SMapper: A Multi-Modal Data Acquisition Platform for SLAM Benchmarking

Pedro Miguel Bastos Soares[1,*], Ali Tourani[1,2], Miguel Fernandez-Cortizas[1],
Asier Bikandi-Noya[1], Holger Voos[1,3], and Jose Luis Sanchez-Lopez[1] [*†‡§¶]

October 13, 2025

Abstract

Advancing research in fields such as Simultaneous Localization and Mapping (SLAM) and autonomous navigation critically depends on the availability of reliable and reproducible multimodal datasets. While several influential datasets have driven progress in these domains, they often suffer from limitations in sensing modalities, environmental diversity, and the reproducibility of the underlying hardware setups. To address these challenges, this paper introduces *SMapper*, a novel open-hardware, multi-sensor platform designed explicitly for, though not limited to, SLAM research. The device integrates synchronized LiDAR, multi-camera, and inertial sensing, supported by a robust calibration and synchronization pipeline that ensures precise spatio-temporal alignment across modalities. Its open and replicable design allows researchers to extend its capabilities and reproduce experiments across both handheld and robot-mounted scenarios. To demonstrate its practicality, we additionally release *SMapper-light*, a publicly available SLAM dataset containing representative indoor and outdoor sequences. The dataset includes tightly synchronized multimodal data and ground truth trajectories derived from offline LiDAR-based SLAM with sub-centimeter accuracy, alongside dense 3D reconstructions. Furthermore, the paper contains benchmarking results on state-of-the-art LiDAR and visual SLAM frameworks using the *SMapper-light* dataset. By combining open-hardware design, reproducible data collection, and comprehensive benchmarking, *SMapper* establishes a robust foundation for advancing SLAM algorithm development, evaluation, and reproducibility. The project's documentation, including source code, CAD models, and dataset links, is publicly available at `https://snt-arg.github.io/smapper_docs/`.

1 Introduction

Collecting reliable high-quality data is a critical task in robotics research, driving progress in autonomous navigation, Simultaneous Localization and Mapping (SLAM), and enhancing robots' perceptual and situational awareness [1]. Such data instances not only provide the foundation for developing novel algorithms against precise ground truth, but also serve as the benchmark for comparing diverse approaches, ensuring that robotic systems can operate robustly in diverse environments [2, 3]. In particular, progress in SLAM critically depends on access to multimodal data collected across dissimilar scenarios, where evaluation under varying motion dynamics, lighting conditions, and structural layouts is essential for achieving robustness, accuracy, and generalization [4, 5].

During the past decade, several influential and practical datasets, such as KITTI [6], EuRoC [7], and TUM RGB-D [8], have significantly shaped the field of SLAM and served as evaluation benchmarks for many developed systems. However, they often suffer from limitations in sensing modalities, the diversity of environments, and the availability of reproducible hardware setups for further data collection. As a result, the fixed nature of existing data instances

---

[*1]Automation and Robotics Research Group (ARG), Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, L-1359 Luxembourg, Luxembourg. pedro.bastos@ext.uni.lu (PM.BS.); ali.tourani@uni.lu (A.T.); miguel.fernandez@uni.lu (M.FC.); asier.bikandi@uni.lu (A.B.); holger.voos@uni.lu (H.V.); joseluis.sanchezlopez@uni.lu (JL.SL.);

[†2]Institute for Advanced Studies (IAS), University of Luxembourg, L-4365 Esch-sur-Alzette, Luxembourg.

[‡3]Faculty of Science, Technology and Medicine, University of Luxembourg, L-4365 Esch-sur-Alzette, Luxembourg.

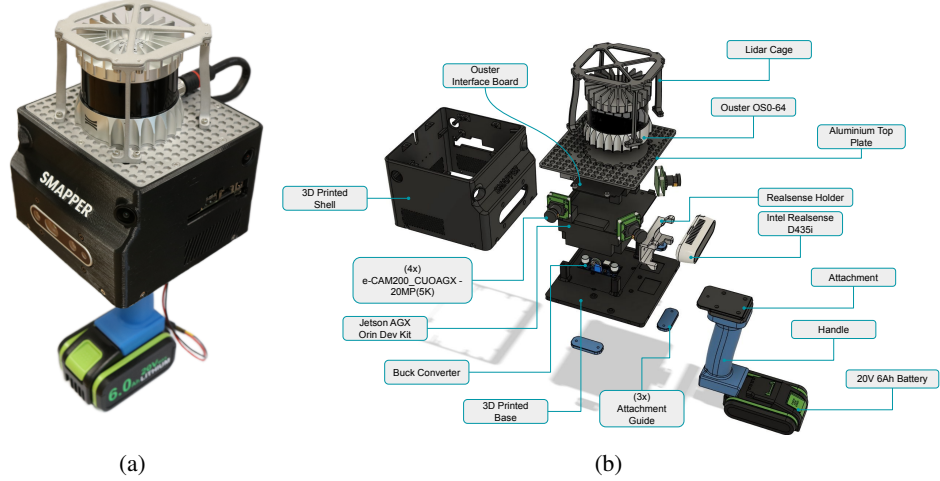arXiv:2509.09509v2 [cs.RO] 10 Oct 2025

Figure 1: Overview of the *SMapper* platform: **(a)** the fully assembled physical prototype, **(b)** the system overview highlighting the integrated sensors, components, and design elements tailored for diverse SLAM scenarios.

constrains researchers and industries, making it challenging to replicate sensor configurations or extend them to new scenarios.

To address such challenges, multimodal sensing has emerged as a reliable solution to advance SLAM data collection and processing [9]. While unimodal sensors such as visual sensors (*e.g.,* monocular, stereo, RGB-D, or event cameras), Light Detection And Ranging (LiDAR), and Inertial Measurement Units (IMUs) provide the fundamental measurements for SLAM, integrating them into multimodal configurations adds complementary data streams that enable more robust, accurate, and reliable system development [10]. Additionally, these sensing modalities are often complemented by external systems, such as Motion Capture System (MCS) or GPS, which provide accurate ground truth reference trajectories for SLAM evaluation and benchmarking [11]. However, most existing efforts do not release both the *collected datasets* and the *underlying hardware design*, leaving a gap in reproducibility and extensibility for the SLAM community.

