Monocular ORB-SLAM3 Evaluation for Multi-Altitude VTOL UAV Mapping

Ming-Jyun Yang [1], Jyun-Ping Jhan [2], Ru-Shan Tang [3]

[1] Graduate Institute of A.I. Cross-disciplinary Technology, NTUST, Taiwan – justbeyourself@mail.ntust.edu.tw
[2] Graduate Institute of A.I. Cross-disciplinary Technology, NTUST, Taiwan – jpjhan@mail.ntust.edu.tw
[3] Systems Development Center, NCSIST, Taiwan – randydalph@gmail.com

Keywords: UAV, Monocular ORB-SLAM3, Multi-Altitude Mapping, VTOL, Trajectory Alignment

Abstract

Reliable visual localization is essential for long-range VTOL UAV mapping in GNSS-degraded environments. This paper presents a quantitative evaluation framework for monocular ORB-SLAM3 using a 66.48 km multi-altitude UAV mission and aerial-triangulation-derived camera poses as reference data. The workflow associates SLAM and reference trajectories by image key, applies Sim(3)-based metric alignment, corrects coordinate-axis inconsistency, and refines attitude by a global rotation offset, enabling full-mission and segment-level comparison in a common metric frame. The evaluation covers four altitude segments, namely 100, 150, 200, and 250 m AGL, under three protocols: No-Loop (NL), With-Loop Global Slice (GS), and With-Loop Local Re-Sim(3) (LR). For the full mission, the proposed alignment achieves a 3D position RMSE of 7.41 m over 5330 matched frames and substantially reduces the geometric deformation observed in the S+T baseline. Segment-level results show a strong altitude dependency in the isolated NL runs, with 3D RMSE decreasing from 22.95 m at 100 m to 5.49 m at 250 m. Among the three protocols, LR consistently yields the best segment-level position accuracy, reaching 4.00, 8.26, 3.94, and 3.92 m at 100, 150, 200, and 250 m, respectively. Long-range analysis further shows that the trajectory remains globally bounded, while cumulative 3D endpoint drift increases from 0.35 m at 50 m to 10.66 m at 25.6 km. These results indicate that ORB-SLAM3 can support large-scale trajectory estimation for UAV mapping, but its evaluated quality depends strongly on alignment, segmentation, and evaluation strategy.

1. Introduction

Reliable localization is a fundamental requirement for unmanned aerial vehicle (UAV) applications in mapping, navigation, and autonomous inspection. This requirement becomes even more critical when Global Navigation Satellite System (GNSS) signals are degraded, intermittent, or unavailable. Under such conditions, vision-based motion estimation provides an important alternative because it can infer platform movement directly from onboard imagery without relying entirely on external positioning infrastructure. Among the available approaches, monocular simultaneous localization and mapping (SLAM) is particularly attractive due to its lightweight sensor requirement and its demonstrated effectiveness in visual trajectory estimation (Mur-Artal et al., 2015; Campos et al., 2021). However, when applied to long-range aerial missions, its performance remains difficult to interpret because scale ambiguity, accumulated drift, and trajectory deformation may become significant over extended flight distances. This issue is especially relevant for practical UAV mapping, where stable pose estimation is required not only for navigation continuity but also for maintaining geometric consistency throughout the image acquisition process.

1.1 Background

In aerial mapping scenarios, the above challenges become more pronounced when the mission includes changes in altitude, terrain texture, and viewing geometry. Variations in ground sampling distance (GSD), image footprint, and scene structure directly influence the quality and repeatability of image features, and therefore affect the stability of monocular SLAM estimation. As a result, a trajectory that appears globally reasonable may still contain non-negligible geometric distortion in particular flight segments. This issue is important in mapping-oriented applications because local inconsistency may influence the interpretation of trajectory quality even when the overall path remains visually plausible.
Although monocular SLAM has shown strong performance on established benchmarks and has become a standard solution in visual navigation research, its practical use in large-scale UAV mapping still requires careful evaluation. In particular, when the platform undergoes long continuous flight and experiences changing image scale, the final assessment may depend not only on the SLAM algorithm itself, but also on how the estimated trajectory is aligned to the reference data and how accuracy is interpreted across the mission. For this reason, the problem addressed in this study is not simply whether monocular ORB-SLAM3 can reconstruct a UAV trajectory, but rather how its trajectory quality should be understood in a realistic aerial mapping mission where mission scale, altitude variation, and reference-based evaluation are all involved.

1.2 Related work

Monocular visual SLAM has been extensively studied in robotics and computer vision, and has become one of the most representative approaches for camera-based motion estimation. Among existing systems, ORB-SLAM established a versatile monocular SLAM framework with automatic initialization, relocalization, and loop closing, and demonstrated strong accuracy across diverse benchmark sequences. ORB-SLAM3 further extended this framework to visual, visual-inertial, and multi-map SLAM, showing robust trajectory estimation capability under multiple sensor configurations and operating conditions. Owing to these characteristics, ORB-SLAM3 has become a widely adopted baseline for visual navigation research.

The performance of visual SLAM systems is commonly reported on public benchmarks such as KITTI and EuRoC (Geiger et al., 2012; Burri et al., 2016). These datasets provide standardized image sequences and reliable reference trajectories, and have therefore played an essential role in algorithm development and comparison. However, such benchmark-oriented evaluations do not fully represent the conditions encountered in practical UAV mapping missions, where a single continuous flight may involve substantial altitude variation, changing ground sampling distance, heterogeneous terrain texture, and large viewpoint transitions.

