GrandTour: A Legged Robotics Dataset in the Wild for Multi-Modal Perception and State Estimation

Journal Title
XX(X):1–27
©The Author(s) 2025
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Jonas Frey[1,2*], Turcan Tuna[1*], Frank Fu[3], Katherine Patterson[1], Tinao Xu[1], Maurice Fallon[3], Cesar Cadena[1], Marco Hutter[1]

Abstract

Accurate state estimation and multi-modal perception are prerequisites for autonomous legged robots in complex, large-scale environments. To date, no large-scale public legged-robot dataset captures the real-world conditions needed to develop and benchmark algorithms for legged-robot state estimation, perception, and navigation. To address this, we introduce the GrandTour dataset, a multi-modal legged-robotics dataset collected across challenging outdoor and indoor environments with an ANYmal (Hutter et al. 2016) quadruped equipped with the *Boxi* (Frey et al. 2025) multi-modal sensor payload. GrandTour spans a broad range of environments and operational scenarios across distinct test sites, ranging from alpine scenery and forests to demolished buildings and urban areas, and covers a wide variation in scale, complexity, illumination, and weather conditions. The dataset provides time-synchronized sensor data from spinning LiDARs, multiple RGB cameras with complementary characteristics, proprioceptive sensors, and stereo depth cameras. Moreover, it includes high-precision ground-truth trajectories from satellite-based RTK-GNSS and a Leica Geosystems total station. This dataset supports research in SLAM, high-precision state estimation, and multi-modal learning, enabling rigorous evaluation and new approaches to sensor fusion in legged robotic systems. With its extensive scope, GrandTour represents the largest open-access legged-robotics dataset to date. The dataset is available at https://grand-tour.leggedrobotics.com, on HuggingFace (ROS-independent), and in ROS1/2 formats, along with tools and demo resources.

Keywords

Legged robotics, SLAM, multi-sensor dataset, state estimation, multi-modal perception, computer vision, sensor fusion, autonomy

1 Introduction

Legged robots have emerged as essential platforms for traversing environments inaccessible to wheeled and aerial robots, with applications ranging from disaster response and industrial inspection to exploration of unstructured terrains. This expanded mobility exposes autonomy systems to new challenges in robust state estimation, accurate mapping, scene understanding, and safe navigation.

Despite the well-known fact that enabling autonomy in legged robots requires advanced SLAM and multi-modal perception algorithms, targeted progress on these systems has been hindered by the lack of comprehensive public datasets. Existing datasets, such as KITTI for autonomous driving (Geiger et al. 2013), EuRoC for micro-aerial vehicles (Burri et al. 2016), and TUM VI for visual–inertial odometry (Schubert et al. 2018), have enabled breakthroughs in their respective domains. However, analogous datasets for legged robots in real-world environments remain sparse. Combined, these points highlight the need for legged robot datasets that jointly capture the motion characteristics of legged robots, versatile environments in which legged robots are suitable, and multi-modal sensing.

For legged robots, only a handful of datasets exist, *FusionPortable* (Jiao et al. 2022), *M3ED* (Chaney et al. 2023), *DiTer* (Jeong et al. 2024), *TAIL* (Yao et al. 2024), and *TAIL+* (Wang et al. 2024b), but each is limited either in scale, sensing modalities, or environmental diversity. As a result, researchers often resort to collection of small-scale and non-representative experiment data for specific scenarios, or simulations to develop legged robot state estimation, mapping, and multi-modal perception to action methods, which may not generalize to the complexity of real-world scenarios. To address this gap, we present GrandTour, a versatile, large-scale public dataset featuring a state-of-the-art multi-sensor payload on a quadrupedal robot deployed across a wide range of environments. The dataset is collected using an ANYmal (Hutter et al. 2016) quadruped legged robot equipped with *Boxi* (Frey et al. 2025), a modern and comprehensive sensor payload, yielding an unprecedented set of sensing modalities on a mobile legged robot. GrandTour captures data from three LiDARs, 10 cameras, eight IMUs, seven depth cameras, 12 proprioceptive sensors, and a

[1]ETH Zurich, Switzerland
[2]Max Planck Institute for Intelligent Systems, Germany
[3]Oxford University, UK
[*]Equal Contribution

Corresponding author:
Dr. Jonas Frey and Turcan Tuna, Robotic Systems Lab, ETH Zurich, Zurich 8092, Switzerland.
Email: jonfrey@ethz.ch, tutuna@ethz.ch



Figure 1. GrandTour dataset preview. Top: views of ANYmal traversing diverse environments during GrandTour. Bottom: aerial imagery alongside on-board camera views for six missions of GrandTour (ETH-1, PIL-1, EIG-1, SPX-2, HEAP-1, ARC-2); inset summarizes sensor suite and dataset scale (49 missions, $>10\,\mathrm{km}$, $>5\,\mathrm{h}$).

GNSS/INS module. The sensor configuration is described in detail later in Sec. 3.1 and illustrated in Fig. 2.

As part of the GrandTour's main application domain, the challenges for reliable state estimation, multi-modal perception, and SLAM in legged robots remain significant. Firstly, the inherent dynamics of legged locomotion, such as intermittent contacts, foot slippage, and rapid body orientation changes, introduce disturbances to sensor measurements.

Secondly, legged robots are often deployed in unstructured and harsh field conditions, where poor illumination, occlusions, tight spaces, slippery terrain, airborne particles, and textureless surfaces further degrade sensing. Finally, to cope with these issues, legged platforms utilize complementary exteroceptive sensors (cameras, LiDAR, GNSS) and high-rate inertial/proprioceptive sensing. However, exploiting this suite of sensors is non-trivial due to distinct noise profiles, latencies, fields of view, and failure modes under locomotion-induced and environmental disturbances.

For classical state estimation and SLAM, GrandTour provides benchmarks for LiDAR-inertial, visual-inertial, LiDAR-only odometry, and LiDAR-inertial-visual odometry, enabling systematic evaluation of multi-modal fusion and drift over different terrains in challenging environments. Moreover, GrandTour has already been used in sensor fusion frameworks (Nubert et al. 2025) and the continuous-time state estimation problem (Cao et al. 2025a).

For learning-based methods, the dataset has already been used to improve sim-to-real transfer (Roth et al. 2025), to enable real-to-sim transfer of large-scale environments and neural scene representations (Escontrela et al. 2025), learn physical parameters (Chen et al. 2024a), to train foundation models for navigation (Inglin et al. 2025), benchmark vision language models for scene understanding (Windecker et al. 2025), and support self-supervised representation learning (Patel et al. 2025).

Furthermore, the field deployment of legged robots relies on closed-loop autonomy, where perception and state estimation inform planning and control, while resulting motions affect sensor measurements. Modern approaches emphasize the action-perception feedback loop (Lee et al. 2020; Korthals et al. 2022), and in legged systems, this loop is reinforced by the coupling between proprioception and

exteroception, where gait patterns affect sensing and thus measurement quality (Bloesch et al. 2017b; Kim et al. 2021a).

Consequently, multi-modal perception has emerged as an essential tool for robust autonomy in challenging environments (Chavez-Garcia et al. 2018; Feng et al. 2023; Cui et al. 2023). Such rich multi-modal data enables both perceptive locomotion with action-perception feedback (Miki et al. 2022a, 2024; Kumar et al. 2021) and learned legged state estimation that fuses proprioceptive and exteroceptive cues under locomotion-induced disturbances (Youm et al. 2025; Buchanan et al. 2022).

Beyond its main applications, GrandTour also enables the extraction of a large number of cross-view image pairs under realistic legged robot motion and field conditions, which can directly support DUSt3R (Wang et al. 2024a) and VGGT (Wang et al. 2025) visual-geometry tuning for improved geometry and relocalization in legged robots. Similarly, recent progress on foundation models for robotics (Firoozi et al. 2025) further increases the value of large, well-synchronized, and calibrated datasets. GrandTour provides diverse, synchronized multi-modal legged data with accurate ground truth, making it a strong resource for fine-tuning existing foundation models to improve robustness and transfer to real-world legged deployments.

In summary, the primary contributions of this work are:

- **GrandTour Dataset:** We introduce GrandTour, a large, open-access legged robotics dataset capturing 49 sequences across indoor facilities, urban outdoors, and natural terrains, representing the largest collection of real-world legged robot data to date. GrandTour offers synchronized and calibrated data from an extensive array of sensors (including multi-LiDAR, multi-camera, multi depth-camera, IMUs, GPS, and full proprioception). Each data sequence is accompanied by high-precision ground truth, enabling rigorous benchmarking of SLAM, odometry, and sensor fusion algorithms under realistic conditions.
- **A Comprehensive Localization Benchmark:** We evaluate a comprehensive set of methods on the data of GrandTour and perform in-depth analysis on the performance and the methodologies.
- **Broad Utility and Open-Source Release:** We release GrandTour to the community under an open-source license, along with tools and documentation to facilitate its use. Critically, the GrandTour dataset provides intermediate and processed outputs, such as legged-inertial odometry, desekwed point clouds, LiDAR-inertial odometry, terrain maps, point-cloud maps, and occupancy maps, to facilitate ease of adoption and targeted development by researchers.

2 Related Work

2.1 Popular robotic state estimation datasets

A variety of datasets have shaped the development of SLAM and robotic perception by pairing versatile sensor suites with high-quality ground truth. The KITTI benchmark (Geiger et al. 2013) remains highly influential for stereo, visual odometry, and 3D perception with synchronized cameras, LiDAR, and GPS/INS ground truth. For aerial robots, EuRoC MAV (Burri et al. 2016) established a visual–inertial odometry evaluation using synchronized stereo and IMU data, along with pose ground truth from Vicon and Leica tracking systems. TUM VI (Schubert et al. 2018) extended VI evaluation with a stereo camera, photometric calibration, an IMU, and motion-capture ground truth at sequence start/end. The Newer College dataset (Ramezani et al. 2020) provided handheld LiDAR–visual–inertial recordings on the Oxford campus with centimetre-accurate trajectories obtained by registering mobile LiDAR to a millimetre-accurate TLS map. The Hilti SLAM Challenge datasets (2021–2023) benchmark high-precision SLAM in GNSS-denied industrial settings using calibrated visual/LiDAR/IMU suites with sparse ground truth (Helmberger et al. 2022; Zhang et al. 2023). Together with large-scale datasets such as Oxford RobotCar (Maddern et al. 2017), and Boreas (Burnett et al. 2023).

However, these reference datasets primarily target wheeled, aerial, or handheld platforms and therefore do not capture the locomotion-induced disturbances, proprioception, exteroception coupling, or unstructured terrain that characterize legged autonomy. GrandTour is intended as a legged-robot counterpart with comparable calibration and ground-truth fidelity, extending this lineage of datasets with a tightly synchronized multi-LiDAR/multi-camera/IMU payload, plus full proprioception, enabling missions with loop closures and coverage of indoor, urban, industrial, and natural environments.

2.2 Multi-modal perception datasets

While many earlier datasets already include more than one exteroceptive sensor, ongoing efforts specifically focus on collecting multimodal data. This shift is driven by the fact that no single sensor remains reliable across all environments and failure modes. Combining complementary modalities (e.g., LiDAR geometry with visual/thermal semantics or radar robustness) enables robustness under lighting, adverse weather, sparse texture, and self-similar or dynamic scenes.

M2DGR (Yin et al. 2022) equips a ground robot with fisheye, infrared, and event cameras alongside a 32-beam LiDAR, multiple IMUs, and GNSS receivers, offering indoor and outdoor sequences with ground truth from motion-capture systems and laser trackers. FusionPortable and FusionPortableV2 (Jiao et al. 2022; Wei et al. 2025) extend this concept by mounting a unified multi-sensor payload including LiDAR, RGB cameras, depth sensors, and IMUs across handheld, legged, and wheeled platforms to enable cross-platform benchmarking. The MulRan dataset (Kim et al. 2020) highlights radar–LiDAR fusion for place recognition in urban driving, while M3ED (Chaney et al. 2023) introduces synchronized event cameras, LiDAR, and inertial sensing across wheeled, legged, and aerial robots in natural environments. The MCD dataset (Nguyen et al. 2024b) augments cameras and LiDARs with radar, event cameras, and multiple IMUs, and releases sequences recorded during day, night, and adverse weather, explicitly targeting robustness under challenging conditions.

The Oxford Spires dataset (Tao et al. 2025) was collected around historic landmarks in Oxford using a perception unit with three synchronised global-shutter colour cameras, a large field of view 3D LiDAR, and an inertial sensor, with millimetre-accurate TLS scans providing ground truth for localisation and 3D reconstruction. Highlighting the limitations in existing multi-modal perception datasets, the authors in Soares et al. (2025) introduce the SMapper

Table 1. Comparison of related datasets. Each row lists dataset name, **platform abbreviation(s)**, environment types, diversity of scenarios and sensor coverage. **Platform abbreviations:** A = Aerial, W = Wheeled, L = Legged, C = Car, H = Handheld, U = UGV. The dashed horizontal divider separates datasets that include legged platforms (below) from those that do not (above). TODO !!! add Liang et al. (2025) dataset

Dataset	Platform	Environment	Diversity	Sensor coverage
KITTI (Geiger et al. 2013)	C	City, rural roads	22 sequences; 6 h	Stereo cams, LiDAR, GPS, IMU
Oxford RobotCar (Maddern et al. 2017)	C	Urban	1000 km; year-long; all-weather	6 cams, 2D/3D LiDARs, GPS/INS
Boreas (Burnett et al. 2023)	C	Repeated route	350 km; multi-season (1 yr)	LiDAR, 360° radar, RGB cam, GNSS, IMU
EuRoC (Burri et al. 2016)	A	Indoor machine halls	11 sequences; varying difficulty	Stereo cam, IMU
TUM VI (Schubert et al. 2018)	H	Indoor/outdoor	13 sequences; HDR imagery	Stereo cams, IMU, Vicon GT
Newer College (Ramezani et al. 2020)	H	Campus buildings/quads	6.7 km; buildings & vegetation	LiDAR, stereo RGB cams, IMU
M2DGR (Yin et al. 2022)	U	Indoor/outdoor	36 sequences; campus	6 fisheye, IR & event cams, LiDAR, IMUs, GNSS
MulRan (Kim et al. 2020)	C	City streets	Multi-city; long trajectories	Radar, LiDAR
MCD (Nguyen et al. 2024b)	U, H	Multi-campus indoor/outdoor	59 k labelled scans; 3 campuses	LiDAR, RGB, IMU, UWB
Oxford Spires (Tao et al. 2025)	H	Multi-campus, outdoor	Urban, historic landmarks, visual focus	LiDAR, **3×** RGB cams, IMU
FusionPortable (Jiao et al. 2022)	L, H, U, C	Indoor, campus, urban	17 sequences; multi-platform	LiDAR, RGB, event cams, IMU, GNSS
FusionPortableV2 (Wei et al. 2025)	L, H, U, C	Buildings, campus, urban	27 sequences; 2.5 h; varied motion	LiDAR, FLIR/DAVIS stereo, IMU, GNSS
M3ED (Chaney et al. 2023)	L, W, A	Off-road trails, dense forests	Multi-robot; event-centric	Event cams, RGB, IMU, LiDAR
SubT-MRS (Zhao et al. 2024b)	L, A, W	Vast and diverse	2000 h, 300 km; multi-year; all-weather	LiDAR, fisheye, depth, thermal cams, IMU
TAIL (Yao et al. 2024)	L, W	Granular soil	**14** sequences; wheeled and quadruped	LiDAR, RGB(D) cams, IMU, GNSS, kinematics
TAIL-Plus (Wang et al. 2024b)	L, W	Granular soil	Multi-loop; day/night	LiDAR, RGB-D, RGB, IMU, RTK-GPS
DiTer (Jeong et al. 2024)	L	Parks, slopes, sand	Diverse terrain	LiDAR, RGB-D, thermal cams, IMU, GNSS
DiTer++ (Kim et al. 2024)	L	Park, forest	Day/night	LiDAR, RGB, thermal cams, IMU, GNSS
EnvoDat (Nwankwo et al. 2024)	L, W	Indoor, underground, urban	26 sequences; 13 scenes; fog, rain	RGB, depth, thermal/IR, LiDAR, IMU

handheld data acquisition platform and subsequently the SMapper-light dataset, primarily targeting multi-modal SLAM applications in a diverse range of environments.

Collectively, these datasets highlight a growing focus on fusing LiDAR, RGB, and inertial modalities, often complemented by radar or event sensors, to develop robust perception pipelines across diverse environments and robotic platforms.

2.3 Legged-robot datasets

Compared to other platforms, public legged-robot datasets remain limited. *TAIL* (Yao et al. 2024) targets sandy, deformable terrains using a rotating 3D LiDAR, a forward-facing stereo pair, multiple downward RGB-D cameras, an IMU, and RTK, and logs proprioceptive states (gait/motor/foot status). *TAIL-Plus* (Wang et al. 2024b) adds more sequences with multiple loops and day–night traverses, includes an extra IMU, and was collected on beach analog sites for granular terrains. *DiTer* (Jeong et al. 2024) and *DiTer++* (Kim et al. 2024) focus on legged robots traversing in park-like/forested areas and provide RGB-D, thermal, LiDAR, IMU, and GPS sensing; *DiTer++* further introduces multi-robot coordination and day–night sequences. Multi-platform datasets such as *M3ED* (Chaney et al. 2023) and *SubT-MRS* (Zhao et al. 2024b) include legged platforms where, *M3ED* emphasizes event-camera benchmarks across ground/legged/aerial platforms and *SubT-MRS* stands out for its scale and emphasis on robustness in degraded sensing conditions (e.g., low light, obscurants, and adverse environments), with synchronized LiDAR and multi-camera (including thermal) sensing and long-term collection. However, as a multi-robot dataset spanning aerial, wheeled, and legged platforms, *SubT-MRS* is not primarily optimized around legged-specific sensing and evaluation (e.g., rich proprioception, contact events, or legged-only mission structure). *EnvoDat* (Nwankwo et al. 2024) collects multiple sequences across indoor, outdoor, and underground scenes with ten modalities on quadruped and wheeled platforms, using RGB-D and monocular cameras rather than multi-camera rigs. *FusionPortable* and *FusionPortableV2* (Jiao et al. 2022; Wei et al. 2025) mount a unified multi-sensor payload across handheld, legged, and wheeled platforms; notably, *V2* records Unitree A1 proprioception (joint encoders, foot contacts) in addition to LiDAR, RGB/depth, event, IMU, and GNSS.

Despite these efforts, large-scale datasets focused on legged robots remain scarce. To the best of our knowledge, *GrandTour* is the largest legged-robot dataset to date, filling a unique gap by combining multiple LiDARs, multiple RGB/depth camera bundles, high-grade IMUs, and rich proprioception with dual RTK-GPS and survey-grade ground truth. GrandTour spans more than 49 missions across urban, industrial, and natural settings, including day/night operations and adverse weather conditions. By combining survey-grade total-station measurements with high-end inertial sensors and dual RTK-GNSS, GrandTour achieves millimeter-level accuracy with sub-millisecond synchronization, enabling research in contact-aware perception, terrain-adaptive locomotion, and multi-modal perception that was previously limited by the available legged-robot datasets.

3 System overview

3.1 Sensor suite

GrandTour uses the *Boxi* perception payload (Frey et al. 2025) in combination with the ANYmal quadruped robot. All sensors inside *Boxi* and ANYmal are listed in Table 2. *Boxi* integrates exteroceptive and proprioceptive sensors into a compact, aluminum monoblock housing that is CNC machined to 0.05 mm tolerance, providing excellent thermal stability and mechanical rigidity. *Boxi* includes two high-performance rotating LiDARs (a Livox Mid 360 (`Livox`) for near field and a Hesai XT-32 (`Hesai`) for long range geometric perception), together with ten RGB cameras (mixtures of high dynamic range, global shutter, and rolling shutter models), a ZED2i stereo camera (`ZED2i`), a Sevensense CoreResearch unit (`CoreResearch`), and three TierIV C1 HDR cameras (`HDR`), enabling wide field monocular, stereo, and multi-view imaging. The suite further contains seven IMUs of varying quality, ranging from low-cost MEMS IMUs (`STIM320`, `ADIS`, `AP20-IMU`, `Bosch`, `TDK`, `ZED2i-IMU`) to a $12 000 Honeywell HG4930 IMU unit (`HG4930`), as well as proprioceptive encoders that measure joint positions, velocities, and contact forces. Finally,

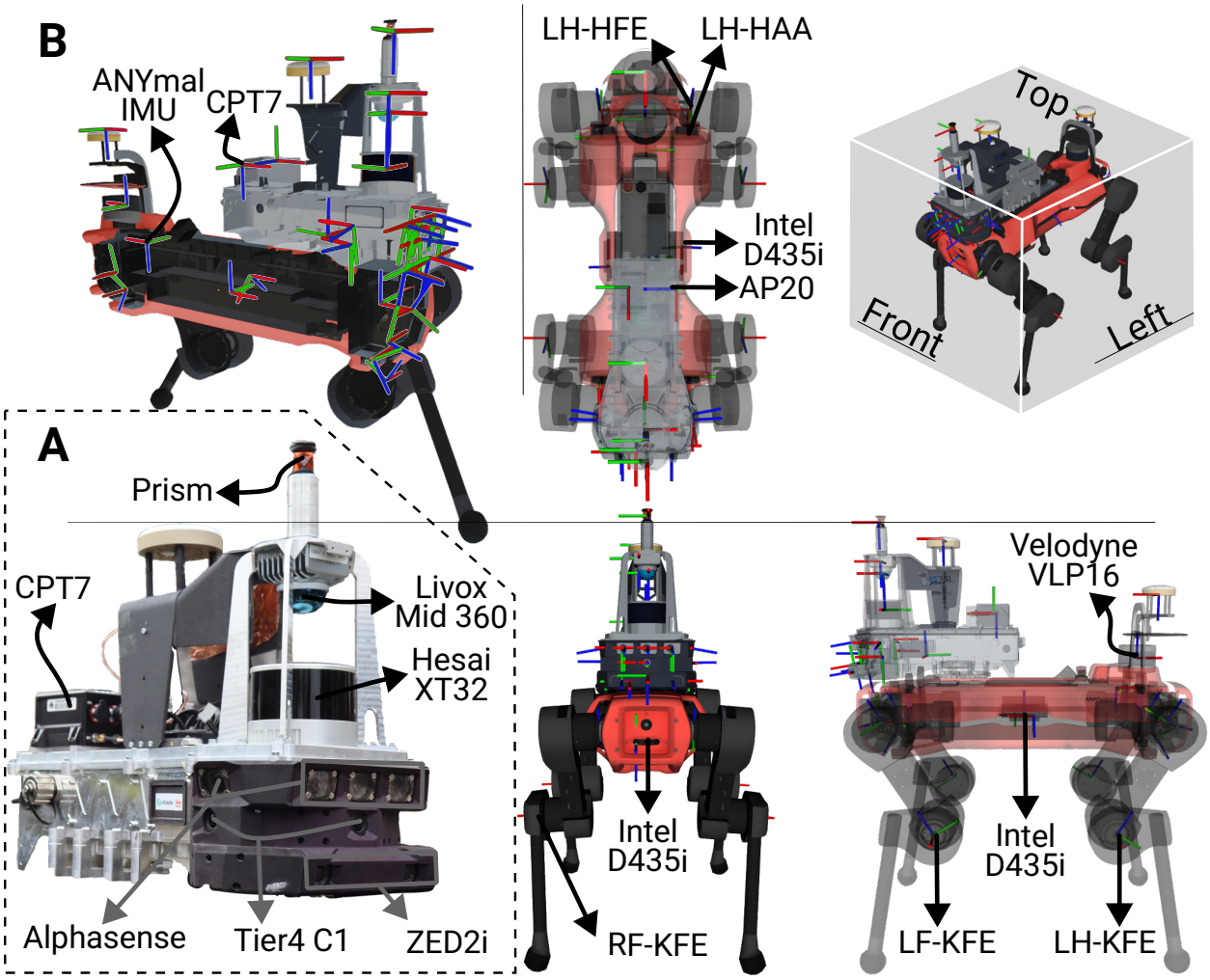


Figure 2. Sensor placement visualization of the entire GrandTour sensor suite. **A)** Shows the spatial relationships of the sensors on the *Boxi* payload, and **B)** shows the sensors of the ANYmal base platform components. Each colored axis represents a sensor or joint frame. For brevity, repeated sensors are not shown (e.g., 6× Intel RealSense D435i depth cameras).

a NovAtel CPT7 dual-antenna RTK GNSS receiver with Inertial Explorer software solution provides global pose estimates, whereas a Leica Geosystems custom AP20 solution enables real-time recording of the position of a Leica Mini Prism with respect to the Leica Geosystems MS60 total station. The entire sensor suite weighs about 7.1 kg and draws roughly 120 W. For brevity, we will refer to each sensor by its acronym.

