ECMD: An Event-Centric Multisensory Driving Dataset for SLAM

Peiyu Chen[1*], Weipeng Guan[1*], Feng Huang[2*], Yihan Zhong[2], Weisong Wen[2], Li-Ta Hsu[2], Peng Lu[1†]

Abstract—Leveraging multiple sensors enhances complex environmental perception and increases resilience to varying luminance conditions and high-speed motion patterns, achieving precise localization and mapping. This paper proposes, ECMD, an event-centric multisensory dataset containing 81 sequences and covering over 200 km of various challenging driving scenarios including high-speed motion, repetitive scenarios, dynamic objects, etc. ECMD provides data from two sets of stereo event cameras with different resolutions (640×480, 346×260), stereo industrial cameras, an infrared camera, a top-installed mechanical LiDAR with two slanted LiDARs, two consumer-level GNSS receivers, and an onboard IMU. Meanwhile, the ground-truth of the vehicle was obtained using a centimeter-level high-accuracy GNSS-RTK/INS navigation system. All sensors are well-calibrated and temporally synchronized at the hardware level, with recording data simultaneously. We additionally evaluate several state-of-the-art SLAM algorithms for benchmarking visual and LiDAR SLAM and identifying their limitations. The dataset is available at https://arclab-hku.github.io/ecmd/.

Index Terms—Event-based Vision, Multi-sensor Fusion, SLAM, Autonomous Driving, Dataset.

I. INTRODUCTION

VISUAL and LiDAR simultaneous localization and mapping (SLAM) achieved notable progress within driving scenarios in recent years. However, they encounter the challenging task of operating robustly under heterogeneous environments, such as varying lighting conditions, low-texture scenarios, repetitive structures, diverse motion patterns, dense dynamic objects, etc. Utilizing novel sensors and integrating multiple sensors can provide a comprehensive perception and enhance the robustness of the entire system [1]–[3]. These motivate us to develop a dataset that integrates novel sensors under realistic and complex driving scenarios, thereby promoting SLAM research.
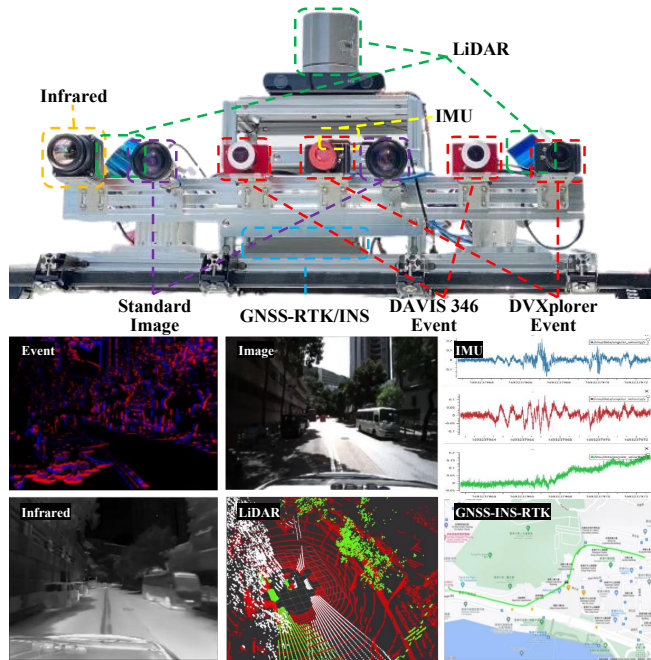

Fig. 1. An overview of the sensor setup and dataset visualization. Above: The sensor suite is mounted on the top of the vehicle. Below: Sensor systems include two sets of stereo event cameras, stereo industrial cameras, an infrared camera, a top-installed LiDAR with two slanted LiDAR, IMU, and GNSS-RTK/INS systems. Each sensor is indicated with the letter box.

