MARS-LVIG dataset: A multi-sensor aerial robots SLAM dataset for LiDAR-visual-inertial-GNSS fusion

Haotian Li[1] , **Yuying Zou[1]**, **Nan Chen[1]** , **Jiarong Lin[1]**, **Xiyuan Liu[1]** , **Wei Xu[1]**, **Chunran Zheng[1]**, **Rundong Li[1]**, **Dongjiao He[1]** , **Fanze Kong[1]**, **Yixi Cai[1]** , **Zheng Liu[1]**, **Shunbo Zhou[2]**, **Kaiwen Xue[2]** and **Fu Zhang[1]**

Abstract

In recent years, advancements in Light Detection and Ranging (LiDAR) technology have made 3D LiDAR sensors more compact, lightweight, and affordable. This progress has spurred interest in integrating LiDAR with sensors such as Inertial Measurement Units (IMUs) and cameras for Simultaneous Localization and Mapping (SLAM) research. Public datasets covering different scenarios, platforms, and viewpoints are crucial for multi-sensor fusion SLAM studies, yet most focus on handheld or vehicle-mounted devices with front or 360-degree views. Data from aerial vehicles with downward-looking views is scarce, existing relevant datasets usually feature low altitudes and are mostly limited to small campus environments. To fill this gap, we introduce the Multi-sensor Aerial Robots SLAM dataset (MARS-LVIG dataset), providing unique aerial downward-looking LiDAR-Visual-Inertial-GNSS data with viewpoints from altitudes between 80 m and 130 m. The dataset not only offers new aspects to test and evaluate existing SLAM algorithms, but also brings new challenges which can facilitate researches and developments of more advanced SLAM algorithms. The MARS-LVIG dataset contains 21 sequences, acquired across diversified large-area environments including an aero-model airfield, an island, a rural town, and a valley. Within these sequences, the UAV has speeds varying from 3 m/s to 12 m/s, a scanning area reaching up to 577,000 m^2, and the max path length of 7.148 km in a single flight. This dataset encapsulates data collected by a lightweight, hardware-synchronized sensor package that includes a solid-state 3D LiDAR, a global-shutter RGB camera, IMUs, and a raw message receiver of the Global Navigation Satellite System (GNSS). For algorithm evaluation, this dataset releases ground truth of both localization and mapping, which are acquired by on-board Real-time Kinematic (RTK) and DJI L1 (post-processed by its supporting software DJI Terra), respectively. The dataset can be downloaded from: https://mars.hku.hk/dataset.html.

Keywords

Dataset, aerial robots, multi-sensor fusion, LiDAR, camera, Simultaneous Localization and Mapping, Global Navigation Satellite System, Inertial Measurement Unit

1. Introduction

Simultaneous Localization and Mapping (SLAM) technologies have played a crucial role in the development of robotic autonomous systems in recent years. Precise and robust state estimation constitutes a critical aspect of robotic autonomy. Hence, the solution of combining SLAM and mobile robots became a popular topic. As for mobile robots, the integration of multi-sensor data for SLAM inherently surpasses the use of a single sensor in complex environments (Xu et al., 2022; Zhu et al., 2022). The advancements in Light Detection and Ranging (LiDAR) technology have resulted in 3D LiDAR sensors becoming more cost-effective, lightweight, and compact. The fusion between Light Detection and Ranging (LiDAR), camera, Inertial Measurement Units (IMUs), and Global Navigation

Satellite System (GNSS) raw measurements has gained significant attention and shown high potential in mobile robot navigation and mobile mapping (Xu and Zhang, 2021; Liu and Zhang, 2021b; Yuan et al., 2023; Cao et al., 2022). This approach prevents the degradation of individual

[1]Department of Mechanical Engineering, The University of Hong Kong, Hong Kong
[2]Huawei Cloud Computing Technical Innovation Dept., Huawei Cloud Computing Technologies Co., Ltd., Gui'an, China

Corresponding author:
Fu Zhang, Mechatronics and Robotic Systems Laboratory, Department of Mechanical Engineering, The University of Hong Kong, HW 7-18, Pokfulam, Hong Kong.
Email: fuzhang@hku.hk

sensors, ensuring accurate state estimation for robust navigation (Liu and Zhang, 2021a; Lin et al., 2020, 2021; He et al., 2023). Furthermore, the combination of LiDAR, camera, and IMU has distinct advantages in handling continuous movements during mobile mapping and can build high-accuracy, high-resolution, and full RGB-colored 3D models of the traveled environment (Lin and Zhang, 2022; Zheng et al., 2022; Qin et al., 2018). Multi-sensor SLAM has the potential to benefit a variety of applications, including geological surveying, construction site reconstruction, and aerial mapping (Liu et al., 2023a, 2023b).

To evaluate the real-world feasibility of SLAM algorithms, datasets that contain realistic scenarios collected by advanced devices are necessary. Meanwhile, the majority of existing aerial datasets primarily concentrate on low altitude, indoor environments (e.g., school campuses) or are absent of LiDAR data, which are not suitable for large-scale and long-duration mapping applications, such as aerial surveying and mapping at high altitudes. In order to augment the variety encompassed within aerial datasets, we introduce a multi-sensor aerial robots SLAM dataset for LiDAR-Visual-Inertial-GNSS fusion (MARS-LVIG dataset), which is the first dataset to use a downward-looking method with both LiDAR and camera sensors from a high altitude (higher than 80 m), primarily for SLAM applications. The downward-looking scanning method brings about significant challenges for SLAM, considering that LiDAR SLAM algorithms are prone to degeneration and the visual SLAM algorithms can easily fail at texture-less flats when facing the ground, while currently providing great value for applications such as aerial survey and mapping. To ensure the diversity of scenarios within our dataset, we collect data across various natural and urban environments. For aerial datasets, particularly in large environments, the collection of ground truth data presents a substantial challenge. In this dataset, we utilized the Real-time Kinematic (RTK) receiver and the DJI L1 to provide ground truth for localization and mapping, respectively.

In summary, our contribution is a multi-sensor aerial robot SLAM dataset, which has three unique features detailed below:

- This dataset includes unique downward-looking data collected by a solid-state 3D LiDAR, a global-shutter RGB camera, IMUs, and a raw message receiver of GNSS from high altitude between 80 m and 130 m;
- This dataset is collected across diversified environments, including an aero-model airfield, an island, a rural town, and a valley. The scanning area covers from 94,000 m^2 to 577,000 m^2 with a maximum path length of 7.148 km;
- This dataset provides both position and mapping ground truth at centimeter-level accuracy using RTK and DJI L1, respectively. These ground truth measurements are provided for the evaluation of SLAM algorithms.

The structure of this paper is as follows. Section 2 discusses the detailed comparison between our dataset and existing datasets. The sensor configuration, as depicted in Figure 1, encompassing a 3D LiDAR, an RGB camera, a raw message receiver of GNSS, an RTK antenna, and a DJI L1, is explicated in Section 3. Section 4 presents the characteristics of our dataset, mainly focusing on the features of different scenarios. We also conduct verification of the data and present the results in Section 5.

2. Related works

Tables 1 and 2 list a selection of recent datasets, incorporating both those collected by terrestrial LiDAR devices and those collected by aerial platforms. There are several datasets with LiDAR collected by ground vehicles such as cars (Geiger et al., 2012; Blanco-Claraco et al., 2014; Jeong et al., 2019) and wheeled robots (Carlevaris-Bianco et al., 2016; Shi et al., 2020; Yin et al., 2021), by handheld platforms (Ramezani et al., 2020; Zhang et al., 2021; Helmberger et al., 2022), and by multi-platform (Jiao et al., 2022).

As for car-platform datasets, the most representative dataset should be the KITTI dataset (Geiger et al., 2012), which is one of the widely used datasets in computer vision research and autonomous driving research. It provides raw data collected by a Velodyne HDL-64E LiDAR, along with two color and two grayscale PointGrey Flea2 video cameras. This sensor setup effectively satisfies the requirements of both visual SLAM and laser SLAM simultaneously. For classic vehicular datasets, including Malaga (Blanco-Claraco et al., 2014) and Complex Urban (Jeong et al., 2019), the selection and arrangement of sensors tailored for autonomous driving demonstrate exemplary design principles in the context of such scenarios.