This paper presents *SMapper*, a novel, multi-sensor, open-hardware platform specifically designed for data collection in SLAM, covering visual, LiDAR, and multi-sensor SLAM variants. The platform integrates synchronized visual, LiDAR, and inertial sensing, and can be replicated by researchers owing to its open hardware specification and detailed technical documentation. Furthermore, we release a publicly available Visual SLAM (VSLAM) dataset to demonstrate the practicality of the platform for benchmarking and evaluation. The dataset includes trajectories derived from visual, inertial, and LiDAR sensing, where the LiDAR data are processed using a high-accuracy LiDAR SLAM system to generate trajectory estimates that serve as the ground truth reference for evaluating visual–inertial SLAM. An overview of the proposed *SMapper* device is presented in Fig. 1, depicting both the fully assembled physical prototype and the inner sensor and component configuration designed for comprehensive data acquisition.

With these, the paper offers the following contributions:

- A compact, open-hardware, and fully synchronized multimodal SLAM data collection platform, integrating LiDAR, IMU, and visual sensors, supporting both handheld and robot-mounted configurations,

- A publicly available multimodal dataset for VSLAM, accompanied by benchmarking results to depict the platform's performance across diverse scenarios; and

- An open-source software suite for sensor calibration, data acquisition, and real-time monitoring, facilitating reproducible and efficient data collection.

The remainder of the paper is structured as follows: Section 2 reviews related works and various dataset collection setups for SLAM. In Section 3, the design and technical specifications of the proposed device are detailed. Section 4.2 outlines the data collection methodology and setup, with benchmarking results and evaluations presented in Section 4. The paper concludes and discusses potential directions for future work in Section 6.

2   Related Works

Many efforts have explored diverse sensing modalities to collect reliable experimental and ground truth data for the SLAM and VSLAM domains [12]. Datasets such as TUM RGB-D [8] and OpenLORIS [13] provide accessible benchmarks through simple setups (*e.g.,* handheld vision sensors or mobile robots), while others utilize custom multimodal platforms. Such platforms tightly integrate complementary sensing modalities, including LiDAR, vision, inertial, and GNSS, to tackle complex challenges in navigation, large-scale mapping, and multimodal fusion benchmarking.

Accordingly, the Newer College Dataset [14] was collected using a hand-carried platform integrating a RealSense camera, a 64-beam Ouster OS-1 3D LiDAR with embedded IMU, and an onboard Intel NUC[1]. It comprises seven outdoor sequences of different durations, capturing a range of trajectories that include loops, spins, and other challenging motion patterns. While offering millimeter-accurate ground truth via a tripod-mounted survey-grade scanner, the setup faces challenges in terms of limited Field of View (FoV) and sensor synchronization. To address these, Zhang *et al.* [15] extended this dataset by incorporating a 128-beam Ouster OS-0 and a four-camera Alphasense kit, enhancing FoV and spatiotemporal resolution[2]. The resulting dataset remains confined to a few outdoor sequences with varying trajectories, motion aggressiveness, and scene complexity. Another recent contribution is the Oxford Spires dataset [16], collected using a custom-built multi-sensor perception unit in combination with a millimetre-accurate map obtained from a Terrestrial LiDAR Scanner (TLS). The dataset comprises 24 indoor and outdoor sequences covering a range of challenging scenarios and environmental conditions[3]. The perception unit integrates a 64-beam Hesai QT64 LiDAR, three global-shutter Alphasense Core fisheye cameras, and a synchronized IMU, all of which are precisely calibrated. Similarly, the Phasma device in the Hilti-Oxford Dataset [17] utilizes a 32-channel PandarXT-32 LiDAR, five infrared cameras, and an IMU. It includes a diverse set of sequences from indoor and outdoor construction sites, capturing a broad spectrum of challenging scenarios. A $Z + F$ Imager laser scanner is used to obtain prior maps of the environments as ground truth. However, this methodology faces high setup complexity and is less suited for dynamic, real-time environments.

PALoc [18] proposed a prior map-assisted framework for ground truth pose generation in LiDAR SLAM[4]. The device, supporting both handheld and robot-mounted configurations, captures dense 6-Degree of Freedom (DoF) trajectories across indoor and outdoor settings, but its reliance on monocular vision and LiDAR limits generalization to multimodal or vision-centric tasks. Inspired by PALoc, FusionPortableV2 [19] introduced a versatile multi-sensor platform for SLAM and autonomous driving data collection[5]. It integrates a 128-beam Ouster OS-1 LiDAR with IMU, two FLIR-BFS monochrome cameras, two DAVIS-346 event cameras, and a dual-antenna GNSS/Inertial Navigation System (INS) sensor. The dataset includes sequences captured in both indoor and outdoor environments, spanning handheld operation as well as deployments on legged robots, wheeled robots, and high-speed vehicles. While FusionPortableV2 enhances sensing richness, it faces high complexity in calibration and data alignment across modalities. In another work, GEODE [20] provides a versatile robotics platform integrating GNSS, motion capture, and multiple 3D laser scanners to support ground truth generation. While designed to address LiDAR degeneracies using diverse sensors, generalizing the platform introduced challenges in maintaining calibration and synchronization, causing increased sensitivity to synchronization drift. MapEval [21] focuses on standardizing evaluation for SLAM point cloud maps, emphasizing metrics for both global geometry and local consistency. It utilizes a simpler multi-sensor platform, comprising a 3D LiDAR and an SBG INS, primarily tailored for LiDAR-based SLAM scenarios. Besides evaluating on FusionPortableV2, Newer College, and GEODE datasets, the authors collected a 15-sequence in-house dataset. However, the platform requires careful adjustment of voxel size in relation to scene complexity and map density.

Table 1 summarizes the surveyed platforms and datasets that have significantly contributed to the field of SLAM datasets. The review of existing data collection platforms exposes a consistent trade-off inherent in their design. On one hand, platforms like FusionPortableV2 offer immense sensor richness but at the cost of high complexity in calibration and data management. On the other hand, simpler setups are more accessible but are often constrained by a limited number of sensor modalities or a narrow field of view. Furthermore, a common challenge across many of these systems is the complexity of the calibration and synchronization pipeline, which remains a significant barrier to practical use. Moreover, despite their value, only a limited number of platforms provide openly available hardware designs or software tools, which restricts reproducibility and hinders broader adoption within the research community.

