Evaluating ORB-SLAM 3 Performance Using a Photogrammetry-Based Reference Trajectory

Leonardo Sales Galvão[1], Daniel Juchem Regner[1], Moacir Wendhausen[1], Tiago Loureiro Figaro da Costa Pinto[1], Armando Albertazzi Gonçalves Júnior[1]

LABMETRO, Department of Mechanical Engineering, Federal University of Santa Catarina, Trindade, Florianópolis, SC, 88040-970, Brazil - leonardo.sales@labmetro.ufsc.br

Keywords: ORB-SLAM 3, SLAM, Reference Trajectory, Photogrammetry, Ground Truth.

Abstract

The robust evaluation of Visual Simultaneous Localization and Mapping (vSLAM) systems is fundamental to their development and deployment. However, this process is often constrained by the reliance on expensive and complex external infrastructure, such as laser trackers or motion capture systems, to provide accurate ground-truth trajectories. This paper introduces a novel and self-contained methodology for the high-fidelity evaluation of stereo vSLAM and stereo-inertial algorithms. Our approach leverages the very same image sequence used by the SLAM algorithm to generate a dense, globally optimized photogrammetric model. The proposed methodology comprises two fundamental steps, the first step consists of validating photogrammetry as a ground truth method. For this purpose, the linear displacement measured by photogrammetry was compared with the displacement of a precision guide, which was benchmarked against a laser interferometer as the standard. Once the reference was validated, the second step assessed the performance of ORB-SLAM 3 on a free trajectory within a complex environment, by directly comparing the SLAM result to the trajectory generated by photogrammetry. The accuracy was then quantified using standard metrics, including Absolute Trajectory Error (APE) and Relative Pose Error (RPE). The results validate our approach as an accessible, low-cost, and reliable alternative for benchmarking vSLAM systems, enabling a way of performance analysis using only the data from the sensor suite under evaluation.

1. Introduction

Visual Simultaneous Localization and Mapping (vSLAM) is a fundamental capability for autonomous robotic mapping and localization in unknown environments (Yarovoi and Cho, 2024). Vision sensors offer low-cost, information-rich environmental data compared to other modalities (Chen et al., 2022). While state-of-the-art algorithms like ORB-SLAM 3 demonstrate robust real-time performance (Campos et al., 2021), their operational reliability requires rigorous quantitative evaluation to ensure real-world suitability (Cort et al., 2024; Nair et al., 2024; Yang et al., 2023).

A fundamental challenge in vSLAM validation is establishing a reference trajectory with a known uncertainty budget that is significantly lower than the error of the system under test. The gold standard typically relies on expensive external infrastructure, such as industrial laser trackers or marker-based motion capture systems, which are mostly constrained to controlled indoor environments (Fontan et al., 2024). To address this issue, emerging alternatives have explored depth sensors (e.g., RGB-D) and LiDAR-based SLAM to establish reference trajectories, including applications in GNSS-denied setups or multi-sensor fusion datasets (Ha et al., 2022; Sturm et al., 2012; Yin et al., 2022). While valuable, these alternatives still require additional specialized hardware.

Standardized public datasets (e.g., KITTI, EuRoC, TUM-VI) have significantly advanced the field (Burri et al., 2016; Geiger et al., 2012; Schubert et al., 2018), yet their ground-truth data is still largely captured using infrastructure-dependent systems. Although photogrammetry is a well-established technique for accurate 3D modelling (James et al., 2017), using it to generate ground truth directly from the vSLAM system's input imagery remains uncommon. Existing studies employing photogrammetric references often rely on entirely separate, dedicated data capture processes (Ascard and Movahedi, 2023).

Thus, this work introduces a self-contained evaluation methodology for stereo vSLAM systems. Our framework leverages the algorithm’s native image sequence—using full-resolution images for global photogrammetric optimization and downscaled versions for real-time SLAM—to generate a high-fidelity ground-truth trajectory. This approach eliminates the need for external hardware, providing an accessible, low-cost validation tool. The methodology is validated using ORB-SLAM 3, quantifying its performance via standard metrics such as ATE and RPE (Grupp et al., 2017). The remainder of this paper is organized as follows: Section 2 reviews the methodology, Section 3 details the experimental procedures and results, Section 4 discusses the findings, and Section 5 concludes the paper.

2. Methodology

This section details the proposed methodology for evaluating ORB-SLAM 3 algorithm using a self-generated photogrammetric ground truth. The process is presented as a sequential workflow, beginning with the acquisition of a single stereo image sequence. It then describes the two parallel pipelines that use this sequence: the generation of the high-fidelity ground-truth trajectory using photogrammetry, and the trajectory estimation performed by the ORB-SLAM 3 algorithm. The section concludes by detailing the procedure for aligning these two trajectories and the quantitative metrics used for their comparison.

2.1 Overall Workflow

The proposed evaluation methodology is structured as a clear, sequential workflow, illustrated in Figure 1. The process begins with a Real-Time Data Acquisition step, where synchronized Stereo System Images and IMU Data are captured using ROS 2. This raw data then serves as the input for two independent, parallel processing pipelines.

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-1147-2026

Pipeline A (Ground-Truth Generation): The captured stereo images are saved in their original resolution and processed offline using the photogrammetry software COLMAP (Schönberger et al., 2016a, 2016b). This pipeline generates a Dense Map of the environment and, most importantly, computes a higher-accuracy camera trajectory that serves as the ground-truth reference.

