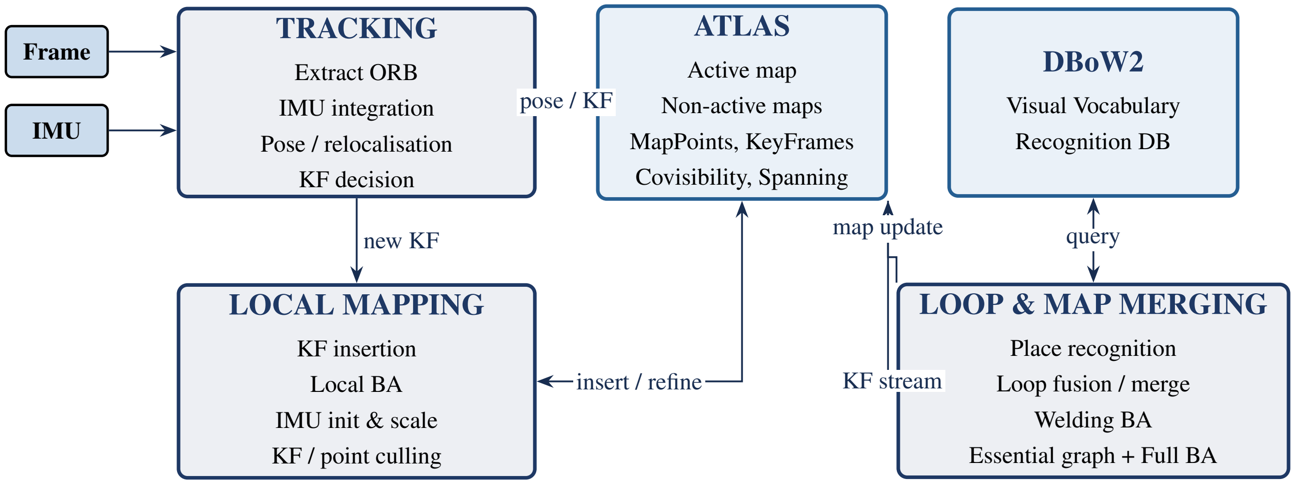


Figure 1.6: ORB-SLAM3 system architecture (adapted from Campos et al. [1]).

This architecture decouples real-time tracking from computationally intensive optimisation. The Atlas maintains one active and multiple inactive maps; on tracking loss, IMU propagation bridges the gap while relocalisation is attempted, and a new map is started if recovery fails, with prior maps retained for later merging.

ORB-SLAM3 uses a three-stage initialisation process: monocular structure-up-to-scale estimation, IMU-based alignment to recover scale and gravity, and full visual-inertial optimisation with bias refinement. This enables rapid metric convergence, with reported scale error reducing from approximately 5% to around 1% within ~ 15 s [1]. On EuRoC, it achieves 3.5 cm RMS ATE in stereo-inertial mode and outperforms VINS-Mono by approximately 2.6× in monocular-inertial configurations. Improvements are largely attributed to stronger mid-term data association via the covisibility graph, which stabilises tracking prior to loop closure.

ORB-SLAM3 also supports fisheye cameras using the Kannala-Brandt model [20], avoiding rectification and preserving wide field-of-view geometry. This improves feature stability at image boundaries, which is important in outdoor environments where peripheral structure contributes significantly to tracking robustness. Place recognition is handled via an extended DBoW2 pipeline, combined with geometric and covisibility verification to reduce false loop closures and improve long-term consistency.

1.3.4 Synthesis and Research Gap

The literature demonstrates that feature-based visual-inertial SLAM, particularly ORB-SLAM3, is a leading approach for low-cost vehicle localisation. However, existing evaluations are largely limited to short-duration dataset-based testing; no study evaluates monocular-inertial SLAM against a production-grade LiDAR baseline under sustained real-world autonomous shuttle operation. This thesis addresses that gap through experimental deployment on the nUWAy platform, forming the basis for the engineering objectives in Section 1.4.

1.4 Project Objectives

This project evaluates monocular-inertial SLAM for sustained autonomous shuttle localisation and quantifies its performance relative to a LiDAR-based baseline under real-world operating conditions.

The central **hypothesis** is that a monocular-inertial ORB-SLAM3 pipeline, deployed on a NVIDIA Jetson AGX Orin with a GMSL fisheye camera and SBG IMU, can achieve competitive localisation accuracy and robustness while satisfying onboard compute constraints.

To ensure an unbiased evaluation across all experiments, a trajectory ATE RMSE of 1.0 m was established as a fixed evaluation criterion prior to data collection. The objectives are as follows:

O1. System integration. Deploy ORB-SLAM3 monocular-inertial SLAM on ROS 2, integrating synchronised GMSL camera and SBG IMU data streams on the Jetson AGX Orin.

O2. Sensor calibration and characterisation. Estimate camera intrinsics and IMU noise parameters via Allan-variance analysis, and verify frame conventions and initial bias, gravity, and scale alignment for outdoor operation.

O3. Trajectory evaluation. Quantify localisation accuracy using ATE RMSE against a GPS-aided inertial reference, evaluated over repeated campus-route trials and segmented ablations.

O4. Robustness and computational performance. Assess tracking stability via failure rate and relocalisation behaviour, and measure runtime performance including CPU/GPU utilisation, power draw, and thermal limits on embedded hardware.

O5. Inertial sensitivity analysis. Evaluate system sensitivity to IMU noise modelling by comparing performance under $\times 10$ and $\times 100$ noise inflation relative to Allan-variance-derived parameters.

Expected outcome and impact. Successful completion of these objectives provides a quantitative benchmark of ORB-SLAM3 on a deployed autonomous shuttle platform. Validation of the hypothesis would demonstrate that a monocular-inertial SLAM system can substitute for LiDAR-based localisation under real operational constraints, indicating per-vehicle sensing cost reductions of approximately a factor of four (Table 1.1) together with reduced onboard power consumption and system-integration complexity. The calibration and sensor-frame verification procedures developed are transferable to future visual-inertial autonomous-vehicle deployments and support scalable low-cost localisation system design.

2. Methodology

2.1 Overview

This chapter specifies the experimental methodology used to evaluate ORB-SLAM3 [1] on the nUWAy4 autonomous shuttle. The algorithm itself is treated as a fixed estimator; the chapter therefore concerns the integration, calibration, configuration, and evaluation procedures required to obtain a controlled comparison against the deployed LiDAR-based Nav2 [21] baseline. The investigation is structured as a single-platform benchmarking study centred on a primary campus-loop recording, with stochastic variation of the estimator addressed through repeated trials and a segment-matched ablation. Section 2.2 describes the platform; Section 2.3 the integration architecture; Section 2.4 the sensor characterisation; Section 2.5 the deployed configuration; Section 2.6 the dataset; and Section 2.7 the evaluation procedure, which constitutes the methodological core of the chapter (Figure 2.1).

The study was conducted on real deployed hardware rather than in simulation. Simulation-based evaluations of ORB-SLAM3 are well established in the literature (e.g. 11); a hardware deployment was selected here to evaluate performance under real operating conditions, including real fisheye optics, real IMU vibration coupling, real lighting and motion, and the dominant fixed-route motion profile of the shuttle, none of which are reproducible in simulation.

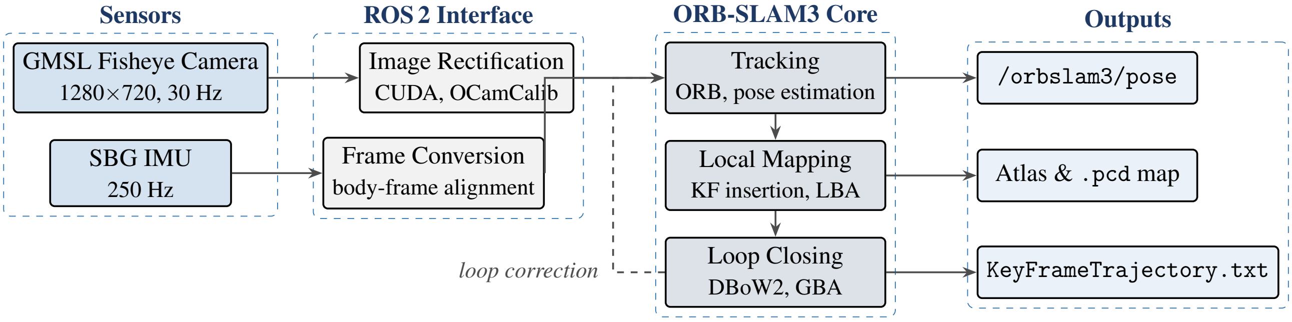


Figure 2.1: Deployed monocular-inertial ORB-SLAM3 pipeline on the nUWAy4 platform.

2.2 nUWAy4 Hardware Platform

All experimental work was conducted on nUWAy4, a second-generation EasyMile EZ10 shuttle operated by the UWA REV programme. The three primary components of the deployed pipeline are summarised in Figure 2.2: onboard compute (Figure 2.2a), the inertial reference (Figure 2.2b), and the visual front-end (Figure 2.2c). Sensing relevant to this investigation comprises a front-facing GMSL2 fisheye camera (Figure 2.2c) with a 143° field of view (rectified to a 1280 × 720 pin-hole equivalent on the GPU before reaching the SLAM pipeline) and an SBG Ellipse-D dual-antenna GNSS/INS (Figure 2.2b) publishing inertial samples and an EKF-fused position estimate at 250 Hz. The existing Velodyne VLP-16 and SICK LD-MRS LiDARs are retained on nUWAy4 only to drive the Nav2 baseline against which the camera-IMU pipeline is benchmarked; they are deliberately excluded from the SLAM input so that the substitution hypothesis can be evaluated without the very modality it seeks to replace. The four corner-mounted SICK LMS-151 units provide the safety-critical emergency-stop function and are not evaluated in this work. Compute is provided by an NVIDIA Jetson AGX Orin Developer Kit (Figure 2.2a) running Ubuntu 22.04 under the JetPack-supplied L4T kernel, with CUDA-enabled OpenCV for hardware-accelerated rectification. Exact dependency versions are reproduced in Appendix B. The IMU is rigidly mounted parallel to the road surface inside the vehicle roof, approximately 1.8 m rearward of the camera optical centre; the sensor positions are shown in Figure 2.3. The resulting coordinate frames are summarised in Table 2.1 and the body-from-camera extrinsic is specified in Equation (2.1).

(a) NVIDIA Jetson AGX Orin Developer Kit.

(b) SBG Ellipse-D GNSS/INS.

(c) GMSL2 fisheye camera (front-facing).

Figure 2.2: Hardware components of the deployed pipeline: Jetson AGX Orin (left), SBG Ellipse-D (centre), GMSL2 fisheye camera (right).

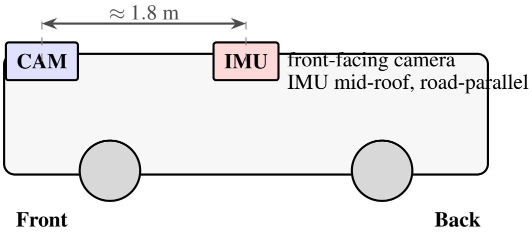


Figure 2.3: Deployed sensor positions on nUWAy4: GMSL2 fisheye camera at the front of the vehicle and SBG Ellipse-D IMU mounted mid-roof, separated by approximately 1.8 m.

Table 2.1: Coordinate frames used in the deployed pipeline (ROS 2 FLU convention [4]).

Frame	Convention	Origin
`map`	FLU [4]	First keyframe pose, gravity-aligned at IMU init
`base_link`	FLU [4]	Vehicle body, coincident with `imu_link`
`camera`	OpenCV optical	1.8 m forward of body origin
SBG IMU native	Sensor-defined (FRD, verified by inspection)	Hardware sensor frame

2.3 Integration Architecture

2.3.1 ROS 2 Middleware

The deployed pipeline runs on ROS 2 Humble, the long-term-support release used across the UWA REV production stack. Two communication primitives were used (Figure 2.4): the topic publisher-subscriber model for continuous sensor streaming (GMSL camera frames and SBG IMU samples), and the service-client model for one-shot operations (map save, parameter queries).

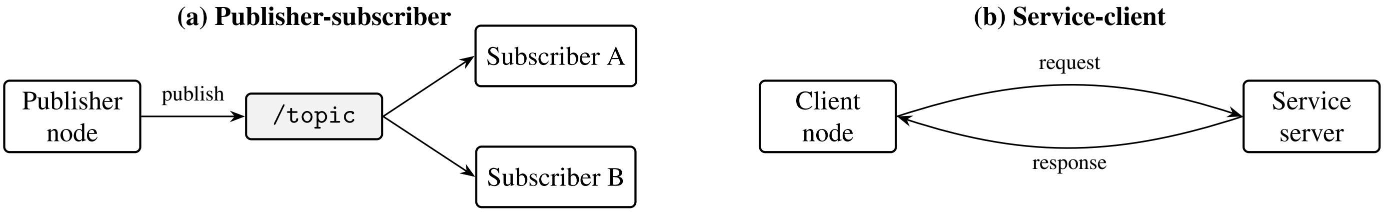


Figure 2.4: ROS 2 communication primitives: topic streaming (a) and service transactions (b).

ROS 2 bags record all topic messages with timestamps to disk (`.mcap`); replay via `ros2 bag play` re-publishes the same sensor stream at identical relative timestamps, and the `-rate` flag permits slower replay. These properties allow multiple SLAM configurations to be evaluated against bit-for-bit identical sensor input from a single drive recording.