In the case of wheeled robots, several notable datasets, such as NCLT (Carlevaris-Bianco et al., 2016), OpenLORIS (Shi et al., 2020), Fusion Portable (Jiao et al., 2022), and M2DGR (Yin et al., 2021), are equipped with a wide range of sensors, including LiDARs and cameras. Fusion Portable dataset (Jiao et al., 2022) incorporates the most recent sensor technologies, such as event cameras. Although the variety of sensors has significantly increased, the operational range of wheeled robots remains confined to the ground. Moreover, a majority of these datasets primarily concentrate on data collection within relatively small areas, such as university campuses. Consequently, the coverage of SLAM scenarios remains inadequate. Compared with the wheeled robots, the payload capacity and moving range of handheld devices (Helmberger et al., 2022; Zhang et al., 2021; Ramezani et al., 2020) is even less, and hence these limitations become more significant.

Furthermore, the moving range of ground vehicles and handheld devices is usually constrained by road and terrain. Specifically, the movement observed in these datasets often lacks vertical translation and rapid rotation (especially for pitch and roll). Therefore, some aerial datasets collected by unmanned aerial vehicles (UAVs) are released, as summarized in Table 2.

Due to the limited payload capacity of UAVs, most aerial datasets are focused on visual and inertial sensors (Burri et al., 2016; Sun et al., 2018; Majdik et al., 2017; Antonini et al., 2018; Delmerico et al., 2019). The EuRoC dataset (Burri et al., 2016) is an indoor UAV dataset for providing visual-inertial localization algorithms on real flight data. UPenn Fast Flight (Sun et al., 2018) aims to present visual-inertial odometry for fast autonomous flight and a UAV
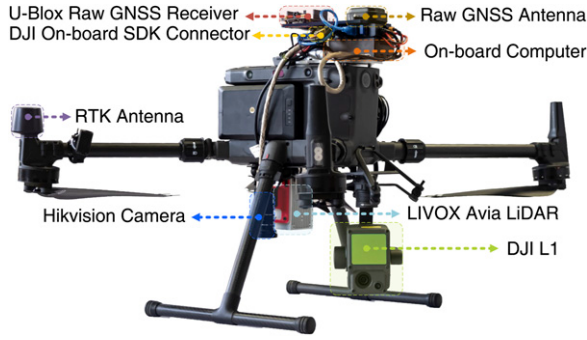

Figure 1. Full view of the data collection device on DJI M300 RTK. For SLAM data acquisition, the Livox Avia LiDAR collects point cloud and IMU data while the Hikvision camera collects RGB images. The LiDAR and camera are installed rigidly on a board, which is then attached to the UAV via rubber-damping balls. The U-Blox receiver and its antenna are set at the top of the UAV to collect raw GNSS measurements. Regarding ground truth data, the UAV's built-in RTK provides position ground truth with its antenna located on the UAV's side. The DJI L1, connected to the UAV through the gimbal, collects the mapping ground truth.

dataset with fast speed is released. As for the Zurich Urban MAV dataset (Majdik et al., 2017), it offers comprehensive real-world datasets in a city. The UZH-FPV dataset (Delmerico et al., 2019) and Blackbird dataset (Antonini et al., 2018) are aiming at providing UAV datasets with both fast speed and agile trajectories. There are also datasets providing valued data and algorithms for downward-looking images from high altitudes like the CLOUD dataset (Patel et al., 2020) and WildNav dataset (Gurgu et al., 2022). Specifically, the CLOUD dataset (Patel et al., 2020) provides a special large-scale dataset across various environments with different light conditions, offering excellent support for aerial visual SLAM research.

Notably, NTU-VIRAL dataset (Nguyen et al., 2022) and GRACO dataset (Zhu et al., 2023) have made pioneering contributions in presenting data of both LiDAR and camera from aerial perspectives. However, certain gaps persist in those datasets with LiDAR data, including low altitude (below 40 m), lack of natural surroundings, and small collection areas as they are all collected on university campuses. At the same time, they only have low-resolution grayscale images and lack of ground-facing data. These gaps have prevented SLAM algorithms from being tested on large-scale high-altitude sequences that are commonly found in aerial mapping applications. There are some datasets collected from aerial vehicles with LiDAR data for environmental and natural research purposes, including the HiWATER dataset (Li et al., 2017), the DALES dataset (Varney and Asari, 2020), and the 3DEP dataset (US Geological

Table 1. Comparison of related ground datasets.

Dataset	Sensors			Ground truth	Platform
	LiDAR	Camera	IMU (Hz)		
KITTI (Geiger et al., 2012)	Velodyne HDL-64E	1392 × 512 (RGB) 1392 × 512 (gray)	10	GPS/INS/RTK	Car
Complex urban (Jeong et al., 2019)	SICK LMS-151 Velodyne VLP-16	1280 × 560 (RGB)	200	GPS	Car
OpenLORIS (Shi et al., 2020)	Hokuyo UTM-30LX	848 × 480 (RGB) 848 × 800 (gray)	400/250/ 200/ 62.5	MoCap	Wheeled robot
M2DGR (Yin et al., 2021)	Velodyne VLP-32C	1280 × 1024 (RGB) 640 × 480 (RGB)	150	GPS/RTK/INS/ Total station	Wheeled robot
The newer college dataset (Ramezani et al., 2020)	Ouster-64	848 × 480 (RGB)	650/100	6DOF ICP	Handheld platform
The newer college dataset extension (Zhang et al., 2021)	Ouster-128	720 × 540 (gray)	200/100	Laser/6DOF ICP	Handheld platform
Hilti (Helmberger et al., 2022)	Livox MID70 Ouster-64	720 × 540 (gray)	800/200/ 100	Laser/MoCap/ Total station	Handheld platform
FusionPortable (Jiao et al., 2022)	Ouster-128	1024 × 768 (RGB)	200	Laser/MoCap/ GPS/RTK	Handheld platform/Wheeled robot/quadruped robot

Table 2. Comparison of related aerial datasets.

| Dataset | Sensors | | Raw GNSS | IMU (Hz) | Ground truth | UAV altitude(m) | Scan area(m^2) | Environments |
	LiDAR	Camera						
EuRoC (Burri et al., 2016)	N/A	752 × 480 (gray)	N/A	200	Laser/MoCap/Total station	≤10	≤200	Indoor: laboratory
UPenn fast flight (Sun et al., 2018)	N/A	960 × 800 (RGB)	N/A	200	GPS	≤10	≤40,000	Indoor: warehouse Outdoor: runway, grassland
Zurich Urban MAV (Majdik et al., 2017)	N/A	1920 × 1080 (RGB)	N/A	10	Photogrammetric	≤15	≤150,000	Outdoor: urban
Blackbird (Antonini et al., 2018)	N/A	Synthetic cameras	N/A	100	MoCap	≤5	≤50	Indoor: apartment, museum, subway
UZH-FPV (Delmerico et al., 2019)	N/A	640 × 480 (gray)	N/A	1000/500	Total station	≤10	≤4800	Indoor: warehouse outdoor: grassland
CLOUD (Patel et al., 2020)	N/A	512 × 384 (gray)	N/A	400	RTK	≤50	≤62,500	Outdoor: grassland, urban, campus
NTU-VIRAL (Nguyen et al., 2022)	Ouster-16	752 × 480 (gray)	N/A	385	Total station	≤10	≤5000	Outdoor: campus
WildNav (Gurgu et al., 2022)	N/A	1920 × 1080 (RGB)	N/A	N/A	GPS	≤120	≤265,000	Outdoor: grassland, forest
GRACO (Zhu et al., 2023)	Velodyne VLP-16	1600 × 1100 (gray)	N/A	125	GPS/INS	≤40	≤60,000	Outdoor: campus
HKU-MaRS dataset	**Livox Avia**	**2448 × 2048 (RGB)**	**ZED-F9P**	**200**	**Position: RTK Mapping: DJI L1**	**80 –130**	**94,000 - 577,000**	**Outdoor:** airfield, island, town, valley

Survey, 2019). They only provide aerial mapping results point cloud like terrain elevation maps instead of raw data from the sensors, making them not suitable for SLAM algorithm development.