Event cameras have low latency (μs-level) and high dynamic range (140 dB compared to 60 dB with standard cameras) properties, which offers great opportunities for visual (VO) and visual-inertial odometry (VIO) in rough terrain, aggressive motions, and high dynamic range (HDR) [4]. Unlike traditional frame-based cameras that directly capture fixed-rate intensity frames, event cameras are motion-activated sensors that capture pixel-wise intensity differences asynchronously in continuous streams. However, the widespread commercialization and implementation of event cameras in robotics are still early due to the expensive cost. In addition, event cameras confront challenges during rapid vibrations and ego-motion, as these conditions generate a substantial quantity of events, leading to intensive computations. Conversely, in cases where minimal relative motion between the event camera and the scene exists, such as under static states, they only provide limited information or even introduce noise [5]. Therefore, we embark on this research

*Equal contribution; †Corresponding Author.

[1]Peiyu Chen, Weipeng Guan, and Peng Lu are with the Adaptive Robotic Controls Lab (ArcLab), Department of Mechanical Engineering, The University of Hong Kong, Hong Kong SAR, China. (e-mail: chenpyhk@connect.hku.hk; wpguan@connect.hku.hk; lupeng@hku.hk).

[2]Feng Huang, Yihang Zhong, Weisong Wen, and Li-Ta Hsu are with the Intelligent Positioning and Navigation Laboratory, Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China. (e-mail: darren-f.huang@connect.polyu.hk; yihan1.zhong@connect.polyu.hk; welson.wen@polyu.edu.hk; lt.hsu@polyu.edu.hk)

effort to explore the inquiry: Are event cameras ready for autonomous driving?

There exist several stereo event-based driving datasets that are worth mentioning and exploring. MVSEC [6] was the first stereo event-based driving dataset proposed for evaluating the localization performance. While MVSEC employs the low resolution of DAVIS346 which limits the feature detection for accurate localization. DSEC [7] offers stereo event streams with a high resolution of 0.31 Megapixels(MP). However, this dataset focuses on computer vision tasks segmentation, depth estimation, optical flow estimation, etc., which is not specifically designed for VO/VIO/SLAM domains. MA-VIED [8] propose a large-scale driving dataset under standard urban scenarios and race track-like loops. The ground-truth trajectory relies on GNSS-RTK, which only ensures high accuracy in open-sky environments and fails to provide high accuracy in GNSS-denied scenarios such as tunnels or densely street areas. Ref. [9] focuses on collecting both stereo event data and stereo intensity images under indoor and urban driving scenes with the ground-truth of GNSS-RTK/INS. Their sequences do not encompass extremely high-speed or repetitive scenarios that could be challenging to VO/VIO/SLAM algorithms.

To address the above drawbacks, we propose ECMD, a dataset procured from diverse challenging driving scenarios with a comprehensive suite of sensors for benchmarking various VO/VIO/SLAM algorithms. To the best of our knowledge, this is the first event-based SLAM dataset specifically focused on densely urbanized driving scenarios. The contributions of our work can be summarized as follows:

1) Our sensor platform consists of various novel sensors shown in Fig.1, including two sets of stereo event cameras with distinct resolutions (640×480, 346×260), an infrared camera, stereo industrial cameras, three mechanical LiDARs (including two slanted LiDARs), a high-quality inertial measurement unit (IMU), and three global navigation satellite system (GNSS) receivers. For the ground-truth, we adopt a centimeter-level position system that combines the GNSS real-time kinematic (RTK) with the fiber optics gyroscope integrated inertial system as GNSS-RTK/INS.

2) ECMD collects 81 sequences covering over 200 kilometers of trajectories in various driving scenarios, including dense streets, urban, tunnels, highways, bridges, and suburbs. These sequences are recorded under daylight and nighttime, providing challenging situations for Visual and LiDAR SLAM, e.g., dynamic objects, high-speed motion, repetitive scenarios, and HDR scenes. Meanwhile, we evaluate existing state-of-the-art visual and LiDAR SLAM algorithms with various sensor modalities on our datasets. Moreover, our dataset and benchmark results are released publicly available on our website.

The remainder of the paper is organized as follows: Section II introduces the related works. Section III presents the sensor setup and sensor calibration. Section IV intro-duces the dataset overview. Section V demonstrates the dataset application. Section VI introduces known issues. The conclusion is given in Section VII.