2.3.2 Pipeline Wrapper

A custom integration node was developed to bridge the GMSL and SBG streams into ORB-SLAM3. The node enforces three properties required for controlled benchmarking. First, camera and IMU samples are paired by nearest-timestamp matching, reducing sensitivity to arrival-order jitter in the ROS 2 message queue. Second, a fixed SO(3) rotation is applied to every IMU sample at ingress (Section 2.5.2), enforcing a single FLU body-frame convention across the visual, inertial, and reference channels. Third, the keyframe trajectory (`KeyFrameTrajectory.txt`) is persisted to disk on node shutdown, and the map point cloud is saved via an explicit ROS 2 service call, ensuring that the offline analysis described in Section 2.7 operates on a deterministic final-state snapshot. The core software components are listed in Table 2.2, and the live ROS 2 topic graph during an ORB-SLAM3 session is shown in Figure 2.5.

Table 2.2: Core software components and rationale.

Component	Role	Rationale
ORB-SLAM3 [1]	Visual-inertial state estimation	Mature open-source library combining tightly-coupled monocular-inertial estimation, multi-map Atlas, and loop closure.
Custom `mono_imu` node	Project-specific launch and synchronisation	Bridges GMSL and SBG topics into ORB-SLAM3 with deterministic time-pairing and body-frame alignment.
SBG ROS 2 driver	Inertial and EKF stream publication	Publishes raw IMU samples and the EKF-fused position estimate from the SBG Ellipse-D at 250 Hz.
GMSL fisheye camera calibration	Intrinsic calibration	Calibrated GMSL2 fisheye intrinsics and OCamCalib rectification model used by the visual front-end.
`allan_variance_ros`	IMU noise characterisation	Standard tool for offline Allan-variance analysis.

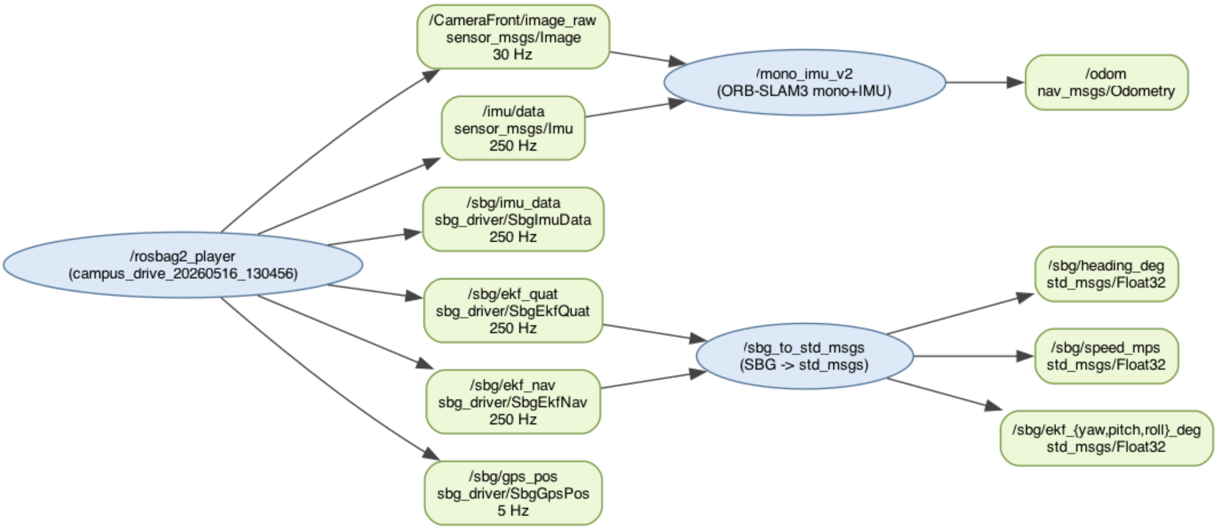


Figure 2.5: ROS 2 topic graph during a live ORB-SLAM3 session on the nUWAy4 platform.

2.4 Sensor Calibration

2.4.1 Fisheye Intrinsic Calibration

Calibration was performed using the MATLAB Camera Calibrator toolbox with an 11×16 inner-corner checkerboard (15 mm squares), captured across the full field of view including the high-distortion periphery (Figure 2.6). A minimum of 20 images were retained following manual screening for motion blur and partial occlusion, with the checkerboard required to occupy at least 20% of the frame in each capture.

An iterative manual outlier-rejection process was applied using per-image reprojection-error analysis; images exhibiting disproportionately high error were removed and the calibration re-run until convergence of the mean reprojection error. The final calibration was accepted based on a mean reprojection error below 1 pixel and visual inspection of the undistorted output using MATLAB's built-in Show Undistorted function.

The resulting intrinsics are used within the GMSL undistortion pipeline to ensure consistency between the rectified imagery and the internal projection model. The numerical calibration results are reported in Chapter 3.

Figure 2.6: GMSL2 fisheye frame before (left) and after (right) calibration and rectification to a pinhole model.

2.4.2 IMU Noise Characterisation

The SBG IMU was characterised by Allan-variance analysis on a 3-hour static recording captured with the vehicle parked level. The `allan_variance_ros2` tool was used to fit characteristic slopes ($-1/2$, 0, $+1/2$) to the log-log $\sigma(\tau)$ curve of each axis, yielding angle/velocity random walk, bias instability, and rate/acceleration random walk parameters for both accelerometer and gyroscope channels. The numerical noise parameters, and the inflation factors applied in the deployed configuration, are reported in Chapter 3 (Section 3.5).

2.5 Pipeline Configuration

2.5.1 Deployed Parameter Rationale

Four parameter groups were modified from ORB-SLAM3 defaults, summarised in Table 2.3; the complete YAML is in Appendix B.

Table 2.3: ORB-SLAM3 parameters modified from defaults for the nUWAy4 deployment. Full YAML in Appendix B.

Parameter	Value	Rationale
`ORBextractor.nFeatures`	3,000	Elevated above default to offset unreliable peripheral features after OCamCalib rectification.
`ORBextractor.iniThFAST`	8	Reduced (default 20) for low-texture road scenes.
`ORBextractor.minThFAST`	3	Reduced (default 7) for the same reason.
`IMU.NoiseGyro`	8.17×10^{-4} $\mathrm{rad\,s^{-1}\,Hz^{-1/2}}$	AVAR ×100 (Section 3.5).
`IMU.NoiseAcc`	$6.58 \times 10^{-3}\ \mathrm{m\,s^{-2}\,Hz^{-1/2}}$	AVAR ×100.
`IMU.GyroWalk`	$9.6 \times 10^{-5}\ \mathrm{rad\,s^{-2}\,Hz^{-1/2}}$	AVAR ×10.
`IMU.AccWalk`	$4.5 \times 10^{-4}\ \mathrm{m\,s^{-3}\,Hz^{-1/2}}$	AVAR ×10.

The body-from-camera extrinsic, in the EuRoC convention used by ORB-SLAM3 [1], is

$$\mathbf{T}_{BC} = \begin{bmatrix} 0 & 0 & 1 & 1.8 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2.1}$$

mapping a point from the camera optical frame into the body FLU frame, with the translation column encoding the 1.8 m longitudinal lever arm. The rotation block reflects the coordinate-axis relationship between the OpenCV optical frame and the FLU body frame; the full matrix was manually constructed from the physically measured sensor separation and verified against the hardware mounting prior to deployment.

2.5.2 IMU Body-Frame Alignment

The SBG Ellipse-D publishes inertial samples in a Forward-Right-Down (FRD) sensor-defined frame, whereas ORB-SLAM3 expects FLU [4]. The `mono_imu` node applies a fixed rotation

$$\mathbf{R}_{\mathrm{FLU}\leftarrow\mathrm{IMU}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{2.2}$$

to every IMU sample at ingress; this is a π-radian rotation about the body X-axis with $\det = +1$, applied identically to accelerometer and gyroscope channels. Correctness was verified by confirming alignment of the post-rotation accelerometer reading with $-Z$ on the parked vehicle, within the ±2% static-gravity tolerance specified below. This corrected a frame-convention error present in the inherited integration path.

2.5.3 Pre-Run Verification Protocol

Each live session was preceded by a four-step verification protocol designed to detect hardware or configuration regressions before recording:

1. IMU and camera publication rates verified to within 5% of 250 Hz and 30 Hz respectively, using `ros2 topic hz`.
2. Accelerometer magnitude verified to within ±2% of $9.81\ \mathrm{m\,s^{-2}}$ on the parked vehicle, confirming both scale and post-rotation orientation.
3. `NaN` and `Inf` values verified absent on both topics over a 10 s idle recording.
4. The YAML $\mathbf{T}_{BC}$ matrix inspected against the physical sensor mounting to guard against silent axis errors.

2.6 Test Data Collection on the UWA Campus

Test data was collected by driving nUWAy4 on the UWA campus and recording the rectified camera stream, SBG IMU samples, and SBG EKF position output into a single ROS 2 bag. Single-bag recording guarantees that ORB-SLAM3 and the SBG reference are evaluated on identical input, eliminating disagreement attributable to differing test conditions. Two recordings were used in this work: a principal campus loop (Chapter 3) and a second-dataset straight-drive segment used to assess robustness under an alternative motion profile (Section 3.7). The two recordings were captured on separate days at similar times of day to preserve comparable lighting conditions. Both routes are shown in Figure 2.7, with the principal loop traced in red and the straight-drive segment in blue.

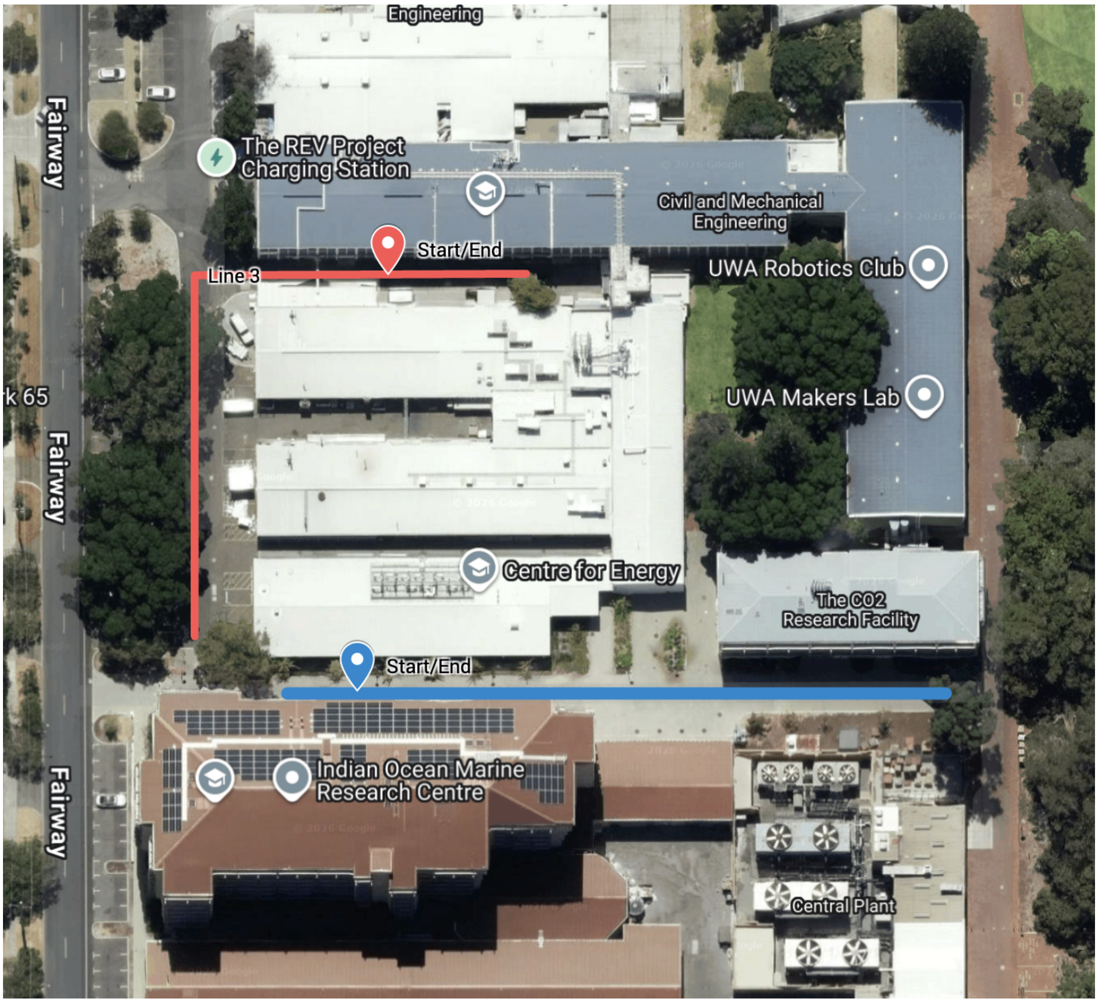


Figure 2.7: Evaluation routes on the UWA campus: principal benchmark loop (red) and straight-drive segment (blue).

Recordings were taken in daylight at vehicle speeds of approximately 1 to 2 $\mathrm{m\,s^{-1}}$ (walking pace). The principal loop traverses a route containing both straight segments and turns; the second dataset follows a predominantly straight corridor segment. Tracking was launched with the vehicle stationary; ORB-SLAM3 then initialised once translational motion commenced and parallax became available. Key properties of both recordings are summarised in Tables 2.4 and 2.5.