Our dataset aims to address the above-mentioned gaps. Specifically, unlike previous aerial SLAM datasets which collect only visual-inertial data, including EuRoC (Burri et al., 2016), UPenn Fast Flight (Sun et al., 2018), Zurich Urban MAV dataset (Majdik et al., 2017), UZH-FPV dataset (Delmerico et al., 2019), Blackbird dataset (Antonini et al., 2018), CLOUD dataset (Patel et al., 2020), and WildNav dataset (Gurgu et al., 2022), we collect hardware-synchronized LiDAR, camera, IMU, and GNSS data for multi-sensor fusion. Different from previous aerial SLAM datasets which contain LiDAR data but were collected in small-scale school campuses at low altitudes, including NTU-VIRAL dataset (Nguyen et al., 2022) and GRACO dataset (Zhu et al., 2023), our dataset includes 21 sequences, captured across a variety of environments including an aero-model airfield, an island, a rural town, and a valley at an altitude of higher than 80 m. In these sequences, the established flight speeds range from 3 m/s to 12 m/s, covering an area of from 94,000 m^2 to 577,000 m^2 in a single flight. Moreover, our sensor package, configured for downward-looking orientation as shown in Figure 2, imposes considerable challenges for SLAM due to the downward-looking viewpoints and large-scale diversified environments. Meanwhile, it holds immense utility for applications such as aerial surveying and mapping, particularly under high-altitude scanning conditions. The distinctive features of our dataset, such as the sensor setup, the higher flight altitudes, the use of a downward-looking scanning method, and unique scenario diversity, set this dataset apart from previous SLAM datasets. Finally, when compared with datasets including the HiWATER dataset (Li et al., 2017), the DALES dataset (Varney and Asari, 2020), and the 3DEP dataset (US Geological Survey, 2019), which are primarily collected for non-SLAM researches, our dataset provides raw Li-DAR, camera, IMU, and GNSS measurements as well as position and mapping ground truth, which can be used for
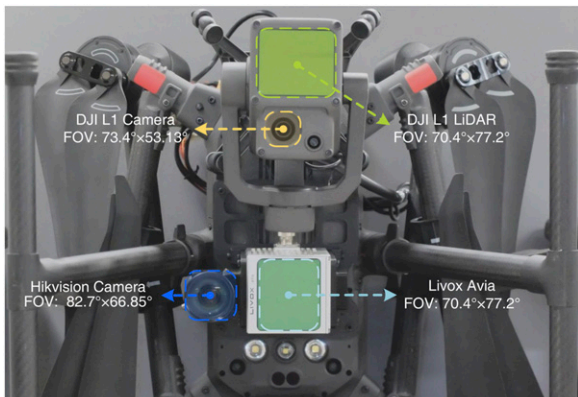

Figure 2. Bottom view of the sensors.

developing and evaluating multi-sensor SLAM algorithms. The complete setup will be described in the following section.

3. System overview

3.1. Sensor setup

A DJI M300 RTK quadrotor UAV[1] is used to carry the customized SLAM sensor suite, including a 3D LiDAR (with built-in IMU), an RGB camera, and a GNSS receiver. The DJI M300 RTK has its own RTK receiver to provide the position only ground truth, and a high-precision aerial mapping system (DJI L1) carried on a gimbal to provide mapping ground truth. All the data from the 3D LiDAR, its internal IMU, and RGB camera are hardware-synchronized and timestamped by the Coordinated Universal Time (UTC). The data from the sensors, except DJI L1, are recorded by a DJI Manifold 2-C on-board computer running with the Robot Operating System (ROS) to facilitate the use of SLAM algorithms. The structure of the devices is shown in Figures 1 and 2. All the Computer-Aided Design (CAD) files of our mechanical design and the sensor drivers are open-sourced on the Web site. A detailed description of the sensors is listed below.

(i) *3D LiDAR* In this work, a Livox Avia LiDAR[2] is placed vertically with its field of view (FoV) facing the ground to collect aerial view data. It is a solid-state 3D LiDAR that operates in a triple-echo non-repetitive scanning mode to obtain the highest density point cloud data at 10 Hz and 72,000 points per frame. The detection range of the LiDAR is up to 190 m @10% reflectivity and 320 m @80% reflectivity in outdoor environments. The LiDAR is connected to the computer through its supporting accessories (Livox Hub). The LiDAR is hardware-synchronized through the Pulse Per Second (PPS) signal triggered by Global Positioning System (GPS). All the point cloud data, including timestamps, the ordinal number of echoes, and the reflectivity are recorded in ROS bags.

(ii) *IMU* In this work, LiDAR's internal IMU (BMI088[3]) outputs angular rate and acceleration measurements at a rate of 200 Hz. The BMI088 has a bias stability of less than 2°/h and a low-temperature coefficient of offset (TCO) below 15 mdps/K. The accelerometer features a low TCO of 0.2 mg/K and low spectral noise of 230 μg/sqrt (Hz). The data of the IMU is sent to the on-board computer through the Livox Hub and recorded in the ROS bags.

(iii) *RGB Camera* A Hikvision CA-050-11UC[4] global-shutter RGB camera equipped with a 5 mm lens is placed under the UAV next to the LiDAR. The camera is connected to the on-board computer by Universal Serial Bus (USB) 3.0 port and is hardware-synchronized with LiDAR and captures RGB

images at 10 Hz. All the camera data, including images and timestamps, are recorded in the ROS bags.

(iv) *Raw GNSS measurement Receiver* The u-blox ZED-F9P receiver[5] is used to receive raw GNSS measurements. The received data is transmitted to the on-board computer via a serial port. All the collected data is recorded in the ROS bags along with others.

(v) *RTK Receiver* The position ground truth provided by the RTK receiver from the DJI M300 RTK UAV is used for the dataset, achieving an accuracy of 1 cm + 1 ppm × D horizontally, and 1.5 cm + 1 ppm × D vertically[1] (D represents the distance between the RTK receiver and the base station. During the data collection of this dataset, the distance is lower than 5 km). The messages including the position information obtained by RTK positioning, the number of the GPS satellites, timing, signal quality, and other relevant data are sent to the on-board computer at 5 Hz by the DJI On-board Software Development Kit (OSDK) connector and recorded in the ROS bags.

(vi) *DJI L1* The DJI Zenmuse L1[6] is a high-precision commercial aerial mapping system, designed for the DJI M300 RTK UAV. The DJI L1 consists of a LiDAR with triple-echo returns, a high-accuracy IMU, and a 1-inch Complementary Metal-Oxide Semiconductor (CMOS) sensor. The DJI L1 is used on our dataset for simultaneous data acquisition with the same orientation as our sensing devices. The data of DJI L1 is post-processed by the DJI Terra[7] software, an advanced RTK-based aerial mapping software for DJI L1, to obtain 3D point clouds at high-accuracy of 10 cm horizontally and 5 cm vertically[6] as a mapping ground truth for the evaluation of SLAM algorithms.