Pipeline B (SLAM Trajectory Estimation): The real-time stream of stereo images and IMU data is fed directly into ORB-SLAM 3. The algorithm uses this information to generate a Sparse Map and the estimated camera trajectory.

Finally, in the Quantitative Comparison stage, the two trajectories are fed into the EVO evaluation tool (Grupp et al., 2017). EVO aligns the trajectories and calculates key performance metrics, including RPE and APE, while also generating plots for visualization.

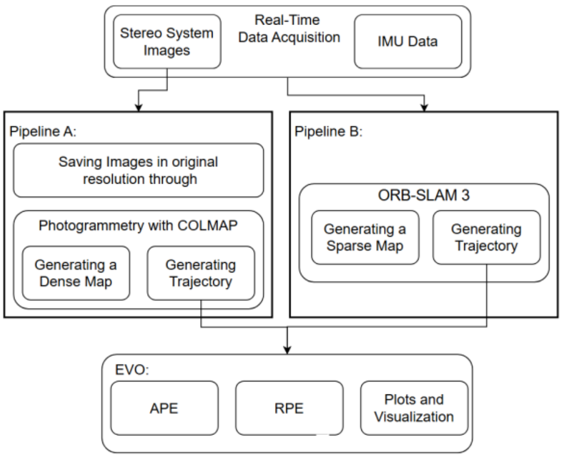

Figure 1. Overview of the proposed evaluation workflow, from data acquisition to quantitative analysis.

2.2 Data Acquisition

The hardware setup consists of a custom-built stereo-inertial camera rig, as shown in Figure 2. The rig utilizes a pair of FLIR Blackfly® S (BFS-U3-50S4C-C) machine vision cameras. These 5.0-megapixel (2448×2048) color cameras are equipped with a 1/1.8" Sony IMX547 CMOS global shutter sensor. This eliminates rolling shutter artifacts (such as skew and 'jello' effects) caused by rapid motion, which is a vital requirement for ensuring the geometric integrity of images used in SLAM and photogrammetry. The cameras are fitted with 3.5mm focal C-mount lenses and are attached on a rigid frame with a fixed baseline of 200mm. To enable stereo-inertial SLAM functionality, the rig is equipped with a high-performance STIM 300 Inertial Measurement Unit (IMU).

The data acquisition and communication process is managed within the Robot Operating System 2 (ROS 2) framework. All image and IMU messages are published over dedicated ROS 2 topics. For real-time trajectory estimation, the ORB-SLAM 3 node subscribes to these topics; for performance reasons, the input images are converted to monochrome and rescaled to 800x600 at 30 frames per second before SLAM processing. Simultaneously, a separate node subscribes to the same raw image topic and saves the unmodified, full-resolution color images to disk at a lower frequency of 4 frames per second. This ensures that the high-resolution, color images saved for photogrammetry correspond precisely to the same moments in time as the low-resolution frames being processed by the SLAM algorithm.

Figure 2. The custom-built stereo-inertial camera rig, showing the FLIR Blackfly S cameras and the STIM 300 IMU.

2.3 Photogrammetric Ground-Truth Generation

The ground-truth trajectory is generated through offline processing of the full-resolution color images that were saved at 4 frames per second. This sequence of images, captured from only one of the cameras in the rig (e.g., the left camera), is processed as a monocular dataset using the open-source COLMAP Structure-from-Motion (SfM) pipeline.

The process begins by providing COLMAP with the pre-calibrated intrinsic parameters for the cameras. Following this, feature extraction is performed on all images. For the crucial feature matching step, various strategies were evaluated, with the VocabTree (Nistér et al., 2006) matching method proving to be the most effective and robust for our dataset. Once features are matched, COLMAP performs an incremental reconstruction, which is then globally refined through a Bundle Adjustment (Granshaw et al., 1980). This final optimization step jointly refines all camera poses and 3D points, minimizing reprojection errors to produce a highly consistent and geometrically accurate sparse reconstruction and camera trajectory.

A critical final step is the metric scale recovery. The trajectory produced by the bundle adjustment is scale-ambiguous, meaning its dimensions are not tied to any real-world units. To resolve this, calibrated patterns with known dimensions were placed in the scene during data acquisition. By measuring the distance between specific points on these patterns in the real world and then measuring the corresponding distance in the reconstructed model, a single, precise scaling factor is calculated. This factor is then applied to the entire model, transforming the trajectory into a true metric reference, ready to be used as the ground truth for comparison against the SLAM output.

2.4 ORB-SLAM 3 Trajectory Estimation

The ORB-SLAM 3 algorithm processes the full, real-time stereo-inertial data stream as it is published by the camera rig. As previously noted, for computational efficiency, the input images are converted to monochrome and rescaled on-the-fly by the SLAM node. This high-frequency, real-time processing occurs in parallel with the separate software node (described in Section 2.2) that subscribes to the same stream but saves the full-resolution images at a lower 4 FPS rate for the offline pipeline.

A crucial aspect for a fair comparison is consistency in calibration, therefore, the exact same images with corresponding resolutions, were used to calibrate camera intrinsic parameters used for the photogrammetry and ORB-SLAM 3. Regarding the algorithm's internal parameters, this work utilizes the optimal ORB feature configuration identified previously. This study evaluates the performance of ORB-SLAM 3 in both its stereo and stereo-inertial modes. The specific mode employed for each experiment will be clearly defined in the corresponding results section.