Table 2.4: Principal benchmark loop recording properties.

Property	**Value**
Duration	$\approx$180 s
Route	UWA campus loop (Figure 2.7)
Lighting	Daylight
Vehicle speed	approximately 1–2 $\mathrm{m\,s^{-1}}$ (walking pace)
Topics recorded	GMSL front (30 Hz), SBG IMU (250 Hz), SBG EKF (250 Hz)
SBG `position_valid`	True for all samples

Table 2.5: Second-dataset straight-drive recording properties.

Property	**Value**
Duration	$\approx$150 s
Route	UWA campus straight-drive segment (Figure 2.7)
Lighting	Daylight
Vehicle speed	approximately 1–2 $\mathrm{m\,s^{-1}}$ (walking pace)
Topics recorded	GMSL front (30 Hz), SBG IMU (250 Hz), SBG EKF (250 Hz)
SBG `position_valid`	True for all samples

2.7 Evaluation Procedure

The evaluation is designed as a controlled comparison between the ORB-SLAM3 monocular-inertial estimate and a deployment-grade reference recorded simultaneously on the same bag. Four elements of that design require justification: the choice of reference, the alignment procedure, the metrics, and the handling of stochastic variation.

2.7.1 Reference Trajectory

The reference is the SBG Ellipse-D EKF position output, which fuses dual-antenna RTK GNSS with the onboard IMU into a single 250 Hz estimate. All 45,089 samples in the principal recording reported `position_valid` = True, with the vehicle parked outdoors before the recording commenced to obtain a stable GPS fix. The SBG EKF is treated as a deployment-grade rather than absolute ground truth: it fuses GPS with an internal IMU and is therefore not independent of an inertial sensor, and accuracy degrades during GPS-denied periods such as those beneath tree canopy or near buildings. A survey-grade RTK GNSS receiver or a tracked retroreflector network would provide a stronger independent reference but was outside the project's scope. The SBG is selected because it is the highest-quality reference physically present on nUWAy4 during operation, providing a deployment-representative onboard reference.

The production nUWAy4 localisation stack uses a Nav2 LiDAR system with a campus-wide 1.25M-point reference map (Figure 2.8). LiDAR-based localisation (such as the NDT or ICP scan-matching approaches reported in comparable autonomous vehicle studies [10]) is capable of sub-decimetre positional accuracy under favourable conditions and would provide a stronger independent reference than the onboard SBG EKF. The ORB-SLAM3 point cloud was visually compared against this LiDAR map as a qualitative topological check: landmark distributions were geometrically consistent with the route geometry, confirming that the monocular vSLAM map correctly captured the campus loop structure at a coarse level.

However, the inconsistency and limited duration of tracking across trials made a rigorous quantitative comparison infeasible within this project. Establishing a precise transformation between the ORB-SLAM3 internal SLAM frame and the LiDAR-map engineering frame requires a sufficient number of accurately associated corresponding poses, which the frequent tracking failures (Section 3.6) did not permit. Consequently, quantitative evaluation was performed against the onboard SBG EKF reference rather than the production LiDAR localisation stack. Should future work achieve consistent and sustained tracking, a frame-to-frame alignment between the ORB-SLAM3 map and the LiDAR point cloud would offer a meaningful secondary accuracy reference that is fully independent of GNSS, particularly in GPS-denied campus environments.

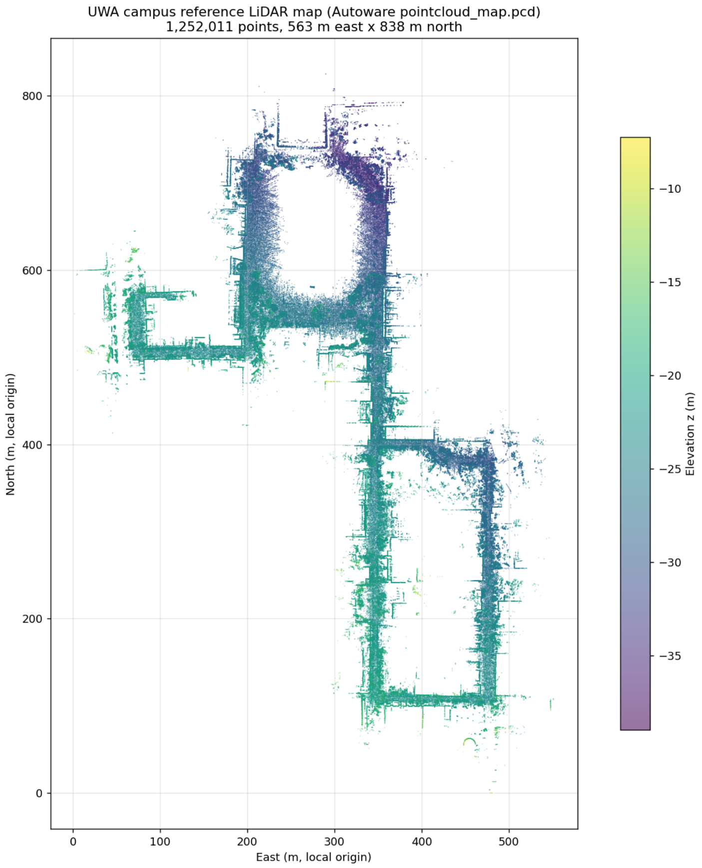


Figure 2.8: UWA campus LiDAR reference map (1.25M points). The benchmark loop occupies the western strip.

2.7.2 Alignment Procedure

Trajectories were aligned using the closed-form Sim(3) transformation of Umeyama [5], which jointly recovers global rotation, translation, and scale. Sim(3) was used in preference to SE(3) because monocular SLAM recovers geometry only up to an unknown scale factor; absorbing scale into the alignment isolates trajectory shape error from absolute-scale error. The recovered scale factor is reported alongside ATE RMSE as a proxy for scale-recovery quality.

2.7.3 Metrics

Trajectory accuracy was quantified using Absolute Trajectory Error (ATE) RMSE and maximum ATE following the definitions of Sturm et al. [22]. ATE evaluates the global consistency of an estimated trajectory by measuring the Euclidean distance between estimated and reference poses after alignment. Unlike Relative Pose Error (RPE), which measures local odometric drift, ATE captures cumulative global error over the full trajectory and is the standard metric for comparing visual and visual-inertial SLAM systems on benchmark datasets.

ATE is computed in three stages. First, the estimated and reference trajectories are spatially aligned, since they occupy arbitrary and independent coordinate frames. In this work, alignment was performed using the closed-form Sim(3) transformation of Umeyama [5] (described in Section 2.7.1), which recovers the optimal rotation, translation, and scale mapping the estimated trajectory onto the reference. Second, at each matched timestamp, the Euclidean displacement between the aligned estimated pose and the reference pose is computed. Third, these per-pose errors are aggregated as the root-mean-square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{p}_{i,\text{est}} - \mathbf{p}_{i,\text{ref}}\right\|^2} \tag{2.3}$$

where $\mathbf{p}_{i,\text{est}}$ and $\mathbf{p}_{i,\text{ref}}$ are the aligned estimated and reference positions at the i-th matched timestamp, and n is the number of matched pairs. The maximum ATE is reported alongside RMSE to capture peak localisation error. An evaluation criterion of 1.0 m ATE RMSE was established prior to data collection as the threshold for deployment-suitable accuracy.

Robustness was characterised by tracking-loss frequency and recovery behaviour, and computational performance through CPU/GPU utilisation and power draw measured by `tegrastats`. Map spatial fidelity was assessed qualitatively by rendering the exported `.pcd` files in top-down view.

2.7.4 Repeatability and Run-to-Run Variation

ORB-SLAM3 is multi-threaded with unseeded RANSAC; repeated trials on the same recording produce slightly different trajectories. This variation was characterised through repeated trials rather than suppressed via a deterministic build, which would not reflect operational behaviour. The evaluation comprised two phases: five distinct configurations (v1–v5) covering monocular-only, varying replay rates, and a ×10 IMU ablation; then a five-trial reproducibility session (v6–v10) of the deployed configuration on the same recording. The v3 long run additionally serves as a controlled baseline for the inflation-factor comparison (Section 3.5), isolating the inflation effect from segment difficulty.

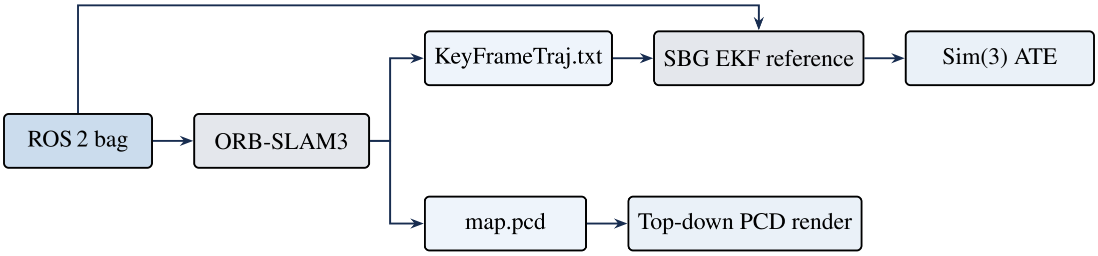


Figure 2.9: Offline evaluation pipeline: bag replay, ORB-SLAM3, Sim(3) alignment, ATE computation.

2.8 Methodological Assumptions and Limitations

Four assumptions bound the validity of the results. The SBG EKF is taken as a deployment-grade reference, not independent ground truth. The IMU mounting is assumed rigid within each recording, with vibration treated as unmodelled noise. ROS 2 header timestamps are assumed adequate for offline evaluation; no sub-second clock-offset correction is applied. Results characterise the nUWAy4 mounting geometry specifically; generalisation to other vehicle classes is outside scope.

Algorithmic limitations. Monocular scale recovery requires translational excitation; monocular-inertial bootstrap introduces an initialisation transient during which the active map may be rebuilt, a known ORB-SLAM3 property. Run-to-run variation from unseeded multi-threaded RANSAC is characterised in Section 2.7.4.

Sensor and calibration limitations. OCamCalib pre-rectification to a 1280×720 pin-hole image reduces feature persistence at the image periphery; native Kannala-Brandt8 support is identified in Chapter 5 as the primary follow-on change. Sub-second camera-IMU clock offsets are not corrected at runtime.

Environmental limitations. Testing was conducted under daylight conditions with sufficient ambient illumination for stable feature extraction and tracking. Weather and lighting conditions were assumed reasonably consistent between trials on the same recording day; significant variation in illumination, glare, shadows, rain, or low-light operation was outside the scope of evaluation. The results therefore characterise monocular vSLAM performance under favourable outdoor illumination, and robustness to adverse conditions remains unquantified.

Evaluation limitations. SBG EKF reference accuracy and the absence of ORB-SLAM3-to-LiDAR-map alignment are discussed in Section 2.7.1.

3. Results

Results are presented in three stages. **Sensor characterisation** (Sections 3.1–3.2) verifies that the camera and IMU meet the accuracy requirements for ORB-SLAM3 initialisation. **Trajectory evaluation** (Sections 3.3–3.5) quantifies localisation accuracy against the 1.0 m ATE RMSE criterion. **Deployment characterisation** (Sections 3.6–3.8) assesses run-to-run reproducibility, sensitivity to motion profile, and onboard resource utilisation.

3.1 Camera Calibration Results

Intrinsic calibration of the front-facing GMSL fisheye camera was performed using a minimum of 20 checkerboard observations spanning the full field of view. Iterative per-image outlier rejection converged to a mean reprojection error of 0.14 pixels (Table 3.1).

Table 3.1: Camera intrinsic calibration results.

Parameter	Calibrated Value
Focal length ($f_x = f_y$)	321.5625 px
Principal point (c_x)	640 px
Principal point (c_y)	360 px
Mean reprojection error	≈ 0.14 pixels
Output resolution	1280×720 (rectified pinhole)

The 0.14-pixel reprojection error is well within the < 0.5 pixel threshold required for reliable ORB feature triangulation [1]. Calibrated intrinsics are adopted over manufacturer-nominal values for all experiments in this chapter.

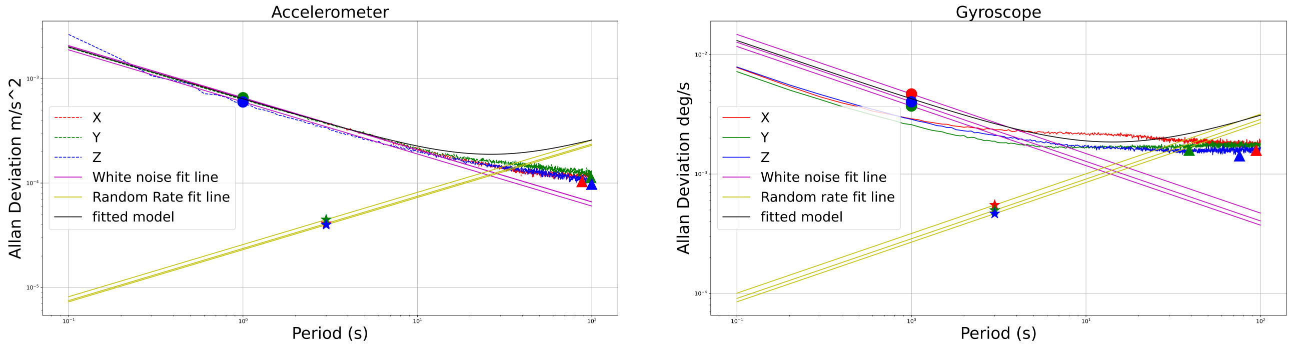
3.2 IMU Allan Variance Characterisation

Allan Variance analysis of the SBG IMU over a 3-hour static recording yielded the noise-density and random-walk parameters in Table 3.2; the corresponding Allan-deviation curves are shown in Figure 3.1.