To bridge these gaps and achieve a balance between affordability, reproducibility, and sensor fusion, we introduce *SMapper*: a compact, tightly synchronized multimodal data collection platform tailored for SLAM benchmarking as well as broader robotics research. Unlike prior efforts limited by narrow sensor modalities or complex calibration pipelines, *SMapper* integrates a multi-camera setup with a wide overlapping FoV, enabling robust perception across

---

[1] https://ori-drs.github.io/newer-college-dataset/stereo-cam/
[2] https://ori-drs.github.io/newer-college-dataset/multi-cam/
[3] https://dynamic.robots.ox.ac.uk/datasets/oxford-spires/
[4] https://github.com/JokerJohn/PALoc
[5] https://fusionportable.github.io/dataset/fusionportable_v2/

Table 1: Various multimodal devices and platforms designed for collecting SLAM datasets. Terrestrial LiDAR Scanner (TLS) and Motion Capture System (MCS) denote common sources of ground truth data.

Device/Dataset	Year	Ground Truth	Sensor Modality				Platform		Environment	Openness	
			Vision	Depth	LiDAR	IMU	Handheld	Robot	Variant	Hardware	Tools
TUM RGB-D [8]	2012	MCS	✓	✓	✗	✓	✓	✓	indoor	✗	✓
OpenLORIS [13]	2020	LiDAR + MCS	✓	✓	✓	✓	✗	✓	indoor	✗	✓
Newer College [14]	2020	LiDAR	✓	✓	✓	✓	✓	✗	outdoor	✗	✗
Zhang et al. [15]	2021	LiDAR	✓	✗	✓	✓	✓	✗	outdoor	✗	✗
Hilti-Oxford [17]	2022	Laser Scanner	✓	✓	✓	✓	✓	✗	in/outdoor	✗	✗
PALoc [18]	2024	MCS	✓	✗	✓	✓	✓	✓	in/outdoor	✗	✓
FusionPortable [19]	2024	Laser Scanner	✓	✓	✓	✓	✓	✓	in/outdoor	✗	✓
GEODE [20]	2024	LiDAR + MCS	✓	✓	✓	✓	✓	✓	in/outdoor	✗	✓
MapEval [21]	2025	Laser Scanner	✗	✗	✓	✓	✓	✓	in/outdoor	✗	✓
Oxford Spires [16]	2025	TLS	✓	✗	✓	✓	✓	✓	in/outdoor	✗	✗
SMapper (ours)	2025	LiDAR	✓	✓	✓	✓	✓	✓	in/outdoor	✓	✓

diverse environments. The platform supports both handheld and robot-mounted use cases, designed for ease of calibration and extensibility, facilitating the collection of high-quality, reproducible datasets. Furthermore, we release the open-hardware design, open-source calibration and acquisition software, and a set of publicly available dataset instances to demonstrate the platform's practicality for SLAM benchmarking.

3 System Overview

As shown in Fig. 1, *SMapper* integrates various components and sensors within a modular structure that supports both *handheld* use (via a detachable handle mounted to the base) and *direct mounting* for ground robots. The sensors and companion computer are rigidly housed in a custom-designed, 3D-printed base, topped with an aluminum plate that serves as the mounting platform for the LiDAR and its protective cage. The complete system weighs ~ 2.5kg (or 1.7kg without the handle and battery). It features a compact form factor of approximately 15cm × 15cm × 38.4cm (or 15cm × 15cm × 19.2cm without the handle and battery), enabling convenient single-handed operation. The *SMapper* device is equipped with an auxiliary onboard NVIDIA Jetson AGX Orin Developer Kit computer. This high-performance embedded computer enables efficient sensor data synchronization, real-time data recording and processing, and potential on-device inference. It is powered by an NVIDIA GPU with 2048 NVIDIA CUDA cores and a 12-core Arm Cortex-A78AE 64-bit CPU, making it well-suited for demanding robotics applications.

3.1 Sensors

Table 2 summarizes the specifications of the computing unit and sensors integrated into the *SMapper* device. According to the table, the device simultaneously captures data from various sources, supporting multimodal data acquisition. The platform includes a 64-beam Ouster OS0 3D LiDAR, offering a 100-meter range and an ultra-wide 360×90° coverage. Designed for short-range, high-precision perception, this sensor captures fine geometric details, making it well-suited for mobile robotics and widely adopted in LiDAR-based SLAM research. The device captures visual data from two sources: a forward-facing Intel RealSense D435i camera for recording RGB-D data and four synchronized e-CAM200 CUOAGX electronic rolling shutter cameras. The RealSense is mounted with a 0-degree pitch angle. The e-CAM200 cameras are particularly designed as a multi-vision solution for the Jetson computer, offering synchronization and consistent exposure for accurate multi-source perception. Integrating two types of visual sensors enables the platform to support a broad spectrum of robotic perception tasks, from depth-based analysis to multi-view geometry, as well as simplifying the benchmarking of diverse visual processing pipelines. While the FoV of the RealSense camera is 69° × 42°, the strategically placed e-CAM200 cameras extend the coverage to approximately 270° × 66°, offering a wide field suitable for situational awareness tasks. This configuration also produces an overlapping FoV of 30° among front and side e-CAM200 cameras.

Moreover, *SMapper* utilizes the built-in IMUs of the LiDAR and RealSense camera, providing three-axis gyroscope and accelerometer measurements with the angular velocity up to 400 Hz and linear acceleration up to 200 Hz for more robust pose estimation while recording data.

Table 2: An overview of the hardware, including sensors and the embedded computer employed in the *SMapper* device.

Hardware	Type	Rate	Characteristics
LiDAR	Ouster OS0	10/20 Hz	64 Channels, 100m Range FoV: $360° \times 90°$ Resolution: 1024×64
Cameras	$4\times$ e-CAM200 CUOAGX	30 Hz	Rolling shutter, RGB FoV: $90° \times 66°$ Resolution: 2K Synchronized
	Intel Realsense D435i	30 Hz	Global shutter, RGB-D FoV RGB: $69° \times 42°$ FoV Depth: $87° \times 58°$ Resolution: 2K Synchronized
IMU	LiDAR IMU	100 Hz	3-axis Gyroscope 3-axis Accelerometer
	Camera IMU	400 Hz	3-axis Gyroscope 3-axis Accelerometer
Onboard Computer	Jetson AGX Orin Developer Kit	1.3 GHz (Processor)	CPU: 12-core ARM GPU: NVIDIA 2048-core Memory: 64GB Vision/DL Accelerator

3.2 Synchronization Procedure

As an essential performance aspect of heterogeneous sensing, temporal synchronization ensures that measurements from different modalities are aligned within a shared time frame. This alignment is crucial for consistent data fusion, particularly in dynamic environments where even slight temporal offsets can result in significant errors in perception, mapping, and localization. Although *SMapper* does not guarantee full hardware synchronization across all sensors, it enforces a shared global clock, allowing users to apply different synchronization policies depending on the needs of their application. Accordingly, all data streams are time-stamped at the time of acquisition and corrected for sensor-specific readout delays, ensuring temporal coherence suitable for multi-view fusion.