The final outputs of this real-time process are the estimated camera trajectory and an associated sparse point cloud. It is important to note that this SLAM process itself does not save the raw images; it only processes the data stream. The estimated trajectory is the key output that forms the basis for the quantitative comparison against the photogrammetric ground truth trajectory.

2.5 Trajectory Alignment and Comparison

The estimated trajectory from ORB-SLAM 3 and the ground-truth trajectory from the photogrammetric reconstruction are initially produced in different, arbitrary coordinate systems. A direct numerical comparison is therefore not possible. To resolve this, the two trajectories must first be aligned into a common coordinate frame. This is achieved by estimating a rigid body transformation matrix belonging to the SE(3) group, according to the method presented in (Salas et al., 2015).

Although physical markers were present in the environment and could theoretically be used to compute this transformation between the photogrammetric and SLAM coordinate systems, we deliberately opted to use the open-source Python package EVO (Grupp et al., 2017) to estimate the SE(3) alignment. Because our primary objective is to benchmark the SLAM trajectory using the photogrammetric model as ground truth, utilizing established SLAM evaluation tools ensures that the alignment process remains fully automated and standardized. This approach avoids reliance on manual coordinate extraction from markers and guarantees that our evaluation pipeline is directly comparable to other studies within the SLAM research community. The EVO tool computes several standard metrics, of which this work focuses on two primary ones: APE and RPE.

APE evaluates the global consistency of the entire trajectory. It computes the direct difference between each pose in the estimated trajectory and its corresponding pose in the ground truth after alignment. The result is typically summarized as a single value, the Root Mean Square Error (RMSE), which provides a comprehensive measure of overall accuracy.

RPE evaluates the local consistency, or drift, of the trajectory. It measures the error in the relative transformation over fixed time intervals or distances, essentially assessing how well the estimated motion between consecutive camera poses matches the ground truth motion.

In addition to these quantitative metrics, EVO also provides qualitative results through visualizations, including top-down views of the aligned trajectories and graphs of the error over time. These visualizations will be presented in the following section to support the numerical results.

3. Methodology Experimental Setup and Results

This section presents the experimental validation of our proposed photogrammetric evaluation methodology. The primary objective is to demonstrate that the generated trajectory serves as a reliable and high-fidelity ground-truth reference for the quantitative analysis of vSLAM systems. To achieve this, the framework is validated by applying it to the well-established ORB-SLAM 3 algorithm as a case study. We will first describe the experimental setup and datasets, then present the quantitative results forming the core evidence, and conclude with a qualitative analysis of the trajectory plots.

3.1 Experimental Procedures

To validate the proposed evaluation methodology, two primary experiments were conducted. The first is a laboratory validation experiment, designed to verify the metric accuracy of the photogrammetry pipeline. The second is a real-world application experiment, demonstrating the complete workflow for evaluating ORB-SLAM 3 using the photogrammetry as ground truth.

All experiments utilized the software and hardware: real-time SLAM processing was performed on the Avell notebook (Ubuntu 22.04, ROS 2 Humble), while the offline photogrammetric reconstruction was processed on the high-performance desktop using COLMAP v3.8. In both experiments, a fundamental data capture procedure was executed, managed by ROS 2. The ORB-SLAM 3 node subscribed to the high-frequency image stream and processed it in real-time (after conversion to low-res monochrome). Simultaneously, a separate software node subscribed to the exact same image topic. This 'saver' node was configured to write the unmodified, full-resolution, color images to disk, but throttled to a lower frequency of 4 FPS. Crucially, every image saved by this node retains its original header timestamp from the ROS 2 message. This makes the photogrammetry image set a time-stamped subset of the complete image stream processed by SLAM, providing the vital link between the two datasets.

3.1.1 Experiment 1: Laboratory Validation with Linear Guide

The first experiment was conducted in a temperature-controlled (20°C) laboratory. As illustrated in Figure 3, the setup includes a high-precision Linear Guide and a Renishaw XL-80 laser interferometer. The procedure was structured to establish a reliable ground-truth trajectory. First, an initial free-hand movement of the camera rig was performed off the guide to capture varied viewpoints of reference patterns, providing the necessary data for the photogrammetric scaling.

Following this, the vision system was mounted on the guide's carriage, and the laser interferometer was zeroed. After a 5-second stationary pause to capture a stable starting point, the carriage was moved manually along the guide. Once the laser registered a displacement of 3.0 meters, the carriage was held stationary for another 5-second pause to ensure sufficient image capture at the end position. The objective was to process the saved 4 FPS images to quantify the errors of the resulting photogrammetric trajectory in two key areas. First, to assess the scale accuracy by comparing the total displacement against the 3.0-meter reference measured by the laser. Second, to evaluate the trajectory's rectitude by quantifying its deviation from the adjusted linear path. This test served as a fundamental quantification of the photogrammetric pipeline's suitability as a reliable ground-truth reference.