While *Boxi* itself already offers a versatile set of sensors, in addition to *Boxi* , we record sensor and actuator information from ANYmal, namely, six Intel D435i depth cameras (`ANYmal-Depth`), a Velodyne VLP16 LiDAR (`VLP16`), an IMU (`ANYmal-IMU`), and 12 joint-encoders (`ANYmal-JE`), commanded velocity, and multiple intermediate processing outputs (elevation maps, state estimation, hidden states of locomotion policy). All sensors are rigidly mounted on the robot with known extrinsic transforms and share a common timing source, utilizing different synchronization methods to ensure temporal alignment. Further details are provided in Sec. 3.2, and the sensor positioning is illustrated in Fig. 2.

3.2 Temporal and Spatial Calibration

Precise intrinsic and extrinsic calibration is fundamental for effective multi-modal sensor fusion. The calibration methodology for GrandTour builds upon the comprehensive procedures developed for the *Boxi* payload, which are thoroughly documented in (Frey et al. 2025).

3.2.1 Time Synchronization Given the heterogeneous nature of the sensors, multi-source time synchronization methods were used, and we aimed to achieve a time synchronization formulation for the GrandTour dataset. The time-synchronization accuracy of the sensors ranges from ns for the IEEE1588v2 PTP protocol to ms for our custom software-based kernel timestamping solution, with the exception of the USB buffer-reliant `ZED2i` RGB-D camera.

In short, the NovAtel SPAN CPT7 GNSS receiver serves as the time grandmaster, and all Precision Time Protocol (PTP)-enabled devices synchronize against this time source (Jetson, NUC, Livox, Hesai, and CoreResearch, as mentioned in Table 2). The Jetson within *Boxi* serves as a Network Time Protocol (NTP) time source for the ANYmal robot. All sensors within ANYmal are guaranteed to be well time-synchronized to its internal compute module by the manufacturer. This compute module synchronizes itself to Jetson within *Boxi* with the NTP protocol. All LiDARs are synchronized through IEEE 1588v2 PTP protocol and therefore, have significant time synchronization accuracy. All IMUs (except the `ZED2i-IMU`) are hardware time-synced and were verified with a custom-developed time alignment

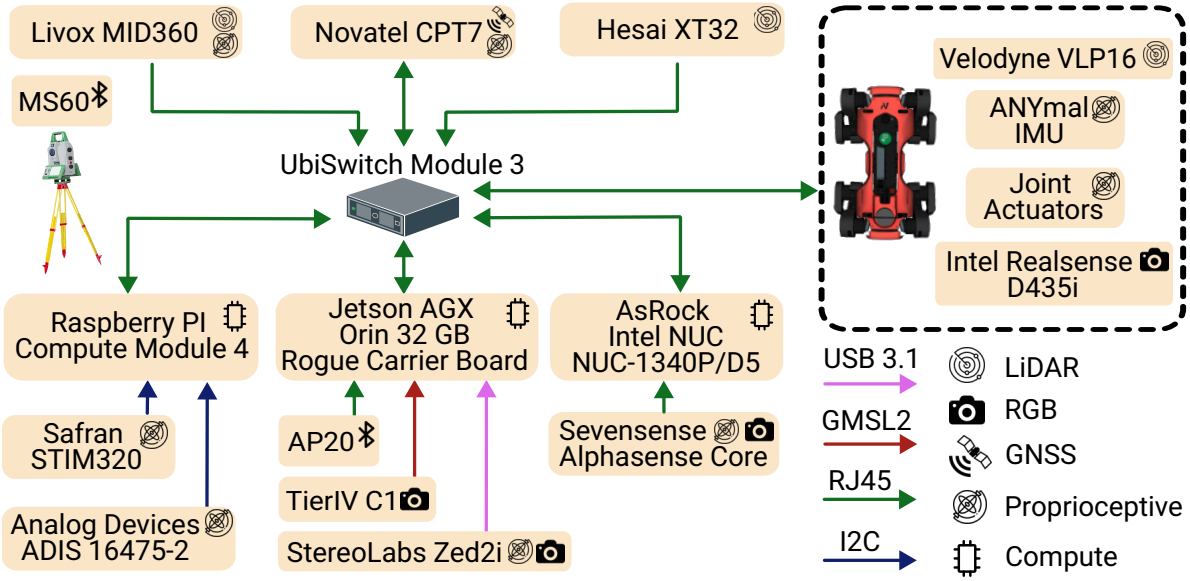


Figure 3. System architecture and sensor interfaces of the combined Boxi–ANYmal platform. All compute units (Jetson AGX Orin, Intel NUC, Raspberry Pi) are connected to the UbiSwitch Ethernet device, and the sensors are connected to their respective compute platforms through various connectivity, such as USB 3.1, GMSL2, RJ45, and I²C links.

	Name	Description	Time Synchronization†
LiDARs			
(Livox)	Livox Mid-360	vFoV: 59°, Range: 0.1 m, 40 m, 10 Hz, Accuracy: ±0.02 m	6.5 W (avg)
(Hesai)	Hesai XT-32	vFoV: 31°, Range: 0.05 m, 120 m, 10 Hz, Accuracy: ±0.01 m	10 W (avg)
(VLP16)	Velodyne VLP-16	vFoV: 30°, Range: 0.5 m, 100 m, 10 Hz, Accuracy: ±0.03 m	IEEE 1588V2 PTP
Cameras			
(5× CoreResearch)	Sevensense CoreResearch	1440×1080 @ 10 fps, FoV: 126°×92.4°, 71.6 dB, GS, 1.6 MP	IEEE 1588V2 PTP
(3× HDR)	TierIV C1	1920×1280 @ 30 fps, FoV: 120°×80.0°, 120 dB, RS, 2.5 MP	GMSL2 Interface
(ZED2i)	Stereolabs ZED2i	1920×1080 @ 15 fps, FoV: 110°×70°, 64.6 dB, RS, 4 MP	USB Buffer
(6× ANYmal-Depth)	Intel RealSense D435i	848×480 @ 15 fps, FoV: 87°×58°, Range: 0.2 m–10 m	IEEE 1588V2 PTP
Reference			
(AP20)*	Modified Leica AP20	MS60 Total Station Bluetooth Communication & Synchronization	1 ms (3σ)
(Prism)	Mini-Prism GRZ101	Static Accuracy ±2 mm (3σ), hFoV 360°, vFoV ±30°	—
(CPT7)	Novatel Span CPT7	Dual RTK GPS, Intertial Explorer, TerraStar-X	IEEE 1588V2 PTP
	Multiband GPS Antenna	Novatel 2×3GNSSA-XT-1	1 W (each)
Proprioceptive			
(HG4930)	Honeywell HG4930	100 Hz, ±400 ° s^{-1}, ±20 g_{grav}, IMU of (CPT7)	HW Trigger w.r.t (CPT7)
(STIM320)	Safran STIM320	500 Hz, ±400 ° s^{-1}, ±5 g_{grav},	free-running Kernel-stamped
(AP20-IMU)	Proprietary IMU	200 Hz	HW Trigger w.r.t (AP20)
(ADIS)	ADIS16475-2	200 Hz, ±500 ° s^{-1}, ±8 g_{grav}	free-running IIO-stamped
(Bosch)	Bosch BMI085	200 Hz, ±2000 ° s^{-1}, ±16 g_{grav}, IMU of (CoreResearch)	HW trigger w.r.t (CoreResearch)
(TDK)	TDK ICM40609	200 Hz, ±2000 ° s^{-1}, ±32 g_{grav}, IMU of (Livox)	HW trigger w.r.t (Livox)
(ZED2i-IMU)	Proprietary IMU	45 Hz, ±1000 ° s^{-1}, ±8 g_{grav}	HW trigger w.r.t (ZED2i)
(ANYmal-IMU)	Proprietary IMU	400 Hz, ±1000 ° s^{-1}, ±8 g_{grav}	SW stamped w.r.t (VLP16)
(12× ANYmal-JE)	ANYmal-D Joint Encoders	12, SW trigger w.r.t (ANYmal-IMU)	—

* (AP20) requires a Leica Geosystems MS60 Total Station.
† Relative to the sensor host platform.

Table 2. Specifications and settings of all the sensors available in the GrandTour data set are provided. GS=GlobalShutter, RS=RollingShutter, SW=Software, HW=Hardware.

tool, ensuring time alignment sub-1 ms, which is significantly lower than the period of every IMU. [4]

3.2.2 Timestamp and Exposure Time of Cameras The rolling shutter HDR cameras are triggered and hardware-timestamped at the midpoint of the exposure time (fixed 11 ms exposure) of the middle row within the image (per-row exposure times can be computed).

The CoreResearch global shutter camera is PTP synchronized, provides auto-exposure for all cameras, allows for exposure synchronization of the front-facing monocular stereo pair, and adjusts the exposure trigger signal for each camera such that the center of the exposure time aligns across all attached cameras and timestamps images at the middle of the exposure time. The ZED2i camera is connected via a USB serial buffer; therefore, only unreliable timestamps are available and without individual line read-out timings.

3.2.3 Calibration of Boxi

Calibration chain: As detailed in (Frey et al. 2025), we accurately calibrate the intrinsic and extrinsic parameters of the 10 cameras into a single, metrically consistent bundle. This accurate camera bundle then enables all subsequent joint

[4] The STIM320 required a software defined time-alignment based on angular velocities due to an time-independent offset, introduced by the communication between the kernel and the driver.

calibrations with the remaining sensors, including the prism, LiDARs, and IMUs.

Camera intrinsic and extrinsic calibration: For camera intrinsic and extrinsic calibration, we implemented a camera bundle calibrator inspired by Kalibr (Furgale et al. 2013; Rehder et al. 2016), which enables simultaneous calibration and data collection with real-time feedback to guide the data collection process. This is done in two phases. During the intrinsic calibration phase, the operator moves an AprilGrid (Olson 2011) throughout each camera's field of view until sufficient observations with adequate image plane coverage are acquired to estimate camera intrinsics, as illustrated in Fig. 20 of (Frey et al. 2025). Given the moderate field of view of all camera lenses, we adopt the pinhole projection model. For lens distortion, we employ the Kannala–Brandt (Kannala and Brandt 2006) equidistant model for the wide-angle fisheye `CoreResearch` and `HDR` cameras, and the rad-tan distortion model for the `ZED2i` cameras.

In the subsequent extrinsic calibration phase, the calibration target is placed statically at multiple locations around *Boxi* , or the payload itself is reoriented to different static poses. The calibration program automatically detects these static intervals and uses only these observations to optimize the inter-camera extrinsic parameters. This restriction is necessary as the 10 camera shutters do not trigger simultaneously. To guide the operator during data collection, the system periodically computes and reports the online covariance estimates of both intrinsic and extrinsic parameters, indicating when sufficient excitation has been achieved.

Camera to IMU calibration: First, we use the Allan variance tool (Buchanan 2021) to calibrate the IMU intrinsics. We then use Kalibr (Furgale et al. 2013) to calibrate the extrinsics of all IMUs with respect to the front-facing global shutter `CoreResearch` cameras. We follow the best practices outlined in (Furgale et al. 2013) for exposure time, camera frame rate, motion, and lighting. The specific parameters are detailed in (Frey et al. 2025).

LiDAR to camera extrinsic calibration: To efficiently perform the LiDAR–camera extrinsic calibration, we employ a variant of the intensity-alignment method DiffCal (Fu et al. 2023). In contrast to plane-based calibration, which requires numerous observations spanning diverse target orientations, the intensity-matching formulation is specifically data efficient and utilizes even partial observations of the calibration target within the LiDAR point cloud.

This property is particularly crucial for the `Hesai` sensor, whose narrow field of view often captures only fragments of the calibration pattern. Supporting partial observations enables us to collect calibration data near the cameras, where the camera pose estimates are more accurate. In contrast, the non-repeating scanning pattern of the `Livox` produces dense coverage through the non-repeating scanning pattern, enabling high-quality calibration even when the full target is not visible in each individual scan. On top of the LiDARs on *Boxi* , the `VLP16` LiDAR and the six `ANYmal-Depth` cameras provide further complementary observations. The complementary nature of the LiDAR and depth sensing in the GrandTour dataset is shown in Fig. 4.

Each LiDAR on *Boxi* is individually calibrated against the five `CoreResearch` cameras. Using the calibrated extrinsic parameters between the cameras, the detected calibration target poses from each camera are fused and jointly calibrated with respect to each LiDAR, without refining the intra-camera extrinsic parameters.

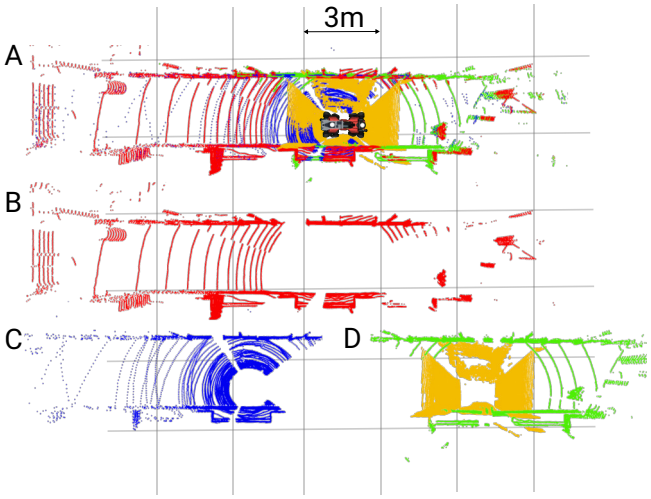


Figure 4. Complementary observations in GrandTour from the three LiDARs and the `ANYmal-Depth` cameras are shown with a grid. **A)**: The combined view of the scene with color-coding. **B)** `Hesai` LiDAR measurements in red. **C)** `Livox` LiDAR measurements in blue. **D)** ANYmal sensors, `VLP16` in green and `ANYmal-Depth` in yellow are shown.

Camera to prism calibration Due to the lack of existing methods, we developed a custom calibration procedure, to calibrate the front-facing `CoreResearch` cameras to the Leica GRZ101 mini prism. Using the MS60 total station position measurements and the detections of a static calibration target, we solve for a 9-degree-of-freedom state: the prism position relative to the reference camera (3-degree-of-freedom), and the 6-degree-of-freedom pose of the calibration target with respect to the total station. To ensure accurate target detection, the camera is positioned within 1 m of the calibration target. *Boxi* is held statically at various poses around the target using a tripod while the total station tracks the prism. Sufficient samples are collected to excite all translational and rotational degrees of freedom, ensuring the observability of the 9-degree-of-freedom optimization state. We validate the calibration by checking the consistency of prism estimates across all cameras, transforming them to the front-center camera frame using the known intra-camera extrinsic transform. The routine yields 3 mm cross-camera consistency. Furthermore, we validate the calibration accuracy by correlating real-world multi-modal data, as demonstrated in Fig. 5. Importantly, we have validated this overlay during a frame of motion. The exceptional alignment observed is the result of precise LiDAR motion compensation, combined with accurate camera–LiDAR extrinsics and camera intrinsics, which enable each motion-distorted LiDAR point to be correctly projected into the corresponding camera frame.

3.2.4 Boxi To ANYmal calibration The *Boxi* payload is mounted to ANYmal via a rigid, machined aluminum interface that utilizes precision locating pins and a preloaded bolt pattern, allowing for assembly until metal-to-metal contact is achieved. The extrinsic transformation between the base of *Boxi* and the base of ANYmal is obtained from the

CAD assembly and validated during integration by projecting the `VLP16` point clouds into the camera bundle frames using the chained extrinsic transformations. We have further verified the extrinsic calibration between the LiDARs of GrandTour by employing a map-to-map point cloud registration method under the locally no-drift assumption.

4 Dataset overview

The dataset comprises 49 distinct missions spread across a wide variety of natural and man-made environments. Each mission spans 3 min to 7 min and covers a distance of several hundred meters. The complete list of missions is provided in Table 3, along with the tags for each mission and its statistics.

The recorded data exhibit a large variance in environmental factors, spanning day and night, sunshine, cloudy weather, and rain, as well as challenging terrains such as sand, snow, and gravel. Fig. 7 summarizes the environments and other tags to facilitate dataset exploration.

Furthermore, the dataset webpage[5] offers a straightforward way to explore and access the data, featuring search functionality, preview videos for each mission, and a technical overview of the sensor setup.In addition to raw sensor measurements, the dataset includes intermediate outputs such as robot pose and velocity estimates, motion-compensated point clouds, accumulated point cloud maps[6], and occupancy maps, as shown in Sec. 6.

4.1 Data format and storage

The GrandTour dataset is available in two primary formats to maximize user accessibility and compatibility with existing workflows.

First, to ensure broader accessibility beyond the robotics community, we provide the dataset on HuggingFaceHub[7] in both *Zarr* and PNG formats. We allow for downloading individual missions, as well as subsets of sensor information, for ease of use. For each sensor and mission, the full extrinsic and intrinsic calibration is available within the *.zarr* files. This format effectively decouples the data from the ROS ecosystem, allowing computer vision researchers, machine learning practitioners, and others without robotics expertise to easily access and process the dataset using familiar tools and libraries. We provide a variety of minimalist Python scripts and notebooks to showcase loading, synchronization, and the use of different sensor streams. The storage hierarchy in the HuggingFace platform is provided in Fig. 8. The data can be easily fetched from HuggingFace through the Python API.

```python
from huggingface_hub import snapshot_download
hugging_face_data_cache_path = snapshot_download(
    "leggedrobotics/grand_tour_dataset",
    allow_patterns="*hdr_front*",
    repo_type="dataset"
)
```

Second, to facilitate rapid usage for robotics purposes and ensure backward compatibility, the dataset is available in ROS Bag (`.bag`) format. The mission is structured into multiple bag files in a minimalist fashion, with each file corresponding to a specific stream of messages (e.g., LiDAR, IMU, image) for a particular mission. This formatting decision allows users to download only the data they need, compared to substantially large `.bag` files that contain multiple sensor streams in an uncompressed format.

Furthermore, the bag files are compressed using the LZ4 lossless compression algorithm to reduce disk space usage. In addition, as ROS-1 reached end-of-life, the GrandTour dataset is accompanied by type-safe conversion scripts[8] that allow users to automatically convert the `.bag` files to CRF compressed `.mcap` files, which are seamlessly interpretable by the ROS-2 ecosystem. In this regard, we provide compatibility guarantees for our converted data with the LTS version of ROS2 Jazzy and Ubuntu 24.04, utilizing a Docker environment.

Each mission is divided into 34 individual ROS bags, from which users can choose. Users can inspect the contents of the bag files through the Kleinkram GUI web interface. These ROS-based formats preserve the original metadata from the deployment and can be integrated seamlessly with existing robotics software stacks.