The synchronization driver implements three selectable timestamping modes:

- `TIME_FROM_ROS` (default): Timestamps are generated directly from the Robot Operating System (ROS) system clock when frames are received. Owing to the low-latency, event-driven architecture of the driver, this mode provides sufficiently accurate timestamps for most use cases with minimal configuration effort.

- `TIME_FROM_TSC`: In this mode, timestamps are derived from the hardware Timestamp System Counter (TSC) of the NVIDIA Jetson platform. The TSC provides a stable, high-resolution clock source that reduces jitter compared to the ROS clock, offering improved accuracy without requiring network synchronization. This mode is beneficial for experiments demanding tighter temporal consistency while maintaining a lightweight setup.

- `TIME_FROM_PTP`: For the highest fidelity, the system supports Precision Time Protocol (PTP). The Jetson Orin AGX is configured as the PTP grandmaster clock, while the Ouster LiDAR operates as a PTP slave synchronized via `linuxptp`. In this mode, the driver converts raw TSC values into the PTP-synchronized time domain, ensuring that all sensors share a common, high-precision reference clock. This mode enables sub-millisecond alignment across modalities, which is essential for demanding SLAM and sensor fusion applications.

It should be added that a primary design decision in *SMapper* is to provide accurately timestamped raw sensor streams rather than pre-bundled synchronized ROS messages. This choice reflects the fact that modern SLAM frameworks implement their own synchronization strategies, making them specifically optimized to work with raw data. Pre-bundling at the driver level would reduce flexibility in applying custom synchronization policies and significantly increase storage requirements due to redundant data duplication. By aligning raw data streams to a high-precision standard clock, *SMapper* ensures temporal consistency while offering a compatible and storage-efficient
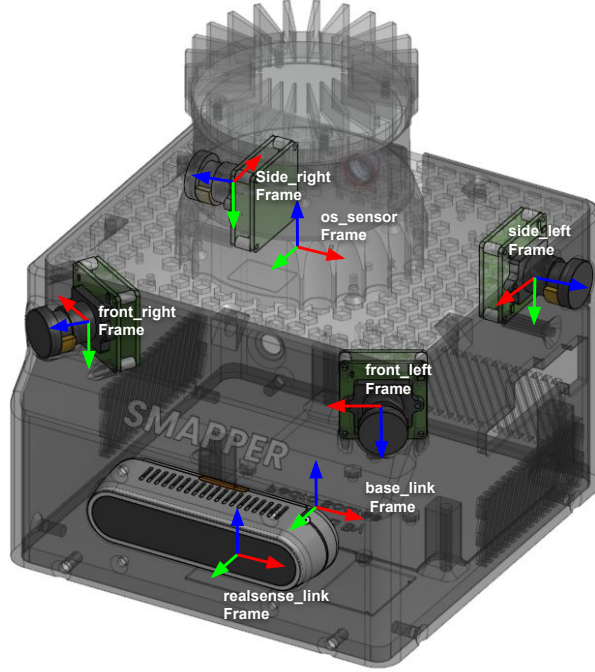
Figure 2: Coordinate frames of the *SMapper* device, containing the spatial configuration of the cameras, LiDAR, and IMU sensors. The device contains the following coordinate systems: the base, RealSense D435i, e-CAM200 cameras, LiDAR, LiDAR IMU, and RealSense IMU.

data format, maximizing its utility for diverse SLAM pipelines.

3.3 Calibration Procedure

Sensor calibration is essential to the practical usability and accuracy of *SMapper*, as precise spatio-temporal calibration forms the foundation for reliable data fusion and downstream SLAM tasks. To address this, we developed a comprehensive IMU-centric calibration pipeline, equipped with automation tools to streamline the process, and complemented by multi-faceted validation to ensure the reliability of the results. As illustrated in Fig. 2, which depicts the spatial arrangement of the onboard sensors and their coordinate frames, the base module, cameras, LiDAR, and IMU are each associated with distinct reference frames. Defining these configurations is essential for accurate sensor fusion and consistent pose estimation in multimodal perception tasks. In this paper, we demonstrate both the manual `Kalibr`-based procedure (§3.3.1) and our automated calibration framework (§3.3.2), highlighting how the latter simplifies the process while maintaining accuracy and repeatability.

3.3.1 Manual Calibration

Accordingly, a robust, multi-step pipeline was developed based on the widely used `Kalibr` toolbox [22], which uses the high-frequency IMU as a standard "*bridge*" to link the different sensor modalities. The calibration pipeline consists of the following steps:

1. *IMU noise characterization*: A 20-hour static dataset was recorded using the Ouster and RealSense IMUs independently. The data were processed with the `allan_variance_ros` tool [23] to estimate gyroscope and accelerometer noise density (white noise) and bias instability (random walk). These parameters are the critical inputs for the `Kalibr` optimization process and subsequent state estimation algorithms.

2. *Camera intrinsic and extrinsic calibration*: This stage estimates each camera's internal parameters (intrinsics) and its rigid 3D pose relative to a reference IMU (extrinsics). A rigid 6×6 AprilTag grid (80×80 cm^2) [24] was used as the calibration target, with data collected by manually moving the device to capture the board from diverse viewpoints. The `Kalibr` toolbox [22] was employed to jointly optimize the camera intrinsics (focal length, principal point, distortion) and the camera-to-IMU transformations. For the four e-CAM200 cameras, extrinsics were computed relative to the Ouster IMU. In contrast, for the RealSense sensor, both its

Table 3: Mean reprojection errors from the `Kalibr` extrinsic calibration, reported in *pixels*.