Table 3.2: SBG Ellipse-D IMU noise parameters from Allan-variance analysis (degree-based units; SI equivalents in Table 2.3).

Parameter	Value (X, Y, Z)	Units
Accel. VRW (Velocity Random Walk)	0.00066, 0.00066, 0.00060	$\mathrm{m\,s^{-1}\,Hz^{-1/2}}$
Accel. BI (Bias Instability)	0.00010, 0.00011, 0.00010	$\mathrm{m\,s^{-2}}$
Accel. RW (Random Walk)	0.00004, 0.00004, 0.00004	$\mathrm{m\,s^{-3}\,Hz^{-1/2}}$
Gyro. ARW (Angle Random Walk)	0.00468, 0.00371, 0.00402	$\mathrm{deg\,s^{-1/2}}$
Gyro. BI (Bias Instability)	0.00158, 0.00158, 0.00142	$\mathrm{deg\,s^{-1}}$
Gyro. RRW (Rate Random Walk)	0.00055, 0.00050, 0.00047	$\mathrm{deg\,s^{-2}\,Hz^{-1/2}}$

All six parameters are within manufacturer-stated specifications. The VRW and BI values confirm that the SBG Ellipse-D meets the noise-floor requirements for ORB-SLAM3 inertial initialisation. An empirical ×100 inflation on white-noise terms and ×10 on random-walk terms was applied in the deployed YAML; the rationale is examined in Chapter 4.



(a) Accelerometer ($\mathrm{m\,s^{-2}}$).

(b) Gyroscope ($\mathrm{deg\,s^{-1}}$).

Figure 3.1: Measured Allan deviation of the SBG IMU from a 3-hour static recording. Slope intercepts yield the parameters in Table 3.2.

3.3 ORB-SLAM3 Pipeline Validation

The integration of the GMSL camera, SBG IMU, and ORB-SLAM3 `mono_imu` node was verified before trajectory evaluation. IMU initialisation completed within ≈20 s of motion onset across all trials, with finite optimiser outputs (gravity vector, biases, scale) sustained for the full duration of each active map. The normalised gravity estimate converged to within 0.05 of $[0,0,-1]^\top$ in each component [4]. Loop-closure events were detected on re-traversal of previously mapped segments, confirming that the Atlas multi-map back-end was operating as intended. These observations establish that the pipeline was correctly integrated; subsequent sections evaluate its localisation accuracy.

3.4 Spatial Consistency of SLAM Maps

Map files (`.pcd`) were exported at shutdown for six representative configurations (mono-only, v1–v5); v6–v10 share the same ×100 configuration and are covered in Section 3.6. Top-down landmark clouds are shown in Figure 3.2; landmark counts are in Table 3.3.

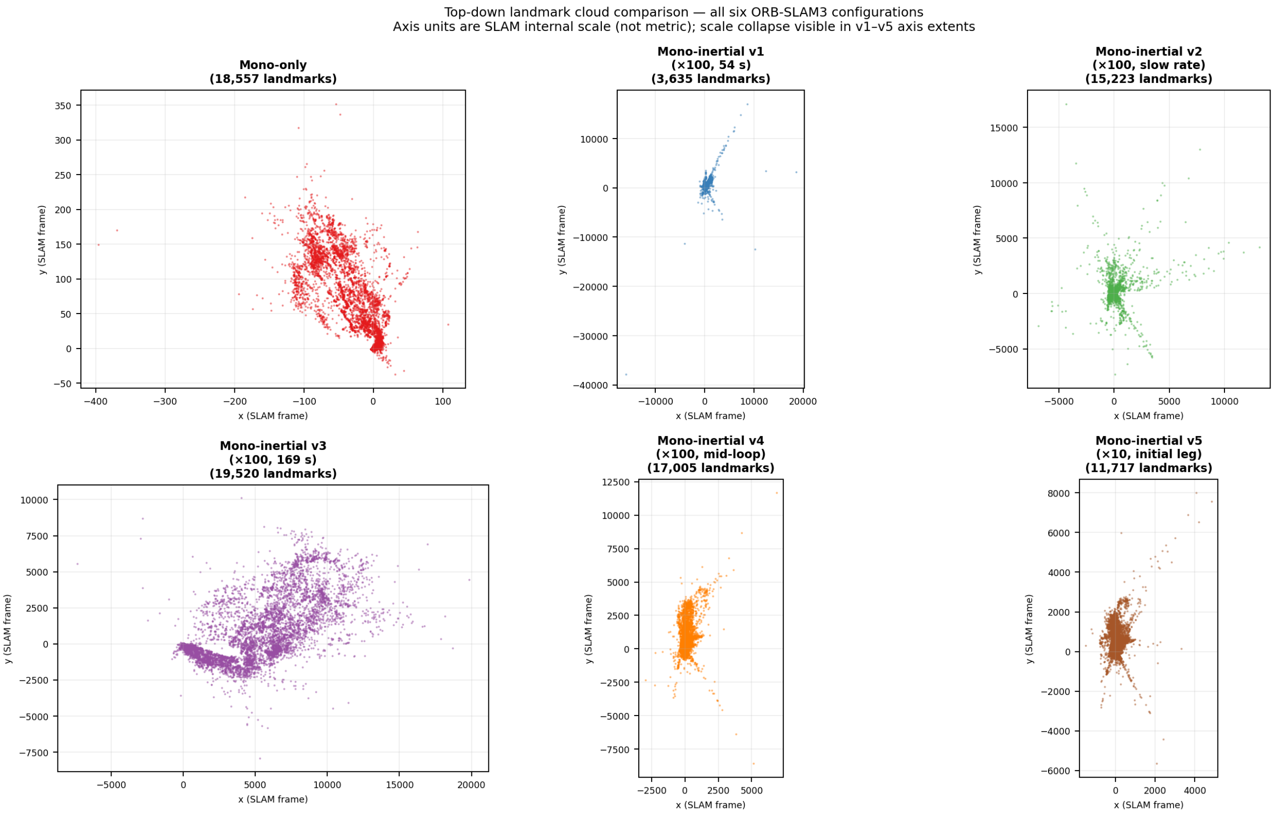


Figure 3.2: Top-down landmark clouds for six representative configurations. Axis units are SLAM-internal scale and are not metric-comparable across panels.

Table 3.3: Landmark counts per configuration.

Configuration	Landmarks
Mono-only	18,557
v1 (×100, 54 s)	3,635
v2 (×100, slow rate)	15,223
v3 (×100, 169 s)	19,520
v4 (×100, mid-loop)	17,005
v5 (×10, initial leg)	11,717

All configurations recover a consistent global loop topology, confirming that scale ambiguity does not prevent structural map recovery. Landmark counts range from 3,635 (v1, 54.5 s) to 19,520 (v3, 169 s) and scale directly with tracking duration, as expected given the fixed ORB feature density per keyframe. Axis extents in Figure 3.2 are in SLAM-internal units and are not metric-comparable across panels; the variation in extent reflects reconstructed coordinate scale, not physical map size.

3.5 Trajectory Accuracy Against Reference

Eight configurations were evaluated offline from the same ROS 2 bag using Sim(3) alignment [5]; results are in Tables 3.4 and 3.5 (Figures 3.3–3.9).

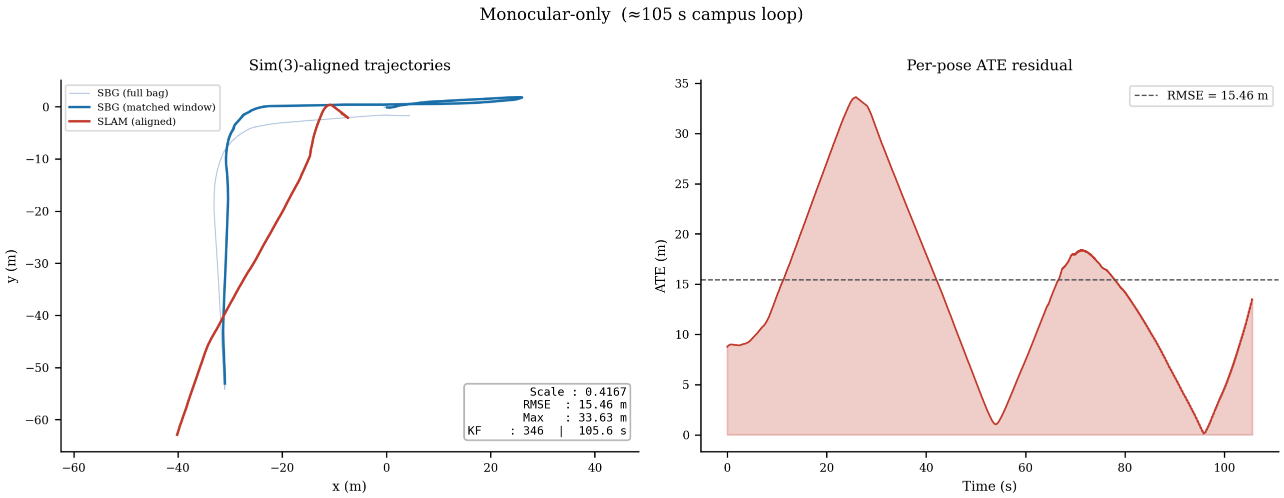


Figure 3.3: Monocular-only trajectory (≈105 s): Sim(3)-aligned path (left) and ATE residual (right).

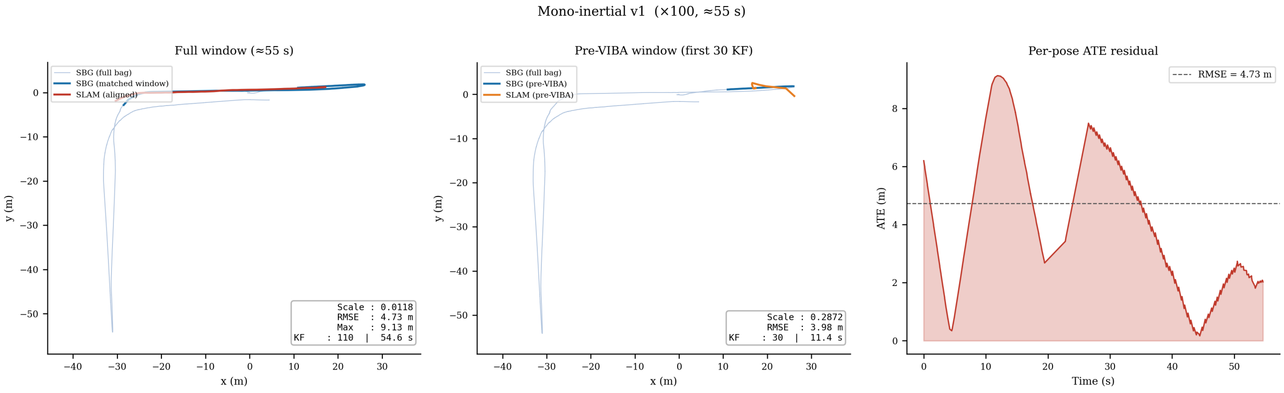


Figure 3.4: v1 trajectory (≈55 s, ×100): Sim(3)-aligned path (left), pre-VIBA window (centre), ATE residual (right).

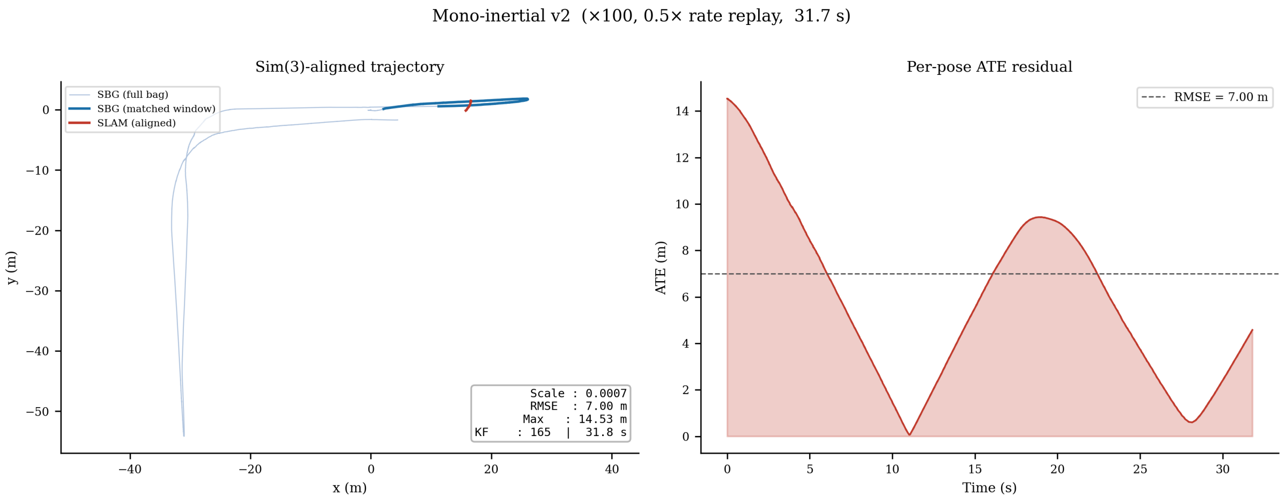


Figure 3.5: v2 trajectory (31.7 s, ×100, 0.5× replay): Sim(3)-aligned path (left) and ATE residual (right).