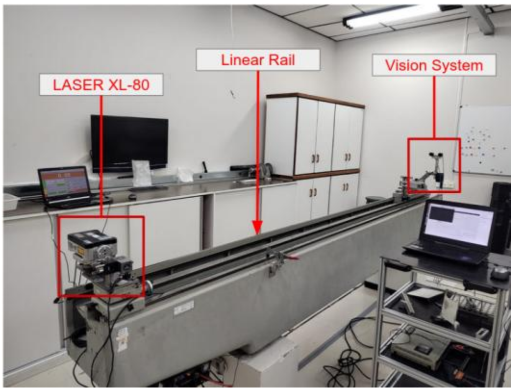

Figure 3. Setup with the Linear Guide and Renishaw XL-80 laser.

3.1.2 Experiment 2: Application in Corridor Environment

The second experiment demonstrated the methodology's practical application in a larger, more complex scenario. As depicted in the floor plan in Figure 4, a tripod was placed at a central location to serve as a distinct start and end reference point for the trajectories. The data capture performed using the stereo-inertial mode of ORB-SLAM 3, involved two paths: first, the operator walked the shorter "Secondary Loop", returning to the tripod. Second, the operator walked the much larger "Main Loop" around the building's central atrium, again concluding the trajectory at the tripod.

During this experiment, the operator walked the two routes shown in Figure 4. This single run generated the two distinct datasets required for the analysis: the real-time, stereo-inertial trajectory as estimated by ORB-SLAM 3, and the corresponding set of full-resolution images saved for later photogrammetric processing. The objective was to apply the complete 'capture-to-evaluation' workflow, where the resulting SLAM trajectory is quantitatively compared against the ground truth photogrammetric trajectory. The results of this comparison are detailed in Section 3.2.2. The final step, therefore, involved using the EVO evaluation tool to quantitatively compare the ORB-SLAM 3 estimate against this photogrammetric reference, thereby demonstrating the complete 'capture-to-evaluation' workflow.

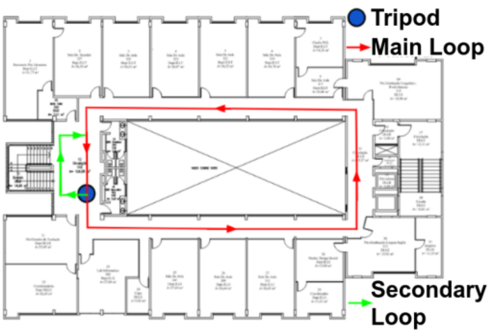

Figure 4. Floor plan of the corridor showing the 'Main Loop' and 'Secondary Loop'

3.1.3 Metric Scale Recovery

A critical step in both experiments was the application of the correct metric scale to the photogrammetric reconstructions. While Experiment 1 utilized a high-precision laser interferometer for displacement verification, the primary method for scale calibration across the scenarios relied on physical reference distances measured with a traceable calibrated tape measure. As demonstrated in Figure 5, multiple redundant reference distances were recorded within the physical environment.

Because scale accuracy is central to this evaluation methodology, it is necessary to quantify the uncertainty of these manual measurements. The uncertainty budget for the calibrated tape measure (used in Experiment 2) comprises the manufacturer's specified tolerance (e.g., typical Class II accuracy of ±1.3 mm at 5 m), reading error (estimated at ±1.0 mm), and thermal expansion effects (considered negligible in a temperature-controlled indoor setting). The combined standard uncertainty for a 5 m reference measurement is therefore conservatively estimated at approximately ±1.6 mm, yielding a relative uncertainty of 0.032%. In contrast, the Renishaw XL-80 laser interferometer used in Experiment 1 provides a precision of ±0.5 µm/m, rendering its contribution to the scale error virtually negligible.

The relationship between these known real-world distances and their corresponding unscaled distances in the COLMAP 3D model was used to compute the metric scale. We computed a single, global scaling factor derived from the median of the redundant physical measurements to ensure the final ground-truth trajectory was in a true metric scale. We recognize that applying a global scaling factor assumes a uniform scale across the entire trajectory, which may not hold true in complex geometries due to non-uniform scale drift. While integrating distance measurements directly into the global Bundle Adjustment (BA) optimization is theoretically more robust, the post-processing global scale factor was adopted to maintain a self-contained pipeline without requiring custom modifications to the core COLMAP algorithm. To mitigate the risks of local scale variations, the reference measurements were distributed across the capture area, though we acknowledge the uniform scale assumption as a limitation for significantly larger environments.

Figure 5. The calibrated tape measure being used to measure a reference distance in the environment.

3.2 Quantitative Results

This section presents the quantitative results, which are organized in two parts. First, the validation results from the linear guide experiment are presented to establish the metric accuracy and quality of our photogrammetric ground truth. Second, the validated methodology is applied to evaluate the ORB-SLAM 3 performance in the corridor experiment.

The core output of our ground-truth pipeline is a high-fidelity 3D reconstruction from which the clean, globally optimized camera trajectory is extracted. Figure 6 provides a visualization of this output from the linear guide experiment, showing the sparse 3D point cloud of the environment generated by COLMAP, with the

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-1147-2026 |

final ground-truth trajectory overlaid in red. Before this trajectory can be confidently used as a reference, its accuracy and quality must be independently verified. The following section provides this validation.

Figure 6. Visualization of the photogrammetric ground-truth trajectory (red) overlaid on the sparse 3D model of the linear guide environment generated by COLMAP.