Camera	*Reference IMU*	*Reprojection Error (Mean ± Std)*
Front Left	Ouster OS0	0.34 ± 0.46
Front Right	Ouster OS0	0.34 ± 0.34
Side Left	Ouster OS0	0.43 ± 0.52
Side Right	Ouster OS0	0.35 ± 0.38
RealSense	RealSense D435i	0.64 ± 0.60

Table 4: Comparison of extrinsic parameters of the primary camera array: `CAD` model vs. `Kalibr` estimates with Euler angles and differences.

Camera	*Source*	*Translation* $[x, y, z]$ (m)	*Euler Angles* $[X, Y, Z]$ (°)	*Position* **Diff.** (m)	*Angular* **Diff.** (°)
Front Left	CAD	[0.046, 0.057, 0.087]	[134.8, -45.4, -44.8]	0.018	1.98
	Kalibr	[0.061, 0.064, 0.085]	[133.6, -44.6, -46.0]		
Front Right	CAD	[0.046, -0.057, 0.087]	[45.4, -90.0, 0.0]	0.013	1.04
	Kalibr	[0.057, -0.063, 0.085]	[44.6, -89.9, 1.2]		
Side Left	CAD	[-0.057, 0.054, 0.087]	[180.0, 0.0, -90.0]	0.035	0.95
	Kalibr	[-0.060, 0.088, 0.080]	[179.2, -0.7, -89.7]		
Side Right	CAD	[-0.057, -0.054, 0.087]	[0.0, -90.0, 90.0]	0.025	1.02
	Kalibr	[-0.064, -0.077, 0.079]	[0.7, -89.5, 89.3]		

RGB camera and integrated IMU were calibrated against the Ouster IMU, yielding consistent transformations across modalities.

3. *Assembling the full transformation tree*: The final step integrates all calibration results into a unified transformation tree for the device. The camera-to-IMU transformations estimated by `Kalibr` are chained with manufacturer-provided extrinsics (*e.g.,* the known offset between the LiDAR and IMU frames), generating the precise 3D pose of every sensor relative to the `base_link` frame.

These calibrated sensor streams, including camera feeds, LiDAR point clouds, and inertial measurements, are published during data collection using the *SMapper* device. Additional technical details, documentation, and setup instructions for calibration procedures are publicly available at `https://snt-arg.github.io/smapper_docs/`.

3.3.2 Automated Calibration

The manual `Kalibr`-based calibration workflow is often complex, repetitive, and prone to errors. To overcome these challenges, we developed the `smapper_toolbox`, a Python command-line utility that fully automates the process. The framework is publicly available in `https://github.com/snt-arg/smapper_toolbox`. Configured through a simple `YAML` file, the toolbox provides a Dockerized environment for executing each stage of the pipeline, including `ROS2-to-ROS1` bag conversion for compatibility with `Kalibr`, execution of intrinsics, and the generation of a ready-to-use `ROS2` launch file that publishes the device's static transforms. The output contains a complete Transformation Tree (TF) for the device, ensuring that the calibrated sensor poses can be readily integrated into downstream SLAM pipelines. This automation streamlines calibration into a robust, accessible, and repeatable process, representing a key contribution of *SMapper*.

3.3.3 Quantitative Validation

Three quantitative and qualitative checks were performed to validate the calibration results. First, the **reprojection error** reported by `Kalibr` was used to assess the internal consistency of the optimization. As shown in Table 3, all cameras achieved low mean reprojection errors, indicating a high-quality solution. Second, the calibrated extrinsics were compared against the nominal sensor placements from the `CAD` model. As shown in Table 4, the slight deviations between design specifications and estimated values confirm that the results are physically plausible. Finally, a visual validation was conducted using a custom point cloud colorization tool, which projects LiDAR points onto the image

7

<div align="center">(a)   (b)</div>

Figure 3: Qualitative calibration validation using point cloud colorization with the `front-right` camera. (**a**) raw camera image; (**b**) LiDAR point cloud colored by the projected image. While the alignment is generally consistent, minor misalignments are visible at building edges, reflecting residual calibration errors.

Table 5: The characteristics of the collected dataset, titled *SMapper-light*. All sequences are stored as ROS bag files in the `.mcap` format, containing synchronized data streams from all onboard sensors.

Scenario	*Instance*	*Length*	*Duration*	*Size*	*Description*
Indoor	IN_SMALL_01	16.4m	01m 29s	5.7 GB	Single-room environment
	IN_MULTI_01	50.2m	06m 46s	13.5 GB	Multi-room linear trajectory
	IN_MULTI_02	84.3m	07m 07s	13.0 GB	Multi-room with loop closure
	IN_LARGE_01	204.3m	09m 30s	57.9 GB	Large-scale indoor with loop
Outdoor	OUT_CAMPUS_01	120.6m	04m 57s	36.4 GB	Urban campus linear path
	OUT_CAMPUS_02	141.2m	05m 15s	37.8 GB	Urban campus circular path
Total		617m	35m 04s	164.3 GB	

planes of the nearest cameras. The resulting colored point cloud (Fig. 3) shows strong geometric–visual alignment, with only minor residual offsets of a few centimeters, highlighting both the accuracy of the calibration and opportunities for further refinement.

4 Benchmarking

To validate the practicality of the platform for SLAM research, we collected representative datasets across diverse environments and conducted benchmarking experiments using state-of-the-art algorithms. The primary goal of these experiments is not only to demonstrate the platform's compatibility with established SLAM pipelines but also to evaluate its ability to generate reliable, high-fidelity datasets for assessing accuracy, robustness, and semantic understanding in real-world scenarios.

4.1 Baselines

The benchmarking was conducted using representative frameworks from both LiDAR- and vision-based SLAM approaches. For LiDAR SLAM, we employed *GLIM* [25], a state-of-the-art LiDAR–IMU odometry system known for its efficiency and accuracy, along with *S-Graphs* [26, 27], which extends LiDAR-based SLAM by integrating semantic 3D scene graphs directly from point clouds. For visual SLAM, we utilized *ORB-SLAM3* [28], a widely recognized framework that supports various sensor configurations, and *vS-Graphs* [29], the vision-based version of *S-Graphs* that augments the reconstructed map with structured 3D scene graph representations. The generated map quality is analyzed quantitatively to understand how different systems perform in scenarios with varying sensor configurations and environmental challenges.

| (a) IN_SMALL_01 | (b) IN_MULTI_01 | (c) IN_MULTI_02 |
| (d) IN_LARGE_01 | (e) OUT_CAMPUS_01 | (f) OUT_CAMPUS_02 |

Figure 4: Sample instances of *SMapper-light* dataset scenarios.