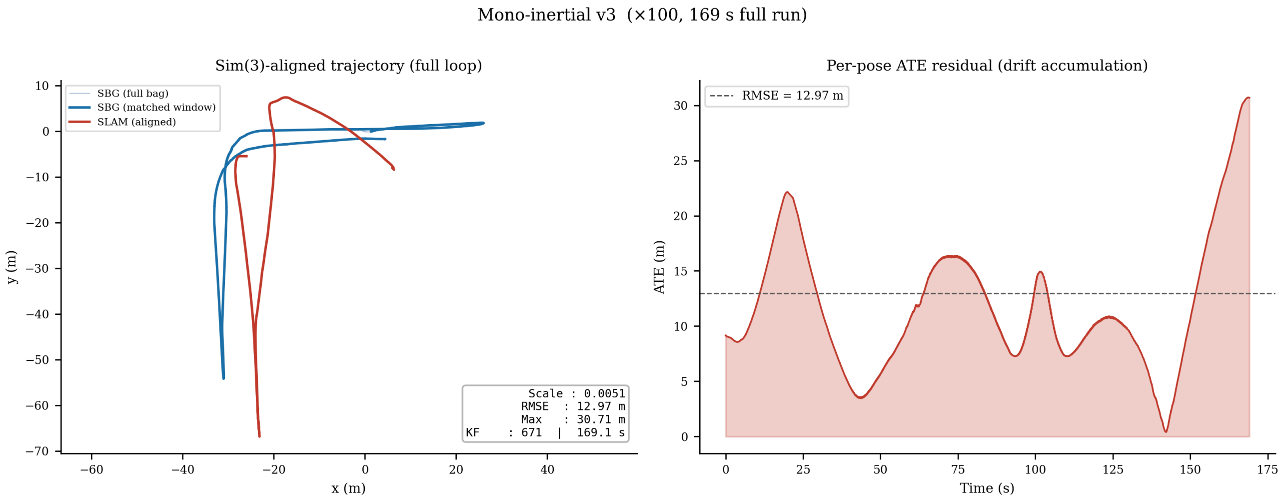


Figure 3.6: v3 trajectory (169 s, ×100): Sim(3)-aligned path (left) and ATE residual (right).

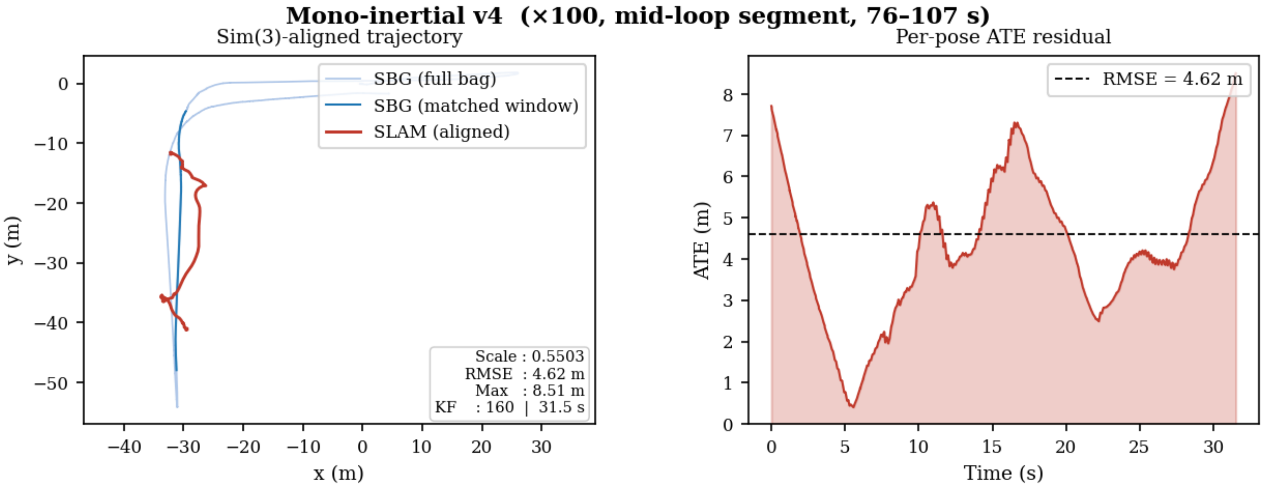


Figure 3.7: v4 trajectory (76–107 s segment, ×100): Sim(3)-aligned path (left) and ATE residual (right).

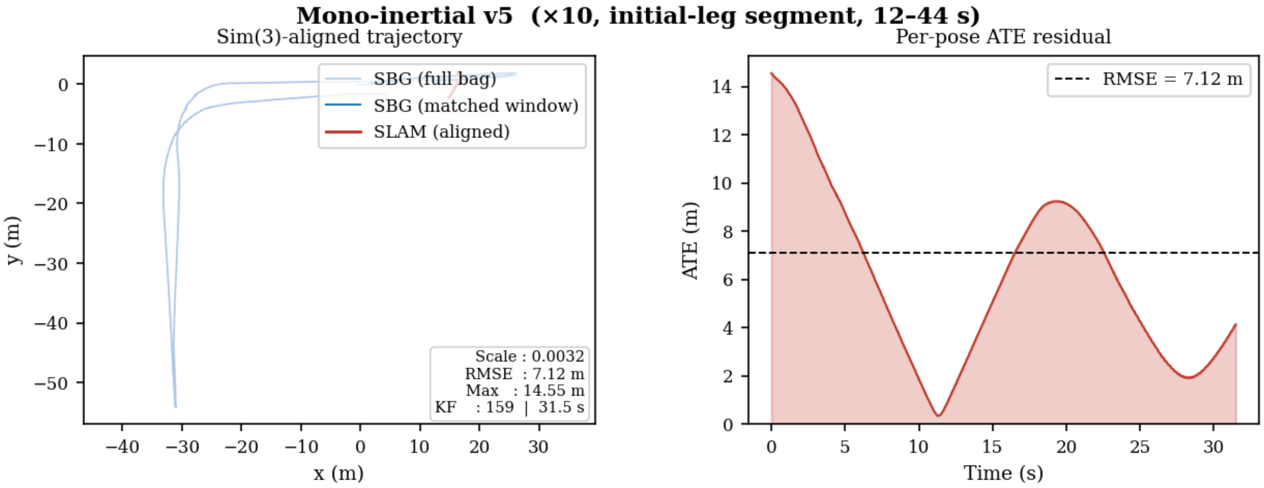


Figure 3.8: v5 trajectory (12–44 s segment, ×10): Sim(3)-aligned path (left) and ATE residual (right).

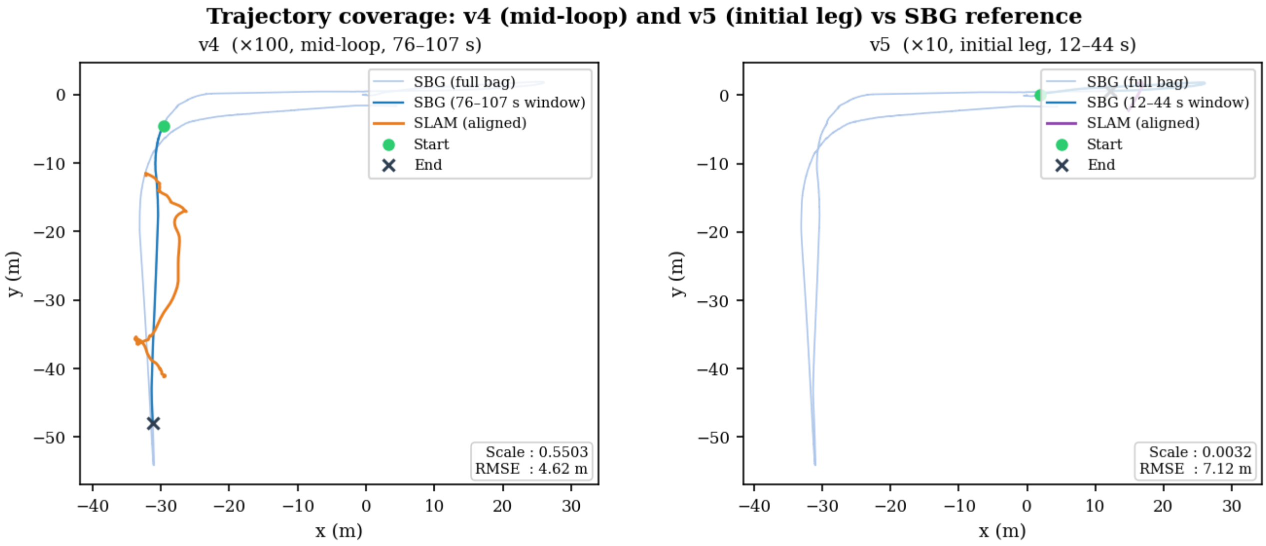


Figure 3.9: Coverage windows of v4 (mid-loop, left) and v5 (initial leg, right) against the SBG reference.

Table 3.4: ATE results for all configurations (Sim(3)-aligned to SBG EKF reference [5]). Scale near unity indicates metric recovery; values below 0.03 indicate collapse.

Configuration	Window	KF	Scale	RMSE (m)	Max (m)
Monocular-only	105.6 s	346	0.4167	15.46	33.63
Mono-inertial v1 (×100 inflation, full)	54.5 s	110	0.0118	4.73	9.15
Mono-inertial v1 (pre-VIBA window)	12 s	30	0.3197	3.95	≈7
Mono-inertial v2 (×100 inflation, slow-rate replay)	31.7 s	165	0.0007	7.01	14.52
Mono-inertial v3 (×100 inflation, long run)	169.0 s	671	0.0051	12.98	30.72
Mono-inertial v4 (×100 inflation, mid-loop straight)	31.4 s	160	0.5508	4.62	8.46
Mono-inertial v5 (×10 inflation, initial leg)	31.4 s	159	0.0032	7.12	14.51
Reproducibility sample (same ×100 configuration as v3):					
Mono-inertial v6 (×100, replicate of v3)	170.6 s	658	0.0030	16.94	37.30

Table 3.5: Segment-matched ablation: v3 (×100) windowed to the v4 and v5 segments for comparison on identical geometry.

Configuration	Window	KF	Scale	RMSE (m)	Max (m)
v3 windowed to v4 segment (76–107 s, ×100)	31.2 s	136	0.0038	1.81	4.32
v3 windowed to v5 segment (12–44 s, ×100)	**31.0 s**	**124**	**0.1373**	**3.59**	**12.15**
v5 on same segment (×10 ablation)	31.4 s	159	0.0032	7.12	14.51

Across campus-loop configurations, inertial fusion reduces ATE RMSE by approximately 3× relative to monocular-only (15.46 m to 4.73 m, v1), yet scale collapses to 0.0118, well below unity. The pre-VIBA window of v1 (first 30 KF, 12 s) retains scale 0.3197 and RMSE 3.95 m, indicating that scale observability degrades progressively as visual-inertial bundle adjustment replaces the accumulated inertial constraint. The longest runs, v3 (169.0 s, RMSE 12.98 m) and its replicate v6 (RMSE 16.94 m), sustain scale below 0.006. All seven campus-loop ×100 trials yield scale below 0.03, a threshold below which the Sim(3) fit absorbs metric error into the recovered scale factor; RMSE for these runs therefore reflects both geometric drift and scale collapse (Table 3.4). Reduced replay rate (v2, 0.5×) does not improve accuracy: scale collapsed to 0.0007 (the lowest of any configuration) and RMSE was 7.01 m despite a shorter evaluation window, confirming that lower apparent speed degrades rather than assists inertial initialisation.

Scale recovery is strongly segment-dependent. v4, evaluated on a 31 s mid-loop straight (76–107 s), recovers scale 0.5508 and RMSE 4.62 m (Figure 3.7), the highest scale of any ×100 trial. The segment-matched ablation (Table 3.5) isolates the geometry effect: v3 windowed to the same segment yields scale 0.0038 and RMSE 1.81 m, confirming that the v4 result is attributable to its favourable initialisation geometry and not to a configuration difference. The inflation-factor ablation compares v3 windowed (×100, initial-leg 12–44 s, RMSE 3.59 m) with v5 (×10, identical segment, RMSE 7.12 m): reducing inflation by 10× approximately doubles trajectory error on this segment, despite ×10 lying within the 3–10× range recommended by Campos et al. [1] for EuRoC sequences.

3.6 Reproducibility Across Trials

Five additional trials (v6–v10) replayed the campus recording through the deployed ×100 configuration to characterise run-to-run variability, since ORB-SLAM3 uses multi-threaded unseeded RANSAC and is therefore non-deterministic across trials on identical input. Individual trajectory plots and the complete per-trial ATE data are provided in Appendix C; key outcomes are summarised here.

Across all seven ×100-configuration trials (v1, v3, v6–v10), no trial achieved complete loop closure (180 s). The two longest-coverage runs, v3 (169.0 s) and v6 (170.6 s), each reached approximately 94% of the campus loop before terminating and differ by ≈4 m RMSE (12.98 m versus 16.94 m) on identical input; the remaining five trials terminated between 50 and 85 s with no consistent stopping point. This spread is consistent with thread-interleaving in the multi-threaded MAP estimator altering keyframe selection on each run. The ≈4 m replicate spread across the two longest runs establishes that single-trial ATE values are not stable enough to characterise system performance.