In addition, SLAM evaluation is often performed after a single global alignment between the estimated trajectory and the reference trajectory. While this practice is useful for measuring overall agreement, it does not always reveal how well local geometric structure is preserved after loop closure or global optimization. For long-range UAV mapping applications, this distinction is important because an apparently well-aligned global trajectory may still contain non-negligible deformation in individual flight segments. Therefore, a more explicit evaluation framework is required to distinguish overall consistency from segment-level geometric behaviour under realistic aerial mapping conditions.

1.3 Objective

Based on the above considerations, this study aims to establish a quantitative evaluation framework for monocular ORB-SLAM3 in a realistic long-range UAV mapping mission. Rather than proposing a new SLAM backend, this work focuses on how the estimated trajectory should be assessed when a high-precision photogrammetric reference is available.

The analysis is conducted on a continuous 66.48 km mission acquired under four flight altitudes, namely 250 m, 200 m, 150 m, and 100 m above ground level. This multi-altitude design enables the effects of scale variation, viewpoint change, and long-range drift to be examined within a single dataset. To provide a metric baseline for comparison, a reference trajectory derived from aerial triangulation is adopted as ground truth, and the refined interior orientation parameters obtained from the photogrammetric process are used to maintain geometric consistency between the reference workflow and the SLAM estimation.

On this basis, the study compares three evaluation conditions, namely No-Loop, With-Loop Global Slice, and With-Loop Local Re-Sim(3). These settings are intended to distinguish pure odometric behaviour, globally optimized trajectory consistency, and segment-level geometric quality. Through this design, the study seeks to clarify how flight altitude, repeated trajectory revisits, and evaluation protocol jointly affect the interpretation of monocular ORB-SLAM3 accuracy in large-scale UAV mapping applications.

2. System and Dataset

This section describes the experimental platform, dataset, and reference data used in this study. It first introduces the UAV platform and imaging system employed for data acquisition, then summarizes the flight mission and study area, and finally presents the photogrammetric workflow used to generate the reference trajectory. Since the subsequent evaluation relies on the comparison between SLAM-estimated poses and photogrammetric reference data, particular attention is given to the sensor configuration and reference-data preparation.

2.1 UAV Platform and Imaging System

The dataset used in this study was acquired using a Vertical Take-Off and Landing (VTOL) UAV platform configured for long-range aerial mapping. The onboard system included an NVIDIA Jetson Orin NX module serving as the central computing and data-logging unit, together with an industrial camera for image acquisition. To support subsequent post-processing, the flight controller and the onboard computer were synchronized so that image timestamps could be associated with the corresponding telemetry records. This hardware configuration provided the visual and navigation data used in both the SLAM processing and the photogrammetric reference workflow.
The imaging sensor was a FLIR Blackfly S industrial camera (model FFY-U3-16S2M-S) equipped with a monochrome global-shutter sensor. The camera operated at a resolution of 1440 × 1080 pixels, with a pixel size of 3.45 μm, a focal length of 1.8 mm, and an image acquisition rate of 16 Hz. Figure 1 shows the UAV platform and onboard imaging system used for data acquisition.

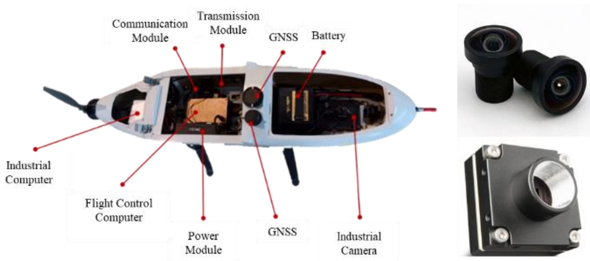

Figure 1. VTOL UAV and onboard imaging system.

2.2 Flight Mission and Study Area

The experimental mission was conducted in the Baguashan region, Taiwan. The study area includes fragmented urban grids, subtropical forest, and agricultural land, and exhibits local elevation differences exceeding 100 m. The full flight trajectory covered 66.48 km and was executed at a cruise speed of approximately 85 km/hr. As a result, the dataset contains substantial variation in terrain relief, land-cover texture, and image footprint within a single continuous flight mission.

To investigate the effect of observation scale under a consistent operational setting, the mission was designed with a stepped descent profile comprising four altitude segments: 250 m, 200 m, 150 m, and 100 m above ground level (AGL). These altitude changes were executed within the same sortie rather than through separate independent flights. As shown in Figure 2, this design introduced systematic variation in flight height along the mission while preserving overall mission continuity. Accordingly, the subsequent analysis could compare trajectory behaviour across altitude segments within a common mission framework.

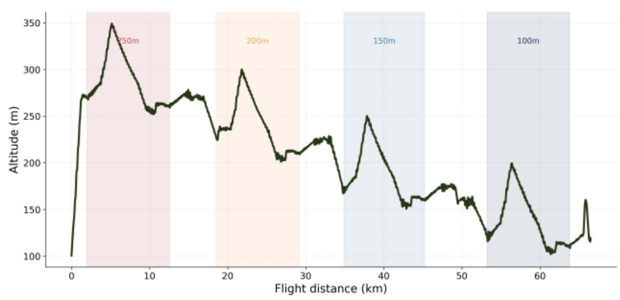

Figure 2. Flight altitude profile of the 66.48 km multi-altitude mission, showing the stepped descent segments at 250, 200, 150, and 100 m AGL.