4.2 Data Collection

Data collection was conducted using the handheld configuration while traversing university campus buildings and sparsely populated areas at varying walking and turning speeds. The resulting dataset encompasses a diverse range of scenarios, from complex indoor environments with varied architectural layouts to spacious outdoor spaces. It should be noted that the manual operation by a human operator introduced natural vibrations caused by pedestrian motion. The collected sequences serve as representative samples to demonstrate the platform's capabilities, while the primary aim of this paper is to introduce the device itself rather than to establish a large-scale SLAM dataset. The characteristics of the collected dataset instances are summarized in Table 5, while representative examples are illustrated in Fig. 4.

The dataset is publicly available under the title *SMapper-light* for research and reproducibility purposes at `https://huggingface.co/datasets/snt-arg/smapper-light`. It contains six `rosbag` sequences, each containing synchronized feeds from cameras, IMU, and LiDAR. For ground truth generation, we adopt a LiDAR-based SLAM approach executed in offline mode. While this could be seen as a potential limitation when benchmarking LiDAR-based methods, it is important to note that offline processing enables the accumulation of dense point clouds and achieves sub-centimeter accuracy (errors less than 3cm). This results in a reliable 6-DoF ground truth trajectory while producing dense 3D maps included with the dataset, which is sufficient for benchmarking both visual and multimodal SLAM pipelines. Moreover, the dense point clouds generated during the process are released as part of the dataset, providing researchers with high-fidelity 3D reconstructions of the environments, which can be employed for tasks beyond SLAM, such as scene graph generation or cross-modal learning.

4.3 Experimental Results

Benchmarking of the collected dataset is performed through qualitative and quantitative experiments. The qualitative evaluations (§4.3.1) involve visual inspections of the reconstructed map and trajectories, while the quantitative assessments assess (§4.3.2) the practicality of the dataset for visual SLAM. It should be noted that our objective is not to identify which SLAM framework performs best, but rather to showcase the applicability of *SMapper-light* in the SLAM domain.

4.3.1 Qualitative Experiments

Fig. 5 depicts the qualitative analysis of the various SLAM frameworks (including LiDAR-based and visual variants) used in the paper. Accordingly, the visual reconstructions of the environments highlight the strengths and limitations of the employed methodologies. In this regard, LiDAR-based approaches present denser maps owing to their highly reliable LiDAR trajectories. Among them, *S-Graphs* further improves the reconstructed maps through geometric

(a) IN_SMALL_01 on ORB-SLAM 3.0   (b) IN_MULTI_02 on ORB-SLAM 3.0   (c) OUT_CAMPUS_02 on ORB-SLAM 3.0

(d) IN_SMALL_01 on vS-Graphs   (e) IN_MULTI_02 on vS-Graphs   (f) OUT_CAMPUS_02 on vS-Graphs

(g) IN_SMALL_01 on *S-Graphs*   (h) IN_MULTI_02 on *S-Graphs*   (i) OUT_CAMPUS_02 on *S-Graphs*

(j) IN_SMALL_01 on GLIM   (k) IN_MULTI_02 on GLIM   (l) OUT_CAMPUS_02 on GLIM
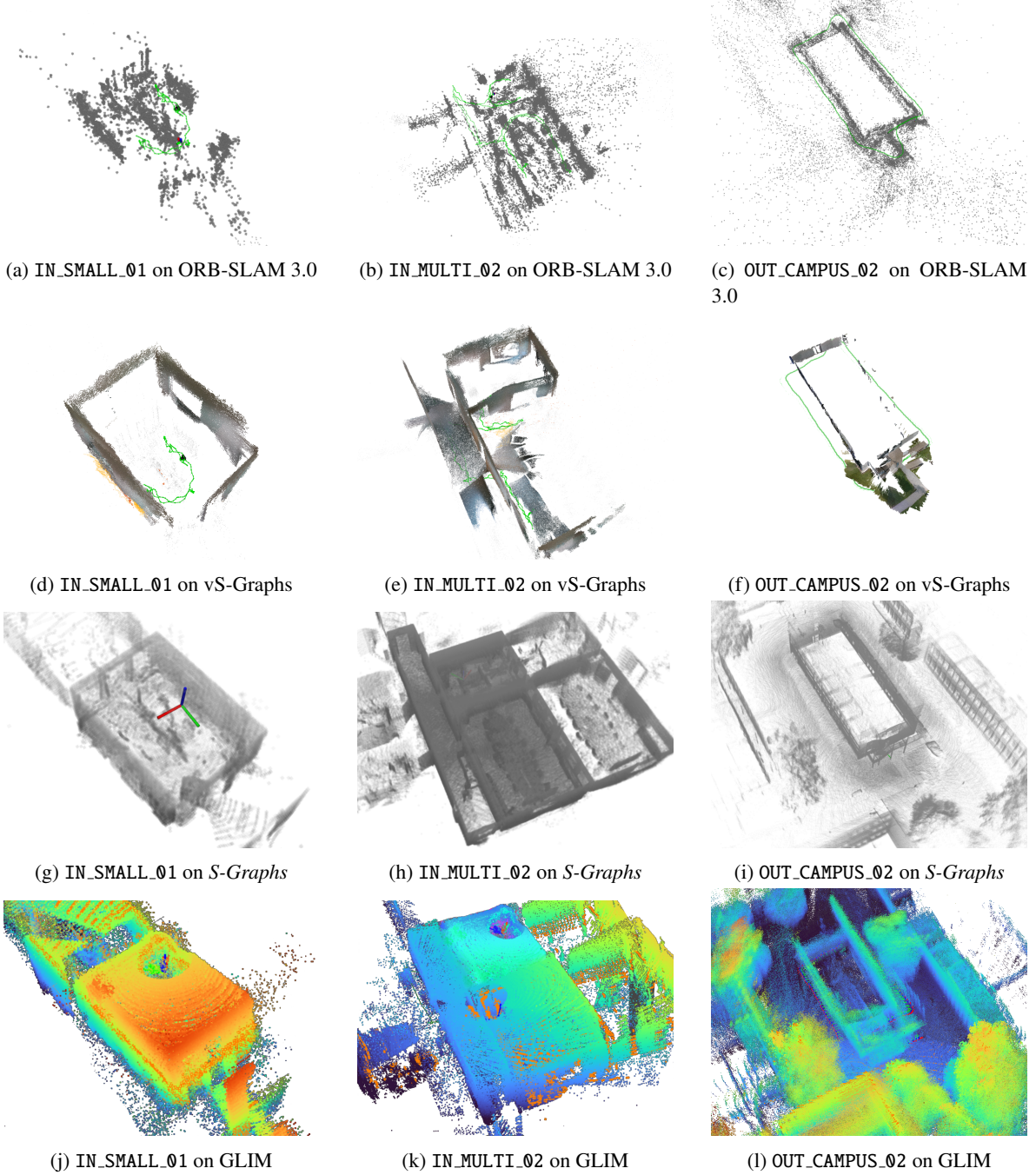
Figure 5: Qualitative results of SLAM benchmarking across selected sequences using various SLAM pipelines, including LiDAR-based (*S-Graphs* [26] and GLIM [25]) and visual SLAM (ORB-SLAM 3.0 [28] and vS-Graphs [29]).