3.2.1 Photogrammetry Validation

The data from the 3-meter linear guide experiment provides the primary validation for our methodology. The results are summarized in Table 1. Quantitative comparison of 3-meter linear displacement.Table 1, which compares the SLAM and photogrammetric trajectories against the high-precision laser interferometer reference. The data presents quantify two distinct types of error: the Trajectory Scale Error (TSE), which measures the accuracy of the total 3-meter displacement, and the rectitude error. This rectitude is quantified by the deviation metrics (MnD, MxD, and SD) which were calculated with respect to a mathematically fitted line representing the ideal linear path.

	Reference	Photogrammetry	SLAM
TSE (mm)	3000.0	2950.5	2902.7
TSE (%)	0.0	-1.6	-3.2
MnD (mm)	0.0	0.5	6.1
MxD (mm)	0.0	1.5	45.5
SD (mm)	0.0	0.3	5.6

Table 1. Quantitative comparison of 3-meter linear displacement.

As shown in Table 1, the laser confirmed a 3000.0 mm displacement. The photogrammetric method recorded this displacement with only a -1.65% scale error, while the SLAM trajectory had a larger error of -3.2%. These errors can be significantly affected by the manual definition of the displacement.

More importantly, the Table 1. Quantitative comparison of 3-meter linear displacement. quantifies the straightness of the trajectories. The photogrammetric method shows a maximum deviation from a fitted line of only 1.5 mm and a standard deviation of 0.3 mm. In contrast, the ORB-SLAM 3 trajectory exhibits a maximum deviation of 45.5 mm, with a standard deviation of 5.6 mm. The photogrammetry SD of the reconstruction is only 5.4% of the SLAM SD.

This quantitative difference in rectitude is also clearly visible in Figure 7. The photogrammetry trajectory (left) is sparse due to the 4 FPS capture rate but forms a clean, straight line. The ORB-SLAM 3 trajectory (right) is much denser but displays significant local noise and deviation from the linear path.

Figure 7. Visual comparison Experiment 1. Photogrammetry trajectory and ORB-SLAM 3 trajectory.

This visual and quantitative analysis is further detailed in the polar deviation plots. Figure 8 shows the deviations for the photogrammetric trajectory. These results visually show all points are contained well within a radius corresponding to the 1.5 mm maximum deviation. Conversely, Figure 9 shows the deviations for the SLAM trajectory, revealing a much wider spread that directly reflects its 45.5 mm maximum deviation, with numerous points plotted far from the centre.

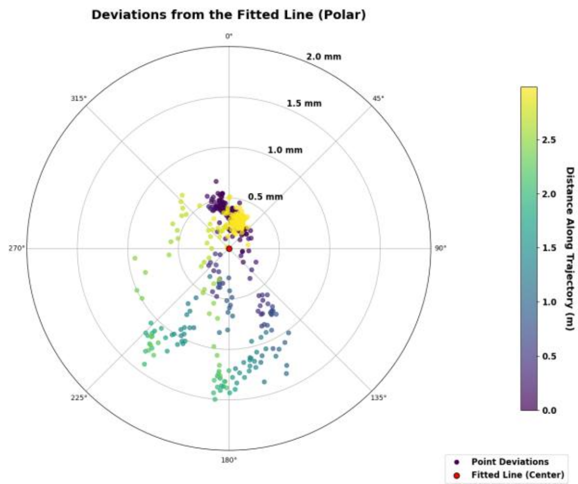

Figure 8. Polar plot of rectitude deviations for the Photogrammetry trajectory.

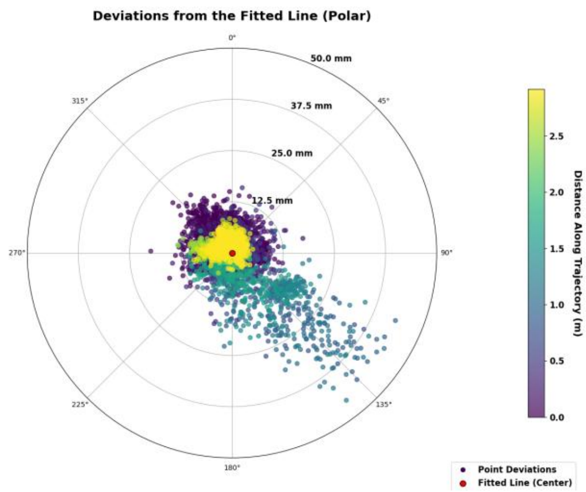

Figure 9. Polar plot of rectitude deviations for the ORB-SLAM 3 trajectory.

To complete the validation, we performed a quantitative comparison using the EVO tool. This time, our validated photogrammetric trajectory was treated as the ground truth and the ORB-SLAM 3 trajectory (from the same Experiment 1) was treated as the estimate. The results for APE and RPE, including their mean and RMSE statistics, are in Table 2.

Metric	RMSE (m)	Mean (m)	Max (m)
APE	1.4050	1.0788	3.8231
RPE	0.0694	0.0531	0.2659

Table 2. APE/RPE Analysis of ORB-SLAM 3 vs. Photogrammetric Ground Truth (Experiment 1).

The SLAM trajectory exhibited a significant APE, with an RMSE of 1.405 m and a maximum error of 3.823 m, despite traveling trajectory on the linear guide experiment. This error accumulation (drift) is visualized in Figure 10. The color map clearly shows the error is low (blue) at the beginning of the trajectory and progressively grows, reaching its maximum (red) at the end of the path.