2.3 Reference Trajectory Generation

To establish a high-precision metric baseline for validating the monocular SLAM trajectory, the collected aerial imagery was processed through a ~~simultaneous~~ self-calibration bundle block adjustment (BBA). The reference trajectory was generated through Structure-from-Motion (SfM) processing in Agisoft

Metashape (Agisoft LLC, 2026), with absolute coordinates established using ground control points (GCPs). The workflow estimates the exterior orientation parameters (EOPs) of each camera pose while simultaneously optimizing the interior orientation parameters (IOPs).

The IOPs are calibrated through self-calibration BBA and are subsequently applied in ORB-SLAM3. To constrain the reference model in an absolute metric frame, eight Ground Control Points (GCPs) surveyed by Real-Time Kinematic Global Navigation Satellite System (RTK-GNSS) were incorporated into the adjustment. These GCPs served as metric anchors for the photogrammetric solution, allowing the resulting EO parameters to be used as ground truth for the comparative analysis. Figure 3 shows the reference trajectory and the RTK-GNSS-measured GCPs used in the photogrammetric workflow.

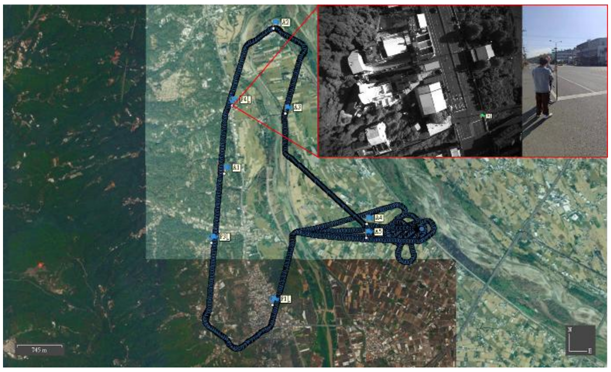

Figure 3. UAV trajectories and RTK-measured GCPs (e.g., A2, P3, P1). The enlarged view shows the field RTK measurement process.

3. Methodology

This chapter describes the evaluation framework adopted in this study for assessing monocular ORB-SLAM3 performance in a long-range UAV mapping mission. The methodology integrates a photogrammetric reference workflow with SLAM-based trajectory estimation and subsequent metric evaluation. The main processing steps include temporal synchronization, trajectory association, metric alignment, and protocol-based analysis.

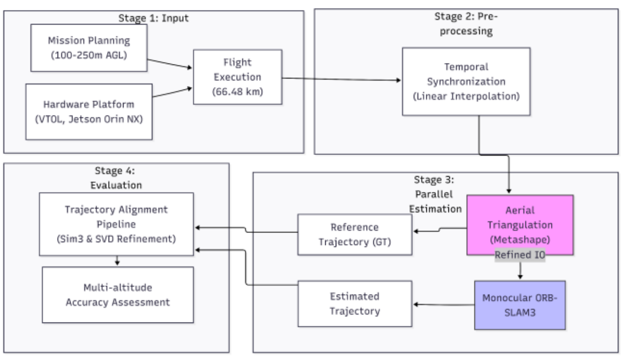

Figure 4. Overall Framework for Monocular SLAM Evaluation

3.1 System Architecture

The proposed framework is organized into four stages, as summarized in Figure 4. The first stage corresponds to the mission and hardware setup used for data acquisition. The second stage performs temporal synchronization between onboard telemetry and imagery. The third stage estimates two trajectory products in parallel, namely the aerial-triangulation-based reference trajectory and the ORB-SLAM3 estimated trajectory. The final stage aligns the SLAM output to the reference coordinate frame and evaluates its accuracy using multiple protocol settings and quantitative metrics.

More specifically, the framework begins with the acquisition of synchronized image and telemetry data from the UAV platform. These data are then prepared for two parallel processing paths: the photogrammetric workflow used to generate the reference trajectory and the monocular ORB-SLAM3 workflow used to estimate the camera trajectory. After both trajectories are obtained, they are associated by common image keys and transformed into a common reference frame through the proposed alignment pipeline. The aligned results are finally analyzed under multiple evaluation protocols and accuracy metrics in order to examine global consistency, local geometric behaviour, and altitude-dependent performance. This workflow links data preparation, trajectory estimation, alignment, and evaluation within a common reference framework.

3.2 Temporal Alignment

To support offline evaluation, the visual and telemetry data were temporally synchronized before the photogrammetric and SLAM workflows were compared. The onboard system recorded MAVLink telemetry at 10 Hz and image data at 16 Hz. Because these two data streams were not sampled at the same frequency, linear interpolation was applied to estimate the UAV pose corresponding to each image timestamp. This synchronized dataset served as the geometric prior for the subsequent aerial triangulation process, thereby improving temporal consistency between the reference trajectory and the image sequence used for SLAM evaluation.

The interpolation was performed using the two telemetry records immediately preceding and following a given image timestamp. Let t_a and t_b denote the timestamps of the surrounding telemetry records, and let t_{img} denote the image timestamp. The corresponding interpolated pose $P(t_{img})$ is obtained by linear interpolation:

$$P(t_{img}) = P(t_a) + \frac{t_{img} - t_a}{t_b - t_a} \cdot \big(P(t_b) - P(t_a)\big) \qquad (1)$$

where $P(t)$ denotes the interpolated 6-degree-of-freedom pose at time t, including both position and attitude components. In the present framework, this interpolation step belongs to the preprocessing stage and provides the temporal linkage between the raw flight telemetry and the image-based reference workflow.