reasoning to identify semantic structures, such as walls and rooms, and integrate them into the optimization process. In contrast, visual SLAM methodologies generate sparser reconstructions due to their reliance on visited visual features. However, these approaches can capture visual and appearance-driven information to augment the maps, as can be seen in vS-Graphs outputs, which extend this capability by incorporating semantic validation during map reconstruction. Overall, these observations confirm that the proposed dataset encompasses diverse scenarios, enabling various SLAM frameworks to showcase their respective mapping and optimization capabilities. This also validates

10

Table 6: Evaluation results on the collected dataset using Root Mean Square Error (RMSE) error in *meters* and Standard Deviation (STD). Best results in each metric are boldfaced.

Instance	RMSE		STD	
	ORB-SLAM 3.0	*vS-Graphs*	*ORB-SLAM 3.0*	*vS-Graphs*
IN_SMALL_01	**0.059**	0.072	**0.015**	0.024
IN_MULTI_01	0.347	**0.260**	0.198	**0.189**
IN_MULTI_02	0.123	**0.117**	**0.050**	0.056
IN_LARGE_01	**0.353**	0.381	0.215	**0.178**
OUT_CAMPUS_01	0.474	**0.434**	**0.206**	0.251
OUT_CAMPUS_02	**0.318**	0.534	**0.213**	0.390

the suitability of the *SMapper* device for data collection and benchmarking purposes.

4.3.2 Quantitative Experiments

Table 6 presents the quantitative evaluation results obtained from the *SMapper-light* dataset. In this regard, the LiDAR-based SLAM trajectory generated by *S-Graphs* serves as the ground truth reference, against which the visual SLAM approaches are benchmarked. The analysis focuses on two primary metrics: the Root Mean Square Error (RMSE) and the Standard Deviation (STD) of the trajectory alignment, both expressed in meters.

According to the table, both *ORB-SLAM 3.0* and *vS-Graphs* demonstrate competitive accuracy across the indoor and outdoor sequences. *ORB-SLAM 3.0* generally performs better in shorter or less complex trajectories, benefiting from stable feature tracking and effective loop closure detection. In contrast, *vS-Graphs* relies heavily on the reliable recognition of environment-driven semantic entities, and fast rotations or missed detections can lead to mapping inconsistencies and reduced accuracy. Nevertheless, when semantic entities are correctly identified, particularly in multi-room and visually complex scenarios, integrating semantic validation within its optimization pipeline enables more coherent and structured reconstructions.

Overall, the relatively low RMSE and STD values across diverse environments confirm that the collected dataset provides well-calibrated, temporally aligned, and reliable multimodal measurements for visual SLAM benchmarking. This further supports the use of the proposed dataset as a practical resource for evaluating and comparing SLAM frameworks under realistic conditions.

5 Discussions

In general, data collection using *SMapper* is tailored to a balance between "portability" and "stability." While handheld usage of the device enables flexible capture in confined spaces, robot-mounted data collections ensured smoother trajectories at the cost of mobility. The *SMapper-light* dataset (summarized in Table 5) comprises 615m of trajectories across indoor and outdoor environments. It serves as a **proof of concept** to demonstrate the versatility of *SMapper* and the practicality of its data acquisition and processing pipelines. Rather than aiming for large-scale coverage, the sequences were intentionally designed to capture a variety of structural layouts and environment-driven conditions that represent typical SLAM evaluation scenarios. Incorporating longer trajectories, dynamic environments, or specialized configurations remains open to other researchers to be explored.

Additionally, an essential aspect of *SMapper* lies in its **open-hardware** nature, which promotes reproducibility and facilitates collaborative extensions. The released `CAD` models, calibration tools, and firmware configurations allow researchers to replicate or modify the system for specific sensing setups, bridging the gap between dataset usage and generation. This can even lead to a wide range of research opportunities, from studying cross-modal calibration to developing robust SLAM frameworks.

In terms of ground truth generation, the proposed LiDAR-based offline solution in *SMapper-light* primarily supports the evaluation of visual and visual–inertial SLAM systems and is not directly applicable to LiDAR-based SLAM methods. Generating accurate trajectories is often challenging because conventional approaches frequently rely on costly equipment, such as terrestrial laser scanners. To address this, several alternative strategies can be explored within the *SMapper* framework. For instance, augmenting the environment with fiducial markers placed at known positions enables pose estimation through multi-camera tracking, albeit at the cost of increased setup effort and potential visual artifacts in the dataset. Another direction is the integration of Building Information Models (BIMs), where available, to provide structural priors for trajectory estimation. Although such solutions are beyond the scope

of this paper, they can establish a foundation for extending the dataset toward benchmarking multi-sensor SLAM pipelines.

6 Conclusion

This paper presented *SMapper*, an open-hardware, multi-sensor platform designed to support reproducible research in SLAM and related fields. The device integrates synchronized LiDAR, multi-camera, and inertial sensing, supported by a robust calibration and synchronization pipeline that ensures accurate spatio-temporal alignment across modalities. Its fully open and replicable design allows researchers to extend its capabilities and deploy it across handheld or robot-mounted configurations. To validate the platform, we additionally introduced *SMapper-light*, a publicly available SLAM dataset containing representative indoor and outdoor sequences. The dataset includes tightly synchronized multimodal data, ground truth trajectories generated through offline LiDAR-based SLAM with sub-centimeter accuracy, and dense 3D reconstructions. Benchmarking experiments with state-of-the-art LiDAR and visual SLAM frameworks demonstrated the practicality and utility of the platform for evaluating diverse approaches.