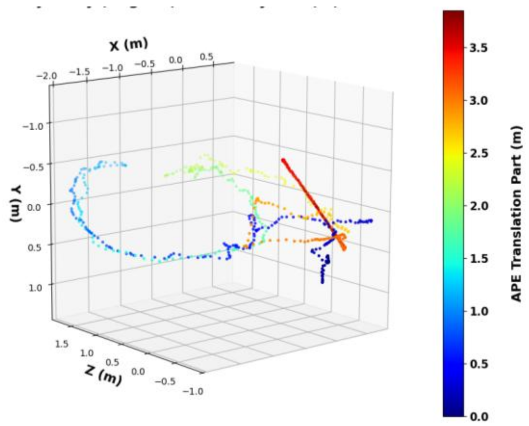

Figure 10. SLAM trajectory (Experiment 1) coloured by Absolute Pose Error (APE) in meters.

The RPE analysis, shown in Figure 11. The RPE, which measures local error, is visibly higher (yellow/red points) at the very beginning and end of the trajectory. This corresponds to the 5-second stationary pauses in the experimental procedure (Sec. 3.1.1), a challenging scenario where the SLAM system can accumulate drift error while stationary. Our methodology's ability to capture and visualize this specific SLAM behaviour demonstrates its high sensitivity and effectiveness as an evaluation tool.

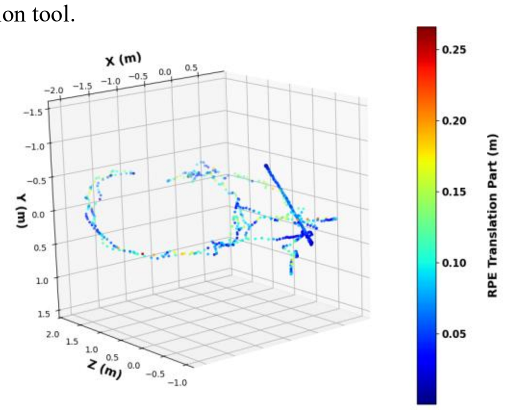

Figure 11. SLAM trajectory (Experiment 1) colored by Relative Pose Error (RPE) in meters.

3.2.2 SLAM Evaluation

Having established the accuracy and high quality of the photogrammetric ground truth in the previous section, this methodology was applied to evaluate the ORB-SLAM 3 (stereo-inertial) trajectory from the corridor experiment. The quantitative results for both the "Secondary Loop" and the 50 m "Main Loop" are presented in Table 3, summarizing the key error metrics computed by EVO.

Metric	RMSE (m)	Mean (m)	Max (m)
APE	9.9907	9.5761	15.2803
RPE	0.1362	0.1153	0.3242

Table 3. APE/RPE Analysis of ORB-SLAM 3 vs. Photogrammetric Ground Truth (Experiment 2).

The results for the "Main Loop", a trajectory of over 50 m, show a final APE of 9.991 m RMSE. This metric quantifies the total accumulated global drift over the entire run. The RPE, which measures the local, frame-to-frame translational drift, was 0.136 m RMSE.

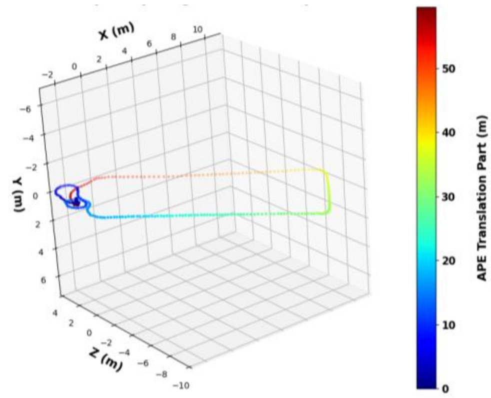

Figure 12. Visualization of Absolute Pose Error (APE) for the Experiment 2 trajectory.

In addition to the quantitative metrics, a qualitative analysis of the trajectory plots provides further insight into the SLAM system's behaviour. Figure 12 visualizes the APE along the 'Main Loop' trajectory. The color gradient, from blue (low error, 5.78 m) to red (high error, 15.28 m), clearly shows how the global drift accumulates over the path. The error is lowest at the beginning of the run and progressively increases, reaching its maximum value as the operator returns to the starting point, which aligns with the quantitative data in Table 3.

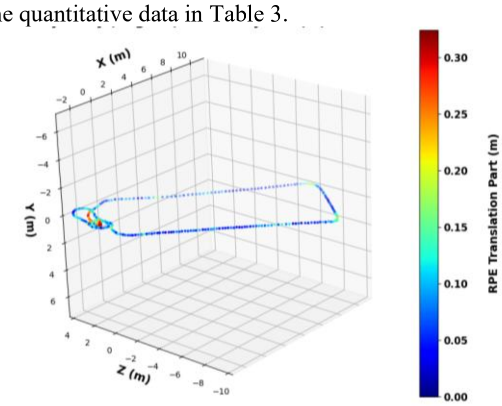

Figure 13. Visualization of Relative Pose Error (RPE) for the Experiment 2 trajectory.