3.7 Robustness on Alternative Dataset

A second recording (a predominantly straight drive with a U-turn and reverse traverse) was evaluated with the deployed YAML and `mono_imu_v2` binary unchanged to probe robustness outside the training route. The recording was made on a different day under different lighting; the route provides significantly less rotational excitation than the campus loop. This constitutes a robustness probe rather than a controlled ablation, since route, day, and lighting all differ simultaneously. Per-trial outcomes are in Table 3.6.

Table 3.6: Straight-drive robustness trials ($n = 5$, ×100 configuration). Active-map switches are Atlas reinitialisation events.

Run	KF	Coverage	VIBA passes	Active-map switches
t1	181	37 s	2	15
t2	168	35 s	1	16
t3	165	34 s	0	17
t4	-	Crash	-	- (segfault mid-run)
t5	156	32 s	2	15

Four of five trials terminated within 32–37 s; one segfaulted mid-run following a relocalisation event. Mean coverage was 34.5 s ($\sigma = 2.1$ s), compared with 90.4 s ($\sigma \approx 55$ s) on the campus loop (Figure 3.10). The markedly narrow straight-drive distribution ($\sigma = 2.1$ s versus $\sigma \approx 50$ s) is inconsistent with stochastic initialisation failure and instead indicates a near-deterministic failure mode driven by insufficient feature parallax during extended forward translation. Each trial also produced 15–17 active-map resets, reflecting repeated Atlas reinitialisation events that the system cannot recover from on this motion profile.

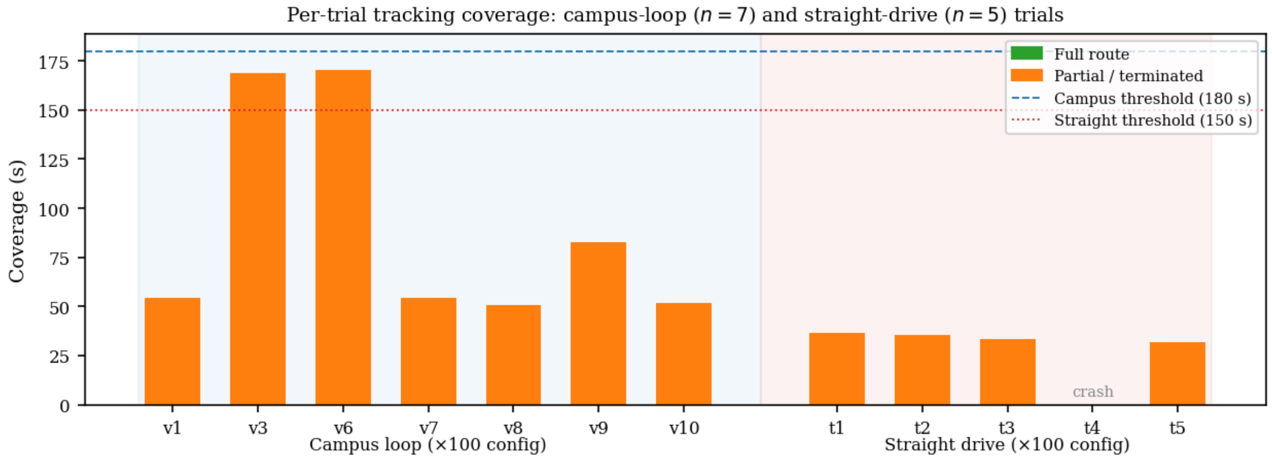


Figure 3.10: Per-trial tracking coverage: campus-loop ($n = 7$) and straight-drive ($n = 5$) trials, both at $\times 100$ config. All bars are partial (orange); no trial reached either full-route threshold. Dashed lines mark full-route durations: 180 s (campus loop) and 150 s (straight drive). "crash" denotes a mid-run segfault.

3.8 Computational Performance on the Jetson AGX Orin

Resource utilisation was logged via `tegrastats` across three independent campus-loop trials run at 1× (real-time) replay rate, yielding 1,147 samples over 574 s. The three trials were run independently of the v1–v10 ablation series: each replayed the principal campus-loop recording with the deployed monocular-inertial configuration. The measurement boundary encompasses the full deployed software stack: the ORB-SLAM3 `mono_imu` node, ROS 2 bag replay, DDS middleware, and Pangolin visualisation (Table 3.7).

Table 3.7: System resource utilisation: `tegrastats`, 1,147 samples over three independent campus-loop trials at 1× replay rate.

Metric	**Mean**	**Peak**
CPU utilisation (12-core avg)	30.1%	66%
GPU utilisation (`GR3D_FREQ`)	35.2%	92%
RAM usage	8.0 GB	8.9 GB
CPU power	3.77 W	7.65 W
GPU + SoC power	3.32 W	4.02 W

Mean CPU and GPU utilisation were 30.1% and 35.2% respectively; GPU peak reached 92% during bundle-adjustment bursts. Total power draw remained below 7.7 W against a 15 W TDP, and RAM was 8.0 GB mean (8.9 GB peak). The Jetson AGX Orin is not the performance bottleneck for the deployed workload; tracking failures observed in Sections 3.6 and 3.7 are therefore of algorithmic rather than computational origin.

4. Discussion

The results of Chapter 3 are evaluated against the 1.0 m ATE RMSE criterion and the research hypothesis of Section 1.4.

Hypothesis evaluation

The strong-form hypothesis is not supported. No campus-loop trial achieved complete loop closure; the two longest runs, v3 (169.0 s) and v6 (170.6 s), each reached approximately 94% of the 180 s route before tracking was lost. The best extended-coverage ATE RMSE is 12.98 m (v3), exceeding the 1.0 m criterion by more than an order of magnitude. The best segment-restricted result, 4.62 m (v4, 31 s mid-loop), misses the criterion by 4.6× and, as the segment-matched ablation demonstrates, is attributable to the initialisation geometry of that specific segment rather than to reliable metric recovery.

Scale collapse and IMU inflation

Sim(3) scale factors below 0.03 across all campus-loop ×100 trials indicate near-total loss of the inertial metric anchor (Tables 3.4 and C.1). A recovered scale of 0.003 maps one metre in SLAM-internal coordinates to approximately 330 m in the reference frame; plotted trajectories appear geometrically plausible because Sim(3) alignment absorbs this discrepancy into the scalar, not the shape. Without ×100 white-noise inflation, tracking fails to initialise on this platform. Two non-exclusive explanations are consistent with the data. First, configuration-level inflation reduces the inertial residual weight in the optimiser by 100×, directly degrading scale observability (a structural consequence of the parameter choice). Second, static AVAR characterisation from a bench recording may not capture the in-motion vibration spectrum at the roof mounting; the inflation then functions as a necessary model-mismatch correction rather than a deliberate accuracy trade-off. In both cases the ×100 inflation is a symptom of an underlying constraint and is the principal factor distinguishing this deployment from the conditions under which ORB-SLAM3 achieves centimetre-level RMSE [1].

Reproducibility and stability

Of seven ×100 trials on identical input, no trial completed the full 180 s campus loop. The two longest runs, v3 (169.0 s) and v6 (170.6 s), differ by ≈4 m RMSE despite covering near-identical segments of the same recording. Partial trials terminate without a consistent stopping point, consistent with the non-deterministic frame selection that arises from thread interleaving in the multi-threaded RANSAC initialiser. The ≈4 m replicate spread means that individual ATE figures are not stable enough to characterise system performance; reliable deployment would require an external monitoring or recovery mechanism to detect and restart failed tracking sessions.

Motion profile sensitivity

The straight-drive coverage distribution ($\sigma = 2.1$ s, $n = 4$) is far narrower than the campus-loop spread ($\sigma \approx 50$ s, $n = 7$), indicating a near-deterministic failure mode rather than stochastic initialisation noise. Each straight-drive trial produced 15–17 active-map resets (Table 3.6), reflecting repeated Atlas reinitialisation driven by insufficient feature parallax during extended forward translation. Importantly, the campus loop (which provides more rotational excitation than typical shuttle operation) produced no complete-loop runs; even the two longest campus-loop trials terminated approximately 11 s before completing the route. The operational motion profile is adversarial for monocular vSLAM scale recovery independently of route familiarity. This behaviour is consistent with the known degeneracy of monocular-inertial systems under low-parallax forward motion [10].

Scope of conclusions

Four bounds apply to the conclusions drawn from these results. The SBG EKF reference fuses GNSS with the onboard IMU (Section 2.7.1); reported errors therefore reflect agreement with a deployment-grade reference, not independent absolute ground truth. The evaluation spans a single campus-loop recording and one straight-drive segment, and generalisation to other routes is not established. Fisheye pre-rectification crops the peripheral field of view, potentially understating performance achievable with native-projection fisheye models. Finally, Sim(3) alignment absorbs scale error into a scalar factor; the sub-0.03 scale value is the primary failure indicator, not visual trajectory agreement.

Comparison with prior work

The 12–17 m extended-coverage RMSE reflects three compounding factors absent from standard benchmarks: a monocular-only sensor configuration, extended forward-motion segments that preclude reliable scale initialisation, and the AVAR mismatch requiring the $\times 100$ inflation. Campos et al. [1] report centimetre-level RMSE on EuRoC under stereo-inertial conditions with physically characterised IMUs, conditions substantially more favourable than those present here. Paul et al. [11] report 1.94 m ATE RMSE for monocular ORB-SLAM in a simulated parking-lot environment, approximately $7\times$ below this study's best extended-coverage result; however, that evaluation uses synthetic imagery, no physical IMU noise, and a short controlled trajectory, making direct comparison unreliable. Their reported $\approx$87% ATE reduction from monocular to stereo under ideal simulation conditions represents an optimistic upper bound on the improvement achievable through stereo migration under the constraints of the nUWAy4 platform.

Design implications

Two independent root causes are identified. A calibration-deployment gap exists because the static AVAR noise model derived from a stationary bench recording does not capture the in-motion vibration spectrum at the roof mounting; the $\times 100$ inflation is the operational consequence. An algorithmic–platform mismatch compounds this: monocular-inertial scale initialisation requires rotational excitation and non-planar parallax that straight-line shuttle transit does not consistently provide. These causes are orthogonal, and addressing either in isolation is insufficient to reach the 1.0 m criterion.

In order of expected impact, three remediations are identified. Migration to stereo-inertial operation simultaneously resolves both scale collapse and forward-motion degeneracy by providing a fixed physical baseline that makes metric depth observable from image geometry alone, independent of IMU input. Adoption of native Kannala-Brandt fisheye projection within the SLAM back-end would preserve the peripheral features discarded by the current OCamCalib rectification crop, increasing feature density during straight segments. Incorporation of a brief rotational excitation manoeuvre at startup (such as a steering sweep before entering service) would provide the parallax and acceleration needed to constrain the gravity direction and velocity state before straight-line operation begins. Joint camera-IMU calibration is addressed in Section 5.2.

5. Conclusions and Future Work

5.1 Conclusions

This project evaluated whether a monocular visual-inertial ORB-SLAM3 pipeline, integrated on the nUWAy4 autonomous shuttle bus and running on an NVIDIA Jetson AGX Orin, could meet a pre-specified 1.0 m ATE RMSE localisation criterion suitable for substituting the vehicle's LiDAR Nav2 baseline. The four project objectives were addressed in full: the pipeline was integrated on ROS 2 with synchronised GMSL camera and SBG IMU inputs (O1); Allan-variance characterisation and frame-convention correction were completed (O2); trajectory accuracy was quantified across twelve trials including a segment-matched ablation (O3); and computational performance was measured across three independent runs (O4).

The strong-form hypothesis is not supported. Best extended-coverage ATE RMSE was 12.98 m (v3, 169 s), exceeding the 1.0 m criterion by a factor of 13. No campus-loop trial achieved complete loop closure; the two longest runs terminated approximately 11 s before completing the 180 s route. Sim(3) scale factors as low as 0.0030 indicate that the inertial metric anchor was not reliably established under the deployed conditions. All five straight-drive trials terminated within 37 s, consistent with a near-deterministic failure mode driven by insufficient feature parallax during extended forward translation. These results place monocular vSLAM substantially below the performance envelope required for production autonomous shuttle localisation in this configuration.

Two independent root causes are identified. First, a calibration-deployment gap: the static AVAR noise model required $\times 100$ inflation to sustain tracking under on-vehicle vibration, degrading scale observability as a direct consequence of the inflated inertial residual weighting. Second, an algorithmic-platform mismatch: monocular-inertial scale initialisation requires rotational excitation and non-planar parallax that the shuttle's predominantly straight-line route does not consistently provide. These causes are orthogonal; addressing either in isolation is insufficient. This analysis extends the findings of Paul et al. [11], who report a $7\times$ improvement from monocular to stereo under idealised simulation conditions, by identifying the specific physical mechanisms that must be resolved on a real deployed platform.

The computational platform is not the limiting factor. The Jetson AGX Orin sustained real-time processing at 30.1% mean CPU, 35.2% mean GPU, and $\approx$7 W, confirming that the embedded hardware presents substantial headroom for more complex visual-inertial pipelines including stereo and native fisheye configurations. This finding is significant: it establishes that the primary barrier to deployment is sensor-algorithm compatibility rather than compute capacity, and that the same hardware is capable of supporting the next-generation configuration identified as necessary.

5.2 Future Work

Future investigations are structured in order of expected impact on the identified root causes.