3.2. Sensor calibration

In this dataset, the extrinsic parameters among the LiDAR, the camera, the antenna phase centers of RTK and the raw GNSS receiver are obtained from and released with the CAD design files (see Figure 3). Meanwhile, since the LiDAR's internal IMU sensor we used is rigidly connected to the LiDAR, the extrinsic parameter between the IMU and the LiDAR is calibrated and obtained by the manufacturer and is considered to be fixed during the data collection. The
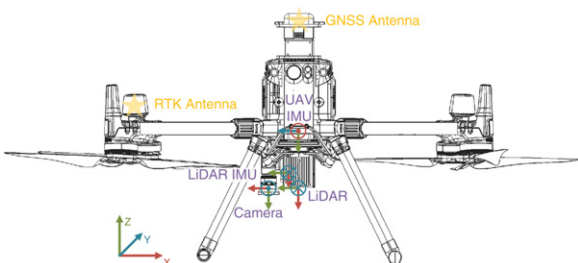

Figure 3. Relative positions and coordinates of sensors.

intrinsic parameter of the camera is calibrated by the chessboard method (Zhang, 2000). The extrinsic parameters between LiDAR and the camera are critical for multi-sensor fusion and point cloud colorization. Hence, to achieve a more accurate extrinsic parameter than that provided by CAD files, we also employed a targetless calibration method developed by (Liu et al., 2022; Yuan et al., 2021) (see Figure 4). Such a method is suitable for our proposed scenarios where calibration is needed before each data collection while the calibration target is not feasible to carry. In each calibration, the initial extrinsic parameter is obtained from the CAD model. As illustrated in Figure 4 (a)–(b), the image edge features and LiDAR depth-continuous edge features are extracted separately, then the extrinsic is optimized by minimizing the edge features' re-projection errors. Moreover, this method remains efficient where the whole calibration completes within a minute while achieving 0.1° and 1-cm-level of accuracy (see Figure 4 (d)–(g)). Both the results of calibration using our method and the raw data used for calibration are available on the dataset Web site.

3.3. Time synchronization

This dataset uses the UAV's on-board RTK for sensor hardware synchronization. Specifically, the recommended minimum specific GPS/Transit data (GPRMC) received by the RTK receiver is transmitted from the DJI OSDK connector to the on-board computer via serial port. At the same time, the PPS signal received by the RTK receiver is sent to an ATmega microchip, which is set to hardware interrupt mode and has the capability to generate synchronized 1-Hz and 10-Hz pulse signals as shown in Figure 5. The 1-Hz pulse signal is sent to the LiDAR in order to maintain GPS synchronization. Simultaneously, the corresponding GPRMC information is sent to the on-board computer. The Livox ROS driver combines the GPRMC information with the matched point clouds triggered by the PPS signal and assigns their timestamps. The 10-Hz pulse signal is sent to trigger the camera for capturing images at a rate of 10 Hz and it also uses the GPRMC to assign the timestamps. As the 10-Hz point cloud message and the 200-Hz IMU message are already synchronized internally by LiDAR itself, thus the timestamp of the point cloud message, IMU message, and image message are all aligned with the UTC time of GPS. Finally, the raw GNSS receiver is also synchronized with the UTC time of GPS.

3.4. Data format

The information contained within the ROS bag, along with their corresponding message types and frequencies, are listed in Table 3. As for the LiDAR sensor, it is important to note that, in order to obtain raw data of the LiDAR, the LiDAR's ROS message type is *CustomMsg*, whose content is defined by the LiDAR manufacturer (Livox). To balance
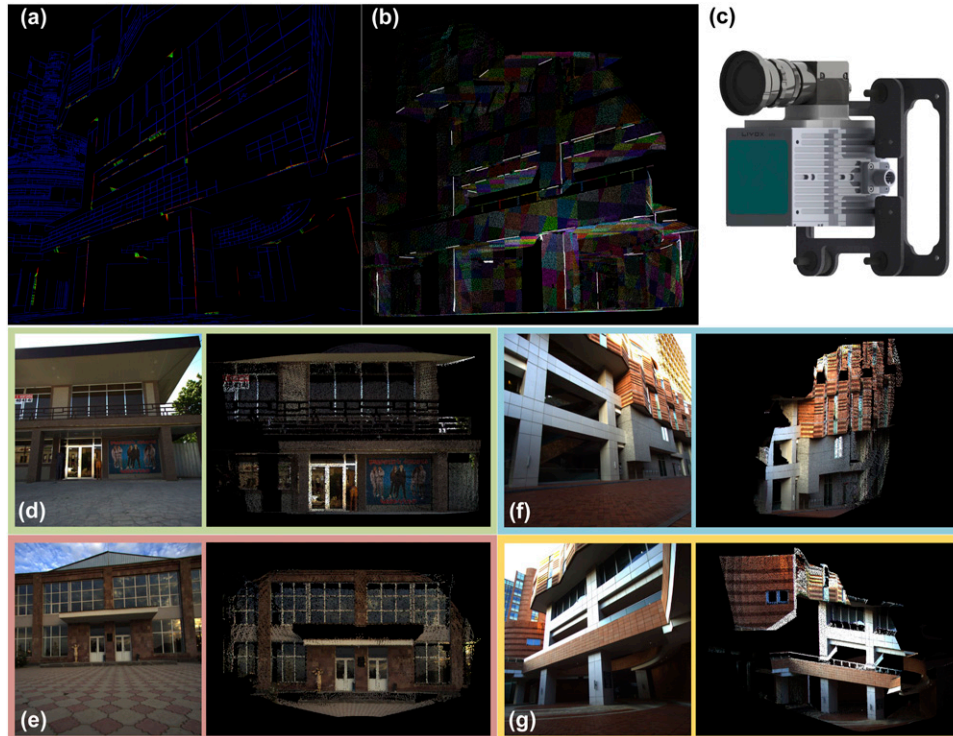
Figure 4. (a) Edge features correspondence matching. Blue edges are extracted from the image and the red edges are extracted from the LiDAR point cloud. The green lines indicate that two features are matched. (b) LiDAR depth-continuous edges (white lines) are extracted using adaptive voxelization. The LiDAR point cloud is first segmented into planes with adaptive sizes, then an edge feature is extracted by the intersection of adjacent two planes. (c) CAD model of our camera-LiDAR sensor suite. (d)–(g) The calibration scene in Armenia (d)–(e) and Hong Kong (f)–(g) and their corresponding colorized point clouds using the calibrated extrinsic parameters.
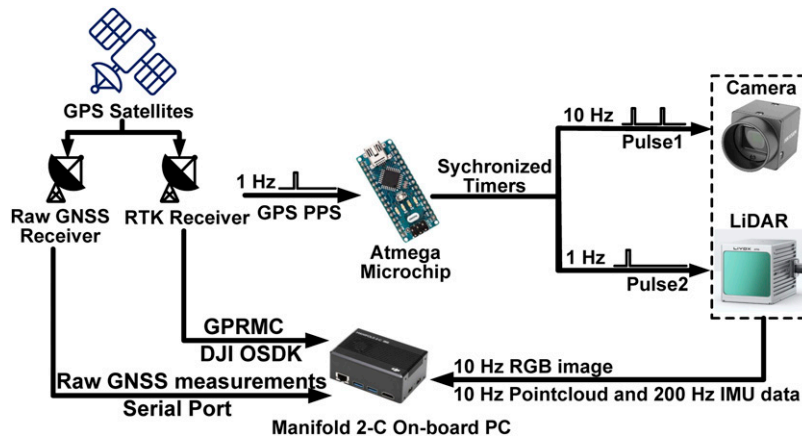

Figure 5. Diagrammatic illustration of the process involved in synchronizing the on-board sensor systems.

image quality and file size for large scene data packets, the camera's image data message type is *CompressedImage*. To ensure a high-quality evaluation of ground truth, information regarding the RTK connection status, time-synchronized trigger message sources, and GPS satellite conditions have been recorded into the ROS bag as well. In terms of the raw GNSS message receiver, to facilitate its application in the field of SLAM, we also utilize the ROS data format to record the receiver's raw observational

information, including raw measurements, broadcast ephemeris, GNSS solutions, and broadcast ionospheric parameters. Among all ROS drivers, the LiDAR and RTK receiver drivers employ their manufacturer-released ROS drivers, while the camera and GNSS receiver ROS drivers are open-sourced along with our dataset.

The data from the DJI L1 was acquired using its own system and saved directly onto its flash card. Both original data files and reconstruction results including 3D point clouds processed

Table 3. Message information of sensors and devices. A dash ("-") signifies that there is no fixed frequency for these messages.