The bag files are stored on ETH Zurich-based data servers and accessed through our in-house-built data tooling, called Kleinkram[9] (Püntener et al. 2025), a lightweight storage solution. Kleinkram is developed by the Robotic Systems Lab (RSL) and offers both a web and a command-line interface. The Kleinkram web interface provides a sortable list of all missions, along with the convenience of a web GUI, while the Kleinkram CLI (available via the Python package index[10] is recommended for systematic downloads and scripting. Kleinkram CLI can be easily installed with

```
pip3 install kleinkram
```

```
klein download --project GrandTourDataset
               --mission 2024-10-01-12-00-49
               ↪  --dest /tmp 'hdr_front'
```

4.2 Ground Truth Pose Generation

Satellite Reference For outdoor experiments, we obtain ground-truth poses from a NovAtel SPAN CPT7 inertial navigation system (INS), a dual-antenna GNSS receiver with the integrated `HG4930` Honeywell high-end MEMS IMU. The CPT7 provides real-time pose estimates using the TerraStar-C PRO[11] multi-constellation correction service for Precise Point Positioning (PPP), and refined offline trajectories using Inertial Explorer (IE) (NovAtel 2025a),

[5] https://grand-tour.leggedrobotics.com
[6] https://drive.google.com/drive/folders/1VV83LVWjfVmjx8WXG_QFYENjXSH2GN60?usp=sharing
[7] https://huggingface.co/datasets/leggedrobotics/grand_tour_dataset
[8] https://github.com/leggedrobotics/grand_tour_box/blob/main/box_utils/box_auto/python/box_auto/scripts/special/generizableROS2conversion.py
[9] https://github.com/leggedrobotics/kleinkram
[10] https://pypi.org/project/kleinkram/
[11] https://hexagon.com/company/divisions/autonomy-and-positioning/rtk-from-sky

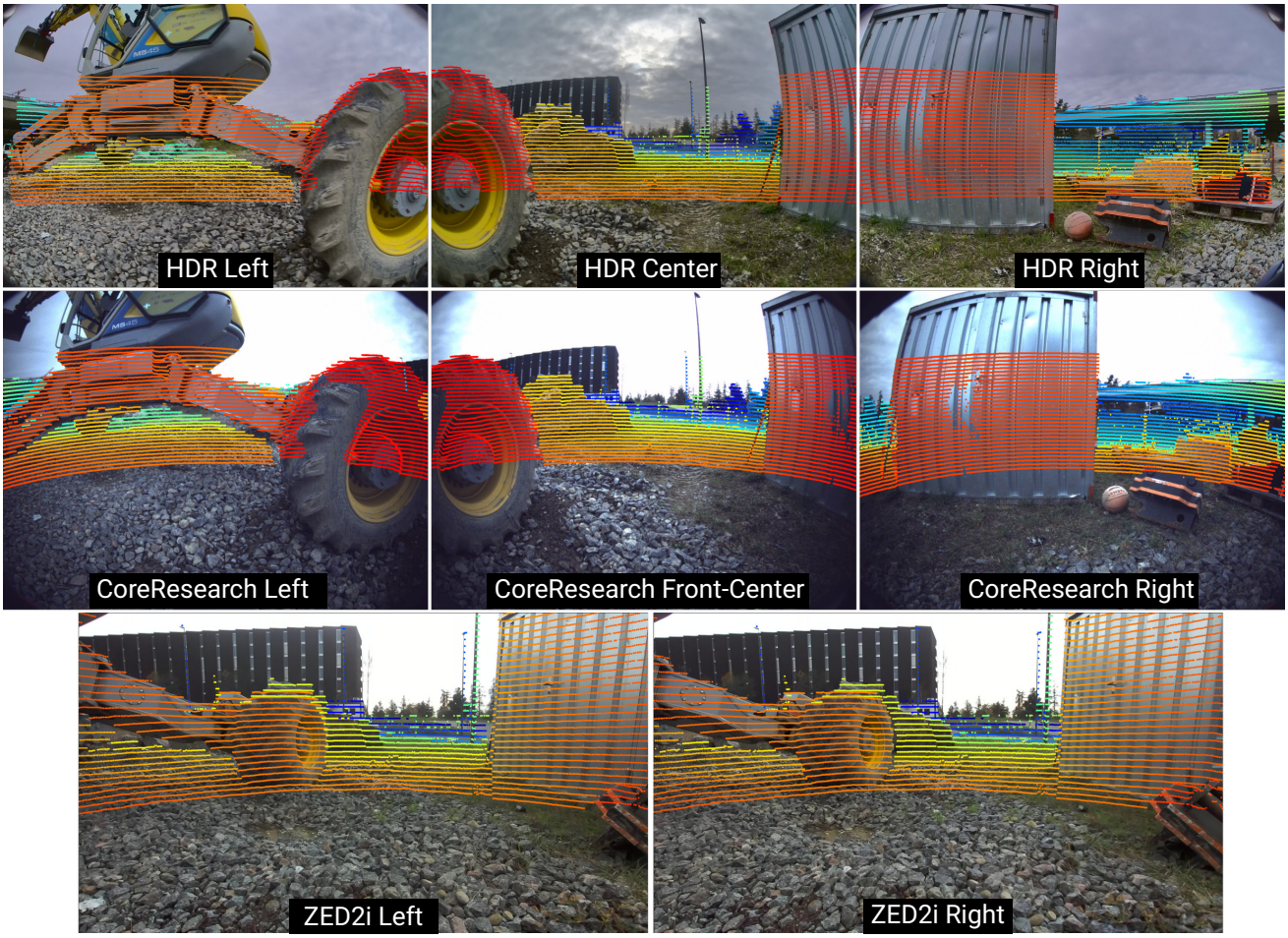


Figure 5. The calibration provided with the GrandTour dataset is validated through the overlay of point clouds from different LiDARs onto all available RGB images of GrandTour. As shown, the point clouds align with the correct visual features in the image. Points are colorized based on their depth in the camera axis; red indicates closer, and blue is further away.

a GNSS/INS post-processing software by NovAtel Inc., in combination with the HxGN SmartNet correction service.[12]

IE employs a Kalman filter that jointly estimates the GNSS states (position, velocity, clock offset, and carrier-phase ambiguities) and the platform states (position, velocity, attitude), along with gyroscope and accelerometer biases.

The two filters are tightly integrated: GNSS-derived measurements bound the drift of the inertial solution, while the INS provides position and velocity estimates that stabilize and constrain the GNSS solution under challenging conditions, mitigating multipath effects and signal outages. In addition to the platform pose, IE outputs linear velocity, IMU bias estimates, associated covariance information, and quality metrics.

We use the tightly coupled (TC) processing modes of IE. In this mode, GNSS and inertial data are fused at the measurement level, enhancing satellite signal utilization under challenging GNSS conditions and mitigating inertial error growth even without a full GNSS position update. NovAtel recommends the tightly coupled solution due to its higher accuracy and faster convergence.

According to the CPT7 performance specifications of IE (NovAtel 2025b), GNSS outages of up to $10\,\mathrm{s}$ result in position RMS errors in the range of $0.01\,\mathrm{m}$ to $0.02\,\mathrm{m}$, velocity RMS below $0.015\,\mathrm{m\,s^{-1}}$, and attitude RMS below $0.01°$.

4.2.1 Robotic Totalstation Reference To improve accuracy and expand beyond open-sky outdoor settings, we follow previous work (Vaidis et al. 2024) and establish ground-truth trajectories using a Leica MS60 robotic total station (TPS). Our Leica MS60 robotic total station provides 3D position measurements at $20\,\mathrm{Hz}$ with a range accuracy of $1.5\,\mathrm{mm}$. Unlike prior work, our TPS measurements are highly accurate in time synchronization with the robot's sensors using a custom Leica Geosystems AP20 autopole (Frey et al. 2025).

In contrast, most other works rely on stationary laser scanners (Tao et al. 2025; Zhao et al. 2024b) to generate highly accurate point maps; however, the collection is expensive, impractical, or even impossible in dynamic environments. Moreover, the ground truth accuracy and its frequency are limited by the measurement characteristics of the onboard LiDAR used for reference against the map. Overall, the achieved accuracy is highly dependent on environmental characteristics (e.g., LiDAR-degenerate environments).

While our TPS measurements are highly accurate, they require a direct line-of-sight to the reflector mounted on the robot and, therefore, are not always available throughout a deployment. This directly motivates the fusion of inertial measurements, GNSS data, and TPS measurements, described

[12] https://hxgnsmartnet.com/

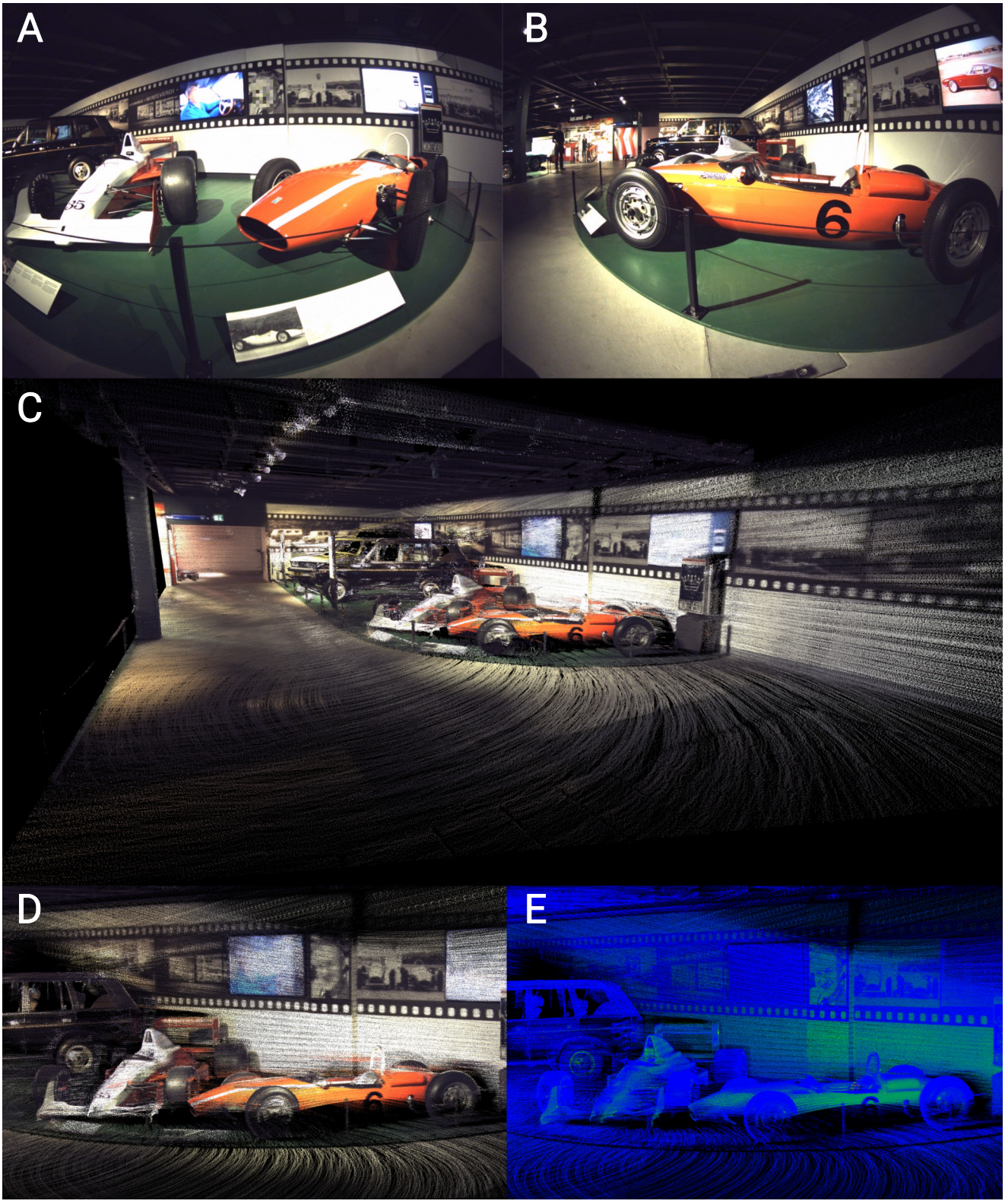


Figure 6. A colorized point cloud map of the HAUS-1 sequence is shown to highlight the accuracy of the sensor calibration. **A)** and **B)** shows the front-center colored `CoreResearch` camera from different moments during the sequence, whereas **C**) shows the colorized point cloud map generated by FAST-LIVO2 (Zheng et al. 2024) using `Hesai` LiDAR and the front-center colored `CoreResearch` camera. **D)** shows a view rendered within the colorized map, and **E)** illustrates the same view on the intensity spectrum of `Hesai` LiDAR.

next, to achieve highly accurate, high-frequency ground-truth pose estimates.

Holistic Fusion Ground Truth Holistic Fusion (HF) (Nubert et al. 2025), a flexible task- and setup-agnostic robot localization and state estimation library based on GTSAM (Dellaert 2012), enables the fusion of TPS position measurements, IE post-processed poses,

Codename	Collection date	Duration[s]	Walked Distance[m]	MS60 Coverage[%]	GNSS	Tags
ETH-1	2024-10-01 11:29:55	333	204.729	74.934	Yes	Outdoor, Pavement, Cars, People, Rain, Cloudy
ETH-2	2024-10-01 11:47:44	418	210.591	67.846	No	Indoor, People, Landmarks, Dynamic Objects
ETH-3	2024-10-01 12:00:49	464	214.495	6.519	Yes	Urban, Outdoor, Pavement, Crosswalk, Cars, People, Rain, Cloudy
SPX-1	2024-11-02 17:10:25	374	147.114	75.748	Yes	Industrial, Outdoor, Metal, Stairs, People, Dawn
SPX-2	2024-11-02 17:18:32	383	184.128	68.168	Yes	Industrial, Outdoor, Metal, Stairs, People, Dawn
SPX-3	2024-11-02 17:43:10	213	101.313	82.302	Yes	Industrial, Outdoor, Metal, People, Night
ICE-1	2024-11-02 21:12:51	251	56.253	91.71	Yes	Indoor, Ice, People
SNOW-1	2024-11-03 07:52:45	172	36.533	47.39	Yes	Search & Rescue, Mountain, Outdoor, Snow, Sunlight
SNOW-2	2024-11-03 07:57:34	261	173.927	99.433	Yes	Search & Rescue, Mountain, Outdoor, Snow, Sunlight
SNOW-3	2024-11-03 08:17:23	236	145.253	99.872	Yes	Search & Rescue, Mountain, Outdoor, Snow, Sunlight
EIG-1	2024-11-03 13:51:43	429	219.702	70.360	Yes	Search & Rescue, Mountain, Outdoor, Gravel/Dirt, Trail, Pavement, Metal, Stairs, People, Tracks, Sunlight
EIG-2	2024-11-03 13:59:54	348	198.286	72.319	Yes	Search & Rescue, Mountain, Outdoor, Gravel/Dirt, Trail, Pavement, Metal, Stairs, People, Tracks, Sunlight
GRI-1	2024-11-04 10:57:34	455	266.747	58.072	Yes	Urban, Outdoor, Grass, Pavement, Stairs, Cars, People, Sunlight
CYN-1	2024-11-04 12:55:59	510	297.702	23.971	Yes	Search & Rescue, Forest, Mountain, Outdoor, Gravel/Dirt, Trail, Pavement, Sunlight
CYN-2	2024-11-04 13:07:13	314	164.594	49.308	Yes	Search & Rescue, Forest, Mountain, Outdoor, Gravel/Dirt, Trail, Pavement, Sunlight
HIL-1	2024-11-04 16:05:00	430	326.08	72.191	Yes	Forest, Outdoor, Mountain, House
RIV-1	2024-11-04 16:52:38	537	204.016	64.089	No	Forest, Water, Outdoor, Gravel/Dirt, Sand, Trail, Cars, Sunlight
PIL-1	2024-11-11 12:07:40	946	470.525	21.009	Yes	Mountain, Outdoor, Pavement, Cobblestone, Stairs, People, Sunlight
PIL-2	2024-11-11 12:42:47	409	273.567	98.661	Yes	Mountain, Outdoor, Gravel/Dirt, Trail, Stairs, People, Sunlight
ROOT-1	2024-11-11 14:29:44	340	177.51	69.225	Yes	Search & Rescue, Forest, Mountain, Outdoor, Gravel/Dirt, Trail, Sunlight
HAUS-1	2024-11-11 16:14:23	337	170.493	66.008	No	Indoor, Pavement, Cars, People
HÖB-1	2024-11-14 11:17:02	517	289.299	39.166	Partially	Forest, Outdoor, Trail, Cloudy
HÖB-2	2024-11-14 12:01:26	445	197.271	27.944	Partially	Forest, Outdoor, Trail, Cloudy
HEAP-1	2024-11-14 13:45:37	290	162.041	86.737	Yes	Water, Industrial, Outdoor, Gravel/Dirt, Grass, Trail, Cloudy
KÄB-1	2024-11-14 14:36:02	465	205.374	79.346	Yes	Forest, Outdoor, Grass, Trail, Cloudy
KÄB-2	2024-11-14 15:22:43	383	185.596	41.099	Yes	Forest, Outdoor, Gravel/Dirt, Trail, Cloudy
KÄB-3	2024-11-14 16:04:09	475	364.575	48.3	Yes	Outdoor, Gravel/Dirt, Trail, Pavement, People, Cloudy
TRIM-1	2024-11-15 10:16:35	465	267.811	72.489	Yes	Forest, Outdoor, Gravel/Dirt, Trail, People, Cloudy
ALB-1	2024-11-15 11:18:14	282	107.976	41.31	Yes	Forest, Outdoor, Gravel/Dirt, Trail, People, Cloudy
ALB-2	2024-11-15 11:37:15	508	206.572	65.356	Yes	Forest, Outdoor, Gravel/Dirt, Mud, Trail, Cloudy
ALB-3	2024-11-15 12:06:03	209	77.971	88.959	Yes	Forest, Outdoor, Gravel/Dirt, Mud, Trail, Cloudy
LMB-1	2024-11-15 14:14:12	528	204.999	58.058	Yes	Forest, Outdoor, Grass, Trail, Cloudy
LMB-2	2024-11-15 14:43:52	244	100.881	18.822	Yes	Forest, Outdoor, Gravel/Dirt, Trail, Cloudy
LEE-1	2024-11-15 16:41:14	529	331.839	71.635	Partially	Underground, Industrial, Indoor, Outdoor, Pavement, People, Cloudy
ARC-1	2024-11-18 12:05:01	433	247.878	71.266	Yes	Search & Rescue, Industrial, Urban, Indoor, Outdoor, Pavement, Cobblestone, Metal, Stairs, Cloudy
ARC-2	2024-11-18 13:22:14	424	136.222	27.346	Yes	Search & Rescue, Water, Industrial, Urban, Indoor, Outdoor, Pavement, Stairs, Cloudy
ARC-3	2024-11-18 13:48:19	713	331.974	27.467	Partially	Search & Rescue, Industrial, Urban, Indoor, Outdoor, Grass, Pavement, Stairs, Cloudy
ARC-4	2024-11-18 15:46:05	397	160.04	20.256	Yes	Search & Rescue, Industrial, Urban, Indoor, Outdoor, Grass, Pavement, Stairs, Cloudy
ARC-5	2024-11-18 16:59:23	391	173.303	42.052	Yes	Search & Rescue, Industrial, Urban, Indoor, Outdoor, Grass, Pavement, Cobblestone, Dawn
ARC-6	2024-11-18 17:13:09	349	103.493	62.913	No	Search & Rescue, Industrial, Indoor, Smoke, Pavement
ARC-7	2024-11-18 17:31:36	322	157.244	35.778	Yes	Search & Rescue, Industrial, Urban, Indoor, Outdoor, Smoke, Pavement, Metal, Stairs, Night
LEICA-1	2024-11-25 14:57:08	310	152.656	80.195	Yes	Urban, Indoor, Outdoor, Pavement, Cars, People, Sunlight
LEICA-2	2024-11-25 16:36:19	422	208.171	46.744	Yes	Indoor & Outdoor, Warehouse, People, Industrial
SBB-1	2024-12-03 13:15:38	502	311.25	99.492	Yes	Industrial, Outdoor, Gravel/Dirt, Tracks, Sunlight
SBB-2	2024-12-03 13:26:40	271	149.167	73.16	Yes	Industrial, Outdoor, Gravel/Dirt, Tracks, Sunlight
CON-1	2024-12-09 09:34:43	388	263.267	75.999	No	Water, Construction, Indoor, Outdoor, Gravel/Dirt, Woodplanks, People, Cloudy
CON-2	2024-12-09 09:41:46	384	266.356	61.221	No	Water, Construction, Indoor, Outdoor, Gravel/Dirt, People, Cloudy
CON-3	2024-12-09 11:28:28	796	188.934	23.632	No	Construction, Indoor, Outdoor, Stairs, People, Cloudy
CON-4	2024-12-09 11:53:11	743	468.315	67.106	No	Construction, Outdoor, Mud, Stairs, People, Cloudy

Table 3. Overview of all missions available in the GrandTour dataset is provided.

Table 4. Performance characteristics of Inertial Explorer. All values are within 2σ variation.

GNSS Outage [s]	Position RMS[m]		Velocity RMS[$\mathrm{m\,s^{-1}}$]		Attitude RMS[°]	
	Horiz.	Vert.	Horiz.	Vert.	Roll/Pitch	Heading
0 s	0.01	0.02	0.015	0.01	0.003	0.01
10 s	0.01	0.02	0.015	0.01	0.003	0.01
60 s	0.11	0.05	0.017	0.01	0.004	0.014

and HG4930 IMU measurements. HF uses a factor graph formulation, where the state variables of the robot and the relative context variables, such as the alignment transforms, are estimated as a MAP estimate:

$$\mathcal{X}^\star = \arg\max_{\mathcal{X}} p(\mathcal{X}|\mathcal{Z}) \propto p(\mathcal{Z}|\mathcal{X})p(\mathcal{X}). \quad (1)$$

Here, $\mathcal{X}$ denotes the full set of robot states and dynamic state variables $\{{}^{I}\mathcal{X}_{N_I}, {}^{G}\mathcal{X}_{N_G}, {}^{R}\mathcal{X}_{N_R}\}$, with ${}^{I}\mathcal{X}_{N_I}$ denoting the robot state in the world frame, while ${}^{G}\mathcal{X}_{N_G}$ and ${}^{R}\mathcal{X}_{N_R}$ denote global (ECEF)- and reference (MS60) frame alignment-states.

The key element in this formulation is the alignment transform context variable, which actively aligns the two global fixed frames while ensuring the 6-DoF pose is retrieved through the integration of IMU measurements, regardless of the availability of the IE post-processed pose. Despite this formulation, periods without both MS60 position measurements and IE post-processed unary poses necessitate IMU-only dead reckoning, resulting in drift accumulation. Therefore, we restrict the ground-truth time span to segments with IE pose availability or MS60 measurements.