3.3 SLAM Configuration

Monocular ORB-SLAM3 was employed to estimate the camera trajectory from the image sequence (Campos et al., 2021). To maintain geometric consistency between the SLAM estimation and the photogrammetric reference workflow, the refined interior orientation parameters derived from the bundle block adjustment were adopted as the intrinsic calibration for ORB-SLAM3. These parameters include the focal length, principal point, and lens distortion coefficients used to define the camera model for trajectory estimation. The current implementation adopts a pinhole camera model throughout the evaluation.

In the current implementation, 2,000 ORB features are extracted per frame across 8 pyramid levels with a scale factor of 1.15. The FAST thresholds are set to 15 and 7 for the initial and minimum detection stages, respectively. These parameters were used as the operational configuration in all experiments and were not independently optimized in this study.

3.4 Trajectory Alignment Pipeline

Because monocular SLAM operates in an arbitrary and unscaled coordinate system, the estimated trajectory cannot be compared directly with the photogrammetric reference trajectory. A multi-step alignment pipeline was therefore applied to transform the ORB-SLAM3 output into the reference coordinate frame. In the present framework, the pipeline consists of four main steps: frame-wise trajectory association, Sim(3)-based metric alignment, fixed coordinate-axis correction, and global rotation refinement.

First, the SLAM trajectory and the reference trajectory were associated by common image keys. In the implementation, the ORB-SLAM trajectory file was parsed into image-name keys, and the reference metadata were organized using the same naming convention. The evaluation then retained only the overlapping image keys shared by both trajectories, ensuring that all subsequent alignment and metric computations were performed on corresponding camera poses.

Second, metric scale and spatial translation were recovered using the Umeyama algorithm (Umeyama, 1991). Given a set of matched SLAM positions $P_{SLAM,i}$ and reference positions $P_{GT,i}$, the optimal scale factor s, rotation R_{pos}, and translation t_{pos} were estimated in a least-squares sense:

$$P_{GT,i} = s \cdot R_{pos} \cdot P_{SLAM,i} + t_{pos} \tag{2}$$

This transformation maps the SLAM trajectory into the metric coordinate domain of the reference data.

Third, a fixed axis transformation was applied to correct the coordinate convention of the raw SLAM rotations. In the present implementation, a constant transformation matrix M_{fix}, corresponding to a $\pi/2$ rotation about the Y-axis, was applied to the raw SLAM rotations $R_{SLAM,i}$:

$$R_{fixed,i} = M_{fix} \cdot R_{SLAM,i} \tag{3}$$

This step ensures that the SLAM rotations are expressed in a convention consistent with the reference trajectory before further refinement.

Fourth, a global rotation offset R_{off} was estimated to refine the attitude agreement between the aligned SLAM trajectory and the reference trajectory. After applying R_{pos} and M_{fix}, the rotation offset was obtained by solving an orthogonal Procrustes problem over the matched frame set. The covariance matrix M is constructed as

$$M = \sum_{i=1}^{N} R_{GT,i} \cdot \left(R_{pos} \cdot R_{fixed,i}\right)^T \tag{4}$$

Applying singular value decomposition (SVD) to M, such that $M = U \Sigma V^T$, the optimal rotation refinement is given by

$$R_{off} = UV^T \tag{5}$$

The final aligned position and rotation for each matched frame are then expressed as

$$P_{aligned,i} = s \cdot R_{pos} \cdot P_{SLAM,i} + t_{pos} \tag{6}$$

$$\boldsymbol{R_{aligned,i}} = \boldsymbol{R_{off}} \cdot \boldsymbol{R_{pos}} \cdot \boldsymbol{M_{fix}} \cdot \boldsymbol{R_{SLAM,i}} \tag{7}$$

where $P_{SLAM,i}$ and $P_{GT,i}$ denote the SLAM-estimated and reference 3D positions of the i-th matched frame, respectively; s is the recovered scale factor; R_{pos} is the rotation estimated during position alignment; t_{pos} is the corresponding translation vector; M_{fix} is the constant axis-correction matrix; $R_{SLAM,i}$, $R_{fixed,i}$ and $R_{aligned,i}$ denote the raw, axis-corrected, and final aligned rotations; R_{off} is the global rotation refinement estimated from SVD; and N is the number of matched frames used for alignment.

For comparison, a simplified scale-and-translation baseline was also evaluated. This baseline, referred to as the S+T baseline, first flips the raw SLAM trajectory along the Z axis, then estimates a single vertical scale factor from a selected reference point pair with sufficient elevation difference, and finally computes a translation vector from the centroids of the selected alignment frames. Unlike the proposed method, this baseline does not apply any rotation optimization and therefore preserves the raw SLAM orientation. It is included to illustrate the effect of using scale recovery and translation only, without full rotational refinement.

3.5 Evaluation Protocols

To examine how trajectory coverage and alignment scope affect the interpretation of SLAM accuracy, three evaluation protocols were defined in this study: No-Loop (NL), With-Loop Global Slice (GS), and With-Loop Local Re-Sim(3) (LR). These protocols were designed to distinguish segment-wise drift in isolated runs, full-mission global constraints, and segment-level local geometric quality under a common evaluation framework.

The No-Loop (NL) protocol evaluates trajectories estimated separately from the altitude-specific image subsets. In this setting, ORB-SLAM3 was executed independently for each altitude segment using only the images belonging to that altitude range. Because these altitude-specific subsets do not form a complete loop over the study area, the resulting trajectories represent segment-level SLAM behaviour without full-mission loop closure.