Type	Topic name	Message type	Rate (Hz)
Ground truth:	/dji_osdk_ros/rtk_connection_status	std_msgs/UInt8	5
	/dji_osdk_ros/rtk_info_position	std_msgs/UInt8	5
	/dji_osdk_ros/rtk_info_yaw	std_msgs/UInt8	5
	/dji_osdk_ros/rtk_position	sensor_msgs/NavSatFix	5
	/dji_osdk_ros/rtk_velocity	geometry_msgs/Vector3Stamped	5
	/dji_osdk_ros/rtk_yaw	std_msgs/Int16	5
	/dji_osdk_ros/time_sync_fc_time_utc	dji_osdk_ros/FCTimeInUTC	1
	/dji_osdk_ros/time_sync_gps_utc	dji_osdk_ros/GPSUTC	1
	/dji_osdk_ros/time_sync_nmea_msg	nmea_msgs/Sentence	25
	/dji_osdk_ros/time_sync_pps_source	std_msgs/String	1
Camera:	/left_camera/image/compressed	sensor_msgs/CompressedImage	10
IMU:	/livox/imu	sensor_msgs/Imu	200
LiDAR:	/livox/lidar	livox_ros_driver/CustomMsg	10
GNSS:	/ublox_driver/ephem	gnss_comm/GnssEphemMsg	-
	/ublox_driver/glo_ephem	gnss_comm/GnssGloEphemMsg	-
	/ublox_driver/iono_params	gnss_comm/StampedFloat64Array	-
	/ublox_driver/range_meas	gnss_comm/GnssMeasMsg	10
	/ublox_driver/receiver_lla	sensor_msgs/NavSatFix	10
	/ublox_driver/receiver_pvt	gnss_comm/GnssPVTSolnMsg	10
	/ublox_driver/time_pulse_info	gnss_comm/GnssTimePulseInfoMsg	1

by DJI Terra software (Version 3.6.7) are provided within this dataset. The data types are listed as follows:

```
DJI L1 data:

    Original data:
    *.CLC *.CLI *.CMI *.IMU *.LDR
    *.RTB *.RTK *.RTL *.RTS *.JPG

    Reconstruction Results:
    *.pcd *.las *.ply *.pnts *.s3mb
```

4. Dataset character

To increase the diversity of our dataset, we collected data from four distinct scenarios. The data is organized into five groups of sequences according to their scenarios and characters, namely, HKairport, HKisland, AMtown, AMvalley, and Featureless. Except for the Featureless sequence, data within a sequence is obtained at three different cruising speeds while following the same set of waypoints and a consistent altitude. Choosing three different cruising speeds aims to create differentiation in SLAM difficulty levels within the dataset. The choice of altitude for each sequence is determined by the specific scenario, ensuring that the altitude is suitable and challenging. The path planning configuration is carefully designed to make coordinated turns and maintain heading along the route. The routes are continuous without pauses at waypoints, ensuring that the trajectories are smooth and the data is uniformly distributed. The trajectories can also represent the flight patterns of both multi-rotor and fixed-wing aircraft. This approach makes the data applicable to a wide range of situations. Additionally,

considering the initialization requirements of certain algorithms, the UAV hovers at two positions before each flight, one is at the 30 m altitude after take-off, and the other is at the start position of the route. Detailed information about the path settings and ROS bags in each sequence is presented in Table 4. A visualization and illustration of waypoints related to our configuration for each scenario are presented in Figure 6.

4.1. Aero-model airfield

The data sequences named HKairport and Featureless in this scenario are obtained from the Hong Kong Model Engineering Club[8] in Yuen Long, situated at 22.4164°N, 114.0428°E, Hong Kong, China. The scenario features an outdoor 20-m by 200-m concrete runway designated for aero-model aircraft operation, as well as an 800-square-meter concrete helipad reserved for flying model helicopters, with the total surveyed area surpassing 154,000 square meters. During the acquisition of this scenario, the UAV maintained a consistent altitude of 80 m. The large site presented considerable challenges for LiDAR SLAM. Throughout the data collection process, the GNSS base station for high-accuracy RTK positioning provided by the Hong Kong government[9] is employed. The flight velocities of the HKairport sequence are set to 3 m/s, 6 m/s, and 9 m/s.

For the data sequence named Featureless, after flying the established flight trajectory, which is the same as that used above in the HKairport sequence, with a speed of 6 m/s and altitude of 80 m, we further incorporated a slightly low altitude (20 m) manual flight over the runway to create scenarios where LiDAR is degenerated and is

Table 4. Features of each scenario. The ROS bags are organized into sequences according to their scenarios and characters. `HKairport`, `HKisland`, `AMtown`, `AMvalley`, and featureless sequences are collected, respectively, in the aero-model airfield in Hong Kong, an island in Hong Kong, a rural town in Armenia, a valley in Armenia. Those sequences with GNSS suffix are containing raw GNSS measurements. A dash ("-") signifies that only manual flights are conducted in this data collection.

Sequence	Cruising altitude (m)	Cruising speed (m/s)	Path length (km)
HKairport01	80	3	2.040
HKairport02	80	6	
HKairport03	80	9	
HKairport_GNSS01	80	3	
HKairport_GNSS02	80	6	
HKairport_GNSS03	80	9	
HKisland01	90	3	1.846
HKisland02	90	6	
HKisland03	90	9	
HKisland_GNSS01	90	3	
HKisland_GNSS02	90	6	
HKisland_GNSS03	90	9	
AMtown01	80	4	5.109
AMtown02	80	8	
AMtown03	80	12	
AMvalley01	130	4	4.304
AMvalley02	130	8	
AMvalley03	130	12	
Featureless_GNSS01	80	6	6.690
Featureless_GNSS02	80	6	2.456
Featureless_GNSS03	-	-	7.148

texture-less that allows users to test their GNSS fusion SLAM algorithms (see Table 4, Seq. `Featureless_GNSS01–03`).

4.2. Island

The data sequence in this scenario named `HKisland` is collected on Kai Pei Chau in Cape D'Aguilar situated at 22.2057°N, 114.2597°E, Hong Kong, China. The site comprises a large island measuring approximately 243 m by 147 m and a smaller island measuring approximately 150 m by 109 m, with a survey area exceeding 94,000 square meters. During the acquisition of this scenario, the UAV's altitude remained at a constant 90 m. Collecting data in this area posed significant challenges for both LiDAR SLAM and visual SLAM due to the flight route passing over the sea and the varying height of the island. Besides, the change of the wind speed above sea brings larger attitude changes to the aircraft. The data collection process utilized the GNSS base station provided by the Hong Kong government and the flight speeds are set to 3 m/s, 6 m/s, and 9 m/s as well.

4.3. Rural town

The data sequence in this scenario named `AMtown` is obtained from Urtsadzor, situated at 39.9218°N and 44.8229°E, within the Ararat province of Armenia. This particular sequence encompasses a diminutive village spanning approximately 859 m by 464 m, resulting in a surveyed expanse surpassing 411,000 square meters. During the acquisition of this scenario, the UAV's altitude remained at a constant 80 m. During data acquisition, the DJI D-RTK2[10] Mobile GNSS Station is used as the RTK base station, and flight velocities are set to 4 m/s, 8 m/s, and 12 m/s.

4.4. Valley

The data collected in this scenario named `AMvalley` is acquired from Azizken Forest, situated at 39.9486°N and 44.8735°E, in the Ararat province of Armenia. The valley area has approximate dimensions of 932 m by 470 m, resulting in a surveyed area that exceeds 577,000 square meters. During the acquisition of this scenario, the UAV's altitude remained at a constant 130 m from the take-off point. The topographic undulations of the valley present a significant challenge for SLAM algorithms. Throughout the process, the DJI D-RTK2 Mobile GNSS Station is used as the RTK base station, and flight velocities are also set to 4 m/s, 8 m/s, and 12 m/s.

