FusionPortable: A Multi-Sensor Campus-Scene Dataset for Evaluation of Localization and Mapping Accuracy on Diverse Platforms

Jianhao Jiao\textsuperscript{2,*}, Hexiang Wei\textsuperscript{2,*}, Tianshuai Hu\textsuperscript{2,*}, Xiangcheng Hu\textsuperscript{2,*}, Yilong Zhu\textsuperscript{2}, Zhijian He\textsuperscript{2}, Jin Wu\textsuperscript{2}, Jingwen Yu\textsuperscript{2,5}, Xupeng Xie\textsuperscript{2}, Huaiyang Huang\textsuperscript{2}, Ruoyu Geng\textsuperscript{2}, Lujia Wang\textsuperscript{2,4}, Ming Liu\textsuperscript{1,2,3}

Abstract—Combining multiple sensors enables a robot to maximize its perceptual awareness of environments and enhance its robustness to external disturbance, crucial to robotic navigation. This paper proposes the FusionPortable benchmark, a complete multi-sensor dataset with a diverse set of sequences for mobile robots. This paper presents three contributions. We first advance a portable and versatile multi-sensor suite that offers rich sensory measurements: 10Hz LiDAR point clouds, 20Hz stereo event camera images, high-rate and asynchronous events from stereo event cameras, 200Hz inertial readings from an IMU, and 10Hz GPS signal. Sensors are already temporally synchronized in hardware. This device is lightweight, self-contained, and has plug-and-play support for mobile robots. Second, we construct a dataset by collecting 17 sequences that cover a variety of environments on the campus by exploiting multiple robot platforms for data collection. Some sequences are challenging to existing SLAM algorithms. Third, we provide ground truth for the decouple localization and mapping performance evaluation. We additionally evaluate state-of-the-art SLAM approaches and identify their limitations. The dataset, consisting of raw sensor measurements, ground truth, calibration data, and evaluated algorithms, will be released.

I. INTRODUCTION
A. Motivation
Multi-sensor fusion for robust perception is fundamental to various robotic applications. Different sensors can complement each other, and thus the system’s perception capability is enhanced with sensor fusion. Over the past decades, research on multi-sensor SLAM has made substantial progress. High-quality open datasets, which are collections of multisensor data and provide a suite of benchmark tools, significantly contribute to this advancement. On one hand, these datasets can waive inhibitive requirements on budget and workforce, such as system integration calibration and field operations. On the other hand, they investigate the advantages and limitations of current SLAM solutions and elaborate design practical, but challenging sequences [1], [2]. Several of them also introduce novel sensors and indicate future research opportunities [3]. Researchers can easily develop, validate, and rank their algorithms with others, thus accelerating the breakthroughs. However, existing datasets were mostly collected with a single data collection platform or simplified sensor configuration. Researchers may only utilize limited sensors to develop algorithms that has a risk of overfitting to a benchmark. Hence, we consider that a desirable dataset should fulfill the following four requirements.

1) Various sensors are required, making it possible to explore novel approaches to utilize them jointly.
2) Algorithm evaluation should be fairly conducted on various mobile robots. These robots perform different motion patterns that may challenge several SLAM algorithms’ assumptions.
3) Sequences have to cover from room-scale (meter-level) to large-scale (kilometer-level) environments to evaluate algorithms’ scalability.
4) Ground-truth trajectories and 3D maps are required to evaluate algorithms’ localization and surface reconstruction accuracy, respectively.

B. Contributions
There appears to be an absence of compatible public datasets that satisfy these requirements, motivating us to propose a new SLAM benchmark.

This paper proposes the FusionPortable benchmark, a novel multi-sensor dataset with a set of sequences from diverse environments. Our contributions are presented three-fold. First, a portable and versatile multi-sensor device is elaborately manufactured. Two RGB frame cameras are mounted on the left and right side, one high-frequency and high-precision IMU is mounted internally, and one RTK-GPS is installed on the top position. Moreover, thanks to current progress in sensory technology, both novel event cameras and high-resolution 3D LiDAR are available. Thus, we also integrate them with our sensor rig and investigate their performance. All these sensors are mounted on the same rigid aluminum-alloy-based parts. Thus, their spatial relation has a tiny dynamic deviation. The complete device has its own clock synchronization unit, processor, and battery, thus self-contained. Since its size, weight, and extensibility (see Fig. 1) are satisfying, we advance that it would be a plug-and-play support to various mobile robots.