Stereo-inertial migration. The highest-priority intervention is migration to a stereo-inertial configuration. A fixed stereo baseline makes metric scale and depth directly observable from visual geometry alone, independently of motion profile, thereby eliminating both identified root causes simultaneously. The nUWAy4 platform has available mounting locations for a second GMSL2 camera, making this a hardware-tractable upgrade. Campos et al. [1] report that ORB-SLAM3 in stereo-inertial mode achieves RMSE below 4 cm on EuRoC, compared to the decimetre-level errors observed in monocular-inertial mode; the transition on this platform would be expected to yield a commensurate improvement, removing the scale-recovery dependency on rotational excitation.

Native fisheye projection. Simultaneously, the OCamCalib pre-rectification pipeline should be replaced with a native Kannala-Brandt fisheye projection model within the ORB-SLAM3 back-end. The current rectification crop reduces the effective field of view and discards peripheral features that are particularly valuable during straight-line segments. Native fisheye support is implemented in ORB-SLAM3 and requires only a recalibration and YAML update; the expected benefit is increased feature density and more stable tracking on forward-motion-dominated segments.

In-motion IMU calibration. The $\times 100$ inflation factor required to sustain tracking indicates a systematic mismatch between the bench AVAR characterisation and the in-motion vibration spectrum at the roof mounting. A dedicated in-motion calibration procedure, capturing IMU data during normal shuttle operation and refitting the AVAR model, would quantify whether the inflation can be reduced. Reducing the inflation factor directly improves scale observability, as demonstrated by the segment-matched ablation comparing $\times 100$ and $\times 10$ configurations.

Environmental robustness evaluation. All trials in this project were conducted under favourable daytime conditions. A systematic evaluation across the full range of environmental conditions encountered on the UWA campus route is a necessary precondition for operational deployment. This should include low-light and dawn conditions, direct sun glare from the western corridor, heavy shadows under dense tree canopy, and wet-pavement surface reflections. Camera-based SLAM systems are known to degrade significantly under these conditions [10], and demonstrating acceptable tracking stability across this range is required before monocular vSLAM can be considered a viable replacement for the LiDAR baseline.

Longer-term priorities. Joint online temporal calibration to resolve camera-IMU clock offsets under in-motion vibration, a deterministic startup manoeuvre to pre-condition scale initialisation before entering straight-line service, and multi-session map-merging trials over a statistically sufficient run count to characterise route-level localisation repeatability are identified as longer-term research priorities once the primary sensor configuration has been validated.

References

[1] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multi-map SLAM. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021. doi: 10.1109/TRO.2021.3075644.

[2] Raúl Mur-Artal, José M. M. Montiel, and Juan D. Tardós. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015. doi: 10.1109/TRO.2015.2463671.

[3] Miroslav Matejček and Mikuláš Šostronek. New experience with Allan variance: Noise analysis of accelerometers. In *2017 Communication and Information Technologies (KIT)*, pages 1–4. IEEE, 2017. doi: 10.23919/KIT.2017.8109457.

[4] Tully Foote. REP-103: Standard units of measure and coordinate conventions. ROS Enhancement Proposal, 2010. Available: `https://www.ros.org/reps/rep-0103.html`.

[5] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991. doi: 10.1109/34.88573.

[6] Peide Wang. Research on comparison of LiDAR and camera in autonomous driving. In *Journal of Physics: Conference Series*, volume 2093, page 012032, 2021. doi: 10.1088/1742-6596/2093/1/012032.

[7] Yingwei Li, Adams Wei Yu, Tianjian Meng, Ben Caine, Jiquan Ngiam, Daiyi Peng, Junyang Shen, Yifeng Lu, Denny Zhou, Quoc V. Le, Alan Yuille, and Mingxing Tan. DeepFusion: Lidar-Camera deep fusion for multi-modal 3D object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17182–17191, 2022.

[8] Jun Zhu, Hongyi Li, and Tao Zhang. Camera, LiDAR, and IMU based multi-sensor fusion SLAM: A survey. *Tsinghua Science and Technology*, 29(2):415–429, 2024. doi: 10.26599/TST.2023.9010010.

[9] Mohammad O. A. Aqel, Mohammad H. Marhaban, M. Iqbal Saripan, and Napsiah Bt. Ismail. Review of visual odometry: Types, approaches, challenges, and applications. *SpringerPlus*, 5(1):1897, 2016. doi: 10.1186/s40064-016-3573-7.

[10] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, Jose Neira, Ian

Reid, and John J. Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016. doi: 10.1109/TRO.2016.2624754.

[11] Sayandip Paul, C. Hemanth Kumar, and Arunkumar Bongale. Enhancing autonomous navigation: A visual SLAM approach. *Journal of Physics: Conference Series*, 2748(1):012008, 2024. doi: 10.1088/1742-6596/2748/1/012008.

[12] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. In *2011 IEEE International Conference on Computer Vision (ICCV)*, pages 2564–2571. IEEE, 2011. doi: 10.1109/ICCV.2011.6126544.

[13] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *European Conference on Computer Vision (ECCV 2006)*, pages 430–443. Springer, 2006. doi: 10.1007/11744023_34.

[14] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008. doi: 10.1016/j.cviu.2007.09.014.

[15] Alexey Merzlyakov and Steve Macenski. A comparison of modern general-purpose visual SLAM approaches. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9190–9197. IEEE, 2021. doi: 10.1109/IROS51168.2021.9635827.

[16] Yunpeng Li, Noah Snavely, and Daniel Huttenlocher. Location recognition using prioritized feature matching. In *European Conference on Computer Vision (ECCV 2010)*, pages 791–804. Springer, 2010. doi: 10.1007/978-3-642-15555-0_57.

[17] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment — A modern synthesis. In Bill Triggs, Andrew Zisserman, and Richard Szeliski, editors, *Vision Algorithms: Theory and Practice*, pages 298–372. Springer, 2000. doi: 10.1007/3-540-44480-7_21.

[18] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33(1):1–21, 2017. doi: 10.1109/TRO.2016.2597321.

[19] Tong Qin, Peiliang Li, and Shaojie Shen. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018.

[20] Juho Kannala and Sami S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340, 2006. doi: 10.1109/TPAMI.2006.153.

[21] Steve Macenski, Francisco Martin, Ruffin White, and Jonatan Ginés Clávero. The Marathon 2: A navigation system. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[22] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 573–580, 2012.

A. Full Allan Variance Results

The complete Allan Variance characterisation output for the SBG IMU, as computed by the `allan_variance_ros2` tool from a static recording on the nUWAy4 shuttle bus, is reproduced below.

ACCELEROMETER

```
X Velocity Random Walk:  0.00066 m/s/sqrt(s)  0.03932 m/s/sqrt(hr)
Y Velocity Random Walk:  0.00066 m/s/sqrt(s)  0.03946 m/s/sqrt(hr)
Z Velocity Random Walk:  0.00060 m/s/sqrt(s)  0.03595 m/s/sqrt(hr)
X Bias Instability:      0.00010 m/s^2  1339.587 m/hr^2
Y Bias Instability:      0.00011 m/s^2  1470.735 m/hr^2
Z Bias Instability:      0.00010 m/s^2  1263.960 m/hr^2
X Accel Random Walk:     0.00004 m/s^2/sqrt(s)
Y Accel Random Walk:     0.00004 m/s^2/sqrt(s)
Z Accel Random Walk:     0.00004 m/s^2/sqrt(s)
```

GYROSCOPE

```
X Angle Random Walk:  0.00468 deg/sqrt(s)  0.28084 deg/sqrt(hr)
Y Angle Random Walk:  0.00371 deg/sqrt(s)  0.22257 deg/sqrt(hr)
Z Angle Random Walk:  0.00402 deg/sqrt(s)  0.24145 deg/sqrt(hr)
X Bias Instability:   0.00158 deg/s  5.68962 deg/hr
Y Bias Instability:   0.00158 deg/s  5.69579 deg/hr
Z Bias Instability:   0.00142 deg/s  5.10503 deg/hr
X Rate Random Walk:   0.00055 deg/s/sqrt(s)
Y Rate Random Walk:   0.00050 deg/s/sqrt(s)
Z Rate Random Walk:   0.00047 deg/s/sqrt(s)
```

Deployed ORB-SLAM3 IMU YAML values (white-noise terms inflated ×100, random-walk terms inflated ×10): the ×100 figure is above the 3-10× range recommended by Campos et al. [1] on EuRoC sequences, but the segment-matched v5 ablation reported in Section 3.5 shows that reducing it to ×10 approximately doubles the ATE on the same 31 s segment of the campus loop. The deployed values are therefore retained as the empirically-correct configuration for the nUWAy4 mounting geometry.

```
IMU.NoiseGyro: 8.17e-04   # rad/s/sqrt(Hz)
IMU.NoiseAcc:  6.58e-03   # m/s^2/sqrt(Hz)
```

```
IMU.GyroWalk:  9.6e-05    # rad/s^2/sqrt(Hz)
IMU.AccWalk:   4.5e-04    # m/s^3/sqrt(Hz)
IMU.Frequency: 250.0
```

B. ORB-SLAM3 Configuration YAML

The full ORB-SLAM3 configuration file used in the experiments of Chapter 3 is reproduced below. It was saved on the Jetson AGX Orin at `colcon_ws/src/orbslam3_ros2/config/gmsl_front_pinhole_imu_mo`

```
%YAML:1.0
# Production mono-inertial config for nUWAy4 GMSL front camera + SBG Ellipse-D.
# Lever arm restored to 1.8 m (vehicle X) per rig measurement. fx=321.5625
# matches the OCamCalib-rectified pinhole built in mono_imu_v2.cpp.
Camera.type: "PinHole"
Camera.width: 1280
Camera.height: 720
Camera.fx: 321.5625
Camera.fy: 321.5625
Camera.cx: 640.0
Camera.cy: 360.0
Camera.k1: 0.0
Camera.k2: 0.0
Camera.k3: 0.0
Camera.p1: 0.0
Camera.p2: 0.0
OCamCalib.sensor_width: 1920
OCamCalib.sensor_height: 1080
OCamCalib.a0: 321.5625
OCamCalib.a2: -0.0008
OCamCalib.a3: 0.0
OCamCalib.a4: 0.0
OCamCalib.cx: 983.7693
OCamCalib.cy: 521.1453
OCamCalib.c: 0.998
OCamCalib.d: 0.0426
OCamCalib.e: -0.0428
Camera.fps: 30.0
Camera.RGB: 0
Camera.bf: 0.0
ORBextractor.nFeatures: 3000
ORBextractor.scaleFactor: 1.2
```

```
ORBextractor.nLevels: 8
ORBextractor.iniThFAST: 8
ORBextractor.minThFAST: 3
IMU.NoiseGyro: 8.17e-03
IMU.NoiseAcc: 6.58e-02
IMU.GyroWalk: 9.6e-04
IMU.AccWalk: 4.5e-03
IMU.Frequency: 250
# Tbc maps camera optical frame (X=right, Y=down, Z=fwd) into FLU body frame
# (X=fwd, Y=left, Z=up). Translation column is camera position in body frame:
# 1.8 m forward of IMU along vehicle X, no lateral or vertical offset assumed.
# The matching FRD->FLU axis flip on each IMU sample is in mono_imu_v2.cpp;
# do not edit one in isolation from the other.
Tbc: !!opencv-matrix
  rows: 4
  cols: 4
  dt: f
  data: [ 0.0,  0.0, 1.0, 1.8,
         -1.0,  0.0, 0.0, 0.0,
          0.0, -1.0, 0.0, 0.0,
          0.0,  0.0, 0.0, 1.0]
Viewer.KeyFrameSize: 0.05
Viewer.KeyFrameLineWidth: 1
Viewer.GraphLineWidth: 1
Viewer.PointSize: 2
Viewer.CameraSize: 0.08
Viewer.CameraLineWidth: 3
Viewer.ViewpointX: 0.0
Viewer.ViewpointY: -0.7
Viewer.ViewpointZ: -1.8
Viewer.ViewpointF: 500.0
```

C. Reproducibility Trial Details

Five independent replays of the campus-loop bag through the deployed ×100 configuration (v6–v10) are presented below. All trajectories are Sim(3)-aligned to the SBG Ellipse-D EKF reference using the Umeyama least-squares method [5]. Table C.1 summarises the per-trial ATE metrics; Figure C.1 provides an overview of all five aligned paths and ATE residuals; Figures C.2–C.6 show each trial individually.

Table C.1: Reproducibility trials v6–v10 (×100 configuration). No trial reached the 180 s full-loop threshold; v6 (170.6 s) is the longest-coverage run.