II. RELATED WORKS

Currently, several event-based datasets combined with various sensors have been released for VO/VIO/SLAM domains, utilizing handheld devices or a variety of robotics platforms. DAVIS240C [10], TUM-VIE [11], VECtor [12], and HKU-dataset[1] were collected by handheld / head-mounted devices under indoor environments. M2DGR [13] utilizes ground robots to collect a multi-sensor dataset with an event camera under large-scale scenes, while the event streams exhibit large noises. FusionPortable [14] proposes multi-sensor campus-scene datasets with stereo event cameras on diverse platforms (handheld, quadruped robot, and UGV). Moreover, there exist specialized event-based datasets such as UZH-FPV [15] and GRIFFIN [16], which are targeted for flying robots.

Moreover, a number of event-based datasets are published under large-scale driving scenarios for computer vision. These autopilot datasets offer more realistic and challenging conditions, including high-speed scenarios, repetitive situations, and HDR scenes compared to datasets collected from handheld devices. The first dataset catering to driving recordings using an event camera is DDD17 [17], as well as the follow-up DDD20 [18], for studying the end-to-end driving application incorporating diverse vehicle control data. HATS [19], CED [20], Ref. [21], and Ref. [22] published their event-based datasets for the computer vision task of object classification, image reconstruction, and vision place recognition in driving scenarios. MVSEC [6] is a pioneering cross-modal dataset with stereo event and image cameras, as well as LiDAR. However, a limitation of MVSEC resides in the utilization of low-resolution event cameras (346×260) with a compact baseline of 10 cm, coupled with the imprecision of the ground-truth derived from GNSS or LiDAR-SLAM. DSEC [7] proposed an event-based dataset whose scenarios are similar to KITTI [23], providing higher resolution stereo event (640×480) and image, LiDAR, and IMU under various illumination conditions. M3ED [24] encompasses high-resolution event cameras (1280×720) and covers three different robotics platforms: driving, flight, and legged robot. However, both DSEC and M3ED datasets are primarily utilized for computer vision fields, such as optical flow estimation, segmentation, and disparity estimation, rather than specifically for localization or mapping problems. Besides, they do not provide sufficient challenges for SLAM, as the majority of these datasets were collected in rural or suburban areas with relatively low-lying structures, light traffic, and less dynamic objects. ViViD++ [25] focuses on diverse vision sensors for handheld and driving platforms, including event, thermal, and standard cameras. MA-VIED [8] proposes a comprehensive driving dataset that encompasses race track-like loops, maneuvers, and standard driving

---

[1]https://github.com/arclab-hku/Event_based_VO-VIO-SLAM

TABLE I. Comparison with other event-based datasets in driving scenarios, regarding terrain and sensor types. ECMD provides the most extensive sensor configuration and urban scenarios.

Dataset	Terrain	Event	Infrared	Image	LiDAR	GNSS	IMU	GT Pose
		Resolution [MP]						
DDD17 [17]	Urban	0.09×1	✗	0.09×1	✗	✓	✗	✗
N-CARS [19]	Urban	0.07×1	✗	0.07×1	✗	✗	✗	✗
MVSEC [6]	Suburban	0.09×2	✗	0.36×2	VLP-16	✓	✓	GNSS/LiDAR-SLAM
CED [20]	Urban	0.09×1	✗	0.09×1	✗	✗	✗	✗
Ref. [21]	Urban	0.31×1	✗	✗	✗	✗	✗	✗
DDD20 [18]	Urban	0.09×1	✗	0.09×1	✗	✓	✗	✗
Brisbane-Event-VPR [22]	Suburban	0.31×1	✗	0.31×1	✗	✗	✗	✗
Ref. [27]	Suburban, Urban	0.31×1	✗	0.31×1	✗	✗	✗	✗
DSEC [7]	Suburban	0.31×2	✗	1.56×2	VLP-16	✓	✓	GNSS-RTK
ViViD++ [25]	Urban	0.31×1	0.33×1	2.46×1	OS1-64	✓	✓	GNSS-RTK
M3ED [24]	Forest, Urban	0.92×2	✗	1.02×3	OS1-64	✗	✓	LiDAR-SLAM/GNSS-RTK
MA-VIED [8]	Urban, Race track-like loops	0.31×1	✗	2.30×1	✗	✓	✓	GNSS-RTK
Ref. [9]	Urban, Indoor	0.31×2	✗	3.14×2	OS1-128	✓	✓	GNSS-RTK/INS
ECMD	Suburban, Urban, Dense City	0.09×2 0.31×2	0.33×1	2.30×2	VLP-16 Lslidar C16 HDL-32E	✓	✓	GNSS-RTK/INS