The combination of total station measurements, post-processed INS solution, and high-grade IMU delivers ground truth trajectories that outperform typical GNSS-only and LiDAR-only ground truth generation solutions. This is illustrated in Fig. 9, where the HF output is shown to follow absolute position measurements when available and to complement them with IE estimates and the IMU when the line of sight is blocked. When compared against the raw TPS position measurements, ATE for HF is a mean

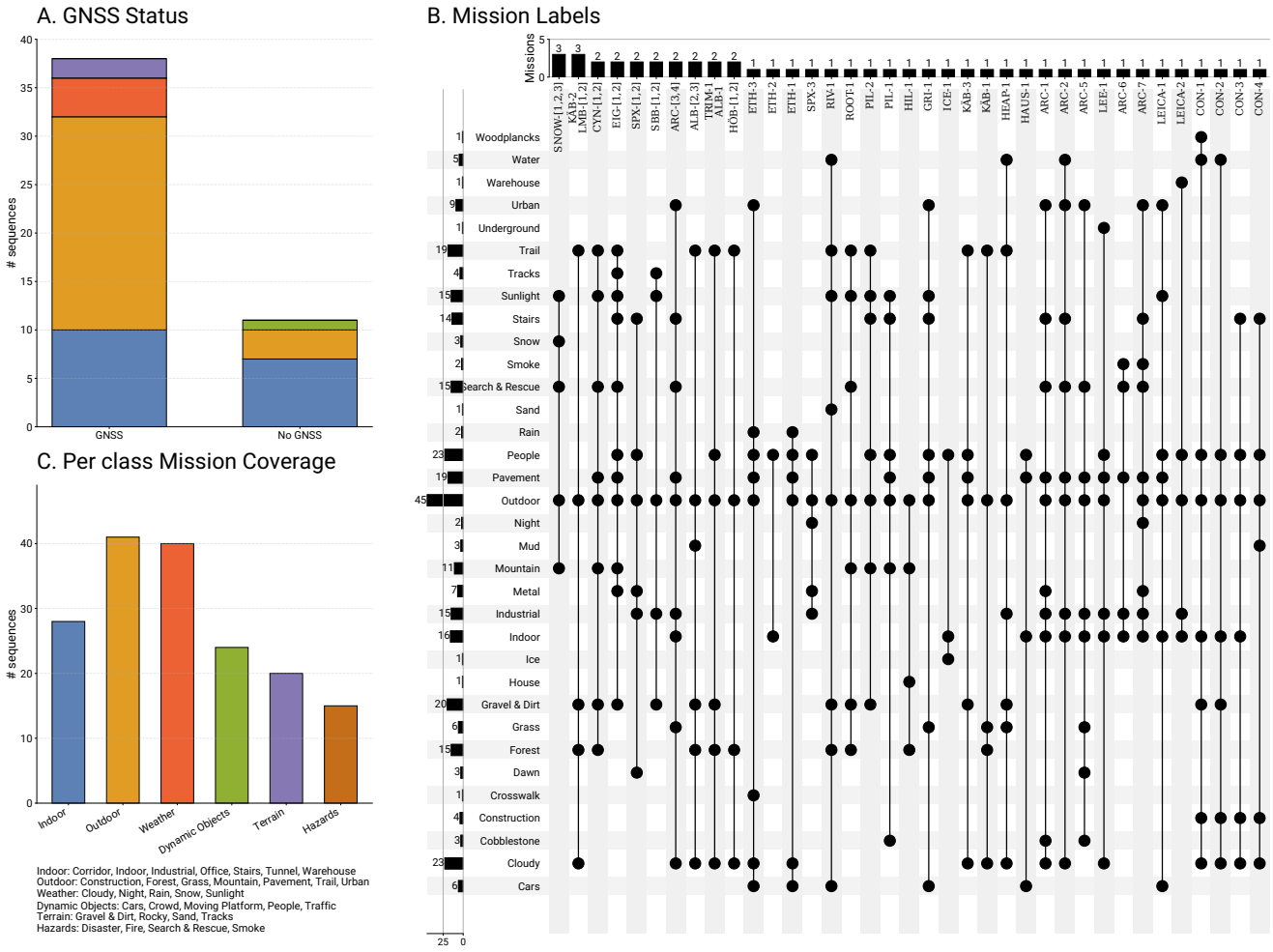


Figure 7. GNSS coverage and semantic diversity of the dataset. **A)** Number of sequences recorded with and without GNSS. **B)** The most common intersections of semantic labels across all missions, illustrating how weather, environment, terrain, and hazard conditions co-occur. **C)** Number of sequences per high-level label class, where each class aggregates the underlying labels listed below the plot.

of $0.0028\,\mathrm{m}$ with standard deviation (σ) of $0.0020\,\mathrm{m}$ and RMSE $0.0034\,\mathrm{m}$, whereas the IE-TC solution yields a mean ATE of $0.132\,\mathrm{m}$ with σ =$0.0721\,\mathrm{m}$ and RMSE $0.1504\,\mathrm{m}$. For RTE (Δ $0.5\,\mathrm{m}$), HF achieves a mean of $0.0030\,\mathrm{m}$ with σ =$0.0021\,\mathrm{m}$ and RMSE $0.0037\,\mathrm{m}$, compared to IE-TC with mean $0.0073\,\mathrm{m}$, σ =$0.0095\,\mathrm{m}$, and RMSE $0.012\,\mathrm{m}$. Overall, HF can effectively fuse the 6-DoF IE estimation without violating the absolute TPS measurements when available.

5 Data Collection

In the following section, we explain the procedure for deploying ANYmal with the *Boxi* payload to collect a GrandTour mission. The sequence is further illustrated in Fig. 10.

1. **Site Preparation:** Define a safe trajectory for robot teleoperation, specifically selecting diverse sites that present varying challenges. Determine a good-coverage position for the Leica MS60 and the location of a reference prism (used to verify that the MS60 remains stationary throughout the mission).
2. **Start Recording:** Verify the validity of sensor data. Initiate recording and measure the reference prism location at the start of the sequence. If the robot is within the line-of-sight of the total station, establish a lock on the robot.
3. **Operator Instruction:** The teleoperator follows behind the robot and aims to be not in the front-camera view. The operator may physically intervene to ensure safety (e.g., stabilizing the robot when descending stairs or navigating near cliffs). If the total station loses track of the robot, the operator navigates to the next location where a lock can be re-established. In some cases, the robot must be stopped to allow for re-locking, resulting in intermittent pauses within the trajectory.
4. **Mission Completion:** Once the robot reaches the final destination, re-measure the reference prism position to validate station stability. Stop the recording and power down the instruments.

6 Collected Data and Derived Outputs

All available raw sensory data is detailed in Table 5 and derived outputs based on further processing during or after the mission are detailed in Table 6. During the mission, we additionally record leg odometry, which is computed online using the `ANYmal-IMU` data and `ANYmal-JE`. The state estimator responsible for this odometry also estimates feet contact states, as shown in Fig. 11.

Sensor	Tag	HuggingFace	Rosbag	Topic	Recording Info
(CoreResearch)	<front_center>, <left>, <right>	alphasense_<tag>	alphasense	/boxi/alphasense/<tag>/camera_info /boxi/alphasense/<tag>/image_raw/compressed	10 Hz - 1440×1080 - RGB
(CoreResearch)	<front_left>, <front_right>	alphasense_<tag>	alphasense	/boxi/alphasense/<tag>/camera_info /boxi/alphasense/<tag>/image_raw/compressed	10 Hz - 1440×1080 - MONO
(HDR)	<front>, <left>, <right>	hdr_<tag>	hdr_<tag>	/boxi/hdr/<tag>/camera_info /boxi/hdr/<tag>/image_raw/compressed	10 Hz - 1920×1280 - RGB
(ZED2i)	<left>, <right>	zed2i_<tag>_image	zed2i_images	/boxi/zed2i/<tag>/camera_info /boxi/zed2i/<tag>/image_raw/compressed	15 Hz - 1920×1080 - RGB
(ZED2i)	–	zed2i_prop	zed2i_prop	/boxi/zed2i/baro/pressure /boxi/zed2i/magnetic_field	Pressure + magnetic field
(Hesai)	–	hesai_packets	hesai_packets	/boxi/hesai/packets	10 Hz - Raw LiDAR packets
(Hesai)	–	hesai	hesai	/boxi/hesai/points	10 Hz - Hesai XT-32, ∼640,000 points [Single Return]
(Hesai)	–	hesai	hesai	/boxi/hesai/intensity_image /boxi/hesai/range_image	Projected range and intensity images
(Livox)	–	livox_points	livox	/boxi/livox/points	10 Hz - Livox Mid360 non-repetitive LiDAR
(VLP16)	–	velodyne	anymal_velodyne	/anymal/velodyne/points	10 Hz - Velodyne VLP-16 (Integrated in ANYmal)
(VLP16)	–	anymal_velodyne_packets	anymal_velodyne_pac	/anymal/velodyne/packets	10 Hz - Raw LiDAR packets
(ZED2i)	–	zed2i_depth_image	zed2i_depth	/boxi/zed2i/depth/camera_info /boxi/zed2i/depth/image_raw/compressed	15 Hz - 1920×1080 - Stereolabs ZED2i Depth
(ZED2i)	–	zed2i_depth_confidence_i	zed2i_depth	/boxi/zed2i/confidence/image_raw/compressed	15 Hz - 1920×1080 - Stereolabs ZED2i Confidence
(ANYmal-Depth)	<front_upper>, <front_lower>, <rear_upper>, <rear_lower>, <left>, <right>	depth_camera_<tag>	anymal_depth_camera	/anymal/depth_camera/<tag>/depth/camera_info /anymal/depth_camera/<tag>/depth/image_rect_raw	15 Hz - 848x480 - RealSense D435i
(HG4930)	–	cpt7_imu	cpt7_imu	/boxi/cpt7/imu	Honeywell HG4930, 100 Hz
(STIM320)	–	stim320_imu	stim320_imu	/boxi/stim320/imu /boxi/stim320/accelerometer_temperature /boxi/stim320/gyroscope_temperature	Safran STIM320, 500 Hz
(ANYmal-IMU)	–	anymal_imu	anymal_imu	/anymal/imu	-, 400 Hz (integrated in ANYmal)
(AP20-IMU)	–	ap20_imu	ap20_imu	/boxi/ap20/imu	Leica AP20, 200 Hz
(ADIS)	–	adis_imu	adis	/boxi/adis/imu	Analog Devices ADIS16475-2, 200 Hz
(Bosch)	–	alphasense_imu	alphasense_imu	/boxi/alphasense/imu	Bosch BMI085, 200 Hz (Integrated in CoreResearch)
(TDK)	–	livox_imu	livox_imu	/boxi/livox/imu	TDK ICM40609, 200 Hz (integrated in Livox Mid360)
(Prism)	–	prism_position	ap20_prism_position	/boxi/ap20/prism_position	Position relative to Leica MS60 Total Station
(ANYmal-JE)	–	anymal_state_actuator	anymal_state	/anymal/actuator_readings	ANYmal joint readings, 12 actuators
ANYmal	–	anymal_state_battery	anymal_state	/anymal/battery_state	ANYmal battery state readings
ANYmal	–	anymal_command_twist	anymal_command	/anymal/twist_command	Operator-issued twist command

Table 5. Overview of all sensor streams recorded on a GrandTour sequence (excluding derived outputs), showing the sensor name, tag, corresponding HuggingFace dataset identifier, bag name, ROS topics.

Sensor	Tag	HuggingFace	Rosbag	Topic	Recording Info
(Hesai)	–	hesai_undist	hesai_undist	/boxi/hesai/points_undistorted	Motion-compensated point cloud
(Livox)	–	livox_undist	livox_undist	/boxi/livox/points_undistorted	Motion-compensated point cloud
(VLP16)	–	anymal_velodyne_undist	anymal_velodyne_undist	/anymal/velodyne/points_undistorted	Motion-compensated point cloud
(CPT7)	<rt>, <tc>	cpt7_ie_<tag>	cpt7_ie_<tag>	/boxi/inertial_explorer/<tag>/raw /boxi/inertial_explorer/<tag>/navsatfix /boxi/inertial_explorer/<tag>/odometry /boxi/inertial_explorer/<tag>/cpt7_imu_acc_bias /boxi/inertial_explorer/<tag>/cpt7_imu_gyro_drift /tf	Inertial Explorer navigation solution + biases
(TF)	<minimal>, <model>	tf_<tag>	tf_<tag>	/tf /tf_static	Transform tree for sensor frames
DLIO (Chen et al. 2023)	–	dlio	dlio	/boxi/dlio/hesai_points_undistorted /boxi/dlio/lidar_map_odometry /tf	Undistorted point cloud, LiDAR pose in map)
(ZED2i)	–	zed2i_vio	zed2i_vio	/boxi/zed2i_vio/map /boxi/zed2i_vio/odom /tf	ZED2i VO outputs (map + odometry)
Elevation Map (Miki et al. 2022b)	–	anymal_elevation	anymal_elevation	/anymal/elevation_mapping/elevation_map_raw	Elevation mapping output (raw map)
TSIF	–	anymal_state_state_estima	anymal_state	/anymal/state_estimator/anymal_state /anymal/state_estimator/odometry /anymal/state_estimator/pose	ANYmal full state, estimator pose/odometry, contact estimation

Table 6. Overview of all derived outputs of the GrandTour dataset is shown alongside auxiliary data.

In post-processing, the leg odometry is utilized to generate motion-compensated LiDAR point clouds for every LiDAR. Furthermore, we provide Inertial Navigation System (INS) trajectories derived from the CPT7. These include both the real-time solution and a high-precision, tightly coupled solution generated via Inertial Explorer. We also provide trajectory data based on Direct LiDAR-Inertial Odometry (DLIO) (Chen et al. 2023). In addition, sanitization efforts have been taken to ensure the data is ready for downstream applications. Some of these efforts include ensuring that measurements appear in their respective formats as provided by the drivers, rather than logging time and filtering the robot body from the LiDAR point clouds.

7 Dataset applications

In this section, we showcase how the GrandTour supports *1)* State Estimation and localization (Sec. 7.1), *1)* Perception (Sec. 7.2), and *3)* Locomotion & Navigation (Sec. 7.3) applications in the robotics literature.

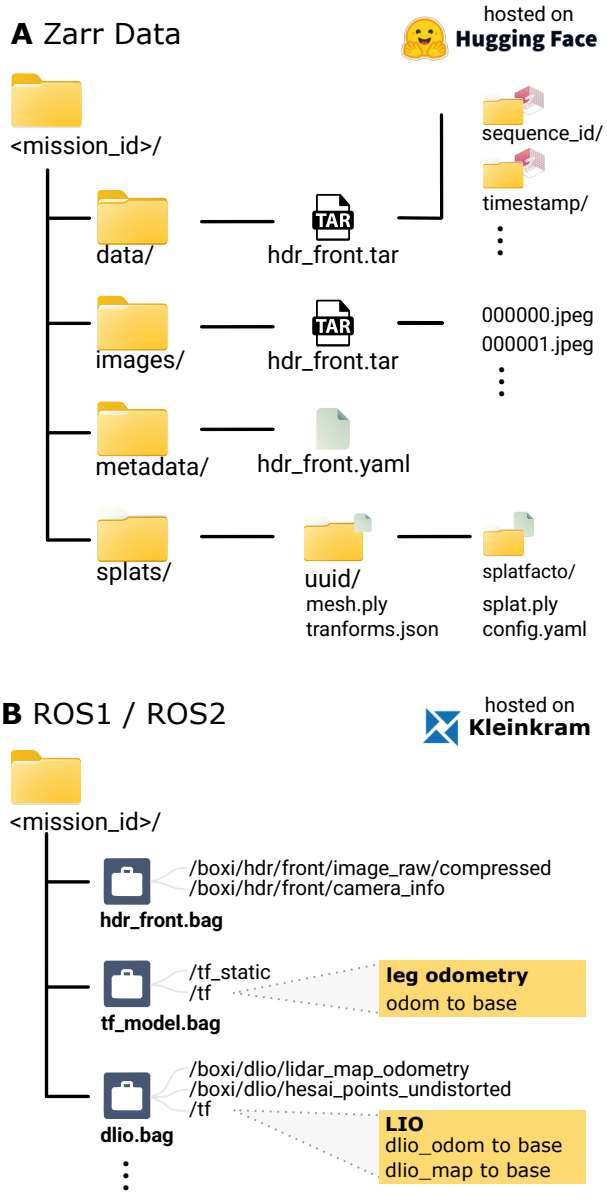


Figure 8. Data format overview. **A**) Zarr layout on HuggingFace. Topics are stored as multi-arrays; for instance, the `hdr_front` contains a timestamp and sequence_id. For images, the sequence_id can be used to retrieve the corresponding `.jpeg` file. Metadata, including intrinsics, extrinsics, and descriptions, is provided both within the Zarr attributes and the YAML file in the metadata folder. Post-processed assets (Gaussian Splats, meshes) are stored in the `splats` folder. Only (depth)-image topics are stored as `.jpeg` or `.png` for compression. **B**) Rosbags are hosted on kleinkram. Topic namespaces distinguish between *Boxi* and *ANYmal* platforms. Sensor intrinsics are embedded in messages, while extrinsics reside in `tf_model.bag`. Leg odometry is included in the `tf` tree by default; derived outputs are separated into distinct bag files.

7.1 State estimation, localization and mapping

Precise state estimation is essential for legged robots operating in real environments. To expose failure modes across diverse motions (e.g., walking, trotting, slipping) and environments (indoor, urban, unstructured terrain), we benchmark a broad set of open-source solutions on six GrandTour sequences: SPX-2, SNOW-2, EIG-1, CON-4, ARC-2, and ARC-7 (previews of these missions are available in Table 3).

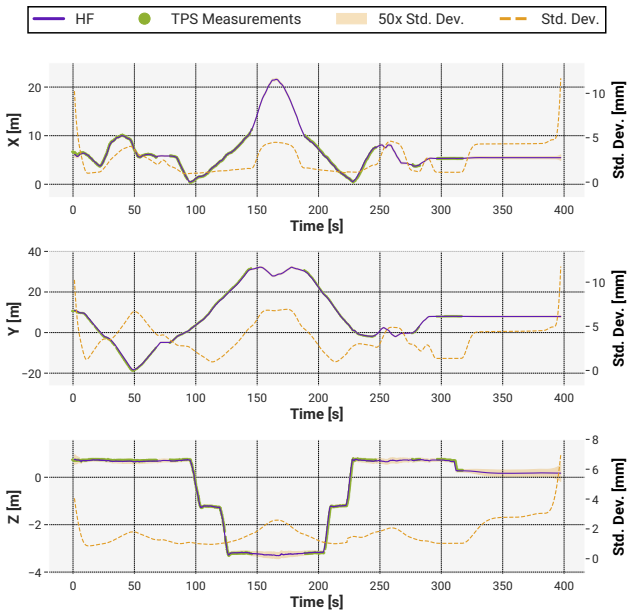


Figure 9. HF-based 6-DoF ground truth generation is shown for the SPX-2 sequence of GrandTour. As expected, when the MS60 line-of-sight is obstructed, the standard deviation of the estimate increases.

- **SPX-2**: On the Jungfraujoch research station, dynamic initialization is needed. Consists of thin structures and moving objects. Presents a loop-closure.
- **SNOW-2**: Fast walking on fully snow-covered terrain. Small-scale loop-closures are present. Limited planar structures.
- **EIG-1**: This sequence illustrates descent from a hill through a train station with long-walks. The large scale environment contains both planar and non-planar features.
- **CON-4**: The longest test sequence, featuring a construction environment, dynamic objects, a large-scale environment, planar features, and a big loop-closure.
- **ARC-2**: This sequence features narrow passages and limited visibility in a rescue worker training center. Featureless walls challenge visual methods, while long-range LiDARs often fail to capture the floor or ceiling.
- **ARC-7**: Starts indoors and showcases completely dark indoor-outdoor passages, significantly challenging visual methods. Limited re-observations.

In total, we evaluated 52 open-source solutions designed to run on robotic systems in real-time and online fashion; these works are presented in Table 7. In our evaluation, we excluded 16 additional works that we attempted to evaluate but were unable to run reliably. To limit speculation on performance, we do not disclose the names of these works. Additionally, we have omitted work that can be applied to multiple baseline solutions, such as information-aware sampling (Petracek et al. 2024; Gelfand et al. 2003), loop-closure (Yuan et al. 2024a; Kim et al. 2021b), offline bundle adjustment (Liu et al. 2023), or degeneracy-awareness (Tuna et al. 2024, 2025; Hu et al. 2025; Hatleskog and Alexis 2024) methods. The evaluated

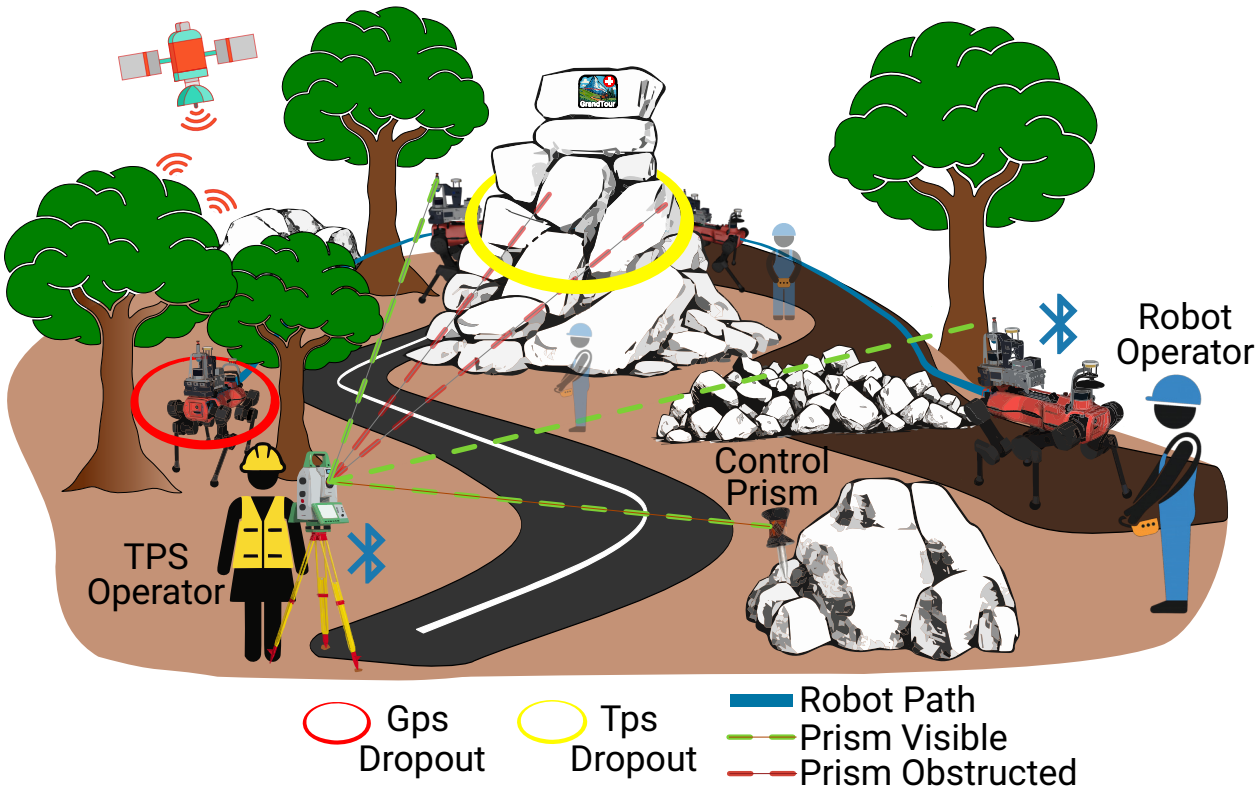


Figure 10. Data collection methodology for the GrandTour dataset. The ANYmal robot equipped with the *Boxi* sensor payload traverses environments with varying GNSS availability (red circle indicates denied regions) and total station visibility constraints (yellow circle). The deployment setup includes a Leica MS60 total station operated by a TPS operator and a separate operator for robot teleoperation. The TPS operator ensures the total station is always tracking the robot, while the teleoperator controls the robot's movement and ensures it traverses the environment safely. The control prism ensures the world frame dictated by the total station is stable and static throughout the mission.

methods span from (multi)-LiDAR odometry (LO), (multi)-LiDAR-inertial odometry and LiDAR-SLAM, (Stereo/Mono)-visual-inertial odometry (VIO), to multi-sensor variants (e.g., LIVO and RGBD-inertial).