Second, we install the sensor rig on various platforms ranging from the handheld mode with a gimbal stabilizer, a quadruped robot, and an autonomous vehicle in performing...
distinguishable motion for the dataset construction. Various structured or semi-structured environments on The Hong Kong University of Science and Technology (HKUST) campus, including the lab, garden, canteen, corridor, escalator, and outdoor road, are examined in the dataset. Also, the collected sequences present several environmental changes caused by external light, moving objects, and scene texture. These issues are challenging to SLAM algorithms.

Third, besides ground-truth poses, we also provide ground-truth maps of most indoor sequences. We consider that measuring the mapping accuracy is crucial for evaluation. We also benchmark several state-of-the-art (SOTA) SLAM systems, including two vision-based methods and four LiDAR-based approaches. To benefit the community, the dataset will be publicly released: [https://ram-lab.com/file/site/multi-sensor-dataset](https://ram-lab.com/file/site/multi-sensor-dataset).

## II. RELATED WORK

There are extensive datasets for robotic perception. Here, we introduce related works with a focus on SLAM.

Several datasets were specifically designed for one type of sensor. Mueggler et al. [3] proposed the event camera dataset for the purpose of overcoming illumination and motion blur issues caused by frame cameras. Pomerleau et al. [1] proposed the point cloud dataset that covers a large spectrum of environmental structures to challenge registration algorithms. Handa et al. [14] promoted the research on RGB-D cameras by publishing the ICL-NUIM dataset.

Complementing vision sensors with inertial measurements, visual-inertial odometry (VIO) approaches can tremendously improve camera tracking accuracy and robustness. Relevant datasets have been reported. Burri et al. [4] presented the EuRoc dataset collected by a micro aerial vehicle (MAV) in an industrial environment and a room. Schubert et al. [5] put forward the TUM VI benchmark by collecting handheld sequences with a careful photometric calibration forwards.

The DARPA challenge has driven the development of autonomous vehicles. Huang et al. [6] presented the MIT DARPA dataset with over 90km sequence. Geiger et al. [7] presented the KITTI driving benchmark where diverse perception tasks are explored. There are other datasets targeting at long-term navigation [15] and urban challenges [9].

Several datasets were collected by handheld devices and other types of ground robots. Ramezani et al. [10] collected the Newer College Dataset with a handheld device. The NCLT dataset [11] facilitated the long-term SLAM research by collecting sequences in a college campus, over 147.4km traverse and 15 months. The M2DGR dataset covers various challenging scenarios such as entering lifts and indoor-outdoor traverse [12] with a ground robot. Zhu et al. [13] proposed a multi-vehicle dataset for event-based perception.

Table I compares existing datasets with our work. In summary, our dataset is more complete from xx aspects: 1) raw and rich sensory measurements; 2) data collection on three different platforms including a legged robot; 3) ground-truth trajectories and 3D maps for algorithm evaluation.

### III. SYSTEM OVERVIEW

This section introduces sensors used in our dataset and how we achieve the spatio-temporal calibration between each sensor. Fig. 1 shows the handheld device equipped with multiple sensors and how it is mounted on three data collection platforms.

#### A. Sensor Configuration

Sensors’ characteristics can be found in Table II. We use the Intel NUC to run sensor drivers, attach timestamps of sensor messages, and record messages into ROS bags on the Ubuntu system. The PC uses an i7 processor, 1TB solid-state drive (SSD), and 64GB DDR4 memory. Below, we provide detailed description of these sensors.

1) **3D LiDARs**: We configure the OS1-128 LiDAR to provide accurate measurements of surrounding environments. This LiDAR has two attractive properties. First, an internal synchronized IMU outputs 100Hz linear accelerations and angular velocities. Second, it additionally outputs depth images, signal images, and ambient images of surroundings.