5. Data validation

For the purpose of data evaluation, we utilized several state-of-the-art multi-sensor fusion algorithms to test our data. The evaluative metric employed was the Root Mean Square Error (RMSE) of Absolute Trajectory Error (ATE), which
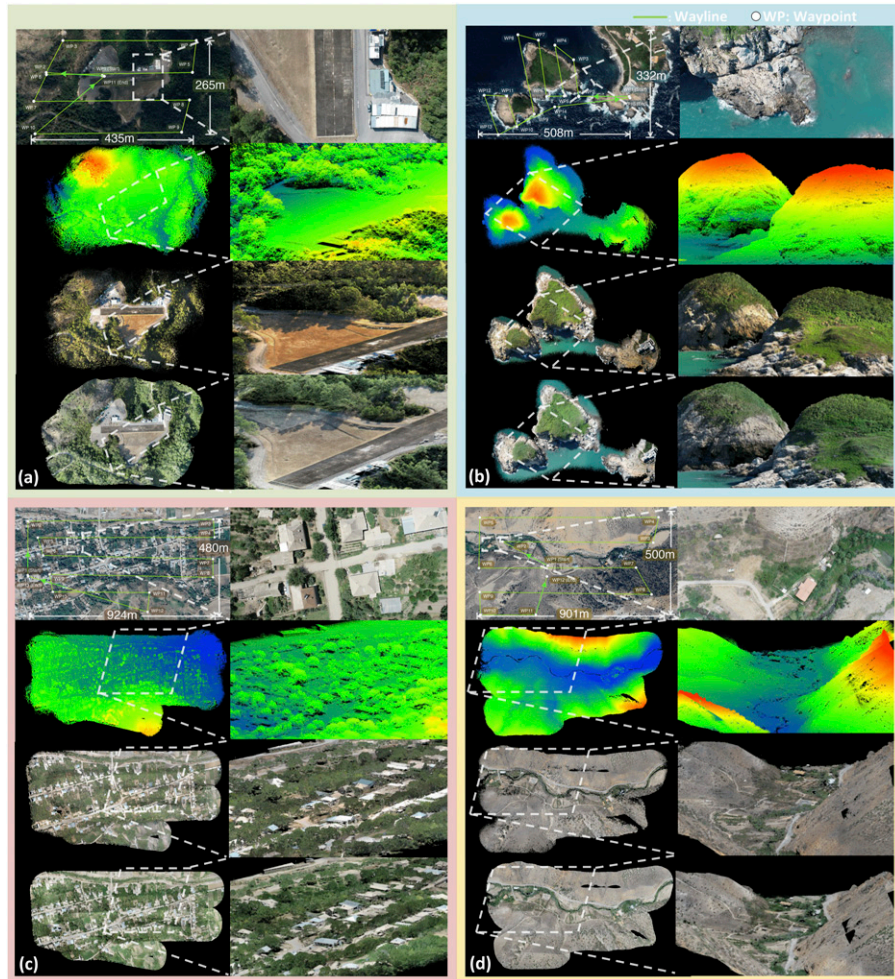
Figure 6. Figures (a)–(d) correspondingly represent the contents of the four distinct scenarios, namely, aero-model airfield, island, rural town, and valley. Each scenario is depicted via a collection of images, which include: a satellite map view with the flight plan route and its dimensions, a height-colored reconstruction result of the FAST-LIO2 (Xu et al., 2022) algorithm, an RGB-colored reconstruction result of the R3LIVE (Lin and Zhang, 2022) algorithm, and ground truth maps reconstructed by DJI L1 and color-rendered through RGB. The left columns show full-view images, while the right columns display corresponding zoomed-in detail images.

was determined by comparing the position trajectories yielded from the multi-sensor fusion algorithms to the RTK position ground truth encapsulated within the dataset. In order to evaluate the results with respect to the trajectories, the ATE drift [%] is calculated by dividing the RMSE of ATE by the length of the flight path (see Table 4). Our choice of multi-sensor fusion algorithms encompassed of different sensor combinations, inclusive of LiDAR-inertial methods (LIO-Livox[11], FAST-LIO2 (Xu et al., 2022)), LiDAR-visual-inertial method (R3LIVE (Lin and Zhang, 2022)), visual-inertial method (ORB-SLAM3 (Campos et al., 2021)), and GNSS-visual-inertial method (G-VINS (Cao et al., 2022)) to validate the dataset.

Additionally, we conducted trials using LIO-SAM (Shan et al., 2020) and LiLi-OM (Li et al., 2021); however, these were not incorporated into the final results due to their incompatibility with our data. Specifically, LIO-SAM demonstrated inadequate support for solid-state LiDARs, as per the original authors' comments, and LiLi-OM proved unsuitable for our

downward-looking data, leading to rapid failure during testing. For all evaluations, we adhered to the default setting for each method. In the alignment of predicted trajectories with the ground truth, the evo tool (Schubert et al., 2018) was employed to produce the ATE results.

The ATE and its drift results are presented in Table 5. The point clouds generated by the algorithms can be seen in Figure 6. LIO-Livox performed well in low-speed scenarios but tended to fail in scenarios involving high-speed or LiDAR degradation. FAST-LIO2 showed robustness across various scenarios but performed suboptimally in those featuring LiDAR degradation, owing to its exclusive use of LiDAR and IMU. R3LIVE demonstrated a distinct improvement over FAST-LIO2 in some scenarios involving LiDAR degradation (e.g., `AMvalley01` and `AMvalley02`). As for ORB-SLAM3, we utilized its Monocular-Inertial mode with loop closure to validate our data. As shown in Table 5, ORB-SLAM3 is working well on slow flights, the accuracy of ORB-SLAM3 is lower than

Table 5. The RMSE of ATE and its drift results of the multi-sensor algorithms as applied to the MARS-LVIG dataset. The optimal odometry outcome for each bag is emphasized in bold typeface. A dash ("-") signifies the failure of the algorithm's execution. The drift is calculated by dividing the RMSE of ATE by the length of the flight path.

|	LIO-Livox		FAST-LIO2		R3LIVE		ORB-SLAM3	
Sequence	ATE (m)	Drift [%]	ATE (m)	Drift [%]	ATE (m)	Drift [%]	ATE (m)	Drift [%]
HKairport01	**0.65**	**0.032**	0.66	0.032	0.68	0.033	3.86	0.189
HKairport02	123.39	6.049	1.14	0.056	**0.82**	**0.040**	9.39	0.460
HKairport03	133.05	6.622	1.81	0.089	**1.12**	**0.055**	-	-
HKairport_GNSS01	0.62	0.030	0.61	0.030	**0.53**	**0.026**	3.35	0.164
HKairport_GNSS02	**2.97**	**0.145**	2.99	0.146	2.98	0.146	-	-
HKairport_GNSS03	-	-	**1.00**	**0.050**	1.41	0.060	-	-
HKisland01	0.75	0.041	**0.64**	**0.034**	0.70	0.037	1.06	0.057
HKisland02	2.20	0.119	2.13	0.115	**2.10**	**0.113**	-	-
HKisland03	-	-	**2.16**	**0.117**	3.93	0.212	-	-
HKisland_GNSS01	2.18	0.118	**2.01**	**0.108**	2.02	0.109	3.33	0.180
HKisland_GNSS02	3.26	0.177	2.18	0.118	**2.11**	**0.114**	29.73	1.610
HKisland_GNSS03	-	-	4.18	0.226	**3.68**	**0.199**	-	-
AMtown01	1.92	0.038	2.28	0.045	**1.44**	**0.028**	93.53	1.830
AMtown02	21.95	0.429	3.24	0.063	**2.10**	**0.041**	-	-
AMtown03	-	-	**2.93**	**0.057**	-	-	-	-
AMvalley01	12.28	0.285	4.54	0.105	**3.82**	**0.088**	32.14	0.746
AMvalley02	-	-	8.12	0.188	**4.47**	**0.103**	50.21	1.160
AMvalley03	-	-	**8.21**	**0.191**	-	-	-	-
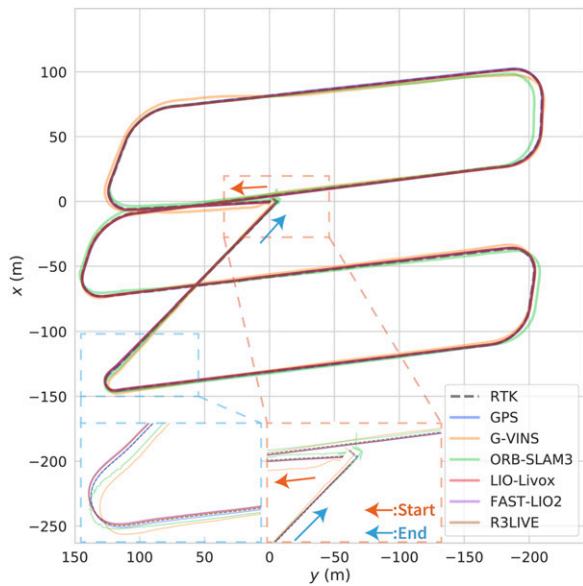
Figure 7. Top view of the trajectories generated by RTK position ground truth, GPS (without RTK) position, and five algorithm-estimated positions of `HKairport_GNSS01` sequence, respectively.

algorithms that use LiDAR-IMU odometry, especially on the *z axis*. Figures 7 and 8 present illustrative results on the `HKairport_GNSS01` scenario.