The authors spent a substantial amount of time adequately tuning and configuring each framework according to its published guidelines to conduct an objective evaluation of the framework through a third-party, i.e., the *baseline* variant. We have disabled loop-closure features for all methods, but have kept online bundle-adjustment and pose-graph-optimization features online. While this is the case, the authors openly invite researchers to submit their *author-tuned* variant to the benchmark that will be available on the dataset website.

We computed error metrics using the `evo` evaluation toolkit (Grupp 2017). Each estimated trajectory is transformed to be expressed in the GrandTour ground-truth reference frame (`prism` frame) using the provided frame conventions and extrinsic calibration.

For trajectory association, rely again on `evo`, this step accounts for differing output rates by matching each estimate timestamp to the closest ground-truth timestamp using threshold $t_{\max_diff}$, set as $t_{\max_diff} \approx 0.5\,\mathrm{median}(\Delta t_{\text{dense}})$, where Δt_{dense} is the timestep of the higher-rate (more densely sampled) reference trajectory. Concretely, for a 20 Hz reference trajectory and a 10 Hz estimator output, $\Delta t_{\text{dense}} = 0.05\,\text{s}$ and thus $t_{\max_diff} = 0.025\,\text{s}$: if the reference timestamps are $\{0.00, 0.05, 0.10, 0.15, \dots\}$ and the estimator timestamps are $\{0.03, 0.13, 0.23, \dots\}$ (no jitter for brevity, just a fixed start offset), `evo` associates $0.03 \to 0.05$ and $0.13 \to 0.15$ with $|\Delta t| = 0.02\,\text{s} \leq 0.025\,\text{s}$, so these pairs are accepted (whereas, e.g., $t_{\max_diff} = 0.01\,\text{s}$ would reject them).

For segment-based relative metrics, we use a path-length increment of $\Delta = 0.5\,\text{m}$, which is selected based on the motion characteristics of a quadruped legged robot. Trajectory alignment to ground truth is performed via Umeyama least-squares (Umeyama 1991) over all correspondences. It is important to note that the Umeyama alignment assumes that a single homogeneous 4×4 transform is sufficient to align the two fixed frames (i.e., the trajectory represents the sensor's pose with respect to a single fixed frame). In addition, we use Evo's `point_relation` functionality via the point-distance pose relation(`-pose_relation point_distance`), i.e., errors are computed as Euclidean distances between corresponding positions and are independent of orientation.

For each mission $s \in \mathcal{S}$ and each method $m \in \mathcal{M}$, we compute two ordinal ranks based on the reported mean errors: $r^{\text{RTE}}_{m,s}$ by sorting $\mu^{\text{RTE}}_{m,s}$ in ascending order, and $r^{\text{ATE}}_{m,s}$ by sorting $\mu^{\text{ATE}}_{m,s}$ in ascending order (lower is better). In case of equal means, ties are broken by the corresponding standard deviation σ (ascending); diverged entries are assigned the worst rank $|\mathcal{M}|$ for the respective metric. The overall rank is then defined as the average of the two metric ranks,

$$r_{m,s} \triangleq \frac{r^{\text{RTE}}_{m,s} + r^{\text{ATE}}_{m,s}}{2}. \tag{2}$$

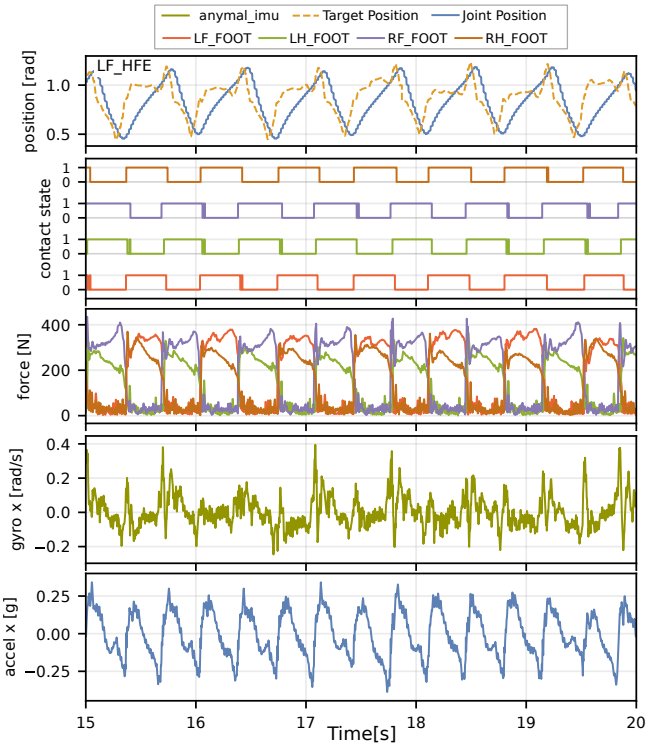


Figure 11. GrandTour dataset example from ANYmal showing synchronized proprioceptive signals over a 5 s window (15 s to 20 s from bag start). From top to bottom: LF_HFE joint position (state-estimator output, SEA-measured position, and commanded/target position), binary foot contact states (LF/RF/LH/RH), corresponding contact force magnitudes $\|\mathbf{f}\|$ (N), and IMU measurements (ω_x in $\mathrm{rad\,s^{-1}}$, a_x in g from `/anymal/imu`.

Finally, the *Summary* column reports the average rank across all missions,

$$\bar{r}_m \triangleq \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} r_{m,s}, \tag{3}$$

where smaller $\bar{r}_m$ indicates better overall performance.

7.1.1 LiDAR Odometry LiDAR odometry can be considered one of the simplest methods for retrieving ego-motion. Often, the methods differ in how the trajectory is represented and how motion compensation is applied to the point cloud. The error metrics of the evaluated open-source solutions are provided in Table 8. These methods have been employed using `Hesai` LiDAR.

As seen in Table 8, Traj-LO achieves the best overall average rank and is the top performer on SPX-2 and EIG-1 (best RTE/ATE), while also leading on SNOW-2 ATE and CON-4 RTE. However, its performance drops substantially on ARC-2, where several methods outperform it. The general performance degradation on the ARC-2 sequence might be explained by the limited visibility in the tight indoor settings. This notion is later strengthened with the utilization of multiple LiDARs in Sec. 7.1.3.

The next best overall (by average rank) are DLO and I^{2}EKF-LO. In particular, I^{2}EKF-LO is best on ARC-7 (both RTE and ATE), while DLO is best on ARC-2 RTE (and second-best ARC-2 ATE). Several methods exhibit mission-dependent weaknesses or missing results (e.g., failures on ARC-2 and/or CON-4), as reflected in their lower summary ranks.

Interestingly, compared to the methods discussed in Sec. 7.1.2, LiDAR-only methods are compatible. However, they are more prone to degradation as there is no redundancy in the system.

7.1.2 LiDAR-Inertial-(Visual) Odometry LiDAR-inertial odometry is currently one of the most widely used methods for robot ego-motion estimation and SLAM. Compared to LiDAR odometry, the addition of high-rate inertial measurements has shown a clear advantage during highly dynamic motions. Furthermore, the availability of observations of the direction of gravity allows stable state estimation. While the advantages are vast, the addition of another modality brings challenges such as time synchronization and intrinsic-extrinsic calibration. The development and improvement of algorithms are still of interest to the robotics community. Under this premise, we have evaluated numerous LIO methods, including LIVO variants. The error metrics of the evaluated open-source LIO solutions are provided in Table 9. These methods have been employed using the `Hesai` LiDAR and the `HF4930` IMU.

Overall, Coco-LIC and FAST-LIVO2 obtain the best average ranks. However, even the top-performing method achieves an average rank of around 4, indicating that there is no single approach that dominates across all missions. Instead, performance is clearly mission-dependent: for example, Voxel-SLAM attains the best ATE on SPX-2, DLIO attains the best ATE on EIG-1, SR-LIO attains the best ATE on CON-4, iG-LIO attains the best ATE on ARC-7, while FAST-LIVO2 attains the best RTE/ATE on ARC-2.

The confined ARC-2 mission is particularly challenging and exposes robustness differences. Several methods exhibit drastic drift, and some methods diverge entirely. This behavior is consistent with the reduced geometric constraint in tight environments where stable registration features are intermittently absent.

Across the other missions, many methods are tightly clustered: for several sequences, differences among the better-performing approaches are on the order of a few millimeters to about a centimeter in RTE and a few centimeters in ATE, often comparable to the reported standard deviations. Consequently, incremental improvements across methods are difficult to claim unambiguously from aggregate ranks alone, and small rank changes can be driven by a single mission. For this reason, in addition to summary statistics, per-mission results (and failure cases) should be emphasized when comparing LIO/LIVO systems.

Every method has a proposed unique advantage; however, these advantages are hard to spot on non-tailored real-world test cases in GrandTour. For example, Point-LIO and RESPLE can operate on a per-point basis rather than a point cloud scan. On the other hand, D-LI Init allows for initialization during motion, a feature that no other methods display. Moreover, solutions like GLIM are featureful and under active development, have many tuning parameters, making it hard to find the optimal configuration for a generally well-performing system.

It is important to note that while two of the top three methods are LIVO, the advantage of vision is ambiguous,

Table 7. Evaluated frameworks and the associated licenses are provided.

Method	Tested Category	Repository Link	License
Adaptive-LIO (Zhao et al. 2024a)	LIO	chengwei0427/Adaptive-LIO	BSD-3[a]
AG-LOAM (Teng et al. 2025)	LO	CR-Robotics/AG-LOAM	MIT
BTSA (Chen et al. 2025)	LIO	thisparticle/btsa	GPL-2.0
CLIC (Lv et al. 2023)	LIVO	APRIL-ZJU/clic	GPL-3.0
Coco-LIC (Lang et al. 2023)	LIVO	APRIL-ZJU/Coco-LIC	GPL-3.0
CT-LO (G3tupup 2022)	LO	G3tupup/ctlo	Unknown
CTE-MLO (Shen et al. 2025)	(M)LO	shenhm516/CTE-MLO	GPL-2.0
D-LI-Init (Xu et al. 2025)	LIO	lian-yue0515/D-LI-Init	GPL-2.0
DLIO (Chen et al. 2023)	LIO	vectr-ucla/direct_lidar_inertial_odometry	MIT
DLO (Chen et al. 2022)	LO	vectr-ucla/direct_lidar_odometry	MIT
DMSA-SLAM (Skuddis and Haala 2024)	LiDAR-SLAM	davidskdds/DMSA_LiDAR_SLAM	MIT
ENVIO (Jung et al. 2022)	VIO	lastflowers/envio	GPL-3.0
FAST-LIMO (Martínez 2023)	LIO	fetty31/fast_LIMO	GPL-3.0
FAST-LIO2 (Xu et al. 2022)	LIO	hku-mars/FAST_LIO	GPL-2.0
FAST-LIO-MULTI (Lee 2024)	(M)LIO	engcang/FAST_LIO_MULTI	GPL-2.0
FAST-LIVO2 (Zheng et al. 2024)	LIVO	hku-mars/FAST-LIVO2	GPL-2.0
Faster-LIO (Bai et al. 2022)	LIO	gaoxiang12/faster-lio	GPL-2.0
GenZ-ICP (Lee et al. 2025a)	LO	cocel-postech/genz-icp	MIT
GLIM (Koide et al. 2024)	LiDAR-SLAM	koide3/glim	MIT
GLO (Su et al. 2025)	LO	robosu12/GLO	GPL-3.0
I^{2}EKF-LO (Yu et al. 2024)	LO	YWL0720/I2EKF-LO	GPL-2.0
iG-LIO (Chen et al. 2024b)	LIO	zijiechenrobotics/ig_lio	GPL-2.0
KISS-ICP (Vizzo et al. 2023)	LO	PRBonn/kiss-icp	MIT
Lightweight-VIO (Choi et al. 2025)	VIO	93won/lightweight_vio	MIT
LIMOncello (Pérez-Ruiz and Solà 2025)	LIO	CPerezRuiz335/LIMOncello	GPL-3.0
LIO-EKF (Wu et al. 2024)	LIO	YibinWu/LIO-EKF	MIT
MAD-ICP (Ferrari et al. 2024)	LO	rvp-group/mad-icp	BSD-3
MA-LIO (Jung et al. 2023)	(M)LIO	minwoo0611/MA-LIO	GPL-2.0
MINS (Lee et al. 2025b)	LIVO	rpng/MINS	GPL-3.0
NV-LIOM (Chung and Kim 2024)	LIO	dhchung/nv_liom	MIT
OKVIS2 (Leutenegger 2022)	VIO	ethz-mrl/okvis2	Custom
Open-VINS (Geneva et al. 2020)	VIO	rpng/open_vins	GPL-3.0
PKO-LO (Choi and Kim 2025)	LO	93won/lidar_odometry	MIT
Point-LIO (He et al. 2023)	LIO	hku-mars/Point-LIO	Custom
PV-LIO (HViktorTsoi)	LIO	HViktorTsoi/PV-LIO	Apache-2.0
R^{2}DIO (Xu et al. 2023)	RGBD	jiejie567/R2DIO	GPL-3.0
R-VIO2 (Huai and Huang 2024)	Mono-VIO	rpng/R-VIO2	GPL-3.0
RESPLE (Cao et al. 2025b)	(M)LO & (M)LIO	ASIG-X/RESPLE	GPL-3.0
RKO-LIO (Malladi et al. 2025)	LIO	PRBonn/rko_lio	MIT
ROVIO (Bloesch et al. 2017a)	Mono-VIO	ethz-asl/rovio	BSD-3
SchurVINS (Fan et al. 2024)	VIO	bytedance/SchurVINS	GPL-3.0
Section-LIO (Meng et al. 2023)	LIO	mengkai98/Section-LIO	Unknown
SLICT2 (Nguyen et al. 2024a)	LIO	brytsknguyen/slict	GPL-2.0
SR-LIO (Yuan et al. 2024b)	LIO	ZikangYuan/sr_lio	GPL-2.0
SuperOdom (Zhao et al. 2021)	LIO	superxslam/SuperOdom	GPL-3.0
Traj-LO (Zheng and Zhu 2024)	LO	kevin2431/Traj-LO	MIT
$\sqrt{\text{VINS}}$ (Peng et al. 2025)	VIO	rpng/sqrtVINS	LGPL-3.0
VINS-Fusion (Qin and Shen 2018)	VIO	HKUST-Aerial-Robotics/VINS-Fusion	GPL-3.0
VINS-RGBD (Shan et al. 2019)	RGBD	Lab-of-AI-and-Robotics/VINS-RGBD	GPL-3.0
Voxel-SLAM (Liu et al. 2024)	LiDAR-SLAM	hku-mars/Voxel-SLAM	GPL-2.0
Voxel-SVIO (Yuan et al. 2025)	VIO	ZikangYuan/voxel_svio	GPL-3.0

[a] License not explicitly tagged by GitHub; BSD-3 inferred from the repository's ROS `package.xml`.

as upon closer inspection, it can be seen that these methods weigh the visual data lower than LiDAR, whereas CLIC, another LIVO method, performs worse, as this method weighs visual observations higher than LiDAR observations.

7.1.3 Multi LiDAR-inertial Odometry This family of methods focuses on utilizing the greater coverage of multiple LiDARs on a robotic system. As discussed in Section 7.1.2, it is common to have a limited field of view with a single LiDAR sensor. When LiDAR observations degrade in an LIO system, the only option is to perform dead-reckoning using the inertial measurements. To prevent this, some methods incorporate multiple LiDAR measurements to increase the field of perception. The error metrics are reported in Table 10.

To employ these methods, we used the `HF4930` IMU for inertial methods and, where possible, used all three LiDARs.

For RESPLE, we used the `Hesai` LiDAR, and the near-field `Livox` LiDARs, whereas for FAST-LIO-MULTI, we used the `Hesai` and `VLP16` LiDARs due to compatibility limitations.

Table 8. Error metrics of the LO methods for the test sequences, per mission (best in **bold**, second best underlined). Summary reports the average rank across the 6 missions, where for each mission the method rank is the mean of its RTE-rank and ATE-rank (lower is better); missing results are assigned the worst rank.

	Summary	SPX-2		SNOW-2		EIG-1		CON-4		ARC-2		ARC-7	
	Avg. rank	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]
Traj-LO	**2.58**	**1.11 (0.77)**	**2.88 (1.26)**	1.28 (1.05)	**1.86 (1.03)**	**1.12 (0.89)**	**3.09 (1.59)**	**1.07 (0.9)**	2.29 (1.07)	17.36 (30.26)	258.43 (105.16)	1.12 (0.82)	2.93 (2.16)
DLO	5.08	1.77 (1.65)	5.74 (2.87)	3.07 (3.05)	6.83 (4.65)	2.49 (2.62)	8 (4.63)	4.11 (4.73)	9.43 (6.01)	**2.91 (4.83)**	13.71 (7.09)	2.01 (5.26)	5.37 (9.58)
I^{2}EKF-LO	5.08	1.22 (1.16)	3.87 (1.96)	4.04 (4)	8.03 (6.75)	4.29 (5.94)	18.6 (13.78)	1.32 (1.04)	3.51 (1.67)	3.69 (7.45)	43.27 (16.48)	**0.85 (0.64)**	**1.64 (1.09)**
CT-LO	5.17	1.55 (1.38)	9.86 (5.33)	1.45 (1.12)	7.89 (2.75)	1.53 (1.33)	10.35 (4.38)	1.39 (1.16)	11.56 (3.78)	4.23 (13.18)	39.87 (17.31)	1.2 (1.19)	4.38 (1.77)
CTE-MLO	5.17	2.12 (2.12)	9 (5.21)	**1.21 (1.04)**	1.99 (1.12)	1.65 (1.6)	3.79 (1.61)	1.1 (1.03)	**2.16 (1)**	11.13 (34.2)	286.67 (115.25)	2.34 (5.65)	7.53 (16.54)
AG-LOAM	6.00	1.59 (1.42)	8.48 (4.07)	3.92 (3.55)	7.69 (5.35)	3.11 (3.37)	7.69 (5.89)	7.47 (9.73)	12.27 (9.33)	3.52 (5.94)	**9.51 (11.39)**	2.07 (4.55)	5.3 (8.55)
GenZ-ICP	6.50	1.51 (1.46)	4.57 (1.98)	26.4 (20.64)	50.39 (40.15)	3.75 (5.95)	6.31 (4.39)	5.57 (10.03)	12.8 (17.82)	3.09 (3.8)	20.65 (8.12)	2.55 (6.73)	3.69 (4.32)
RESPLE-LO	7.00	1.4 (1.3)	5.11 (2.75)	2.01 (3.06)	2.7 (2.5)	2.02 (2.15)	13.27 (3.99)	2.77 (6.34)	4.7 (4.6)	-	-	12.11 (17.52)	12.87 (8.67)
PKO-LO	7.33	2.64 (3.43)	46.23 (22.59)	2.33 (1.86)	10.6 (3.62)	2.42 (2.39)	36.6 (17.62)	2.56 (2.36)	10.43 (5.81)	3.71 (5.54)	20.52 (10.64)	1.86 (2.05)	8.54 (3.5)
GLO	8.67	1.87 (1.75)	6.68 (2.81)	3.84 (3.74)	7.43 (4.06)	3.37 (2.96)	11.64 (5.68)	3.16 (3.07)	24.44 (14.42)	-	-	6.06 (18.9)	9.39 (18.16)
MAD-ICP	9.00	2.21 (2.35)	12.15 (4.42)	18.17 (20.91)	16.64 (13.25)	6.64 (8.69)	13.99 (6.09)	10.43 (14.92)	137.28 (91.17)	6.1 (10)	41.68 (14.32)	1.81 (1.71)	7.05 (2.57)
KISS-ICP	10.58	3.28 (5.6)	9.02 (4.25)	35.78 (21.97)	47.8 (25.96)	7.76 (9.61)	13.28 (6.84)	-	-	21.66 (32.01)	302.37 (135.03)	3.85 (5.66)	7.21 (4.45)

Table 9. Error metrics of the LIO and LiDAR-SLAM methods (including LIVO) for the test sequences, per mission (best in **bold**, second best underlined).