2) **Stereo Frame Cameras**: Two FILR BFS-U3-31S4C global-shutter color cameras are mounted at two sides on the system, facing directly forward. They are synchronized by an external trigger and capture high-resolution images at 20 fps. Their exposure time is set as fixed values to minimize the relative latency. Our experiments show that the average difference in timestamps of these images is below 1ms.
3) Stereo Event Cameras: Two event cameras are also configured. They possess several desirable properties: high temporal resolution, high dynamic range, and low power consumption. The cameras have a 346 × 260 resolution and an internal high-rate IMU output. Event cameras are synchronized using the trigger signal generated from the left camera (master) to deliver sync pulses to the right (slave) through an external wire. But there is no way to synchronize the LiDAR’s laser light, both cameras are equipped with additional infrared filters. For indoor sequences, we manually set and fix the APS exposures, which helps to minimize the latency between cameras. For outdoor sequences, we use auto-exposure to avoid over- or under-exposure.

4) Inertial Measurement Unit: A tactical-grade STIM300 IMU that is rigidly mounted below the LiDAR is employed as the main inertial sensor of the system. It features a high update rate (200Hz) and low noisy and drifting measurements. Its bias Instability is around 0.3°/h.

5) Global Positioning System: We additionally install a ZED-F9P RTK-GPS device on the top of the LiDAR. In outdoor scenes, the GPS is activated and provides accurate latitude, longitude, and altitude readings. But it may sometimes become unstable due to buildings’ occlusion.

B. Sensor Calibration

We carefully calibrate *intrinsics* of individual sensors, *extrinsics*, and overall time latency between sensors in advance. We define the coordinate system of the STIM300 IMU as the *body frame*. We provide calibration data and reports in the dataset website.

1) Clock Synchronization: We use an FPGA to generate an external signal trigger to synchronize clocks of all sensors. This can guarantee data collection across multiple sensors with minimum latency. The FPGA receives a pulse-per-second (PPS) signal from the GPS and outputs 200, 20, 10Hz signal to the IMU, cameras, and LiDAR, respectively. The FPGA switches to use its internal clock to enable the time synchronization in GPS-denied scenes.

2) Stereo Camera Calibration: Intrinsics and extrinsics of our stereo frame and event cameras are estimated using the Matlab calibration toolbox, where the pinhole camera and radial-tangential distortion model are used. We move the sensor suite before a checkerboard to collect a sequence of images. We evenly sample images as the calibration data and manually remove outliers with high reprojection errors.

3) Camera-IMU Extrinsic Calibration: The intrinsics of IMUs are calibrated using the Allen derivation toolbox that estimates the noisy density and random walk for gyroscope and accelerometer measurements. After that, the spatial and temporal parameters of a camera w.r.t. an IMU are obtained by the Kalibr [16]. Our system consists of 4 IMUs: STIM300, ICM20948 in the LiDAR, and two MPU6050 in the DAVIS346 event cameras. Thus, we calibrate the intrinsics of these IMUs, and estimate extrinsics of these sensor pairs: (STIM300, frame cameras), (STIM300, event cameras), (left MPU6050, left DAVIS346), and (right MPU6050, right DAVIS346).

4) Camera-LiDAR Extrinsic Calibration: Given initial extrinsics, we further refine the camera-LiDAR extrinsics. The checkerboard is the calibration target that provides distinctive corners and boundaries for data association. We extend the work proposed by Zhou et al. [17] by improving feature extraction and matching step. We instead extract the outer corners of the board from point clouds and images. The

https://github.com/ori-drs/allan_variance_ros
Fig. 3. Sample sensor measurements. (a)-(d): images captured by the frame camera. (e)-(f): images augmented by positive events (red) and negative events (blue). (h)-(i): 3D point clouds of the LiDAR. The grid size is 10m.

extrinsics are optimized by minimizing the distance of all corresponding corners.

IV. DATASET DESCRIPTION

This section first introduces the overall features of different sequences, which stand as our basic criteria for data collection. Details are then described, including the ground truth estimation method and dataset format.