Future work includes the expansion of *SMapper-light* into a comprehensive dataset covering both handheld and robot-mounted usage across a wide variety of indoor and outdoor environments. Another improvement involves the integration of external ground truth measurements, such as motion capture systems or terrestrial laser scans, to facilitate the creation of high-precision dataset sequences suitable for benchmarking multi-sensor SLAM frameworks.

References

[1] H. Bavle, J. L. Sanchez-Lopez, C. Cimarelli, A. Tourani, and H. Voos, "From slam to situational awareness: Challenges and survey," *Sensors*, vol. 23, no. 10, p. 4849, 2023.

[2] Y. Liu, Y. Fu, F. Chen, B. Goossens, W. Tao, and H. Zhao, "Simultaneous localization and mapping related datasets: A comprehensive survey," *arXiv preprint arXiv:2102.04036*, 2021.

[3] I. Lluvia, E. Lazkano, and A. Ansuategi, "Active mapping and robot exploration: A survey," *Sensors*, vol. 21, no. 7, p. 2445, 2021.

[4] Y. Zhang, P. Shi, and J. Li, "3d lidar slam: A survey," *The Photogrammetric Record*, vol. 39, no. 186, pp. 457–517, 2024.

[5] A. Macario Barros, M. Michel, Y. Moline, G. Corre, and F. Carrel, "A comprehensive survey of visual slam algorithms," *Robotics*, vol. 11, no. 1, p. 24, 2022.

[6] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The international journal of robotics research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[7] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.

[8] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 573–580.

[9] J. Terblanche, S. Claassens, and D. Fourie, "Multimodal navigation-affordance matching for slam," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7728–7735, 2021.

[10] S. Duan, Q. Shi, and J. Wu, "Multimodal sensors and ml-based data fusion for advanced robots," *Advanced Intelligent Systems*, vol. 4, no. 12, p. 2200213, 2022.

[11] H. Sier, Q. Li, X. Yu, J. Peña Queralta, Z. Zou, and T. Westerlund, "A benchmark for multi-modal lidar slam with ground truth in gnss-denied environments," *Remote Sensing*, vol. 15, no. 13, p. 3314, 2023.

[12] A. Tourani, H. Bavle, J. L. Sanchez-Lopez, and H. Voos, "Visual slam: What are the current trends and what to expect?" *Sensors*, vol. 22, no. 23, p. 9297, 2022.

[13] X. Shi, D. Li, P. Zhao, Q. Tian, Y. Tian, Q. Long, C. Zhu, J. Song, F. Qiao, L. Song *et al.*, "Are we ready for service robots? the openloris-scene datasets for lifelong slam," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3139–3145.

[14] M. Ramezani, Y. Wang, M. Camurri, D. Wisth, M. Mattamala, and M. Fallon, "The newer college dataset: Handheld lidar, inertial and vision with ground truth," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4353–4360.

[15] L. Zhang, M. Camurri, D. Wisth, and M. Fallon, "Multi-camera lidar inertial extension to the newer college dataset," *arXiv preprint arXiv:2112.08854*, 2021.

[16] Y. Tao, M. Á. Muñoz-Bañón, L. Zhang, J. Wang, L. F. T. Fu, and M. Fallon, "The oxford spires dataset: Benchmarking large-scale lidar-visual localisation, reconstruction and radiance field methods," *arXiv preprint arXiv:2411.10546*, 2024.

[17] L. Zhang, M. Helmberger, L. F. T. Fu, D. Wisth, M. Camurri, D. Scaramuzza, and M. Fallon, "Hilti-oxford dataset: A millimeter-accurate benchmark for simultaneous localization and mapping," *IEEE Robotics and Automation Letters*, vol. 8, no. 1, pp. 408–415, 2022.

[18] X. Hu, L. Zheng, J. Wu, R. Geng, Y. Yu, H. Wei, X. Tang, L. Wang, J. Jiao, and M. Liu, "Paloc: Advancing slam benchmarking with prior-assisted 6-dof trajectory generation and uncertainty estimation," *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 6, pp. 4297–4308, 2024.

[19] H. Wei, J. Jiao, X. Hu, J. Yu, X. Xie, J. Wu, Y. Zhu, Y. Liu, L. Wang, and M. Liu, "Fusionportablev2: A unified multi-sensor dataset for generalized slam across diverse platforms and scalable environments," 2024. [Online]. Available: https://arxiv.org/abs/2404.08563

[20] Z. Chen, Y. Qi, D. Feng, X. Zhuang, H. Chen, X. Hu, J. Wu, K. Peng, and P. Lu, "Heterogeneous lidar dataset for benchmarking robust localization in diverse degenerate scenarios," 2024. [Online]. Available: https://arxiv.org/abs/2409.04961

[21] X. Hu, J. Wu, M. Jia, H. Yan, Y. Jiang, B. Jiang, W. Zhang, W. He, and P. Tan, "Mapeval: Towards unified, robust and efficient slam map evaluation framework," *IEEE Robotics and Automation Letters*, 2025.

[22] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1280–1286.

[23] R. Buchanan, "Allan variance ros." [Online]. Available: https://github.com/ori-drs/allan_variance_ros

[24] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 3400–3407.

[25] K. Koide, M. Yokozuka, S. Oishi, and A. Banno, "Glim: 3d range-inertial localization and mapping with gpu-accelerated scan matching factors," *Robotics and Autonomous Systems*, vol. 179, p. 104750, 2024.

[26] H. Bavle, J. L. Sanchez-Lopez, M. Shaheer, J. Civera, and H. Voos, "Situational graphs for robot navigation in structured indoor environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9107–9114, 2022.

[27] ——, "S-graphs+: Real-time localization and mapping leveraging hierarchical representations," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4927–4934, 2023.

[28] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE transactions on robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[29] A. Tourani, S. Ejaz, H. Bavle, D. Morilla-Cabello, J. L. Sanchez-Lopez, and H. Voos, "vs-graphs: Integrating visual slam and situational graphs through multi-level scene understanding," *arXiv preprint arXiv:2503.01783*, 2025.