	Summary	SPX-2		SNOW-2		EIG-1		CON-4		ARC-2		ARC-7	
	Avg. rank	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]
Coco-LIC	**4.17**	**0.44 (0.39)**	3.7 (1.53)	**0.41 (0.3)**	**1.06 (0.55)**	**0.4 (0.39)**	4.46 (1.76)	**0.43 (0.33)**	2.4 (1.23)	1.01 (2.59)	8.07 (5.05)	**0.37 (0.3)**	3.81 (3.56)
FAST-LIVO2	4.42	1.05 (0.82)	2.04 (0.97)	1.25 (1.01)	1.43 (0.74)	1.11 (0.88)	3.02 (1.38)	1.18 (0.94)	2.73 (1.35)	**0.7 (0.61)**	**1.1 (0.51)**	1.13 (0.81)	1.22 (0.56)
PV-LIO	7.50	1.14 (0.91)	3.74 (1.64)	1.32 (1)	3.18 (1.23)	1.24 (0.96)	3.15 (1.16)	1.22 (0.96)	4.57 (1.58)	1.05 (0.86)	2.28 (0.93)	0.99 (0.74)	2.02 (0.86)
DLIO	7.67	1.22 (1.04)	2.74 (1.22)	1.51 (1.19)	1.61 (0.85)	1.26 (0.97)	**2.65 (1.39)**	1.43 (1.16)	2.28 (1.06)	1.19 (1.39)	4.26 (1.73)	1.15 (0.9)	3.78 (4.6)
Fast-LIMO	8.17	1.15 (0.99)	5.76 (2.58)	1.16 (0.9)	2.41 (1.31)	1.1 (1.6)	8.75 (3.3)	0.89 (0.74)	2.65 (1.5)	2.49 (3.35)	10.2 (6.94)	0.61 (0.5)	2.31 (1.57)
Voxel-SLAM	8.92	1.03 (0.83)	**2.03 (1.02)**	1.37 (1.13)	1.57 (0.81)	1.34 (1.03)	3.33 (1.4)	3.24 (9.62)	50.87 (72.99)	1.24 (1.43)	8.66 (5.57)	0.91 (0.73)	1.32 (0.65)
Faster-LIO	9.08	1.02 (0.89)	2.68 (1.26)	1.46 (1.35)	7.15 (2.55)	1.3 (1.13)	5.38 (2.73)	1.51 (1.86)	9.04 (3.93)	2.01 (3.49)	5.88 (5.26)	0.94 (0.73)	1.49 (0.65)
SLICT2	9.17	1.11 (1.64)	3.69 (2.3)	1.27 (1.05)	1.37 (0.65)	1.18 (0.88)	4.85 (1.97)	1.19 (0.94)	1.97 (0.89)	-	-	0.9 (0.68)	1.59 (0.61)
iG-LIO	9.25	1.47 (1.24)	4.94 (2.28)	1.16 (1.1)	1.6 (1.5)	8.09 (13.47)	6.3 (11.19)	1.19 (0.92)	2.53 (1.23)	2.66 (8.33)	7.53 (14.53)	0.9 (0.68)	**1.15 (0.61)**
LIMOncello	10.58	1.92 (2.93)	3.23 (1.67)	3.01 (4.39)	1.77 (2.2)	2.39 (3.5)	3.3 (1.95)	2.73 (4.22)	1.88 (1.98)	2.1 (3.34)	1.87 (1.93)	1.86 (3.4)	2.68 (5.78)
FAST-LIO2	11.58	1.07 (0.9)	4.15 (2.26)	1.41 (1.13)	6.47 (2.67)	1.28 (0.95)	4.97 (2.65)	1.54 (1.86)	8.79 (3.83)	3.62 (10.63)	47.63 (16.36)	1.04 (0.72)	2.74 (1.45)
RKO-LIO	11.75	1.22 (1.03)	3.45 (1.7)	2.4 (2.41)	3.4 (3.92)	1.41 (0.99)	4.31 (1.83)	1.39 (1.12)	2.61 (1.45)	3.87 (4.88)	21.54 (15.56)	1.58 (1.14)	6.77 (3.64)
SR-LIO	12.25	1.78 (1.42)	8.35 (5.46)	1.35 (1.05)	1.74 (0.87)	1.14 (0.92)	3.29 (1.52)	1.44 (1.05)	**1.84 (0.87)**	-	-	1.39 (1.06)	3.05 (1.6)
RESPLE	12.42	0.93 (0.83)	3.83 (1.79)	1.28 (1.63)	2.08 (1.63)	1.41 (1.64)	9.7 (4.52)	1.41 (1.99)	4.41 (3.09)	3.61 (4.19)	9.71 (5.01)	6.16 (12.8)	33.2 (10.17)
SuperOdom	14.67	3.9 (12.62)	63.5 (42.63)	1.53 (1.27)	6.64 (6.58)	1.47 (1.25)	6.01 (3.03)	1.64 (2.29)	6.72 (4.17)	2.82 (4.31)	14.61 (9.93)	1.33 (0.98)	2.82 (1.19)
MINS	14.75	0.73 (0.61)	13.7 (4.81)	10.02 (4.13)	213.09 (130.02)	1.1 (1.68)	50.41 (20.14)	0.77 (1.87)	46.01 (20.38)	5.69 (22.9)	68.15 (53.05)	0.72 (0.83)	11.9 (3.55)
D-LI-Init	16.00	1.26 (0.98)	3.92 (1.94)	1.96 (3.53)	2.41 (3.27)	2.46 (2.81)	5.62 (5.04)	1.61 (3.15)	2.35 (2.52)	5.06 (23.39)	121.39 (49.16)	-	-
GLIM	17.92	0.96 (0.79)	4.19 (2.29)	49.65 (28.59)	24.73 (15.16)	1.39 (1.07)	8.87 (4.4)	20.46 (24.31)	18.8 (14.16)	43.99 (37.63)	103.1 (51.41)	4.39 (5.44)	12.21 (5.69)
BTSA	18.00	2.41 (5.63)	30.03 (5.99)	1.66 (3.49)	27.61 (5.49)	1.67 (1.58)	29.86 (4.85)	1.65 (3.34)	31.1 (1.45)	4.21 (9.36)	30.68 (2.12)	2.02 (5.35)	17.29 (8.52)
Section-LIO	18.42	2.27 (2.29)	10.16 (3.85)	1.64 (1.2)	7.27 (2.62)	1.67 (1.2)	7.27 (2.63)	1.83 (1.63)	9.4 (3.19)	-	-	2.3 (1.8)	11.79 (8.64)
CLIC	18.67	4.05 (14.67)	72.73 (69.22)	0.54 (0.42)	1.4 (0.66)	-	-	0.62 (0.59)	6.25 (4.03)	-	-	-	-
Point-LIO	20.67	2.2 (2.74)	20.43 (8.01)	8.47 (11.22)	85.59 (30.25)	6.01 (9.65)	60.07 (25.26)	1.94 (3.81)	25.83 (6)	10.94 (21.75)	49.84 (28.28)	4.7 (11.14)	26.4 (10.8)
DMSA-SLAM	20.75	4.05 (8.01)	48.98 (5.15)	3.82 (4.66)	48.01 (6.41)	2.99 (4.19)	47.64 (6.83)	3.55 (6.96)	44.93 (11.22)	9.1 (18.22)	49.19 (10.06)	3.42 (2.81)	27.98 (13.58)
NV-LIOM	21.08	6.35 (15.86)	33.59 (7.39)	3.84 (5.47)	31.59 (3.57)	3.55 (4.94)	31.15 (6.31)	3.85 (9.32)	33.26 (10.47)	13.4 (24.95)	29.13 (10.74)	4.72 (6.39)	21.21 (9.63)
Adaptive-LIO	22.33	4.5 (11.06)	79.01 (63.65)	18.18 (17.36)	27.83 (11.37)	11.93 (17.78)	437.33 (245.78)	6.93 (13.45)	323.49 (187.48)	11.31 (24.53)	217.2 (90.87)	1.91 (2.49)	4.28 (3.16)
LIO-EKF	22.67	4.87 (4.15)	10.84 (5.63)	7.58 (6.78)	11.86 (7.25)	5.32 (6.18)	65.67 (33.01)	5.05 (5.57)	9.88 (6.58)	-	-	13.12 (10.35)	37.79 (28.82)

Table 10. Error metrics of Multi-LiDAR methods for the test sequences, per mission (best in **bold**, second best underlined). Summary reports the average rank across the 6 missions, where for each mission the method rank is the mean of its RTE-rank and ATE-rank (lower is better); missing results are assigned the worst rank.

	Summary	SPX-2		SNOW-2		EIG-1		CON-4		ARC-2		ARC-7	
	Avg. rank	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]
CTE-MLO	**1.25**	1.77 (1.6)	**3.88 (1.8)**	**1.24 (0.99)**	**2.35 (1.13)**	1.74 (1.79)	**2.67 (1.48)**	**1.01 (0.92)**	**2.04 (0.9)**	**1.85 (2.3)**	**4.3 (3.49)**	**0.91 (0.73)**	**1.37 (0.9)**
MA-LIO	2.83	6.96 (22.75)	44.02 (49.35)	7.28 (10.76)	14.17 (8.28)	**1.38 (1.09)**	4.46 (2.21)	1.47 (1.32)	5.2 (2.03)	2.37 (3.06)	8.87 (2.46)	1.04 (0.92)	2.82 (1.29)
RESPLE-MLIO	2.83	**1.32 (1.42)**	7.17 (3.25)	5 (5.08)	10.37 (5.69)	1.48 (1.49)	13.63 (3.4)	1.92 (3.25)	11.03 (8.6)	3.94 (4.42)	12.81 (4.47)	6.16 (12.8)	33.2 (10.17)
RESPLE-MLO	4.00	1.88 (1.8)	8.67 (3.97)	6.78 (7.34)	11.64 (6.44)	2.68 (2.51)	14.91 (3.82)	1.8 (1.84)	21.69 (17.59)	-	-	7.21 (13.32)	33.44 (11.76)
FAST-LIO-MULTI	4.25	4.96 (3.91)	10.13 (5.44)	9.47 (9.81)	31.16 (22.79)	5.68 (6.03)	33.53 (12.32)	4.57 (3.55)	5.18 (3.49)	-	-	4.22 (4.31)	7.08 (5.37)

As seen in Table 10, CTE-MLO is the most consistent method across missions, achieving the best average rank and the lowest ATE on all reported missions, while also attaining the best (or second-best) RTE in most cases. The remaining approaches are less consistent across environments: MA-LIO attains the best RTE on EIG-1 but exhibits substantially larger errors on other missions (e.g., SPX-2 and SNOW-2), and RESPLE-MLIO achieves the best RTE on SPX-2 but degrades markedly on ARC-7. Finally, RESPLE-MLO and FAST-LIO-MULTI obtain weaker overall ranks, and both fail on ARC-2.

The analysis indicates that multi-LiDAR sensing can improve perception coverage and, consequently, the robustness of odometry estimates, but the accuracy improvement is method- and scenario-dependent rather than uniform across sequences. It should be noted that different LiDAR models may exhibit different noise and coverage characteristics, which, in turn, affect the efficacy of the employed method.

7.1.4 Visual-Inertial Odometry Despite the performance of LIO methods, the relatively large weight and size of LiDARs, and the higher base prices, vision-based ego-motion estimation has recently returned to the forefront. Despite the widespread adoption, technical challenges persist in employing a visual-inertial ego-motion estimation method. Common challenges include intrinsic calibration, sensitivity to lighting, reduced parallax due to environmental factors, and the dependence of observability in the estimation space on the platform's motion. Despite these challenges, VIO remains relevant today, thanks to its compatibility with lightweight robotic systems and decades of well-developed understanding of the technology. Under this premise, we have evaluated

Table 11. Error metrics of the VIO methods (including RGBD) for the test sequences, per mission, reported in cm (best in **bold**, second best underlined).

	Summary	SPX-2		SNOW-2		EIG-1		CON-4		ARC-2		ARC-7	
	Avg. rank	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]
Voxel-SVIO	**1.75**	1.38 (1.13)	18.70 (11.95)	1.40 (1.17)	16.17 (8.97)	**1.25 (0.98)**	**12.25 (8.09)**	**1.34 (1.46)**	**38.90 (22.06)**	**1.02 (1.11)**	**7.69 (3.53)**	**1.23 (1.23)**	14.75 (5.99)
$\sqrt{\text{VINS}}$	3.08	1.90 (1.89)	34.26 (17.32)	1.23 (0.59)	23.07 (12.40)	1.33 (1.72)	37.41 (22.99)	1.91 (3.79)	80.18 (38.00)	4.05 (7.91)	37.83 (39.06)	1.24 (2.20)	**9.96 (3.55)**
OKVIS2	4.25	1.342 (1.35)	28.85 (13.69)	1.196 (0.958)	**6.86 (3.906)**	1.287 (1.58)	30.93 (18.9)	3.06 (5.36)	449.03 (236.35)	1.50 (2.51)	14.64 (4.22)	-	-
Open-VINS	5.42	1.32 (2.66)	42.75 (24.21)	2.86 (0.69)	54.83 (29.97)	2.20 (2.39)	85.58 (36.41)	5.05 (10.16)	189.11 (141.52)	2.48 (3.05)	17.47 (10.27)	-	-
VINS-RGBD	6.00	6.12 (10.32)	117.47 (44.46)	3.07 (2.60)	60.50 (29.28)	3.43 (6.28)	100.30 (42.57)	5.56 (9.91)	432.92 (148.89)	19.45 (106.09)	401.89 (118.22)	4.41 (5.32)	30.60 (17.47)
SchurVINS	6.33	**1.13 (1.04)**	**10.08 (5.18)**	**1.189 (1.01)**	20.25 (9.76)	3.17 (6.67)	171.51 (81.31)	3.72 (30.58)	299.68 (188.51)	-	-	-	-
ENVIO	8.00	-	-	3.71 (2.96)	88.92 (42.71)	3.09 (2.90)	140.43 (70.03)	33.48 (85.95)	2041.95 (803.61)	20.85 (41.52)	172.18 (108.12)	15.14 (36.18)	95.80 (50.00)
VINS-Fusion	8.50	22.31 (29.07)	405.53 (147.03)	24.65 (14.22)	414.18 (147.93)	27.13 (33.98)	411.52 (183.93)	32.56 (55.25)	1069.63 (428.24)	27.41 (54.28)	180.76 (106.39)	11.21 (14.15)	112.21 (52.88)
ROVIO*	8.92	32.08 (36.51)	484.61 (635.29)	9.21 (8.28)	168.60 (73.43)	6.98 (8.87)	190.66 (138.92)	12.92 (21.97)	474.93 (271.68)	19.70 (59.14)	192.08 (204.65)	-	-
R-VIO2*	9.33	-	-	3.05 (4.20)	33.32 (13.02)	13.90 (42.62)	457.80 (399.06)	5.21 (19.48)	110.28 (162.24)	-	-	-	-
R^{2}DIO	9.75	-	-	8.39 (6.34)	1005.50 (491.91)	10.91 (27.47)	947.15 (421.05)	21.08 (95.47)	691.99 (521.18)	10.65 (12.52)	95.29 (26.88)	-	-
Lightweight-VIO	11.17	26.81 (52.60)	760.54 (531.81)	24.23 (16.05)	1091.97 (337.18)	20.06 (17.67)	499.50 (234.45)	-	-	-	-	-	-

recent and well-known Stereo and Monocular VIO pipelines and reported their performance metrics in Table 11. To employ these methods, the `CoreResearch` cameras are utilized: specifically, the front-facing RGB `CoreResearch` camera for monocular methods and the front-facing gray-scale stereo pair for stereo methods. Moreover, the two RGB-D methods are run with the `Zed2i` RGB-D camera. For all methods that require inertial data, the 100 Hz `HG4930` IMU is utilized.

Overall, Table 11 shows that Voxel-SVIO performs consistently well across the evaluated missions, indicating strong robustness to varying scene structure and motion profiles. At the same time, many pipelines achieve comparable relative translation errors across multiple sequences, as expected for odometry systems that primarily optimize local consistency. In contrast, absolute trajectory error varies more widely, reflecting differences in long-horizon drift, sensitivity to degradation of visual constraints, and robustness to estimator inconsistency.

The SPX-2 mission begins with highly dynamic motion, which adversely affects initialization for pipelines that require a short, stationary segment. This is reflected in Table 11 by the diverged results. When initialization succeeds, several methods still maintain competitive local motion estimates, but the drift behavior over the remainder of the mission varies considerably.

The SNOW-2 mission covers a very bright, snowy area, which poses challenges for feature extraction and results in a poor distribution of features across the view. This is visible through the elevated ATE and RTE metrics for this sequence.

The CON-4 mission is a particularly demanding, large-scale trajectory with a large loop, and it accumulates long-term drift effects.

Accordingly, Table 11 highlights the variety of absolute accuracy on CON-4 across methods, where relative errors remain comparable.

Finally, the ARC missions expose the sensitivity of VIO pipelines to illumination and featureless views. As seen in Table 11, changing lighting conditions adversely affect all methods on ARC-2 and ARC-7, where ARC-7 includes an extended very dark interval during which multiple pipelines diverge or fail to produce valid estimates.

These failures are particularly consequential in our evaluation because the benchmarking setup and many methods lack an external reinitialization mechanism or an automated restart strategy; consequently, once a method diverges, it typically remains failed for the remainder of the mission rather than recovering after the adverse segment. These findings emphasize the practical importance of robust exposure control, photometric normalization (or relighting), explicit failure detection, and re-initialization for deployment.

7.1.5 Ablation: Online-Calibration It is well known in the robotics community that the quality of the extrinsic calibration between tightly coupled sensors is crucial for achieving top-quality results. Conversely, retrieving good-quality extrinsic calibration is a tedious process that is often neglected in real-world research, but is deeply considered in the industry. Hence, some solutions include an online-calibration module that estimates (parts of) the LiDAR–IMU extrinsics during operation. The error metrics are reported in Table 12 for both online calibration enabled and disabled cases.

The results are mixed and do not indicate a consistent advantage of enabling online calibration. For Fast-LIMO, online calibration improves some missions (e.g., SPX-2 and ATE on CON-4/ARC-7), but is worse on others (notably SNOW-2 and EIG-1, and ARC-7 RTE). For Faster-LIO, online calibration tends to help ATE more often (including a large gain on ARC-2 ATE), while RTE is frequently similar or slightly worse. For Fast-LIO, enabling online calibration is worse on most missions and metrics, in several cases by a large margin in ATE. This implies either a non-observable state or implementation problems.

A plausible explanation is that online-calibration parameters are not uniformly observable: their identifiability depends on the available motion excitation and on the geometric constraints provided by the observations. When the data weakly constrain some calibration modes, estimating them online can have limited benefit and may correlate with other states. Accordingly, online calibration of LIO systems should be treated as a scenario-dependent feature rather than a guaranteed improvement. Lastly, in real-world robotic systems, sensors are mounted rigidly, and variations in mounting pose due to environmental conditions are often negligible compared to sensor noise (Frey et al. 2025).

7.1.6 Summary: State estimation, localization and mapping Across the evaluated LO, LIO/LIVO, multi-LiDAR, and VIO pipelines (Tables 8 to 11), relative translation errors are often comparable on the easier missions, while absolute trajectory errors separate methods much more strongly. In general, LiDAR-inertial methods are more accurate and drift-stable than pure LO and especially than VIO, with Coco-LIC and FAST-LIVO2 consistently among the strongest, while Traj-LO represents the most competitive pure-LO baseline when conditions are favorable. The large-scale CON-4 mission presents a clear challenge for long-term consistency:

Table 12. Error metrics of methods with and without online calibration feature for a subset of the test sequences, per mission, reported in cm (better of online calibration Yes/No in **bold**).

		SPX-2		SNOW-2		EIG-1		CON-4		ARC-2		ARC-7	
	Online Calib	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]	RTE $\mu(\sigma)$[cm]	ATE $\mu(\sigma)$[cm]
Fast-LIMO	Yes	**1.15 (0.99)**	**5.76 (2.58)**	1.16 (0.90)	2.41 (1.31)	1.10 (1.60)	8.75 (3.30)	**0.89 (0.74)**	**2.65 (1.50)**	**2.49 (3.35)**	10.20 (6.94)	0.61 (0.50)	**2.31 (1.57)**
Fast-LIMO	No	1.32 (1.05)	6.99 (3.62)	**1.01 (0.86)**	**2.11 (1.08)**	**0.89 (0.77)**	**5.79 (1.92)**	**0.89 (0.68)**	2.73 (1.55)	2.71 (3.73)	**9.64 (6.17)**	**0.59 (0.47)**	2.45 (1.54)
Faster-LIO	Yes	1.02 (0.89)	**2.68 (1.26)**	**1.46 (1.35)**	**7.15 (2.55)**	**1.30 (1.13)**	5.38 (2.73)	1.51 (1.86)	**9.04 (3.93)**	2.01 (3.49)	**5.8826 (5.26)**	0.94 (0.73)	**1.49 (0.65)**
Faster-LIO	No	**0.99 (0.82)**	2.85 (1.32)	**1.46 (1.15)**	7.18 (2.48)	1.32 (1.11)	**5.13 (2.66)**	**1.50 (1.88)**	9.25 (3.99)	**2.00 (3.75)**	11.75 (5.08)	**0.89 (0.69)**	1.52 (0.68)
Fast-LIO	Yes	**1.07 (0.90)**	4.15 (2.26)	1.41 (1.13)	6.47 (2.67)	1.28 (0.95)	4.97 (2.65)	1.54 (1.86)	8.79 (3.83)	3.62 (10.63)	47.63 (16.36)	**1.04 (0.72)**	2.74 (1.45)
Fast-LIO	No	1.11 (0.83)	**3.61 (2.13)**	**1.24 (1.03)**	**1.65 (0.88)**	**1.19 (0.87)**	**3.71 (1.55)**	**1.24 (0.96)**	**1.86 (0.80)**	**3.00 (7.47)**	**8.66 (9.73)**	1.17 (0.82)	**2.64 (1.43)**

several methods remain locally reasonable but accumulate substantial drift, with particularly severe degradation visible in multiple VIO pipelines (e.g., ENVIO, VINS-Fusion, OpenVINS) compared to the better LiDAR-inertial solutions. Multi-LiDAR methods can be highly competitive, such as the CTE-MLO, which performs consistently, but the benefit is method-dependent (e.g., MA-LIO and the RESPLE multi-LiDAR variants vary more), reflecting the added sensitivity to calibration/synchronization and configuration. Failure cases are present in every category: VIO shows the most frequent failures on the hardest sequences, while LiDAR-based methods fail less often but exhibit higher error metrics due to visibility and feature-extraction reliability. Within missions, poorer local accuracy generally correlates with higher drift (e.g., weaker LO baselines such as KISS-ICP tend to drift more), but good local performance does not guarantee low accumulated error over long trajectories. Overall, cross-mission robustness, initialization resilience, and explicit failure handling/recovery matter at least as much as incremental metric gains on any single sequence.

Sensitivity to tuning: Across nearly all evaluated methods, performance is highly sensitive to parameter choices, and many methods are susceptible to overtuning to specific motion profiles, sensor characteristics, or environments. In practice, this results in significant performance variability when transferring a configuration between missions, even within the same platform and sensor suite. Notably, this sensitivity is not limited to complex pipelines with many hyperparameters: methods marketed as simple, robust, or adaptive, and often described as approaches that "just work," also exhibit substantial performance swings under modest changes in conditions (e.g., dynamics, lighting, scene structure) or when default settings are used. Overall, our analysis suggests that there is still room for methods that are truly adaptive to the environments in which they are deployed.