The With-Loop protocols are derived from the full-mission ORB-SLAM3 run. In this case, the complete image sequence was processed as a single trajectory. Because the mission revisits the same area multiple times at different flight altitudes, the full-mission run contains repeated traversals that may introduce loop-closure constraints during SLAM processing, although their influence is not necessarily uniform across all trajectory segments.

The With-Loop Global Slice (GS) protocol extracts the altitude-specific segments directly from the full-mission trajectory estimated by ORB-SLAM3. It therefore represents the segment behaviour inherited from the globally optimized full trajectory. In contrast, the With-Loop Local Re-Sim(3) (LR) protocol also starts from the same full-mission run, but each extracted segment is independently re-aligned to the reference data using a segment-specific Sim(3) transformation. This protocol reduces the influence of residual global deformation inherited from the full trajectory and reveals the local geometric quality of each segment more directly.

Through these three protocols, the evaluation framework distinguishes whether an observed performance difference arises primarily from segment-wise drift in isolated runs, from the influence of full-mission global optimization, or from the local geometric properties of the reconstructed segment after re-alignment.

3.6 Accuracy Evaluation Metrics

To quantify the performance of the aligned SLAM trajectory, several metrics were adopted to evaluate absolute position accuracy, attitude consistency, and drift over increasing travel distances. These metrics were selected to support both full-mission assessment and segment-level comparison under the three evaluation protocols introduced in Section 3.5.

3.6.1 Position Accuracy

Absolute position accuracy is evaluated by comparing the aligned SLAM positions $P_{aligned,i}$ with the reference positions $P_{GT,i}$. The primary metric is the Root Mean Square Error (RMSE), computed over all matched frames:

$$RMSE_{3D} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} ||P_{GT,i} - P_{aligned,i}||^2} \tag{8}$$

where N is the number of matched frames used in the evaluation. In addition to the 3D RMSE, separate RMSE values are reported for the X, Y, and Z components in order to distinguish horizontal and vertical error characteristics across different altitude segments. This formulation is consistent with the current implementation, in which the position metrics are summarized as 3D RMSE together with axis-wise RMSE values

3.6.2 Attitude Accuracy

Attitude accuracy is evaluated using both Euler-angle RMSE and geodesic rotation error. Let $R_{aligned,i}$ and $R_{GT,i}$ denote the aligned SLAM rotation and the reference rotation of the i-th matched frame, respectively. The geodesic error represents the angular difference between the two rotations and is defined as

$$\Delta R = R_{aligned,i} R_{GT,i}^{T} \tag{9}$$

and the corresponding geodesic error is computed as

$$\Delta \Phi = deg(\arccos\left(\frac{tr(\Delta R) - 1}{2}\right)) \tag{10}$$

The geodesic RMSE is then computed over all attitude-valid frames. In addition, the RMSE values of yaw, pitch, and roll are reported after converting the aligned and reference rotations to Euler angles under a consistent rotation convention.

3.6.3 Relative Pose Drift

To evaluate drift as a function of traveled distance, KITTI-style subsequence analysis is adopted over a set of predefined path lengths L. For each subsequence, the relative pose error is computed between the aligned SLAM trajectory and the reference trajectory. The mean translation drift is expressed as

$$E_{trans}(L) = \frac{1}{M}\sum_{j=1}^{M} \frac{||\Delta T_j||}{L} \times 100\% \tag{11}$$

where M is the number of valid subsequences of length L, and ΔT_j is the relative translation discrepancy for the j-th subsequence. The corresponding rotation drift is defined as

$$E_{rot}(L) = \frac{1}{M}\sum_{j=1}^{M} \frac{\Delta \Phi_{rel,j}}{L} \times 100 \tag{12}$$

where $\Delta \Phi_{rel,j}$ denotes the relative rotation error of the j-th subsequence in degrees. In the present implementation, these metrics are evaluated over subsequence lengths ranging from 50 m to 25.6 km and reported in the units of percent for translation drift and degrees per 100 m for rotation drift.

In addition to the pose-based KITTI-style drift, a position-only subsequence drift is also computed by comparing the displacement vectors between the aligned SLAM positions and the reference positions over the same subsequence lengths. This complementary measure is included because it isolates translational endpoint drift without the influence of relative attitude error and is used in the subsequent long-range drift analysis.

3.6.4 Windowed Cumulative Drift

To provide a more intuitive description of error growth over distance, a windowed cumulative drift analysis is further performed. For each subsequence length L, the endpoint difference between the aligned SLAM displacement and the reference displacement is accumulated over all valid windows. The mean 3D endpoint drift is defined as

$$D_{3D}(L) = \frac{1}{M}\sum_{j=1}^{M} ||\left(P_{aligned,j}^{end} - P_{aligned,j}^{start}\right) - \left(P_{GT,j}^{end} - P_{GT,j}^{start}\right)|| \tag{13}$$

where M is the number of valid windows of length N. In addition to the 3D drift, the current implementation also reports the corresponding horizontal (XY) and vertical (Z) components. This metric is used to examine how translational drift accumulates with subsequence length and therefore complements the normalized KITTI-style drift reported above.

4. Results and Analysis

This chapter presents the experimental results of the proposed evaluation framework for monocular ORB-SLAM3 in the multi-altitude UAV mapping mission. The analysis is organized from full-mission alignment performance to segment-level comparisons, so that both global consistency and local geometric behaviour can be examined in a unified manner. The results are further interpreted with respect to altitude-dependent image geometry, full-mission constraints, and the effect of trajectory alignment strategy.