scenarios. However, both of these datasets exclusively offer monocular data for each camera type, thereby precluding the possibility of conducting stereo visual SLAM. Ref. [9] introduces a stereo visual localization dataset that exploits both the high-resolution event and standard cameras under indoor and urban scenarios.

Table I. summarizes the differences between our ECMD and other event-based datasets under autonomous driving scenarios. Compared to other datasets, our ECMD offers several advantages: (i) Capture diverse visual data format (RGB image, event stream, and infrared image) from multiple types of vision sensors in varying luminance conditions and urbanized scenarios; (ii) *1kHz-rate* event streams from different resolution event cameras empower in-depth exploration of event-based perception; (iii) Based on our previous work [26], three LiDARs, including two slanted LiDARs, are employed to collect high-rising building structures for LiDAR point cloud maps generation; (iv) We employ a centimeter-level localization system, GNSS-RTK/INS, as ground-truth, enabling a comprehensive evaluation of various SLAM algorithms.

III. SYSTEM OVERVIEW

A. Sensors Setup

The data collection platform is shown in Fig.1. Our sensor suite consists of a multi-camera setup (event camera, industrial camera, and infrared camera) equipped with three LiDARs, high-quality IMU, three GNSS receivers, and GNSS-RTK/INS systems. The specific specifications of each sensor are presented in Table II. An Intel NUC (i7-1260P, 32GB RAM) and an industrial computer (i7-10610U, 32GB RAM) are used to run sensor drivers, and record data into ROS bags on the Ubuntu system.

1) Visual Sensors: Two sets of stereo event cameras with different resolutions, DAVIS436 (346×260) and DVXplorer

(640×480), are configured at a baseline of 30 cm respectively. DAVIS346 produces asynchronous events and intensity frames. In contrast, DVXplorer exclusively generates events, while its resolution surpasses that of DAVIS346, enabling the provision of more intricate scene information. Each event camera is equipped with additional infrared filters to mitigate interference from LiDAR. Two FLIR BFLY-U3-23S3C industrial cameras with a resolution of 1920×1200 are used to capture RGB images at 20 Hz in fixed exposure mode. Forward-facing stereo industrial cameras are installed with a baseline of 30 cm, ensuring fairness by maintaining consistency with the baseline of the stereo event cameras. Hikrobot MV-CI003-GL-N6 infrared camera collects thermal frames at 20 Hz, encompassing a response band of 8-14μm and equipped with a 6.3mm focal length lens.

2) Mechanical LiDAR: We configure three mechanical LiDARs including two slanted LiDARs to collect accurate point clouds of surrounding environments. Velodyne HDL-32E is positioned on the top of the vehicle to capture the surroundings horizontally. Two slanted LiDARs, Lslidar C16 and Velodyne VLP-16, are mounted on the left and right sides of the sensor kit, respectively. This configuration facilitates the thorough recording of architectural particulars relevant to high-rising buildings in urbanized areas and all LiDAR data are collected at 10 Hz.

3) GNSS-RTK/INS Sensor: A tactical-level Xsens-MTI-30 IMU is employed to collect the raw acceleration and angular velocity at 400 Hz. The accurate ground-truth of localization is furnished by a centimeter-level GNSS-RTK/INS navigation system, further details can be found in Section IV-B1.