Technical depth and extendability: There is a corpus of open-source methods that build on top of cornerstone works such as Fast-LIO2 Xu et al. (2022), DLIO (Chen et al. 2023), Vins-Mono (Qin et al. 2018), and LOAM (Zhang and Singh 2017). Unfortunately, many works that build upon these solutions fail to follow up on code quality and implementation, making it increasingly difficult to assess their correctness. Furthermore, as the codebase grows, complexity and now-obsolete parts are retained, leading to misconceptions about what the framework actually does.

Epidemic of evaluation: Nowadays, almost every state estimation and localization work is expected to evaluate their proposed solutions against the state-of-the-art methods on publicly available datasets. However, despite this expectation, there is no standardization of evaluations, and their quality is questionable. One of the most common pitfalls is the configuration and tuning of third-party solutions. The authors do not have an incentive to thoroughly tune and configure the reference works or replicate already reported error metrics on certain datasets. Next, the discrepancy between baseline performance and application-based performance is increasing. There are many solutions that suggest improvement over a baseline solution in specific conditions; however, when backtested, these improvements often result in reduced performance in other environments. Lastly, the evaluation methods and error metrics available to roboticists are outdated. It is quite easy to misinterpret and tamper with error metrics at the centimeter level; furthermore, current trajectory alignment methods do not gracefully support continuous-time or robust re-initialization, which may require re-initialization in a new fixed frame. GrandTour conducted a substantial evaluation of available state-of-the-art solutions as an effort to mitigate some of the points above.

7.2 Perception

Today, GrandTour supports multiple works aimed at improving the perception capabilities of legged robots. While a comprehensive discussion of these contributions is beyond the scope of this section, we highlight several studies and integrated tools. These examples are illustrated in Fig. 12.

LiDAR-based 3D mesh generation: As shown in Fig. 12-(A,E), we have integrated open-source 3D online mesh generation methods ImMesh (Lin et al. 2023) and SLAMMesh (Ruan et al. 2023) to demonstrate the performance of state-of-the-art online mesh generation. These LiDAR-based online mesh generation methods can accelerate the convergence of radiance field methods, might enable better foothold placement for locomotion, or provide better mission planning thanks to increased environmental awareness.

Multi-Modal Elevation Mapping: We integrated MEM (Miki et al. 2022b; Erni et al. 2023), shown in Fig. 12-(H), a GPU-accelerated elevation-mapping framework that allows the fusion of depth measurements into a local or global elevation map representation. Furthermore, visual data such as RGB images, semantic classes, and latent features can be easily integrated as separate layers into the same representation.

WaveMap: As shown in Fig. 12-(D), we integrated WaveMap (Reijgwart et al. 2023), a 3D multi-resolution volumetric mapping framework that fuses depth measurements into a wavelet-based 3D representation. This framework supports query of occupancy, depth rendering, and enables high-rate planning applications.

Geometric Understanding : Accurate time synchronization between LiDAR and cameras enables the generation

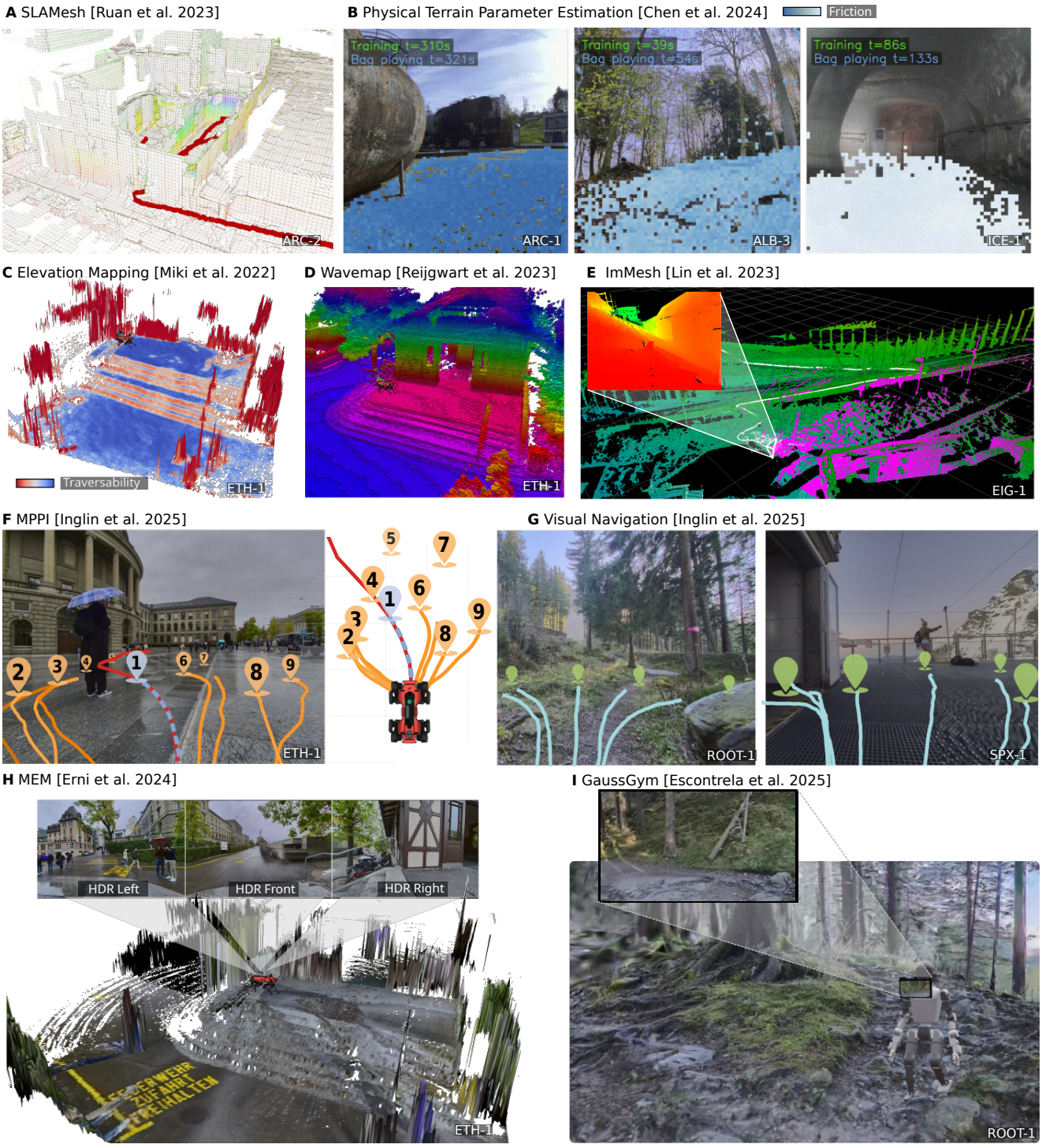


Figure 12. Overview of GrandTour applications for perception, locomotion, and navigation. **(A)** Online mesh generation with SLAMesh (Ruan et al. 2023). **(B)** Prediction of friction parameters from RGB images using training data collected via proprioceptive terrain interaction (Chen et al. 2024a). **(C)** Integration with Elevation Mapping Cupy (Miki et al. 2022b), enabling the assessment of foothold placement scores. **(D)** WaveMap (Reijgwart et al. 2023) 3D probabilistic occupancy mapping integration. **(E)** LiDAR-based online mesh generation and depth rendering with the ImMesh (Lin et al. 2023) meshing pipeline. **(F)** Illustration of MPPI-based path planning (Inglin et al. 2025) and the teleoperated trajectory. **(G)** Less is More (Inglin et al. 2025) visual navigation predictions. **(H)** Projection of camera images onto the map using Multi-modal Elevation Map (MEM) (Erni et al. 2023). **(I)** GaussGym (Escontrela et al. 2025) enables training RGB-based locomotion policies in simulation, providing photorealistic rendering via 3D Gaussian Splatting and mesh reconstructions.

of ground-truth metric geometry. While various datasets exist, only a limited number of datasets provide collections in natural environments. Using the GrandTour dataset, we trained and evaluated monocular depth estimation networks with a specific focus on natural environments. This can be further extended to stereo depth estimation.

Physical Properties Detection: Shown in Fig. 12-(B), (Chen et al. 2024a) enable the estimation of simulation

parameters—corresponding to physical terrain properties—from proprioceptive joint information. These parameters, such as terrain stiffness and friction, can then be used to train a vision-based decoder that predicts the same terrain properties directly from visual input.

Dynamic Point Segmentation: We integrated functionality to filter out dynamic points based on pre-trained image segmentation models. This facilitates more accurate volumetric mapping of environments by directly removing depth measurements associated with dynamic semantic classes. The generated masks are available on HuggingFace.

NerfStudio: We provide converters that enable the direct transformation of the GrandTour dataset into NerfStudio-compatible formats for training neural scene representations (Tancik et al. 2023). GrandTour enables the study of the effects of integrating sparse and dense depth measurements into multimodal neural rendering research.

7.3 Locomotion and Navigation

The GrandTour dataset supports the development of locomotion and navigation capabilities for legged robots. All deployments are collected using the same locomotion controller, and the robot is deliberately and safely teleoperated throughout the environment with specific navigation goals or inspection targets in mind. In the following, we highlight recent projects that directly benefited from the GrandTour dataset or extended it:

NaviTrace: This project (Windecker et al. 2025) uses a subset of the HDR camera images within GrandTour to establish a visual question answering (VQA) benchmark. This benchmark evaluates the capabilities of vision-language models to generate navigation behavior. Specifically, human-annotators perform a key frame selection and annotate the selected key frames with navigation task instructions and an image trace (set of image coordinates) defining the navigation behavior, for different robot embodiments. This dataset allows for evaluating and fine-tuning Vision Language models for navigation.

MPPI Path Planning: The GrandTour dataset facilitates the development of classical motion planners. For example, in previous work, we implemented an MPPI-based path planner that takes as input the robot's local traversability analysis, computed using the MEM framework—along with a goal position, and outputs a set of safe actions that lead the robot toward that goal. This example is shown in Fig. 12-(F) as an example. However, specifying the desired robot behavior requires careful tuning of numerous hyperparameters. Using a recorded reference behavior, we demonstrated that black-box optimization methods can automatically adjust these parameters to minimize the error between the executed trajectory and the MPPI-predicted trajectory. This results in a highly effective, robust path-planning system.

Visual Navigation: The GrandTour dataset can be seamlessly integrated into existing projects, for example, to support the learning of visual navigation costs, which in prior work required collecting specific datasets for testing and benchmarking (Frey et al. 2023). In more recent work, the dataset has been used for imitation learning. A key strength of GrandTour is the availability of dense geometric information around the robot, which enables the generation of paths to arbitrary goals using methods such as MPPI-based planning, as illustrated in Fig. 12-(G). This greatly increases the diversity of available navigation trajectories, enabling state-of-the-art visual navigation performance even when training on a comparatively small dataset. The synthetic paths are available on HuggingFace.

GaussGym: While recent work has shown that even limited data can suffice to learn world models that support training locomotion policies (Li et al. 2025), the dominant paradigm in robotics remains sim-to-real transfer. GaussGym (Escontrela et al. 2025) enables real-to-sim transfer to enhance simulation capabilities by directly integrating deployment environments within the GrandTour into simulation as shown in Fig. 12-(I). This is achieved by neural radiance fields and scene reconstruction techniques. Furthermore, GaussGym (Escontrela et al. 2025) demonstrates how RGB-based visual locomotion polices can be trained fully in simulation.

8 Discussion and limitations

The GrandTour dataset represents an extensive multi-modal data resource for legged robotics research, yet several limitations must be acknowledged to guide future developments. While the dataset features a wide variety of environmental conditions, its geographic scope is currently limited to Switzerland. This regional focus may introduce biases related to specific architectural styles, vegetation, and ambient lighting conditions. Further bias is introduced by the equipment (Total Station, car, toolboxes) and the personnel required for the recording process. The dataset today does not contain revisits of the same environments under varying conditions (different seasons, weather patterns, or times of day). Despite the comprehensive sensor suite, the dataset excludes certain emerging modalities such as radar, FMCW LiDAR, and event cameras. Lastly, the current version of the dataset does not include semantic annotations.

9 Conclusion and future work

In this work, we have presented the GrandTour dataset, a comprehensive multi-modal legged robotics dataset providing diverse sensor modalities, environmental coverage, motion characteristics, and high-precision ground truth. The dataset contains 49 sequences spanning a wide range of environments, illumination conditions, and weather types, exceeding the scope and diversity of existing legged robot datasets.

A distinguishing feature of GrandTour is the availability of centimeter-level ground truth poses through NovAtel's Inertial Explorer solution when GNSS is available, as well as millimeter-level and time-synchronized ground truth positions obtained through Leica MS60 total station measurements and accurate extrinsic calibration. The reference data enables rigorous quantitative evaluation of state estimation and SLAM algorithms, such as visual-inertial odometry, LiDAR-based SLAM, legged-inertial, and hybrid approaches under real-world conditions. Furthermore, the dataset is accompanied by post-processed outputs and open-source software tools to facilitate straightforward data access and analysis. The synchronized multi-modal dataset supports research on sensor fusion, depth estimation, real-to-sim transfer, and many more

applications on the perception side, from training foundation models, to self-supervised, cross-modal learning approaches. Furthermore, the GrandTour can support the development and benchmarking of traversability assessment and navigation.

Looking ahead, we plan to expand GrandTour in several aspects. First, we aim to incorporate additional missions featuring extreme perception challenges, such as those in volcanic scenery and extended-duration deployments, a wider range of environments, and a broader set of robots, featuring bipeds and wheeled-legged robots. Second, we will upgrade the sensor suite to include emerging technologies such as higher-resolution solid-state LiDARs, FMCW LiDARs, radars, and event cameras.

Through these systematic expansions, GrandTour will continue to serve as a valuable benchmark and resource for the robotics and computer vision research communities, bridging the gap between controlled laboratory experiments and real-world deployment scenarios for legged robots.

Acknowledgements

We thank William Talbot for integrating the ADIS 16475 IMU and being an alpha-tester, Julian Nubert for his early support in development, and for running Holistic Fusion on the GrandTour data. Yves Inglin, and Changang Chen for visual navigation; Alejandro Escontrela and Justin Kerr for the integration of GaussGym; Carlos Pérez for assistance in running Limoncello; Victor Reijgwart for help integrating Wavemap; Cyrill Püntener for expertise in network configurations; and Benjamin Krummenacher.

This work was supported in part by Hexagon | Leica Geosystems. We are grateful to Benjamin Müller, Benjamin Schöll, Bernhard Sprenger, Stefan Eisenreich, and Jonas Nussdorfer for their expertise in networking, calibration, time-synchronization, and for their assistance in guiding us with the collection of calibration data. We further thank Leica Geosystems for providing the MS60 total station and a custom AP20 unit, which were essential for obtaining high-precision ground truth measurements. Additionally, we appreciate the advice of Dr. Matej Varga and Dr. Jemil Avers Butt (Department of Civil, Environmental, and Geomatic Engineering, ETH Zurich) on robotic total station operation. Moreover, we acknowledge the support of Hexagon | NovAtel Inc., with special thanks to Ryan Dixon for guidance on the utilization of the NovAtel CPT7 and IE solutions.

We acknowledge that the International Foundation High Altitude Research Stations Jungfraujoch and Gornergrat (HFSJG), 3012 Bern, Switzerland, made it possible for us to carry out our experiment(s) at the High Altitude Research Station at Jungfraujoch. We also thank the custodians, Mrs. Daniela Bissig and Mr. Erich Bissig, guide Ms. Doris Graf Jud (Jungfraubahnen), and Annette Fuhrer (Jungfraubahnen) for their support during the Jungfraujoch trip.

This work would not have been possible without the combined efforts of these partners and collaborators.

Finally, this work is partially funded by the National Centre of Competence in Research Robotics (NCCR Robotics), the ETH RobotX research grant, which is funded through the ETH Zurich Foundation, and the European Union's Horizon 2020 research and innovation program under grant agreement No. 101016970, No. 101070405, and No. 101070596. This work was also supported by the Open Research Data Grant at ETH Zurich.

References

Bai C, Xiao T, Chen Y, Wang H, Zhang F and Gao X (2022) Faster-lio: Lightweight tightly coupled lidar-inertial odometry using parallel sparse incremental voxels. *IEEE Robotics and Automation Letters* 7(2): 4861–4868. DOI:10.1109/LRA.2022.3152830.

Bloesch M, Burri M, Omari S, Hutter M and Siegwart R (2017a) Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback. *The International Journal of Robotics Research* 36(10): 1053–1072.

Bloesch M, Hutter M, Hoepflinger MA, Leutenegger S, Gehring C, Remy CD and Siegwart R (2017b) State estimation for legged robots-consistent fusion of leg kinematics and imu. *Robotics: Science and Systems* 8: 17–24.

Buchanan R (2021) Allan variance ros. URL https://github.com/ori-drs/allan_variance_ros.

Buchanan R, Camurri M, Dellaert F and Fallon M (2022) Learning inertial odometry for dynamic legged robot state estimation. In: *Conference on robot learning*. PMLR, pp. 1575–1584.

Burnett K, Yoon DJ, Wu Y, Li AZ, Zhang H, Lu S, Qian J, Tseng W, Lambert A, Leung KYK et al. (2023) Boreas: A multi-season autonomous driving dataset. *International Journal of Robotics Research* 42(1–2): 33–42.

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

Cao Z, Talbot W and Li K (2025a) Resple: Recursive spline estimation for lidar-based odometry. *IEEE Robotics and Automation Letters* .

Cao Z, Talbot W and Li K (2025b) Resple: Recursive spline estimation for lidar-based odometry. *IEEE Robotics and Automation Letters* 10(10): 10666–10673.

Chaney K, Cladera F, Wang Z, Bisulco A, Hsieh MA, Korpela C, Kumar V, Taylor CJ and Daniilidis K (2023) M3ed: Multi-robot, multi-sensor, multi-environment event dataset. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. pp. 4015–4022.

Chavez-Garcia RO, Fourmy M, Foehn P, Flayols T, Dell'Amico M, Solà J, Besseron X, Lachner J, Stasse O, Chaumette F et al. (2018) Multi-modal sensor fusion for robust and accurate legged robot state estimation. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1364–1371.

Chen J, Frey J, Zhou R, Miki T, Martius G and Hutter M (2024a) Identifying terrain physical parameters from vision-towards physical-parameter-aware locomotion and navigation. *IEEE Robotics and Automation Letters* .

Chen K, Lopez BT, Agha-mohammadi Aa and Mehta A (2022) Direct lidar odometry: Fast localization with dense point clouds. *IEEE Robotics and Automation Letters* 7(2): 2000–2007. DOI: 10.1109/LRA.2022.3142739.

Chen K, Nemiroff R and Lopez BT (2023) Direct lidar-inertial odometry: Lightweight lio with continuous-time motion correction. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 3983–3989. DOI:10.1109/ICRA48891.2023.10160508.

Chen Z, Le Gentil C, Lin F, Lu M, Qiao Q, Xu B, Qi Y and Lu P (2025) Breaking the static assumption: A dynamic-aware

lio framework via spatio-temporal normal analysis. *IEEE Robotics and Automation Letters* 10(12): 12636–12643. DOI: 10.1109/LRA.2025.3623436.

Chen Z, Xu Y, Yuan S and Xie L (2024b) ig-lio: An incremental gicp-based tightly-coupled lidar-inertial odometry. *IEEE Robotics and Automation Letters* 9(2): 1883–1890. DOI:10.1109/LRA.2024.3349915.

Choi S and Kim TW (2025) Probabilistic kernel optimization for robust state estimation. *IEEE Robotics and Automation Letters* 10(3): 2998–3005. DOI:10.1109/LRA.2025.3536294.

Choi S, Park D, Hwang SY and Kim TW (2025) Statistical uncertainty learning for robust visual-inertial state estimation. URL https://arxiv.org/abs/2510.01648.

Chung D and Kim J (2024) Nv-liom: Lidar-inertial odometry and mapping using normal vectors towards robust slam in multifloor environments. *IEEE Robotics and Automation Letters* 9(11): 9375–9382. DOI:10.1109/LRA.2024.3457373.

Cui Y, Chen R, Chu W, Chen L, Tian D, Li Y, Cao D, Yu Y, Chen YY, Yuan J et al. (2023) Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 1861–1885.

Dellaert F (2012) Factor graphs and gtsam: A hands-on introduction. *Georgia Institute of Technology, Tech. Rep* 2(4).

Erni G, Frey J, Miki T, Mattamala M and Hutter M (2023) Mem: Multi-modal elevation mapping for robotics and learning. In: *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 11011–11018.

Escontrela A, Kerr J, Allshire A, Frey J, Duan R, Sferrazza C and Abbeel P (2025) Gaussgym: An open-source real-to-sim framework for learning locomotion from pixels. *arXiv preprint arXiv:2510.15352* .

Fan Y, Zhao T and Wang G (2024) Schurvins: Schur complement-based lightweight visual inertial navigation system. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 17964–17973.

Feng F, Braun J, Titus A, McGill SG, Ott L, Siegwart R and Nieto J (2023) Scale-aware visual-inertial localization over large areas in real-time on a legged robot. *IEEE Transactions on Robotics* 39(4): 2512–2530.

Ferrari S, Giammarino LD, Brizi L and Grisetti G (2024) Mad-icp: It is all about matching data – robust and informed lidar odometry. *IEEE Robotics and Automation Letters* 9(11): 9175–9182. DOI: 10.1109/LRA.2024.3456509.

Firoozi R, Tucker J, Tian S, Majumdar A, Sun J, Liu W, Zhu Y, Song S, Kapoor A, Hausman K, Ichter B, Driess D, Wu J, Lu C and Schwager M (2025) Foundation models in robotics: Applications, challenges, and the future. *The International Journal of Robotics Research* 44(5): 701–739. DOI:10.1177/02783649241281508.

Frey J, Mattamala M, Chebrolu N, Cadena C, Fallon M and Hutter M (2023) Fast Traversability Estimation for Wild Visual Navigation. In: *Proceedings of Robotics: Science and Systems*. Daegu, Republic of Korea, pp. 54–70. DOI:10.15607/RSS.2023.XIX.054. URL https://roboticsproceedings.org/rss19/p054.html.

Frey J, Tuna T, Fu LFT, Weibel C, Patterson K, Krummenacher B, Müller M, Nubert J, Fallon M, Cadena C and Hutter M (2025) Boxi: Design decisions in the context of algorithmic performance for robotics. In: *Proceedings of Robotics: Science and Systems*. Los Angeles, United States, pp. 1–9.

Fu LFT, Chebrolu N and Fallon M (2023) Extrinsic calibration of camera to lidar using a differentiable checkerboard model. In: *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 1825–1831. DOI:10.1109/IROS55552.2023.10341781.