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-63-2026

4.1 Full-Mission Alignment

The first analysis focuses on the full-mission trajectory in order to verify the effectiveness of the proposed alignment pipeline and to establish a global performance baseline. Figure 5 compares the AT reference trajectory, the proposed aligned SLAM trajectory, and the S+T baseline in planimetric view. The proposed alignment shows substantially closer agreement with the reference trajectory, whereas the S+T baseline exhibits clear lateral deformation and orientation inconsistency over the full mission extent. This comparison indicates that scale recovery and translation alone are insufficient to restore the global geometric structure of the monocular SLAM trajectory.

Figure 5. Planimetric comparison of the AT reference trajectory, the proposed aligned SLAM trajectory, and the S+T baseline.

The same tendency can be observed in the elevation profile shown in Figure 6. After the proposed alignment, the SLAM trajectory follows the reference height variation much more closely throughout the mission, while the S+T baseline retains noticeable vertical distortion and scale inconsistency. Quantitatively, the full mission covers 66.48 km and contains 5330 matched frames. Under the default full-mission alignment using all matched frames, the proposed method achieves a 3D position RMSE of 7.41 m. The recovered Sim(3) scale is 1.78, whereas the raw S+T baseline yields a vertical scale of 2.02. These results indicate that full metric alignment requires not only scale recovery but also consistent rotation correction in both position and attitude.

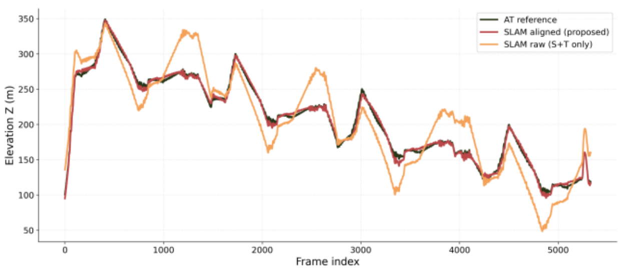

Figure 6. Elevation profile comparison of the AT reference trajectory, the proposed aligned SLAM trajectory, and the S+T baseline.

Overall, the full-mission comparison demonstrates that the proposed alignment pipeline provides a stable basis for subsequent protocol-based evaluation. Therefore, the following analyses are carried out on trajectories that have been transformed into the common reference frame using the proposed alignment procedure.

4.2 Multi-Altitude and Protocol Comparison

This section compares segment-level SLAM accuracy across the four flight altitudes of 100 m, 150 m, 200 m, and 250 m under the three evaluation protocols defined in Section 3.5, namely No-Loop (NL), With-Loop Global Slice (GS), and With-Loop Local Re-Sim(3) (LR). Figure 7 summarizes the 3D position RMSE for all altitude segments.

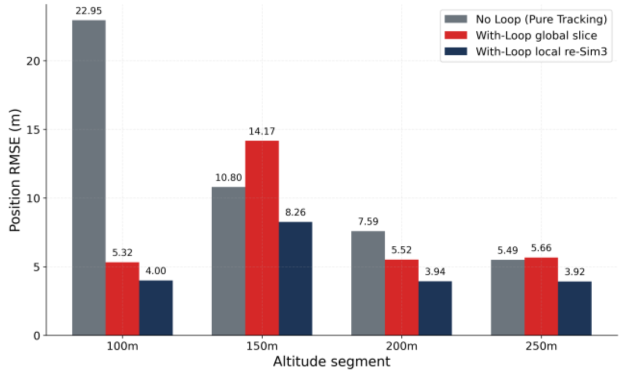

Figure 7. Segment-level 3D position RMSE across altitude segments under the three evaluation protocols.

The No-Loop (NL) results show a clear altitude dependency. The 3D RMSE decreases from 22.95 m at 100 m to 10.80 m at 150 m, 7.59 m at 200 m, and 5.49 m at 250 m. This trend indicates that the isolated altitude-specific runs become progressively more stable as flight altitude increases. A plausible explanation is that higher-altitude imagery yields smoother viewpoint transitions and more stable feature tracking, thereby reducing drift accumulation in monocular estimation. By contrast, the 100 m run exhibits the weakest geometric stability among all isolated segment runs.

The With-Loop protocols reveal a different pattern. At 100 m, both GS and LR are substantially better than NL, with 3D RMSE values of 5.32 m and 4.00 m, respectively. At 200 m and 250 m, the LR protocol again provides the best segment-level accuracy, reaching 3.94 m and 3.92 m, while the GS protocol remains close to 5.52 m and 5.66 m. These results suggest that the full-mission run introduces useful global constraints, but that local re-alignment is more effective for revealing the intrinsic geometric quality of each segment.

A notable exception is observed at 150 m. In this segment, the GS result degrades to 14.17 m, which is worse than the corresponding NL result of 10.80 m, whereas the LR result improves again to 8.26 m. A plausible explanation is that the 150 m portion of the full-mission trajectory is influenced by the descent transition from 200 m to 150 m, during which additional loitering turns were performed. These manoeuvres introduce stronger attitude variation and trajectory curvature into the full-mission optimization, and their residual effect may therefore be inherited by the sliced GS segment. By contrast, the NL result is obtained from the 150 m image subset alone and is less affected by the global deformation associated with the transition phase.