A. Sequences

The collected sequences should cover various environments, lighting conditions, motion patterns, dynamic objects, etc. We categorize major characteristics of our collected sequences as follows:

1) **Location:** Environmental locations are divided into indoors and outdoors. GPS signal is available but sometimes unstable in outdoor environments.

2) **Structure:** Structured environments can mainly be explained using geometric primitives (e.g., offices or buildings), while semi-structured environments have both geometric and complex elements like trees and sundries. Scenarios like narrow corridors are structured but may cause state estimators.

3) **Lighting Condition:** Frame cameras are sensitive to external lighting conditions. Both weak and strong light may raise challenges to visual processing algorithms.

4) **Appearance:** Texture-rich scenes facilitate visual algorithms to extract stable features (e.g., points and lines), while textureless may negatively affect the performance. Also, many events are triggered in texture-rich scenes.

5) **Motion Pattern:** Slow, normal, and fast motion may be performed. Regarding mounted platforms, the handheld device performs arbitrary 6-DoF and jerky motions, the device installed on a gimbal stabilizer conducts 6-DoF but stable motions, the quadruped robot mostly performs planar but jerky motions. In contrast, the vehicle performs planar movements at a constant speed.

6) **Object Motion:** In dynamic environments, several elements are moving while the data are captured. The more time of the data capture, the more deformed the elements will be (e.g., pedestrians or cars) [1]. In contrast, moving objects are few in static environments.

Table [III] summaries key features of each sequence, Fig. 2 shows several scene pictures, and Fig. 3 illustrates sample sensor data. The motion capture room is abbreviated as the MCR in the following sensors.

B. Groundtruth Generation

Most sequences provide ground-truth poses for algorithm evaluation. In several indoor scenes, we also provide ground-truth maps of surrounding environments. The ground truth generation is detailed as follows:

- **Ground-truth maps:** In small- or middle-scale environments, we use the Leica BLK360 laser scanner to record the structure’s high-resolution colorized 3D dense map with millimeter accuracy from multiple locations. Fig. 4 visualizes three examples.

- **Ground-truth poses:** In the motion capture room, we use the OptiTrack system to measure the pose of the center of reflective balls at 120Hz with millimeter accuracy. The OptiTrack is directly connected with the same PC to record poses to minimize the time latency. The extrinsics from the balls’ center to the body frame of the sensor rig are solved by the hand-eye calibration approach. In middle-scale environments that are covered by the ground-truth maps, we employ the NDT-based 6-DoF localization [18] to estimate LiDAR’s poses in a prior map as the ground-truth trajectory. In outdoor environments, we fuse the RTK GPS signal with LiDAR-inertial measurements to obtain accuracy trajectories based on the LIO-SAM [19].

C. Data Format and Post-Processing

Data were collected in the ROS environment. We provide both ROS bags and individual data files for better usage:

1) `env.bag` is the raw rosbag obtained from the data collection process. It can be parsed using ROS tools.

2) `env_ref.bag` is the refined rosbag where sensor data are post-processed with below steps.
TABLE III
SOME STATISTICS AND FEATURES OF EACH SEQUENCE