B. Time Synchronization

We use a Precision Time Protocol (PTP) [29] device to synchronize the clocks of various data collection devices

TABLE II. Sensors specifications

Sensors	Specifications
Event Camera	Inivation DAVIS346 Color (×2), 1000 Hz
	346×260 pixel
	IMU: MPU6150, 6-axis
	baseline: 30 cm
	Inivation DVXplorer (×2), 1000 Hz
	640×480 pixel
	IMU: MPU6150, 6-axis
	baseline: 30 cm
Infrared Camera	Hikrobot MV-CI003-GL-N6, 20 Hz
	640×512 pixel
	H-FOV: 88.5°, V-FOV: 73.2°
Industrial Camera	FLIR BFLY-U3-23S3C(× 2), 20 Hz
	1920×1200 pixel
	H-FOV: 96.8°, V-FOV: 79.4°
	baseline: 30 cm
LiDAR	Velodyne HDL-32E, 10Hz
	H-FOV: 360°, V-FOV: +10.67°∼-30.67°
	32 channel
	100m range
	Velodyne VLP-16, 10Hz
	H-FOV: 360°, V-FOV: +15°∼-15°
	16 channel
	100m range
	Lslidar C16, 10Hz
	H-FOV: 360°, V-FOV: +14°∼-16°
	16 channel
	150m range
IMU	Xsens Mti-30, 400Hz
	Accelerometer in-run Bias Instability 15 µg
	Gyroscope in-run Bias Instability 18°/h
GNSS Receiver	U-Blox ZED-F9P, 1Hz
	EVK-M8T, 1Hz
GNSS-RTK/INS	NovAtel SPAN-CPT [28], 1Hz
	Localization RMSE 5cm

across the sensor network. The PTP ensures time accuracy within nanoseconds. The synchronization device acquires the NMEA [30] output and pulse-per-second (PPS) signal from a u-blox M8T GNSS receiver to align the ROS time of the onboard computers with the GPS time. This enables sensors such as cameras, LiDAR, and IMU to record timestamps based on the synchronized GPS time. Moreover, to achieve time synchronization between different event cameras, the DAVIS346 on the rightmost side is configured as the master device and transmits trigger signal pulses to the remaining slave event cameras sequentially from left to right via external cables.

C. Sensors Calibration

1) IMU Calibration: To calibrate the IMU, we position it on a level surface for a duration of three hours and record the raw measurements. Utilizing the Kalibr toolbox, we can accurately calibrate the random walk and Gaussian white noise of the IMU.

2) Industrial Cameras Calibration: For industrial cameras, we move the sensor platform against the 9×7 checkerboard in the XYZ-axis and collect the sequence of RGB images and IMU. Then intrinsics calibration of industrial cameras is achieved by Kalibr toolbox [31], where the pinhole and radial-tangential camera models are adopted.

3) Event Cameras Calibration: For event cameras, DAVIS346 can produce fixed-rated frames, enabling image-based calibration, while DVXplorer merely produces asynchronous event streams. Therefore, E2Calib [32] [33] is used to achieve image reconstruction from event streams. With the reconstructed checkerboard images in Fig. 2(b), the intrinsics of event cameras could also be calibrated by Kalibr.
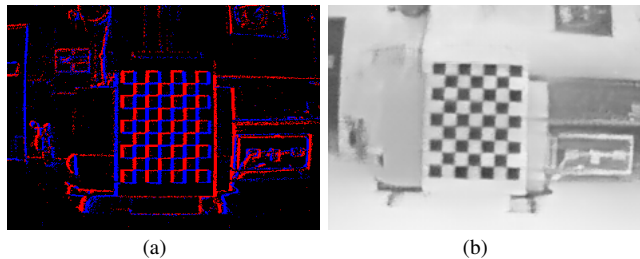

(a)   (b)

Fig. 2. The calibration of event cameras using checkerboard. (a) Raw event streams. (b) Reconstructed images from events.