To complement the position analysis, Table 1 reports the number of evaluated frames and the corresponding geodesic RMSE for each altitude segment. In contrast to the stronger variation observed in position RMSE, the geodesic RMSE remains within a relatively narrow range of approximately 2° to 3.25° across protocols and altitudes. This indicates that the most prominent differences among NL, GS, and LR are reflected more strongly in segment-level positional consistency than in large attitude disagreement. Taken together, these results show that full-mission global constraints do not provide uniform benefits across all altitude segments, and that segment-specific re-alignment remains necessary for isolating local geometric quality.

Altitude (m)	Frames	NL	GS	LR
100	850	3.25	2.19	2.66
150	818	2.19	2.31	2.06
200	837	2.4	2.14	2.41
250	835	2.29	2	2.

Note. Values in the NL, GS, and LR columns are geodesic RMSE in degrees.

Table 1. Segment-level frame counts and geodesic RMSE under the three evaluation protocols.

4.3 Attitude Consistency

In addition to position accuracy, the full-mission trajectory was examined from the perspective of attitude consistency. Figure 8 compares the yaw, pitch, and roll profiles of the aligned SLAM trajectory with those of the AT reference. Overall, the estimated orientation follows the reference trend closely throughout the mission, indicating that the proposed alignment pipeline preserves the dominant rotational behaviour of the UAV over long flight distances.

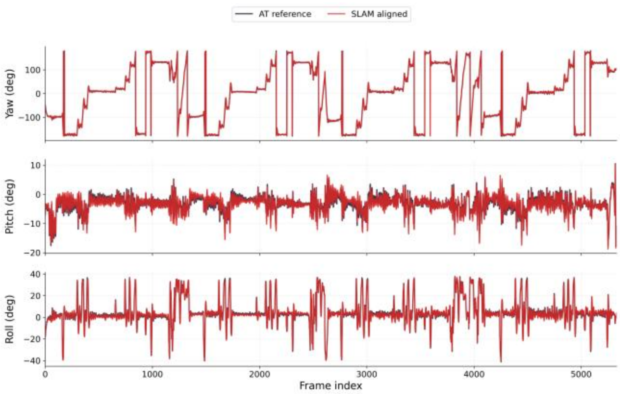

Figure 8. Attitude comparison between the AT reference and the aligned SLAM trajectory.

The agreement is particularly clear in the pitch and roll components, where the aligned SLAM trajectory captures the major variations of the reference solution with only limited local deviation. The yaw component also follows the overall trend well, although slightly larger discrepancies can be observed in some portions of the mission. Quantitatively, the full-mission attitude errors are 1.77° for yaw, 0.79° for pitch, and 0.79° for roll. These values indicate that the aligned trajectory maintains a stable orientation estimate over the complete 66.48 km mission.

The geodesic RMSE of the full-mission solution is 2.11°. This value is consistent with the relatively small Euler-angle RMSE values and confirms that the aligned trajectory maintains good overall rotational agreement with the reference solution. Taken together, these results suggest that segment-level performance variation is reflected more strongly in position consistency than in attitude disagreement.

4.4 Long-Range Drift Characteristics

This section investigates the system performance over extended travel distances by analyzing both normalized subsequence drift and accumulated endpoint error. The drift behaviour is evaluated using KITTI-style relative pose metrics together with windowed cumulative drift, so that both relative and absolute error growth can be examined as the path length increases.

The KITTI-style results indicate that the normalized translation and rotation errors decrease as the subsequence length increases. The mean translation drift decreases from 11.45% at 50 m to 1.36% at 25.6 km, while the mean rotation drift decreases from 16.84 deg/100m to 0.07 deg/100m over the same range. A similar tendency is observed in the position-only subsequence drift, which decreases from 0.63% at 50 m to 0.04% at 25.6 km. These results suggest that although short-range windows are more sensitive to local trajectory noise, the long-range trajectory remains globally constrained and does not exhibit unbounded divergence.

To provide a more intuitive description of error accumulation in metric units, Figure 9 shows the windowed cumulative drift as a function of subsequence length. The 3D endpoint drift increases from 0.35 m at 50 m to 10.66 m at 25.6 km, while the horizontal component increases from 0.23 m to 8.36 m and the vertical component increases from 0.24 m to 5.32 m. The horizontal component is consistently the dominant contributor to the total drift, indicating that long-range error accumulation is mainly expressed in planimetric deviation rather than in height instability.

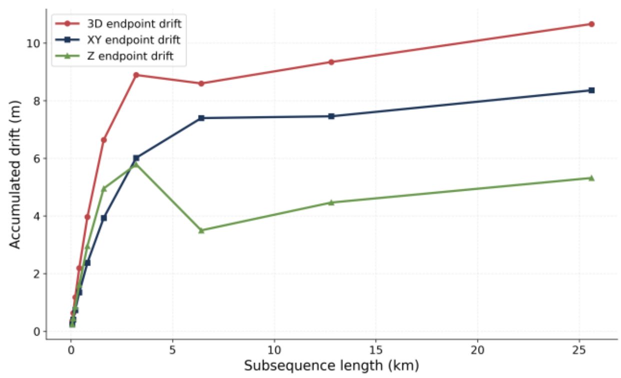

Figure 9. Windowed cumulative drift as a function of subsequence length.

An additional observation is that the cumulative drift does not increase monotonically with distance. In particular, the mean 3D endpoint drift reaches 8.89 m at 3.2 km, decreases slightly to 8.60 m at 6.4 km, and then rises again to 9.34 m at 12.8 km and 10.66 m at 25.6 km. This non-linear behaviour suggests that drift growth is influenced not only by path length but also by the spatial distribution of global constraints inherited from the full-mission trajectory. Table 2 summarizes the representative drift values used in this analysis.