| Platform       | Sequence     | T[s] | D[m] | ||(|τ|)|| (m/s) | Location       | Structure       | Lighting | Texture | Motion      | Object   | GT Pose | GT Map     |
|----------------|--------------|------|------|------|------------|----------------|----------------|----------|----------|-------------|----------|---------|------------|
| Handheld       | canteen_night| 290  | 270  | 0.93 | indoors    | structured     | weak            | rich      | 6-DoF    | static      | 6-DoF    | NDT     | Yes        |
|                | canteen_day  | 230  | 250  | 1.09 | indoors    | structured     | normal          | 6-DoF     | static    | 6-DoF      | NDT      | Yes     |            |
|                | garden_night | 280  | 265  | 0.94 | indoors    | structured     | weak            | rich      | 6-DoF    | static      | 6-DoF    | NDT     | Yes        |
|                | garden_day   | 170  | 173  | 1.02 | indoors    | structured     | normal          | rich       | 6-DoF    | static      | 6-DoF    | NDT     | Yes        |
|                | corridor_day | 572  | 669  | 1.17 | indoors    | structured     | weak            | less       | 6-DoF    | static      | 6-DoF    | NDT     | Yes        |
|                | escalator_day| 315  | 263  | 0.84 | indoors    | structured     | strong          | rich       | 6-DoF    | height changes | dynamic | 6-DoF    | Yes        |
|                | building_day | 599  | 666  | 1.11 | indoors    | structured     | normal          | 6-DoF     | dynamic   | 6-DoF      | NDT      | Yes     |            |
| MCR_slow       | 48           | 50   | 1.03 | indoors | semi-structured | normal          | 6-DoF          | jerky     | static    | OptiTrack   | Yes      |         |            |
| MCR_normal     | 45           | 52   | 1.26 | indoors | semi-structured | normal          | 6-DoF          | jerky     | static    | OptiTrack   | Yes      |         |            |
| MCR_fast       | 34           | 59   | 1.76 | indoors | semi-structured | normal          | 6-DoF          | jerky     | static    | OptiTrack   | Yes      |         |            |
| MCR_slow_00    | 147          | 26   | 0.18 | indoors | semi-structured | normal          | planar         | jerky     | static    | OptiTrack   | Yes      |         |            |
| MCR_slow_01    | 127          | 28   | 0.28 | indoors | semi-structured | normal          | planar         | jerky     | static    | OptiTrack   | Yes      |         |            |
| Quad Cheetah    | MCR_normal_00 | 103  | 48   | 0.54 | indoors    | semi-structured | normal          | planar     | static    | OptiTrack   | Yes      |         |            |
| Robot          | MCR_normal_01 | 95   | 43   | 0.52 | indoors    | semi-structured | normal          | planar     | static    | OptiTrack   | Yes      |         |            |
|                | MCR_fast_00   | 99   | 48   | 0.56 | indoors    | semi-structured | normal          | planar     | static    | OptiTrack   | Yes      |         |            |
|                | MCR_fast_01   | 121  | 90   | 0.83 | indoors    | semi-structured | normal          | planar     | static    | OptiTrack   | Yes      |         |            |
| Apollo Robot    | MCR_normal    | 37   | 116  | 0.52 | indoors    | semi-structured | rich            | planar     | dynamic   | SLAM        | No       |         |            |

T: Total time. D: Total distance traveled. MCR: motion capture room. ||(|τ|)||: Mean linear velocity.

TABLE IV
LOCALIZATION ACCURACY.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Sequence</th>
<th>VINS-Fusion (LC)</th>
<th>A-LIO-LOAM</th>
<th>LIO-Mapping</th>
<th>LIO-SAM</th>
<th>FAST-LIO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handheld</td>
<td>canteen_night</td>
<td>0.409</td>
<td>0.067</td>
<td>0.097</td>
<td>0.063</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>canteen_day</td>
<td>0.691</td>
<td>0.057</td>
<td>0.088</td>
<td>0.053</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>garden_night</td>
<td>0.328</td>
<td>0.567</td>
<td>0.242</td>
<td>0.254</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>garden_day</td>
<td>0.518</td>
<td>0.528</td>
<td>0.097</td>
<td>0.069</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>corridor_day</td>
<td>1.807</td>
<td>0.416</td>
<td>1.755</td>
<td>0.594</td>
<td>1.563</td>
</tr>
<tr>
<td></td>
<td>escalator_day</td>
<td>2.127</td>
<td>0.981</td>
<td>0.346</td>
<td>0.207</td>
<td>4.193</td>
</tr>
<tr>
<td></td>
<td>building_day</td>
<td>12.861</td>
<td>5.380</td>
<td>0.916</td>
<td>0.222</td>
<td>0.146</td>
</tr>
<tr>
<td>MCR_slow</td>
<td>×</td>
<td>0.087</td>
<td>0.042</td>
<td>0.063</td>
<td>0.114</td>
<td></td>
</tr>
<tr>
<td>MCR_normal</td>
<td>0.168</td>
<td>0.328</td>
<td>0.052</td>
<td>0.082</td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td>MCR_fast</td>
<td>×</td>
<td>0.416</td>
<td>0.099</td>
<td>0.117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apollo Robot</td>
<td>MCR_slow_00</td>
<td>0.096</td>
<td>0.120</td>
<td>0.032</td>
<td>0.023</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>MCR_slow_01</td>
<td>0.081</td>
<td>0.054</td>
<td>0.030</td>
<td>0.030</td>
<td>0.051</td>
</tr>
<tr>
<td>Quad Robot</td>
<td>MCR_normal_00</td>
<td>0.094</td>
<td>0.492</td>
<td>0.093</td>
<td>0.042</td>
<td>0.127</td>
</tr>
<tr>
<td>Robot</td>
<td>MCR_normal_01</td>
<td>0.086</td>
<td>0.635</td>
<td>0.390</td>
<td>0.040</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>MCR_fast_00</td>
<td>0.264</td>
<td>4.601</td>
<td>2.405</td>
<td>0.052</td>
<td>0.408</td>
</tr>
<tr>
<td></td>
<td>MCR_fast_01</td>
<td>0.130</td>
<td>8.264</td>
<td>2.210</td>
<td>0.066</td>
<td>1.495</td>
</tr>
</tbody>
</table>