4) Infrared Camera Calibration: Due to infrared cameras solely capturing the temperature rather than the intensity difference, we design a distinct 9×7 checkerboard to make the pattern detectable for infrared cameras. As shown in Fig.3(a), the checkerboard intervals are affixed with aluminum materials, and then using a heating plate to raise the temperature of the checkerboard. Since the superior thermal dissipation of aluminum compared to plastic, a temperature contrast emerges between the two materials, enabling infrared cameras to distinctly capture the lattice shape of the checkerboard, as in Fig.3(b). With the special infrared image of the checkerboard, intrinsic can be calibrated by Kalibr.
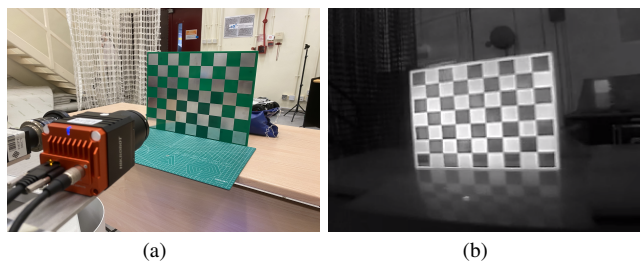

(a)   (b)

Fig. 3. The calibration of the infrared camera using checkerboard. (a) PCB checkerboard. (b) Infrared image.

5) Calibration between Camera and IMU: After completing intrinsics calibration, we move the sensor suite in front of checkerboards along the XYZ-RPY-axis and collect data simultaneously. Subsequently, the extrinsics and the temporal offset between all cameras and IMU could be estimated using Kalibr.

6) Calibration between LiDAR and IMU: For the calibration of mechanical LiDAR, LI-Init [34] is capable of achieving temporal and spatial calibration for LiDAR and IMU without checkerboards or extra devices in Fig.4. We rotate and move the device around the XYZ-axis to ensure sufficient excitation until the data accumulation is completed, thus we acquire the extrinsic transformation between LiDAR and IMU.
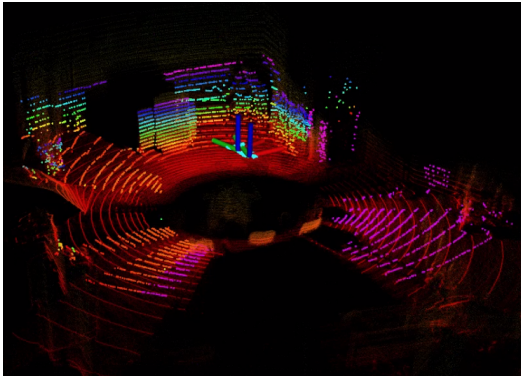

Fig. 4. Calibration between LiDAR and IMU.

IV. DATASET OVERVIEW

Our dataset encompasses a wide range of driving scenes, including urban streets, urban roads, tunnels, highways, bridges, and suburban roads. We have specifically focused on scenarios where visual SLAM algorithms encounter difficulties. These scenarios involve high-speed motion (up to 110 km/h), limited texture, as well as difficult glare conditions in both daytime and nighttime driving. We also targeted situations where LiDAR SLAM encounters limitations, such as long corridors or areas with sparse geometric structures. The complete dataset is partitioned into 81 sequences to facilitate researchers in evaluating their algorithms. Each sequence has an approximate duration of 120 seconds. Additionally, we have retained a few sequences with long duration, lasting approximately 34 minutes, specifically for the evaluation of loop closure in large-scale environments and loop closure scenarios. The summary of sequence types can be found in Table III.

TABLE III. The summary of different sequence types under contrasting lighting conditions

Terrain	Time	Duration [min]	#Sequences
Dense Street	Day, Night	91	19
Urban Road	Day, Night	35.5	19
Tunnel	- - -	6.5	9
Highway	Day, Night	34.3	15
Bridge	Day, Night	10.5	6
Suburban Road	Day, Night	23.5	13
Total	Day, Night	201.3	81

A. Scenarios

1) Dense_Urban_Street: This scenario focuses on low-speed vehicles, around 30km/h, proceeding on highly urbanized areas and urban canyons in Hong Kong with multiple light conditions. The streets are narrow at 10m in width and buildings on both sides of the scene are dense. Meanwhile, the presence of congested traffic and dynamic crowds may produce the degradation of visual or LiDAR localization, such as *Dense_street_day_easy_b*. To evaluate the loop closure performance of SLAM, we remarkably recorded sequences *Dense_street_difficult_circle* and *Dense_street_difficult_loop* where our vehicle was circling in repeated routes.