Length	Pose trans.	Rot.	Pos.-only	3D drift
50m	11.45%	16.84	0.63%	0.35m
400m	10.54%	3.93	0.54%	2.19m
3200m	8.18%	0.54	0.28%	8.89m
25600m	1.36%	0.07	0.04%	10.66m

Note. Rot. denotes rotational error in degrees per 100 m.

Table 2. Compact long-range drift summary at representative subsequence lengths.

Taken together, these results indicate that the proposed evaluation framework captures two complementary aspects of long-range behaviour. The normalized KITTI-style metrics show that the full-mission solution remains globally bounded, while the cumulative endpoint drift reveals how absolute spatial error gradually accumulates with increasing travel distance. This combination provides a more complete description of long-range trajectory quality than either metric alone.

4.5 Alignment Frame Selection

This section evaluates the sensitivity of the full-mission alignment to the number and spatial distribution of the frames used for alignment. Five frame-selection strategies were compared: All Frames, First-1018, Uniform-1018, Random-1018, and Informative-1018. This analysis examines whether the proposed alignment remains stable when only a subset of the matched frames is used.

The results show that the alignment is highly robust to reduced frame density, provided that the selected frames remain spatially distributed across the full trajectory. Compared with the All Frames baseline, the Uniform-1018, Random-1018, and Informative-1018 strategies yield nearly identical position accuracy, with 3D RMSE values of 7.41 m, 7.42 m, and 7.41 m, respectively. Their geodesic RMSE values also remain close to the All Frames result, ranging from 2.04° to 2.11°. This indicates that the alignment quality is only weakly affected by frame reduction when the subset preserves global coverage.

By contrast, the First-1018 strategy produces the worst result, with a 3D RMSE of 8.87 m. Because this strategy uses only the earliest portion of the mission, it does not adequately represent the full spatial extent of the trajectory and is therefore less capable of constraining the global alignment. This comparison indicates that the spatial distribution of the alignment frames is more important than the absolute number of frames. In practical terms, the full-mission alignment does not require all matched frames to achieve stable accuracy, but it does require samples that cover the mission globally. Table 3 summarizes the effect of alignment frame selection on full-mission accuracy.

Strategy	Frames	Pos. (m)	Geo. (deg)
All Frames	5330	7.41	2.11
First-1018	1018	8.87	2.99
Uniform-1018	1018	7.41	2.09
Random-1018	1018	7.42	2.04
Informative-1018	1018	7.41	2.15

Note. Pos. and Geo. denote positional RMSE (m) and geodesic RMSE (deg), respectively.

Table 3. Effect of alignment frame selection on full-mission accuracy.

5. Conclusions

This study presented a quantitative evaluation of monocular ORB-SLAM3 in a long-range multi-altitude UAV mapping mission using aerial-triangulation-derived camera poses as reference data. By combining temporal association, metric trajectory alignment, and protocol-based evaluation, the proposed framework enabled full-mission and segment-level assessment of SLAM behavior under realistic aerial mapping conditions. The results show that the proposed alignment pipeline enables consistent comparison between the SLAM trajectory and the photogrammetric reference, and that scale recovery alone is insufficient to preserve the global geometric structure of the mission trajectory.

The segment-level analysis further demonstrates that flight altitude has a clear influence on monocular SLAM stability. The No-Loop runs show a strong altitude dependency, with lower-altitude segments exhibiting substantially larger position error than higher-altitude segments. In addition, the comparison between the With-Loop Global Slice and With-Loop Local Re-Sim(3) protocols indicates that full-mission global constraints do not benefit all segments uniformly. Instead, local re-alignment provides a more informative representation of the intrinsic geometric quality of individual segments. The long-range drift analysis also shows that the full-mission solution remains globally bounded, while still accumulating measurable absolute spatial error with increasing travel distance.

Taken together, these results indicate that monocular ORB-SLAM3 can provide useful large-scale trajectory estimation for UAV mapping when supported by an appropriate alignment and evaluation framework. More importantly, this study shows that the interpretation of SLAM quality depends strongly on how the trajectory is segmented, aligned, and evaluated, especially in missions involving repeated revisits and altitude transitions.

Future work will focus on extending the present offline framework toward real-time onboard deployment. In particular, a key next step is to investigate whether SLAM-derived trajectory outputs can be provided directly to the flight controller as a positioning source, thereby reducing dependence on GNSS-based position information in degraded or unavailable environments.

Acknowledgements

This work was supported by the National Chung-Shan Institute of Science and Technology under the project entitled “Geographical Image Positioning of Unmanned Aerial Vehicles” (Project No. NCSIST-ACOM-114-172C001). The authors would like to express their sincere gratitude for this support.

References

Agisoft LLC, 2026. Agisoft Metashape Professional, Version 2.3.0. https://www.agisoft.com/downloads/installer/.

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R., 2016. The EuRoC micro aerial vehicle datasets. *International Journal of Robotics Research*, 35(10), 1157-1163.

Campos, C., Elvira, R., Gómez Rodríguez, J.J., Montiel, J.M.M., Tardós, J.D., 2021. ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM. *Information Fusion*, 68, 44-63.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2012. Vision meets robotics: The KITTI dataset. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3354-3361.

Mur-Artal, R., Montiel, J.M.M., Tardós, J.D., 2015. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5), 1147-1163.

Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4), 376-380.