We utilized G-VINS to validate the data including raw GNSS measurement. As for `Featureless` sequences, all selected algorithms, including LIO-Livox, FAST-LIO2, ORB-SLAM3, and R3LIVE, failed due to the low altitude

(20 m) manual flight causing LiDAR and visual degenerations. G-VINS exhibited reliable performance at the start-up of the `Featureless` bags, achieving ATE of 1.73 m, 2.78 m, and 2.54 m for `Featureless_GNSS01`, `Featureless_GNSS02`, and `Featureless_GNSS03`, respectively. Nevertheless, G-VINS fails on all `Featureless` sequences when the UAV starts to make rapid turns, which causes rapid changes in the FoV that can hardly be tracked. For the normal trajectory including raw GNSS measurement, G-VINS works well on `HKairport_GNSS01`, achieving ATE of 3.39 m and drift of 0.166% (see G-VINS results in Figures 7 and 8). As for other datasets in `HKairport_GNSS` sequence which are collected at 6 m/s and 9 m/s, G-VINS performed similar to the `Featureless` sequence, experiencing a reliable performance at the start-up of the GNSS bags but soon fails at rapid turns. As for the datasets in `HKisland_GNSS` sequence, all G-VINS tests fail when UAV is flying over the sea where no stationary salient visual features can be tracked (due to the sea wave moving). Despite this, the raw GNSS measurements and camera data in this sequence were effectively validated through the results, as G-VINS was successfully initiated and operated.

6. Challenges and limitations

To mitigate the effects of high-frequency vibrations on the LiDAR, its internal IMU, and camera, rubber-damping balls are employed as a mounting method between the three sensors and the UAV in our system. This approach
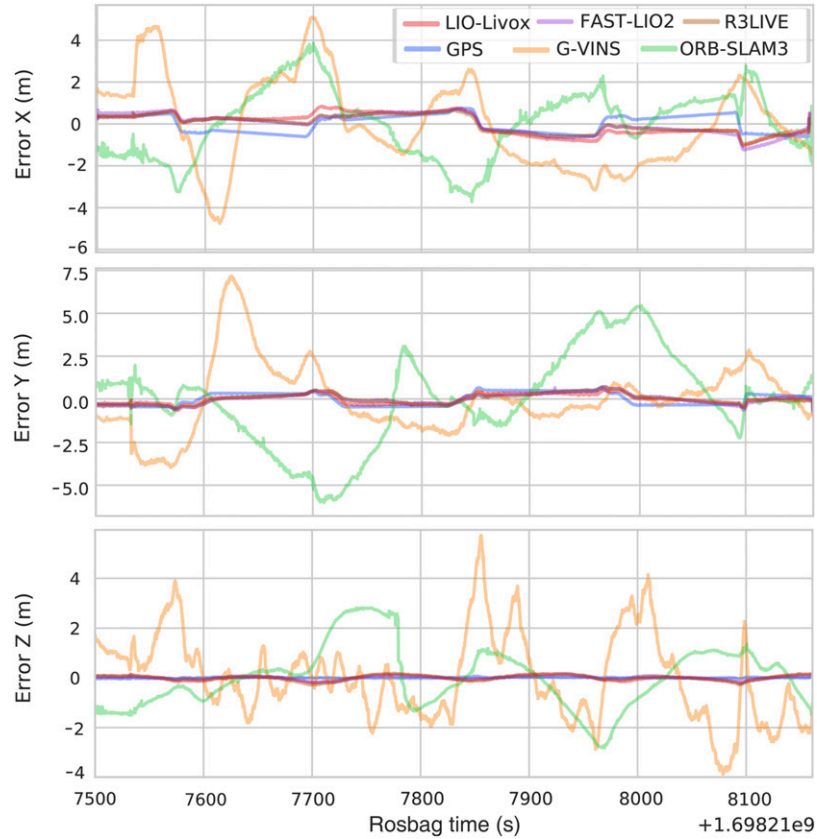
Figure 8. The errors of GPS (without RTK) position and five algorithm-estimated positions concerning RTK position ground truth of `HKairport_GNSS01` sequence, which are illustrated in *X*, *Y*, and *Z* axes, respectively.

effectively reduces IMU noise and eliminates image blur. However, it may introduce errors in the relative position between the sensors and the RTK antenna phase center, which results in potential errors between the estimated position and the ground truth. To assess this error, we utilized the VICON[12] motion capture system. By fixing the UAV on the ground and applying forces to the board where the sensors are installed, we induced the maximum inclination of the sensors. Upon observing that the maximum displacement of the sensor unit is less than 1.5 cm, we can confirm that the impact of the rubber-damping balls on the error of the relative position is very small. Considering the UAV flight altitude, the points measured by the LiDAR are tens of meters away, and the error caused by the 1.5 cm offset (e.g., re-projection to the camera plane) is substantially small (e.g., less than 1 pixel) and can be ignored.

During the data collection process, due to the large scene size and significant variations in environmental illumination, we utilize the camera's built-in auto-exposure feature to ensure high image quality. Yet, the camera message does not provide the exposure time. Fixing the camera exposure time will cause image overexposure or underexposure, thereby imposing significant challenges for both SLAM and point cloud coloring algorithms. After carefully considering the satisfactory results obtained during our validation process, we believe that the impact of exposure variations on the visual SLAM system is acceptable.

7. Conclusion and future work

In this paper, we have presented a comprehensive high-altitude downward-looking dataset that combines the use of a solid-state LiDAR, a global-shutter RGB camera, a raw message receiver of GNSS, and an advanced UAV platform. We collect data in diverse large outdoor environments at a high altitude, aiming to address the absence of high-altitude data and outdoor scenes in existing aerial SLAM datasets, which would pose additional challenges to current SLAM algorithms.

Our future work will be driven by the challenges found in this dataset, we aspire to develop robust SLAM algorithms capable of overcoming these challenges in future research.

G-VINS for their great help from their open-source GNSS receiver driver and detailed description.

ORCID iDs

Haotian Li https://orcid.org/0000-0001-8431-4098
Nan Chen https://orcid.org/0000-0002-1359-7235
Xiyuan Liu https://orcid.org/0000-0002-4925-1498
Dongjiao He https://orcid.org/0000-0003-0131-061X
Yixi Cai https://orcid.org/0000-0003-0321-1164
Kaiwen Xue https://orcid.org/0000-0003-1931-7852

Supplemental Material

Supplemental material for this article is available online.

Notes

1. https://www.dji.com/matrice-300
2. https://www.livoxtech.com/avia
3. https://www.bosch-sensortec.com/products/motion-sensors/imus/bmi088/
4. https://www.hikrobotics.com/en/machinevision/productdetail?id=5762
5. https://www.u-blox.com/en/product/c099-f9p-application-board
6. https://www.dji.com/zenmuse-l1
7. https://www.dji.com/dji-terra
8. https://www.hkmec.com/index.htm
9. https://www.geodetic.gov.hk/en/satref/Net_RTK.htm
10. https://https://www.dji.com/d-rtk-2
11. https://github.com/Livox-SDK/LIO-Livox
12. https://www.vicon.com/

References

Antonini A, Guerra W, Murali V, et al. (2018) The blackbird dataset: a large-scale dataset for uav perception in aggressive flight. *International Symposium on Experimental Robotics*. Springer, 130–139.