2) Urban_Road: This type of scenario records the vehicle traveling at an approximate speed of 60km/h on an expressway in Hong Kong with multiple weather conditions. Compared to the *Dense_Urban_Street* scenario, Urban_Road sequences travel through Hong Kong city at a higher speed, while the buildings are not as tightly packed on either side and the road is more spacious with four lanes. Despite the absence of pedestrians on the road, the scene still includes vehicles overtaking, paralleling, and other situations where the relative motion is not consistent with the absolute motion. The aforementioned discrepancy might pose a challenge for the VIO or LIO system. Moreover, the sequence comprises the vehicle traveling during nighttime in rainy conditions. We record trajectories in rainy situations under nighttime like *Urban_road_night_difficult_rainy_a* which are commonly faced in practical driving scenarios, whereas they are not present in previous datasets.

3) Tunnel: Tunnel scenarios commence with a high-speed vehicle on an open-sky highway, entering an enclosed tunnel without satellite reception. Inside the tunnel, GNSS positioning is unreliable since the satellite signal is completely blocked. Meanwhile, the scenario represents a typical and challenging scene for VIO and LIO systems due to the repetitive and texture-less environments for vision sensors and LiDAR. The sequence collections end after the vehicle exits the tunnel and continues to proceed on the highway for twenty seconds.

4) Highway: The scenario involves vehicles traveling at speeds up to 100km/h on low-texture highways both during the day and night, with sparse buildings alongside the road. High speeds, rapid changes in vehicle speed, repetitive visual scenes, and low-texture environments present significant challenges for autonomous driving. Meanwhile, the vibration of the vehicle body at high-speed motion amplifies the random walk and Gaussian white noise of IMU, thereby diminishing its reliability.

5) Bridge: The motion pattern of vehicles in bridge scenarios resembles that of highways, with vehicles traveling in a straight line at high speed along the bridge. However, this scene differs as there are no buildings on either side of the bridge, only the sea surrounds it. Bridges present scenes with limited texture, and the feature information within these scenes tends to be monotonous and repetitive, which further exacerbates the challenge of achieving accurate localization.