We have three steps to post-process the raw data to generate the env_reg.txt: 1) caused by unperfect IMUs (like the MPU6050), several missing measurements are linearly interpolated; 2) poses provided by the motion capture system are transformed into the body frame with the hand-eye calibration results; and 3) event packages are republished around 1000 Hz for several event-based algorithms [20].

Unrectified RGB images are stored. Events are stored with timestamps, pixel locations, and polarity. IMU measurements are also stored with timestamps, gyroscope measurements, accelerometer measurements, and covariances. Calibration parameters are stored in yaml files.

V. EXPERIMENT

As one of the applications, we can use this dataset to benchmark SOTA SLAM systems. Here, we evaluate several open-source systems with different sensor combinations and methodologies: VINS-Fusion (IMU+stereo frame cameras) [21], ESVO (stereo event cameras) [20], A-LOAM (LiDAR-only) [22], LIO-Mapping (IMU+LiDAR) [23], LIO-SAM (IMU+LiDAR) [19], and FAST-LIO2 (IMU+LiDAR) [24]. Their data loaders are modified to fit our dataset format and also released. We calculate the mean absolute trajectory error (ATE) of estimated trajectories w.r.t. the ground truth. For LiDAR-based systems, we also report the mapping accuracy on two sequences by calculating the mean point-to-point error of algorithms' maps w.r.t. the ground-truth maps.

The quantitative localization results are reported in Table IV. “LC” indicates that the loop closure module is used. “×” means that algorithms fail to finish the sequence. ESVO’s results are not shown here since it cannot finish all sequences. It requires events to be continuously triggered to generate reliable time surface maps for camera tracking. But all these sequences contain textureless scenarios or static motion. Its immediate results on mapping and tracking are shown in the dataset website. VINS-Fusion and FAST-LIO2 fail in some cases since they cannot initialize well at the beginning of the sequence. Without the aid of the IMU, A-LOAM cannot handle jerky and rapid motion and thus performs poorly on two MCR sequences and all sequences on the quadraped robot. Although FAST-LIO2 has a superior real-time performance based on the filter-based state estimator and efficient tree structure, it sometimes has unreliable results on several sequences. Surprisingly, LIO-SAM performs well on all quadraped robot-based sequences, even at large rotated and fast motion. The corridor_day sequence is challenging to all methods, where the scene is textureless and structureless.

We also evaluate the mapping quality of A-LOAM and LIO-SAM on the corridor_day and garden_day sequences. The distance map is Fig. 6. The mean distance is 0.938m and 0.597m respectively. Especially for the corridor mapping, A-LOAM’s map has a large drift on the z-axis.
VI. CONCLUSION

This paper presented the FusionPortable benchmark, a multi-sensor dataset from diverse campus scenes on various platforms. We advanced the self-contained and plug-and-play multi-sensor rig that significantly enhances the perception capability of mobile robots. With the release of this dataset, we intended to challenge current SLAM approaches and encouraged future research. As the future work, we plan to extend this dataset beyond the campus-scale environments.

REFERENCES