Blanco-Claraco JL, Moreno-Duenas FA and González-Jiménez J (2014) The málaga urban dataset: high-rate stereo and lidar in a realistic urban scenario. *The International Journal of Robotics Research* 33(2): 207–214.

Burri M, Nikolic J, Gohl P, et al. (2016) The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research* 35(10): 1157–1163.

Campos C, Elvira R, Rodríguez JJG, et al. (2021) Orb-slam3: an accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics* 37(6): 1874–1890.

Cao S, Lu X and Shen S (2022) Gvins: tightly coupled gnss–visual–inertial fusion for smooth and consistent state estimation. *IEEE Transactions on Robotics* 38(4): 2004–2021.

Carlevaris-Bianco N, Ushani AK and Eustice RM (2016) University of michigan north campus long-term vision and lidar dataset. *The International Journal of Robotics Research* 35(9): 1023–1035.

Delmerico J, Cieslewski T, Rebecq H, et al. (2019) *Are we ready for autonomous drone racing? the uzh-fpv drone racing dataset*. In: 2019 International Conference on Robotics and Automation (ICRA). Montreal, QC, Canada, 2019, pp. 6713–6719, IEEE.

Geiger A, Lenz P and Urtasun R (2012) *Are we ready for autonomous driving? the kitti vision benchmark suite*. In: 2012 IEEE conference on computer vision and pattern recognition. RI, USA, 2012, pp. 3354–3361, IEEE.

Gurgu MM, Queralta JP and Westerlund T (2022) *Vision-based gnss-free localization for uavs in the wild*. In: 2022 7th International Conference on Mechanical Engineering and Robotics Research (ICMERR). Krakow, Poland, 2022, pp. 7–12, IEEE.

He D, Xu W, Chen N, et al. (2023) Point-lio: robust high-bandwidth light detection and ranging inertial odometry. *Advanced Intelligent Systems* 5: 2200459.

Helmberger M, Morin K, Berner B, et al. (2022) The hilti slam challenge dataset. *IEEE Robotics and Automation Letters* 7(3): 7518–7525. DOI: 10.1109/LRA.2022.3183759

Jeong J, Cho Y, Shin YS, et al. (2019) Complex urban dataset with multi-level sensors from highly diverse urban environments. *The International Journal of Robotics Research* 38(6): 642–657.

Jiao J, Wei H, Hu T, 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). Kyoto, Japan, 2022, pp. 3851–3856, IEEE.

Li X, Liu S, Xiao Q, et al. (2017) A multiscale dataset for understanding complex eco-hydrological processes in a heterogeneous oasis system. *Scientific Data* 4(1): 1–11.

Li K, Li M and Hanebeck UD (2021) Towards high-performance solid-state-lidar-inertial odometry and mapping. *IEEE Robotics and Automation Letters* 6(3): 5167–5174. DOI: 10.1109/LRA.2021.3070251

Lin J and Zhang F (2022) R3live: a robust, real-time, rgb-colored, lidar-inertial-visual tightly-coupled state estimation and mapping package. In: 2022 International Conference on Robotics and Automation (ICRA), PA, USA, 2022, pp. 10672–10678. DOI: 10.1109/ICRA46639.2022.9811935

Lin J, Liu X and Zhang F (2020) A decentralized framework for simultaneous calibration, localization and mapping with multiple lidars. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), NV, USA, 2020, pp. 4870–4877. DOI: 10.1109/IROS45743.2020.9340790

Lin J, Zheng C, Xu W, et al. (2021) R2live: a robust, real-time, lidar-inertial-visual tightly-coupled state estimator and mapping. *IEEE Robotics and Automation Letters* 6(4): 7469–7476.

Liu X and Zhang F (2021a) Extrinsic calibration of multiple lidars of small fov in targetless environments. *IEEE Robotics and Automation Letters* 6(2): 2036–2043. DOI: 10.1109/LRA.2021.3061387

Liu Z and Zhang F (2021b) Balm: bundle adjustment for lidar mapping. *IEEE Robotics and Automation Letters* 6(2): 3184–3191.

Liu X, Liu Z, Kong F, et al. (2023a) Large-scale lidar consistent mapping using hierarchical lidar bundle adjustment. *IEEE Robotics and Automation Letters* 8(3): 1523–1530. DOI: 10.1109/LRA.2023.3238902

Liu Z, Liu X and Zhang F (2023b) Efficient and Consistent Bundle Adjustment on Lidar Point Clouds, in *IEEE Transactions on Robotics* 39(6): 4366–4386. DOI: 10.1109/TRO.2023.3311671

Liu X, Yuan C and Zhang F (2022) Targetless extrinsic calibration of multiple small fov lidars and cameras using adaptive voxelization. *IEEE Transactions on Instrumentation and Measurement* 71: 1–12. DOI: 10.1109/TIM.2022.3176889

Majdik AL, Till C and Scaramuzza D (2017) The zurich urban micro aerial vehicle dataset. *The International Journal of Robotics Research* 36(3): 269–273.

Nguyen TM, Yuan S, Cao M, et al. (2022) Ntu viral: a visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint. *The International Journal of Robotics Research* 41(3): 270–280.

Patel B, Barfoot TD and Schoellig AP (2020) *Visual localization with google earth images for robust global pose estimation of uavs*. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France, 2020, pp. 6491–6497, IEEE.

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.

Ramezani M, Wang Y, Camurri M, et al. (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). Las Vegas, NV, USA, 2020, pp. 4353–4360, IEEE.

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

Shan T, Englot B, Meyers D, et al. (2020) Lio-sam: tightly-coupled lidar inertial odometry via smoothing and mapping. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas, NV, USA, 2020, pp. 5135–5142. DOI: 10.1109/IROS45743.2020.9341176

Shi X, Li D, Zhao P, et al. (2020) *Are we ready for service robots? the openloris-scene datasets for lifelong slam*. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France, 2020, pp. 3139–3145, IEEE.

Sun K, Mohta K, Pfrommer B, et al. (2018) Robust stereo visual inertial odometry for fast autonomous flight. *IEEE Robotics and Automation Letters* 3(2): 965–972.

Xu W and Zhang F (2021) Fast-lio: a fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter. *IEEE Robotics and Automation Letters* 6(2): 3317–3324.

US Geological Survey (2019) 3DEP lidar point cloud now available as amazon public dataset. https://www.usgs.gov/news/technical-announcement/usgs-3dep-lidar-point-cloud-now-available-amazon-public-dataset (Accessed 9 Nov 2023).

Varney N and Asari VK and Graehling Q (2020) Dales: a large-scale aerial lidar data set for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, WA, USA, 2020, pp. 186–187.

Xu W, Cai Y, He D, et al. (2022) Fast-lio2: fast direct lidar-inertial odometry. *IEEE Transactions on Robotics* 38(4): 2053–2073.

Yin J, Li A, Li T, et al. (2021) M2dgr: a multi-sensor and multi-scenario slam dataset for ground robots. *IEEE Robotics and Automation Letters* 7(2): 2266–2273.

Yuan C, Liu X, Hong X, et al. (2021) Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments. *IEEE Robotics and Automation Letters* 6(4): 7517–7524. DOI: 10.1109/LRA.2021.3098923

Yuan Z, Wang Q, Cheng K, et al. (2023) Sdv-loam: semi-direct visual-lidar odometry and mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(9): 11203–11220.

Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11): 1330–1334.

Zhang L, Camurri M and Fallon M (2021) Multi-camera Lidar Inertial Extension to the Newer College Dataset. arXiv preprint. *arXiv:2112.08854*.

Zheng C, Zhu Q, Xu W, et al. (2022) Fast-livo: fast and tightly-coupled sparse-direct lidar-inertial-visual odometry. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 2022, pp. 4003–4009. DOI: 10.1109/IROS47612.2022.9981107

Zhu F, Ren Y and Zhang F (2022) *Robust real-time lidar-inertial initialization*. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 2022, pp. 3948–3955, IEEE.

Zhu Y, Kong Y, Jie Y, et al. (2023) Graco: a multimodal dataset for ground and aerial cooperative localization and mapping. *IEEE Robotics and Automation Letters* 8: 966–973.