Furgale P, Rehder J and Siegwart R (2013) Unified temporal and spatial calibration for multi-sensor systems. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 1280–1286.

G3tupup (2022) CTLO: Continuous time lidar odometry. GitHub repository. URL https://github.com/G3tupup/ctlo. Accessed: 2026-01-03.

Geiger A, Lenz P, Stiller C and Urtasun R (2013) Vision meets robotics: The kitti dataset. *The international journal of robotics research* 32(11): 1231–1237.

Gelfand N, Ikemoto L, Rusinkiewicz S and Levoy M (2003) Geometrically stable sampling for the icp algorithm. In: *Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings.* pp. 260–267. DOI: 10.1109/IM.2003.1240258.

Geneva P, Eckenhoff K, Lee W, Yang Y and Huang G (2020) Openvins: A research platform for visual-inertial estimation. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 4666–4672. DOI:10.1109/ICRA40945.2020.9196524.

Grupp M (2017) evo: Python package for the evaluation of odometry and slam. https://github.com/MichaelGrupp/evo.

Hatleskog J and Alexis K (2024) Probabilistic degeneracy detection for point-to-plane error minimization. *IEEE Robotics and Automation Letters* 9(12): 11234–11241. DOI:10.1109/LRA.2024.3484153.

He D, Xu W, Chen N, Kong F, Yuan C and Zhang F (2023) Point-lio: robust high-bandwidth light detection and ranging inertial odometry. *Advanced Intelligent Systems* 5(7): 2200459.

Helmberger M, Morin K, Berner B, Kumar N, Cioffi G and Scaramuzza D (2022) The hilti slam challenge dataset. *IEEE Robotics and Automation Letters* 7(3): 7518–7525.

Hu X, Chen X, Jia M, Wu J, Tan P and Waslander SL (2025) Dcreg: Decoupled characterization for efficient degenerate lidar registration. *arXiv preprint arXiv:2509.06285* .

Huai Z and Huang G (2024) A consistent parallel estimation framework for visual-inertial slam. *IEEE Transactions on Robotics* 40: 3734–3755. DOI:10.1109/TRO.2024.3433868.

Hutter M, Gehring C, Jud D, Lauber A, Bellicoso CD, Tsounis V, Hwangbo J, Bodie K, Fankhauser P, Bloesch M et al. (2016) Anymal-a highly mobile and dynamic quadrupedal robot. In: *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, pp. 38–44.

(HViktorTsoi) K (2023) PV-LIO: A probabilistic voxelmap-based lidar-inertial odometry. GitHub repository. URL https://github.com/HViktorTsoi/PV-LIO. Accessed: 2025-12-23.

Inglin Y, Frey J, Chen C and Hutter M (2025) Less is More: Towards generalizable scalable visual navigation from limited data. *Manuscript in Writing* .

Jeong S, Kim H and Cho Y (2024) Diter: Diverse terrain and multimodal dataset for field robot navigation in outdoor environments. *IEEE Sensors Letters* 8(3): 1–4.

Jiao J, Wei H, Hu T, Hu X, Zhu Y, He Z, Wu J, Yu J, Xie X, Huang H et al. (2022) Fusionportable: A multi-sensor campus-scene dataset for evaluation of localization and mapping accuracy on diverse platforms. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 3851–3856.

Jung JH, Choe Y and Park CG (2022) Photometric visual-inertial navigation with uncertainty-aware ensembles. *IEEE Transactions on Robotics* 38(4): 2039–2052. DOI:10.1109/TRO.2021.3139964.

Jung M, Jung S and Kim A (2023) Asynchronous multiple lidar-inertial odometry using point-wise inter-lidar uncertainty propagation. *IEEE Robotics and Automation Letters* 8(7): 4211–4218. DOI:10.1109/LRA.2023.3281264.

Kannala J and Brandt S (2006) A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE transactions on pattern analysis and machine intelligence* 28: 1335–40. DOI:10.1109/TPAMI.2006.153.

Kim D, Gu Y and Fearing RS (2021a) Legged robot state estimation using invariant kalman filtering and learned contact events. *IEEE Robotics and Automation Letters* 6(2): 2971–2978.

Kim G, Choi S and Kim A (2021b) Scan context++: Structural place recognition robust to rotation and lateral variations in urban environments. *IEEE Transactions on Robotics* 38(3): 1856–1874.

Kim G, Park YS, Cho Y, Jeong J and Kim A (2020) Mulran: Multimodal range dataset for urban place recognition. In: *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, pp. 6246–6253.

Kim J, Kim H, Jeong S, Shin Y and Cho Y (2024) Diter++: Diverse terrain and multi-modal dataset for multi-robot slam in multi-session environments. *arXiv preprint arXiv:2412.05839* .

Koide K, Yokozuka M, Oishi S and Banno A (2024) Glim: 3d range-inertial localization and mapping with gpu-accelerated scan matching factors. *Robotics and Autonomous Systems* 179: 104750. DOI:https://doi.org/10.1016/j.robot.2024.104750. URL https://www.sciencedirect.com/science/article/pii/S0921889024001349.

Korthals T, Kragh M, Christiansen P and Karstoft H (2022) Multi-modal semantic slam for complex dynamic environments. *IEEE Robotics and Automation Letters* 7(2): 2585–2592.

Kumar A, Fu Z, Pathak D and Malik J (2021) Rma: Rapid motor adaptation for legged robots. *arXiv preprint arXiv:2107.04034* .

Lang X, Chen C, Tang K, Ma Y, Lv J, Liu Y and Zuo X (2023) Coco-lic: Continuous-time tightly-coupled lidar-inertial-camera odometry using non-uniform b-spline. *IEEE Robotics and Automation Letters* 8(11): 7074–7081. DOI:10.1109/LRA.2023.3315542.

Lee D, Lim H and Han S (2025a) GenZ-ICP: Generalizable and Degeneracy-Robust LiDAR Odometry Using an Adaptive Weighting. *IEEE Robotics and Automation Letters (RA-L)* 10(1): 152–159. DOI:10.1109/LRA.2024.3498779.

Lee EM (2024) FAST-LIO-MULTI: Multi-lidar version of fast-lio2. GitHub repository. URL https://github.com/engcang/FAST_LIO_MULTI. Accessed: 2025-12-23.

Lee MA, Zhu Y, Srinivasan K, Shah P, Savarese S, Fei-Fei L, Garg A and Bohg J (2020) Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks. *IEEE Transactions on Robotics* 36(3): 582–596.

Lee W, Geneva P, Chen C and Huang G (2025b) Mins: Efficient and robust multisensor-aided inertial navigation system. *Journal of Field Robotics* 42(7): 3252–3284.

Leutenegger S (2022) OKVIS2: Realtime scalable visual-inertial slam with loop closure. *arXiv preprint arXiv:2202.09199* .

Li C, Krause A and Hutter M (2025) Robotic world model: A neural network simulator for robust policy optimization in robotics. *arXiv preprint arXiv:2501.10100* .

Liang J, Song D, Shuvo MNH, Durrani M, Taranath K, Penskiy I, Manocha D and Xiao X (2025) Gnd: Global navigation dataset with multi-modal perception and multi-category traversability in outdoor campus environments. In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 11345–11352.

Lin J, Yuan C, Cai Y, Li H, Ren Y, Zou Y, Hong X and Zhang F (2023) Immesh: An immediate lidar localization and meshing framework. *IEEE Transactions on Robotics* 39(6): 4312–4331. DOI:10.1109/TRO.2023.3321227.

Liu X, Liu Z, Kong F and Zhang F (2023) Large-scale lidar consistent mapping using hierarchical lidar bundle adjustment. *IEEE Robotics and Automation Letters* 8(3): 1523–1530.

Liu Z, Li H, Yuan C, Liu X, Lin J, Li R, Zheng C, Zhou B, Liu W and Zhang F (2024) Voxel-slam: A complete, accurate, and versatile lidar-inertial slam system. *arXiv preprint arXiv:2410.08935* .

Lv J, Lang X, Xu J, Wang M, Liu Y and Zuo X (2023) Continuous-time fixed-lag smoothing for lidar-inertial-camera slam. *IEEE/ASME Transactions on Mechatronics* 28(4): 2259–2270. DOI:10.1109/TMECH.2023.3241398.

Maddern W, Pascoe G, Linegar C and Newman P (2017) 1 000 km of autonomy: The oxford robotcar dataset. *International Journal of Robotics Research* 36(1): 3–15. DOI:10.1177/0278364916679498.

Malladi M, Guadagnino T, Lobefaro L and Stachniss C (2025) A robust approach for lidar-inertial odometry without sensor-specific modeling. *arXiv preprint* arXiv:2509.06593. URL https://arxiv.org/pdf/2509.06593.

Martínez O (2023) fast_LIMO: Fast-limo. GitHub repository. URL https://github.com/fetty31/fast_LIMO. Accessed: 2026-01-03.

Meng K, Sun H, Qi J and Wang H (2023) Section-lio: A high accuracy lidar-inertial odometry using undistorted sectional point. *IEEE Access* 11: 144918–144927. DOI:10.1109/ACCESS.2023.3344037.

Miki T, Lee J, Hwangbo J, Wellhausen L, Koltun V and Hutter M (2022a) Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science robotics* 7(62): eabk2822.

Miki T, Lee J, Wellhausen L and Hutter M (2024) Learning to walk in confined spaces using 3d representation. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 8649–8656.

Miki T, Wellhausen L, Grandia R, Jenelten F, Homberger T and Hutter M (2022b) Elevation mapping for locomotion and navigation using gpu. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 2273–2280.

Nguyen TM, Xu X, Jin T, Yang Y, Li J, Yuan S and Xie L (2024a) Eigen is all you need: Efficient lidar-inertial continuous-time odometry with internal association. *IEEE Robotics and Automation Letters* 9(6): 5330–5337. DOI:10.1109/LRA.2024.3391049.

Nguyen TM, Yuan S, Nguyen TH, Yin P, Cao H, Xie L, Wozniak M, Jensfelt P, Thiel M, Ziegenbein J and Blunder N (2024b) Mcd:

Diverse large-scale multi-campus dataset for robot perception. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 22304–22313.

NovAtel poH (2025a) CPT7. https://novatel.com/products/waypoint-post-processing-software/inertial-explorer. Accessed 05 December 2025.

NovAtel poH (2025b) CPT7. https://hexagondownloads.blob.core.windows.net/public/Novatel/assets/Documents/Papers/SPAN-CPT7-PS/SPAN-CPT7-PS.pdf. Accessed 05 December 2025.

Nubert J, Tuna T, Frey J, Cadena C, Kuchenbecker KJ, Khattak S and Hutter M (2025) Holistic fusion: Task-and setup-agnostic robot localization and state estimation with factor graphs. *arXiv preprint arXiv:2504.06479* .

Nwankwo L, Ellensohn B, Dave V, Hofer P, Forstner J, Villneuve M, Galler R and Rueckert E (2024) Envodat: A large-scale multisensory dataset for robotic spatial awareness and semantic reasoning in heterogeneous environments. *arXiv preprint arXiv:2410.22200* .

Olson E (2011) Apriltag: A robust and flexible visual fiducial system. In: *2011 IEEE International Conference on Robotics and Automation*. pp. 3400–3407. DOI:10.1109/ICRA.2011.5979561.

Patel M, Frey J, Mittal M, Yang F, Hansson A, Bar A, Cadena C and Hutter M (2025) DeFM: Learning foundation representations from depth for robotics. *Manuscript under review* .

Peng Y, Chen C, Wu K and Huang G (2025) $\sqrt{\textbf{vins}}$: Robust and ultrafast square-root filter-based 3d motion tracking. *IEEE Transactions on Robotics* 41: 6570–6589. DOI:10.1109/TRO.2025.3626607.

Petracek P, Alexis K and Saska M (2024) Rms: Redundancy-minimizing point cloud sampling for real-time pose estimation. *IEEE Robotics and Automation Letters* 9(6): 5230–5237. DOI:10.1109/LRA.2024.3389820.

Pérez-Ruiz C and Solà J (2025) Limoncello: Revisited ikfom on the sgal(3) manifold for fast lidar-inertial odometry. URL https://arxiv.org/abs/2512.19567.

Püntener C, Schwabe J, Garmier D, Frey J and Hutter M (2025) Kleinkram: Open robotic data management. URL https://arxiv.org/abs/2511.20492.

Qin T, Li P and Shen S (2018) Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics* 34(4): 1004–1020. DOI:10.1109/TRO.2018.2853729.

Qin T and Shen S (2018) Online temporal calibration for monocular visual-inertial systems. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 3662–3669. DOI:10.1109/IROS.2018.8593603.

Ramezani M, Wang Y, Camurri M, Wisth D, Mattamala M and Fallon M (2020) The newer college dataset: Handheld lidar, inertial and vision with ground truth. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 4353–4360. DOI:10.1109/IROS45743.2020.9340849.

Rehder J, Nikolic J, Schneider T, Hinzmann T and Siegwart R (2016) Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 4304–4311.

Reijgwart V, Cadena C, Siegwart R and Ott L (2023) Efficient volumetric mapping of multi-scale environments using wavelet-based compression. In: *Proceedings of Robotics: Science and Systems*. Daegu, Republic of Korea, pp. 65–76. DOI:10.15607/RSS.2023.XIX.065.

Roth P, Frey J, Cadena C and Hutter M (2025) Learned perceptive forward dynamics model for safe and platform-aware robotic navigation. *arXiv preprint arXiv:2504.19322* .

Ruan J, Li B, Wang Y and Sun Y (2023) Slamesh: Real-time lidar simultaneous localization and meshing. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 3546–3552. DOI:10.1109/ICRA48891.2023.10161425.

Schubert D, Goll T, Demmel N, Usenko V, Stückler J and Cremers D (2018) The tum vi benchmark for evaluating visual-inertial odometry. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1680–1687.

Shan Z, Li R and Schwertfeger S (2019) Rgbd-inertial trajectory estimation and mapping for ground robots. *Sensors* 19(10): 2251.

Shen H, Wu Z, Hui Y, Wang W, Lyu Q, Deng T, Zhu Y, Tian B and Wang D (2025) Cte-mlo: Continuous-time and efficient multi-lidar odometry with localizability-aware point cloud sampling. *IEEE Transactions on Field Robotics* 2: 165–187. DOI:10.1109/TFR.2025.3543142.

Skuddis D and Haala N (2024) Dmsa - dense multi scan adjustment for lidar inertial odometry and global optimization. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 12027–12033. DOI:10.1109/ICRA57147.2024.10610818.

Soares PMB, Tourani A, Fernandez-Cortizas M, Noya AB, Sanchez-Lopez JL and Voos H (2025) Smapper: A multi-modal data acquisition platform for slam benchmarking. *arXiv preprint arXiv:2509.09509* .

Su Y, Shao S, Zhang Z, Xu P, Cao Y and Cheng H (2025) Glo: General lidar-only odometry with high efficiency and low drift. *IEEE Robotics and Automation Letters* 10(4): 3518–3525. DOI:10.1109/LRA.2025.3542704.

Tancik M, Weber E, Ng E, Li R, Yi B, Wang T, Kristoffersen A, Austin J, Salahi K, Ahuja A et al. (2023) Nerfstudio: A modular framework for neural radiance field development. In: *ACM SIGGRAPH 2023 conference proceedings*. pp. 1–12.

Tao Y, Muñoz-Bañón MÁ, Zhang L, Wang J, Fu LFT and Fallon M (2025) The Oxford Spires dataset: Benchmarking large-scale lidar-visual localisation, reconstruction and radiance-field methods. *International Journal of Robotics Research* Also available as arXiv:2411.10546.

Teng H, Wang Y, Chatziparaschis D and Karydis K (2025) Adaptive lidar odometry and mapping for autonomous agricultural mobile robots in unmanned farms. *Computers and Electronics in Agriculture* 232: 110023.

Tuna T, Nubert J, Nava Y, Khattak S and Hutter M (2024) X-icp: Localizability-aware lidar registration for robust localization in extreme environments. *IEEE Transactions on Robotics* 40: 452–471. DOI:10.1109/TRO.2023.3335691.

Tuna T, Nubert J, Pfreundschuh P, Cadena C, Khattak S and Hutter M (2025) Informed, constrained, aligned: A field analysis on degeneracy-aware point cloud registration in the wild. *IEEE Transactions on Field Robotics* 2: 485–515. DOI:10.1109/TFR.2025.3576053.

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. DOI:10.1109/34.88573.

Vaidis M, Shahraji MH, Daum E, Dubois W, Giguère P and Pomerleau F (2024) Rts-gt: Robotic total stations ground truthing dataset. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 17050–17056.

Vizzo I, Guadagnino T, Mersch B, Wiesmann L, Behley J and Stachniss C (2023) Kiss-icp: In defense of point-to-point icp – simple, accurate, and robust registration if done the right way. *IEEE Robotics and Automation Letters* 8(2): 1029–1036. DOI: 10.1109/LRA.2023.3236571.

Wang J, Chen M, Karaev N, Vedaldi A, Rupprecht C and Novotny D (2025) Vggt: Visual geometry grounded transformer. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 5294–5306.

Wang S, Leroy V, Cabon Y, Chidlovskii B and Revaud J (2024a) Dust3r: Geometric 3d vision made easy. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 20697–20709.

Wang Z, Yao C, Ge Y, Shi G, Yang N, Zhu Z, Dong K, Wei H, Jia Z and Wu J (2024b) Are we ready for planetary exploration robots? the tail-plus dataset for slam in granular environments. *arXiv preprint arXiv:2404.13600* .

Wei H, Jiao J, Hu X, Yu J, Xie X, Wu J, Zhu Y, Liu Y, Wang L and Liu M (2025) Fusionportablev2: A unified multi-sensor dataset for generalized slam across diverse platforms and scalable environments. *The International Journal of Robotics Research* 44(7): 1093–1116.

Windecker T, Patel M, Reuss M, Schwarzkopf R, Cadena C, Lioutikov R, Hutter M and Frey J (2025) Navitrace: Evaluating embodied navigation of vision-language models. *arXiv preprint arXiv:2510.26909* .

Wu Y, Guadagnino T, Wiesmann L, Klingbeil L, Stachniss C and Kuhlmann H (2024) Lio-ekf: High frequency lidar-inertial odometry using extended kalman filters. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 13741–13747. DOI:10.1109/ICRA57147.2024.10610667.

Xu J, Li R, Huang S, Zhao X, Qiu S, Chen Z and Zhao L (2023) R2dio: A robust and real-time depth-inertial odometry leveraging multimodal constraints for challenging environments. *IEEE Transactions on Instrumentation and Measurement* 72: 1–11. DOI:10.1109/TIM.2023.3320753.

Xu J, Ma Y, Li Y, Zhang X, Zhou J, Yuan S and Xie L (2025) Dynamic initialization for lidar-inertial slam. *IEEE/ASME Transactions on Mechatronics* : 1–12DOI:10.1109/TMECH.2025.3554878.

Xu W, Cai Y, He D, Lin J and Zhang F (2022) Fast-lio2: Fast direct lidar-inertial odometry. *IEEE Transactions on Robotics* 38(4): 2053–2073. DOI:10.1109/TRO.2022.3141876.

Yao C, Ge Y, Shi G, Wang Z, Yang N, Zhu Z, Wei H, Zhao Y, Wu J and Jia Z (2024) Tail: A terrain-aware multi-modal slam dataset for robot locomotion in deformable granular environments. *IEEE Robotics and Automation Letters* 9(7): 6696–6703.

Yin J, Li A, Li T, Yu W and Zou D (2022) M2DGR: A multi-sensor and multi-scenario slam dataset for ground robots. *IEEE Robotics & Automation Letters* 7(2): 4869–4876.

Youm D, Oh H, Choi S, Kim H, Jeon S and Hwangbo J (2025) Legged robot state estimation with invariant extended kalman filter using neural measurement network. In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 670–676.

Yu W, Xu J, Zhao C, Zhao L, Nguyen TM, Yuan S, Bai M and Xie L (2024) I2ekf-lo: A dual-iteration extended kalman filter based lidar odometry. In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 10453–10460. DOI:10.1109/IROS58592.2024.10801455.

Yuan C, Lin J, Liu Z, Wei H, Hong X and Zhang F (2024a) Btc: A binary and triangle combined descriptor for 3-d place recognition. *IEEE Transactions on Robotics* 40: 1580–1599. DOI:10.1109/TRO.2024.3353076.

Yuan Z, Lang F, Deng J, Luo H and Yang X (2025) Voxel-svio: Stereo visual-inertial odometry based on voxel map. *IEEE Robotics and Automation Letters* 10(6): 6352–6359. DOI: 10.1109/LRA.2025.3568307.

Yuan Z, Lang F, Xu T and Yang X (2024b) Sr-lio: Lidar-inertial odometry with sweep reconstruction. In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 7862–7869. DOI:10.1109/IROS58592.2024.10802314.

Zhang J and Singh S (2017) Low-drift and real-time lidar odometry and mapping. *Autonomous robots* 41(2): 401–416.

Zhang L, Helmberger M, Fu LFT, Wisth D, Camurri M, Scaramuzza D and Fallon M (2023) Hilti-oxford dataset: A millimeter-accurate benchmark for simultaneous localization and mapping. *IEEE Robotics and Automation Letters* 8(1): 408–415. DOI: 10.1109/LRA.2022.3226077.

Zhao C, Hu K, Xu J, Zhao L, Han B, Wu K, Tian M and Yuan S (2024a) Adaptive-lio: Enhancing robustness and precision through environmental adaptation in lidar inertial odometry. *IEEE Internet of Things Journal* .

Zhao S, Gao Y, Wu T, Singh D, Jiang R, Sun H, Sarawata M, Qiu Y, Whittaker W, Higgins I et al. (2024b) Subt-mrs dataset: Pushing slam towards all-weather environments. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 22647–22657.

Zhao S, Zhang H, Wang P, Nogueira L and Scherer S (2021) Super odometry: Imu-centric lidar-visual-inertial estimator for challenging environments. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 8729–8736. DOI:10.1109/IROS51168.2021.9635862.

Zheng C, Xu W, Zou Z, Hua T, Yuan C, He D, Zhou B, Liu Z, Lin J, Zhu F et al. (2024) Fast-livo2: Fast, direct lidar-inertial-visual odometry. *IEEE Transactions on Robotics* .

Zheng X and Zhu J (2024) Traj-lo: In defense of lidar-only odometry using an effective continuous-time trajectory. *IEEE Robotics and Automation Letters* 9(2): 1961–1968. DOI:10.1109/LRA.2024.3352360.