Figure 13 visualizes the Relative Pose Error (RPE), which measures local drift. This plot shows that most of the trajectory has a very low local error (blue/green). However, it also highlights specific areas, particularly during sharp rotations, where the RPE spikes (red dots, max 0.32 m). This illustrates the methodology's potential to not only provide a single global error

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-1147-2026 |

number, but also to help pinpoint specific moments of high translational drift within this experiment.

4. Results Discussions

This section interprets the significance of the results presented in Section 3. First, we analysed the findings from the linear guide experiment, which validate the photogrammetric pipeline itself. Second, we discuss the corridor experiment's results, which required a specific methodological approach to make the comparison between the temporal SLAM and the atemporal photogrammetry possible and meaningful.

The linear guide experiment (Exp 1) provided the fundamental validation. The data in Table 1 is particularly revealing, as it directly compares the geometric quality of the trajectories with the reference linear guide. The results show the ORB-SLAM 3 trajectory had a maximum deviation of 45.5 mm from an adjusted line, while our 4 FPS photogrammetric trajectory had a maximum deviation of only 1.5 mm. This disparity is central to our paper’s argument: ORB-SLAM 3 is optimized for real-time performance, which necessitates compromises in local accuracy to maintain speed. Our method, using COLMAP, is an offline, global bundle adjustment. It processes all images jointly, without time constraints, to find the single most mathematically consistent solution. The globally optimized, even if sparse, trajectory is a far superior and more consistent reference. The full EVO analysis, presented in Table 2 (APE/RPE Results) and visualized in Figure 10 and Figure 11 (APE/RPE 3D Plots), further quantifies the SLAM's drift relative to this superior reference.

The corridor experiment (Exp 2) demonstrates the application of this validated method. However, comparing the SLAM trajectory to the photogrammetric ground truth presents a significant challenge: SLAM is temporal (estimating pose at a given time), while photogrammetry is atemporal (solving for geometry, not time). A direct, timestamp-to-timestamp comparison for APE is therefore impossible, as the raw COLMAP output lacks valid timestamps.

Our methodology initially overcame this by re-introducing a temporal basis. First, the correct chronological order of the photogrammetric poses was re-established by sorting them by their image filenames. Second, synthetic, linear timestamps (e.g., 0.0 to 100.0) were applied to both the ordered photogrammetric trajectory and the SLAM trajectory. While this "synchronization by relative progress" enables an APE calculation, the quantitative evaluation using the evo tool revealed a seemingly high APE (9.991 m in Experiment 2). Rather than indicating inherent SLAM failures, this inflated APE is an artifact of this synchronization strategy combined with a vast discrepancy in trajectory density. ORB-SLAM 3 processes images at 30 FPS, generating a highly dense trajectory, whereas the photogrammetric reference was deliberately captured at 4 FPS. Aligning these vastly different densities using synthetic relative progress forces point-to-point associations that accumulate artificial deviations, inherently inflating the global APE metric.

To resolve this methodological gap and eliminate constant-velocity biases, future iterations must bypass relative progress entirely by storing a reference lookup table during data acquisition, directly mapping original ROS header timestamps to the saved image filenames.

Conversely, the RPE, which evaluates local frame-to-frame consistency, is far less affected by this global timestamp synchronization. As seen in Figure 13, it successfully diagnoses where the SLAM system failed locally, such as during sharp rotations, yielding values (0.136 m RMSE) consistent with literature expectations. Regarding the 4 FPS capture rate, it proved sufficient to prevent reconstruction gaps, and the use of global shutter sensors successfully mitigated motion blur during sudden rotations.

Further limitations include the computationally intensive, offline nature of the COLMAP pipeline and the reliance on manual measurements (the calibrated tape measure) to set the global scale. Furthermore, a fundamental limitation is the lack of strict sensor independence: because both rely on the exact same sensor data, they are susceptible to failing under comparable circumstances.

Lastly, while our photogrammetric method still relies on environmental texture to function, this constraint is mitigated by fully utilizing the sensor's spatial capacity. Although the raw images are captured and saved in color, it is important to acknowledge that the standard feature extraction process in COLMAP (e.g., SIFT) operates on grayscale intensities. Therefore, the primary advantage of the photogrammetric pipeline over the SLAM estimation does not stem from color information, but rather from processing the unmodified, high-resolution (2448 x 2048) imagery. This high spatial resolution allows for the extraction of finer, more reliable geometric details, whereas the real-time SLAM system must actively downscale the input to a low-resolution (800 x 600) format to maintain high-frequency processing speeds.

5. Conclusions

This paper addressed the significant challenge of acquiring accurate ground truth for vSLAM evaluation, which often relies on expensive external measurement infrastructure. We presented and validated a low-cost, self-contained methodology that generates a high-fidelity ground-truth trajectory using offline photogrammetry from the vSLAM system's own input imagery.

The methodology's validity was first established in a laboratory experiment, which used a linear guide with a laser interferometer to evaluate that our photogrammetric trajectory was not only metrically accurate but also significantly cleaner and more reliable than the real-time SLAM output. This validated framework was then successfully applied in a large-scale corridor (Exp 2), proving its effectiveness in quantifying complex, real-world vSLAM performance, including total geometric drift (APE) and local inconsistencies (RPE). A key contribution was solving the temporal for atemporal comparison challenge through synchronization by relative progress.