Run	**Window (s)**	**KF**	**Scale**	**RMSE (m)**	**Max (m)**	**Coverage**
v6	170.6	658	0.0030	16.94	37.30	Near-complete
v7	54.1	247	0.1055	7.14	15.04	Partial
v8	50.8	252	0.0189	9.73	19.82	Partial
v9	82.5	331	0.0045	9.43	21.74	Partial
v10	51.5	249	0.0279	6.06	14.35	Partial

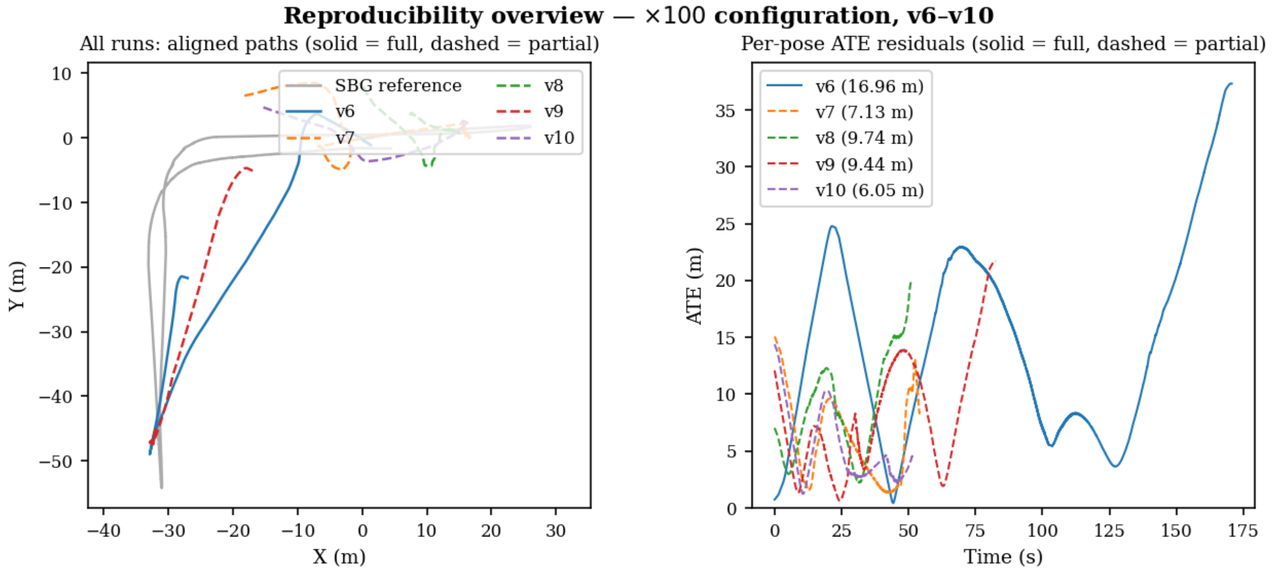


Figure C.1: Overview of reproducibility trials v6–v10 (×100 configuration): all Sim(3)-aligned paths overlaid on the SBG reference (left) and per-pose ATE residuals for all five trials (right). Solid lines indicate near-complete coverage (≈170 s); dashed lines indicate partial coverage.

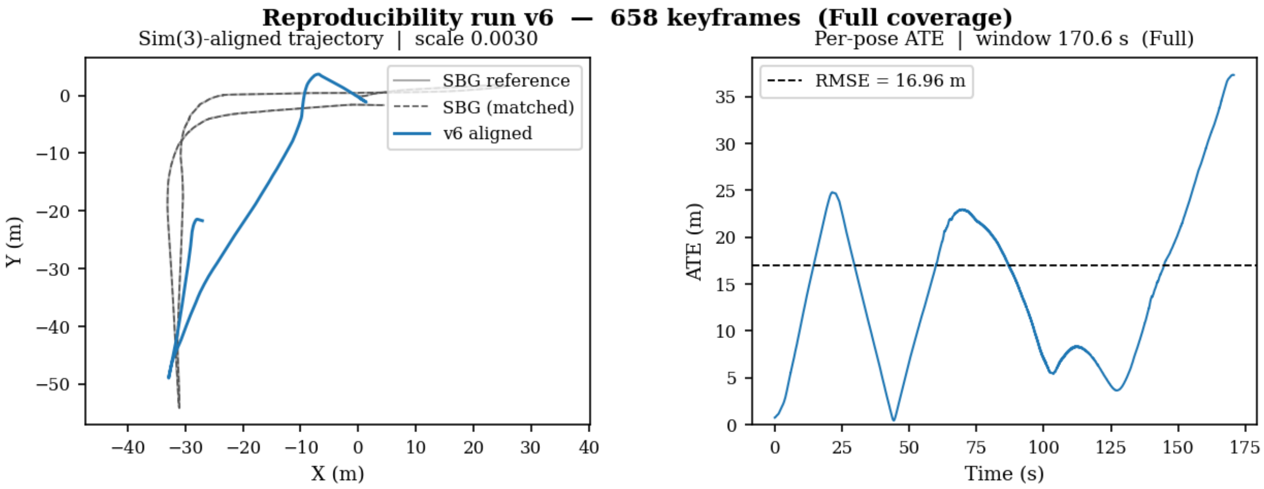


Figure C.2: v6 trajectory (170.6 s, near-complete coverage, ×100): Sim(3)-aligned path (left) and ATE residual over time (right). Scale factor 0.0030 indicates metric scale was not recovered; RMSE 16.94 m.

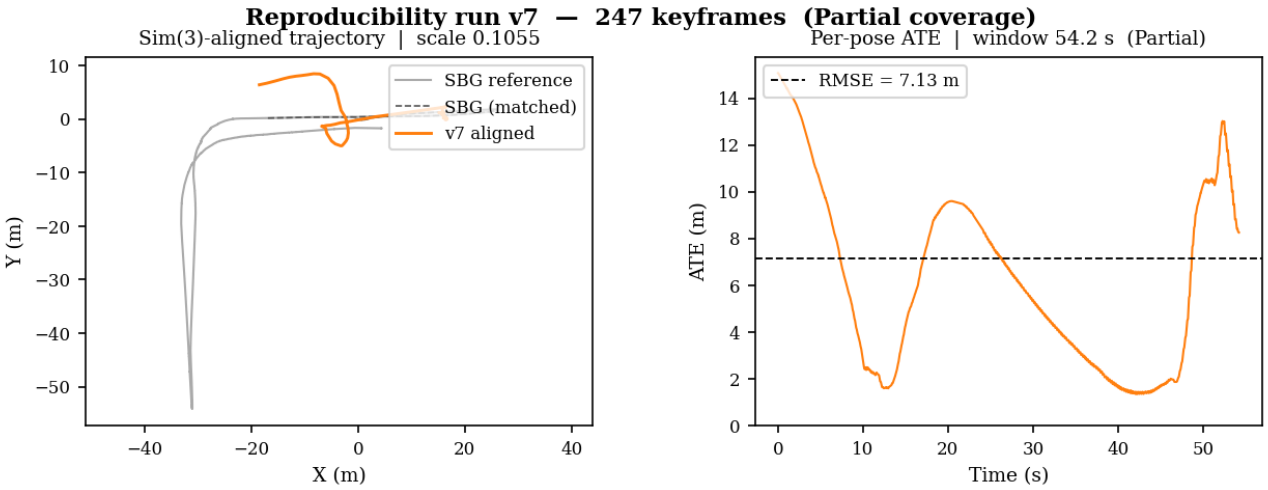


Figure C.3: v7 trajectory (54.1 s, partial, ×100): Sim(3)-aligned path (left) and ATE residual over time (right). Scale factor 0.1055; RMSE 7.14 m.

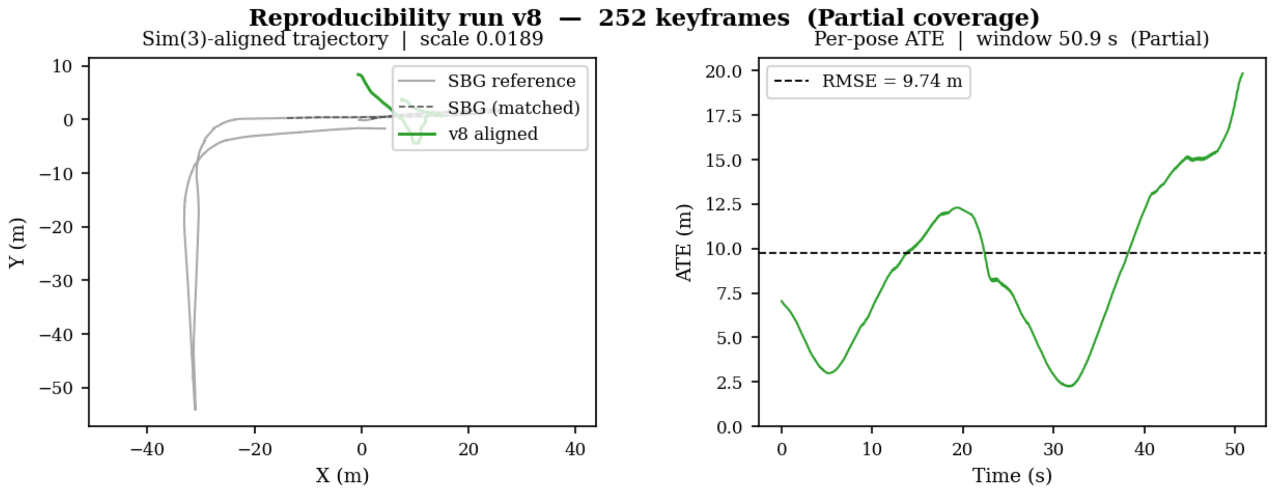


Figure C.4: v8 trajectory (50.8 s, partial, ×100): Sim(3)-aligned path (left) and ATE residual over time (right). Scale factor 0.0189; RMSE 9.73 m.

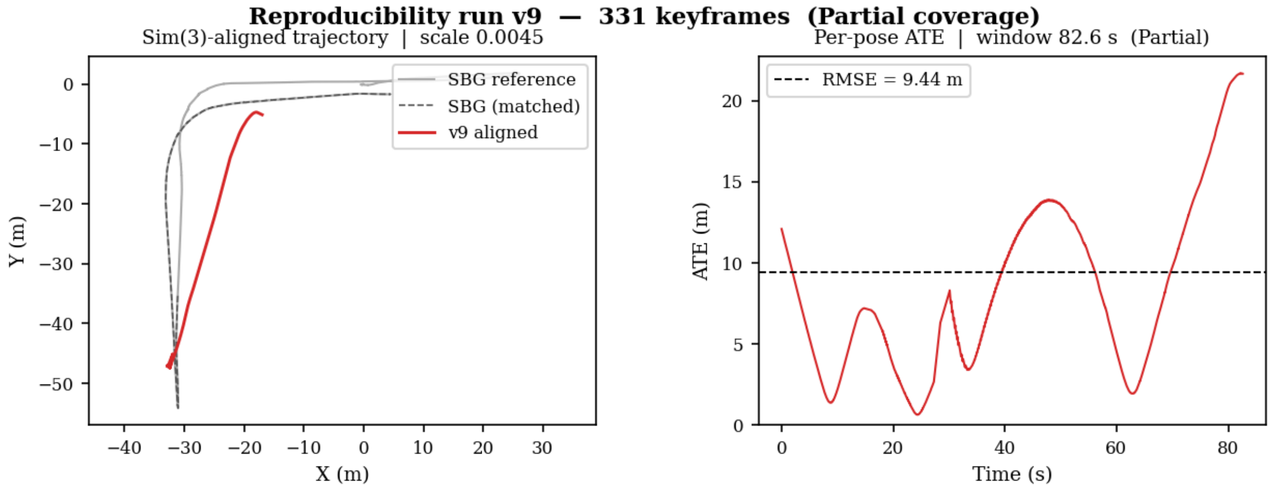


Figure C.5: v9 trajectory (82.5 s, partial, ×100): Sim(3)-aligned path (left) and ATE residual over time (right). Scale factor 0.0045; RMSE 9.43 m.

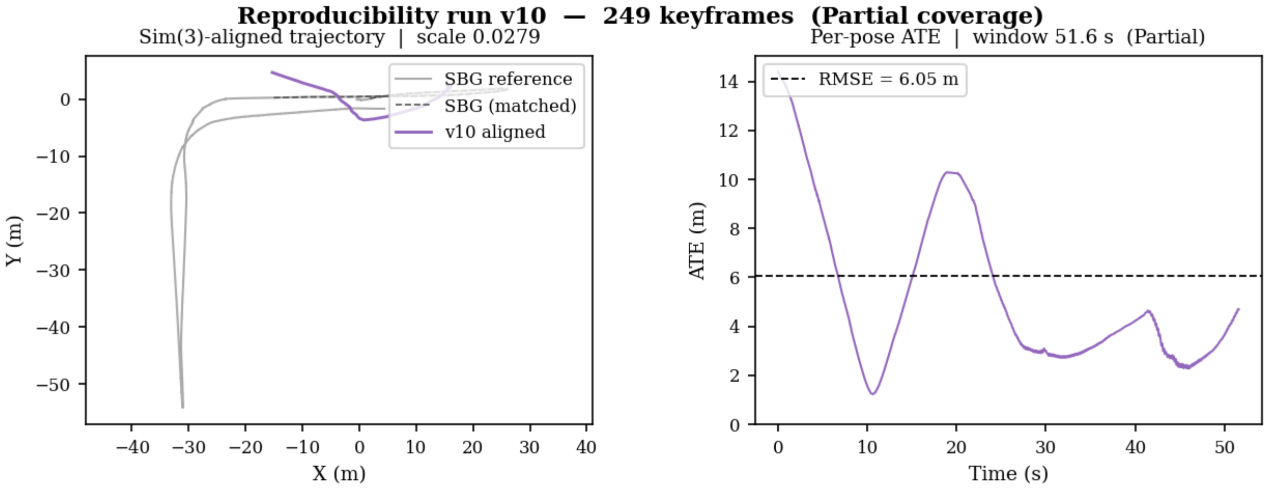


Figure C.6: v10 trajectory (51.5 s, partial, ×100): Sim(3)-aligned path (left) and ATE residual over time (right). Scale factor 0.0279; RMSE 6.06 m.