Based on the validation achieved in the current study, future work will focus on several key areas. First, automating the scale recovery process to reduce human intervention, and investigating faster global optimization techniques to reduce offline processing time. More importantly, building upon the positive RPE evaluation achieved in the corridor experiment which successfully diagnosed local SLAM drift in a moderately complex indoor setting the next logical progression is to stress-test this methodology in more severe, unstructured environments. While the current results robustly validate the core framework, future deployments will target highly degraded "real-world" scenarios. These include environments with poor illumination, dynamic objects, and textureless surfaces that actively challenge feature extraction for both the real-time SLAM and the photogrammetric reconstruction. Ultimately, this leads to our goal of adapting this evaluation framework for underwater

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-1147-2026

environments, which inherently present these extreme challenges for vision-based navigation.

In conclusion, this work provides the robotics and photogrammetry community with an accessible, low-cost, and validated framework for vSLAM evaluation. This framework removes the dependency on expensive, external measurement infrastructure and opens a path for its application in challenging new domains.

Acknowledgements

The authors would like to thank FINEP for its financial support of the VORIS project, the scholarship from the National Council for Scientific and Technological Development (CNPq) of the Ministry of Science, Technology, and Innovation (MCTI), and the Metrology and Instrumentation Laboratory of UFSC (Federal University of Santa Catarina) for providing space for research development.

References

Ascard, M., Movahedi, F., 2023. Assessing the Efficiency of COLMAP, DROID-SLAM, and NeRF-SLAM in 3D Road Scene Reconstruction. Master's Theses in Mathematical Sciences, ISSN 1404-6342.

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

Campos, C., Elvira, R., Rodriguez, J. J. G., Montiel, J. M. M., Tardós, J. D., 2021: ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. IEEE Transactions on Robotics, 37(6), 1874-1890.

Chen, W., Shang, G., Ji, A., Zhou, C., Wang, X., Xu, C., Li, Z., Hu, K., 2022: An Overview on Visual SLAM: From Tradition to Semantic. Remote Sensing, 14, 3010.

Cort, C., Dalal, A., De Ocampo, J., Du, Y., Feng, S., Vela, P. A., 2024. Low-Cost Real-World Benchmark Framework for Closed-Loop Robot Navigation Systems with SLAM. Proceedings of the 2024 IEEE Opportunity Research Scholars Symposium (ORSS), Atlanta, GA, USA, pp. 53-56.

Fontan, A., Civera, J., Fischer, T., Milford, M., 2024. Look Ma, No Ground Truth! Ground-Truth-Free Tuning of Structure from Motion and Visual SLAM. ArXiv, abs/2412.01116.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354-3361.

Granshaw, S., 1980: Bundle Adjustment Methods in Engineering Photogrammetry. The Photogrammetric Record, 10, 181-207.

Grupp, M., 2017. evo: Python package for the evaluation of odometry and SLAM. https://github.com/MichaelGrupp/evo (accessed 30 October 2025).

Ha, S., Li, Q., Yu, X., Queralta, J., Zou, Z., Westerlund, T., 2022. A Benchmark for Multi-Modal Lidar SLAM with Ground Truth in GNSS-Denied Environments. ArXiv, abs/2210.00812.

James, M., Robson, S., Smith, M., 2017: 3-D uncertainty-based topographic change detection with structure-from-motion photogrammetry: precision maps for ground control and directly georeferenced surveys. Earth Surface Processes and Landforms, 42, 1769-1788.

Nair, A. D., Kindle, J., Levchev, P., Scaramuzza, D., 2024: Hilti SLAM Challenge 2023: Benchmarking Single + Multi-session SLAM across Sensor Constellations in Construction. IEEE Robotics and Automation Letters, 9(8), 7596-7603.

Nistér, D., Stewénius, H., 2006. Scalable Recognition with a Vocabulary Tree. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Vol. 2, 2161-2168.

Salas, M., Latif, Y., Reid, I. D., Montiel, J., 2015. Trajectory alignment and evaluation in SLAM: Horn’s method vs alignment on the manifold. Proceedings of the Workshop on The Problem of Mobile Sensors (MoSen'15) at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

Schönberger, J. L., Frahm, J.-M., 2016a. Structure-from-Motion Revisited. Conference on Computer Vision and Pattern Recognition (CVPR).

Schönberger, J. L., Zheng, E., Pollefeys, M., Frahm, J.-M., 2016b. Pixelwise View Selection for Unstructured Multi-View Stereo. European Conference on Computer Vision (ECCV).

Schubert, D., Goll, T., Demmel, N., Usenko, V., Stückler, J., Cremers, D., 2018. The TUM VI Benchmark for Evaluating Visual-Inertial Odometry. Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1680-1687.

Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D., 2012. A benchmark for the evaluation of RGB-D SLAM systems. Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 573-580.

Yang, Y., Xu, B., Li, Y., Schwertfeger, S., 2023. The SLAM Hive Benchmarking Suite. Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, United Kingdom, pp. 11257-11263.

Yarovoi, A., Cho, Y. K., 2024: Review of simultaneous localization and mapping (SLAM) for construction robotics applications. Automation in Construction, 162, 105344.

Yin, J., Li, A., Li, T., Yu, W., Zou, D., 2022: M2DGR: A Multi-Sensor and Multi-Scenario SLAM Dataset for Ground Robots. IEEE Robotics and Automation Letters, 7(2), 2266-2273.