6) Suburban_Road: Suburban road scenarios present complex natural environments characterized by winding and rugged roads, steep slopes, and narrow lanes. The vehicle navigates the serpentine mountain roads at a moderate speed
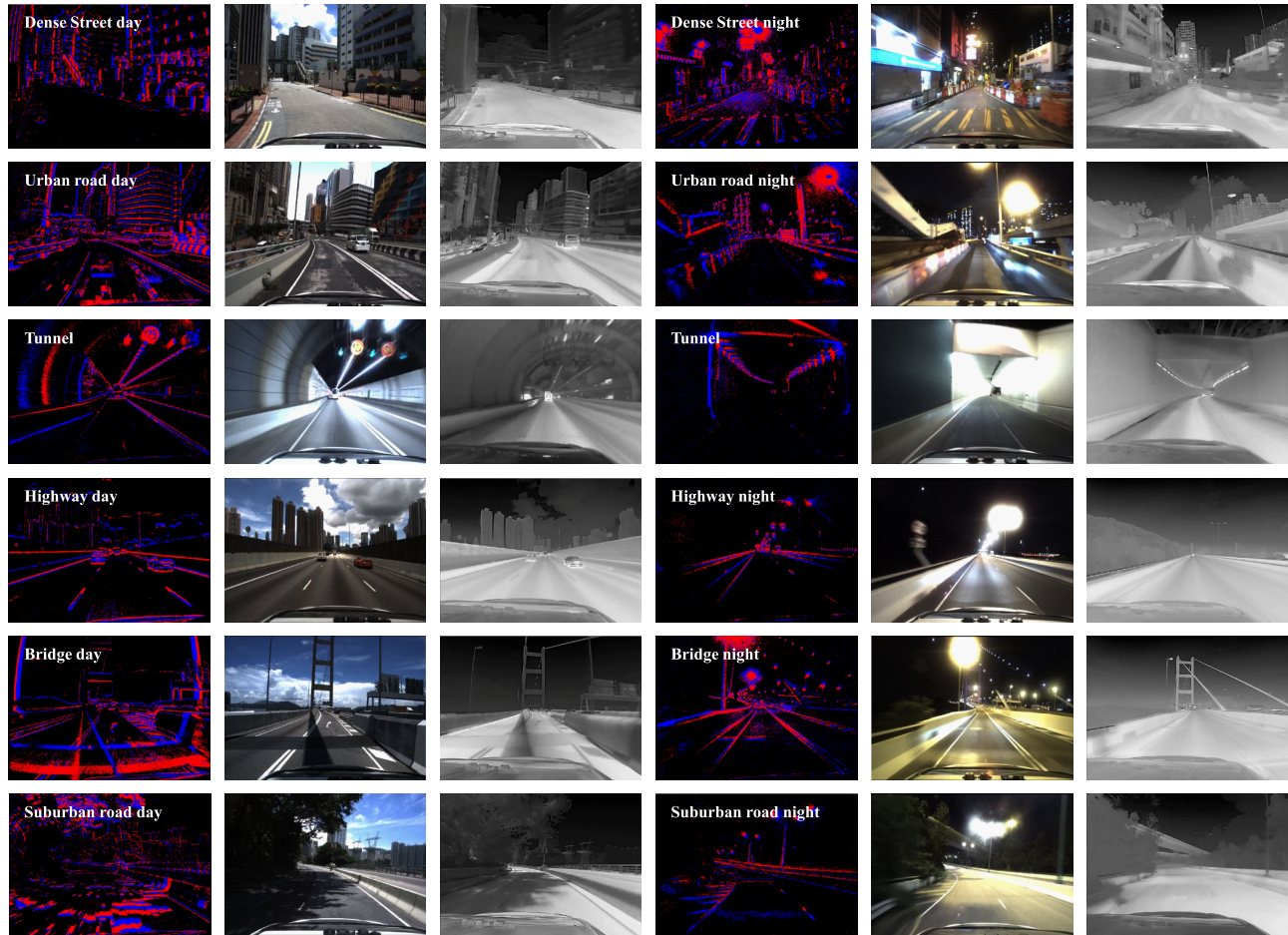
Fig. 5. The visualization of various scenarios including event streams, RGB images, and infrared images.

(approximately 50km/h), with significant altitude changes. The abundant texture information in the mountain road scene facilitates visual algorithms to extract stable features and construct effective constraints.

B. Ground-truth Generation

1) Ground-truth Poses: We obtained the ground-truth positioning from the NovAtel SPAN-CPT [28], a high-performance GNSS RTK/INS integrated navigation system. The ground-truth of most existing event-based driving datasets are derived from LiDAR-SLAM [6] [24], GPS/GNSS [6], GNSS-RTK [7] [25] [24]. The ground-truth derived from LiDAR-SLAM relies on the estimation of vehicle trajectories using LiDAR SLAM which only provides relative trajectories. It is difficult to quantify the accuracy of ground-truth pose, and errors may even exceed ten meters in some cases. The complex environment or the equipment malfunctions may disrupt the satellite reception of GPS/GNSS, thus relying solely on GPS/GNSS for ground-truth pose may lead to significant drift. The GNSS-RTK device can only provide centimeter-level accuracy in the open sky [26] In contrast, our SPAN-CPT can provide continuous high accuracy aided by the internal fiber-optic gyroscopes under high-rise buildings, tunnels, and other environments with weak satellite signals. Furthermore, we post-process the ground-truth positioning from SPAN-CPT using the state-of-the-art NovAtel Inertial Explorer [28] software to maximize the accuracy of the trajectory. For the GNSS positioning benchmark, we provide the WGS84 coordinate data for comparison. For the evaluation of SLAM algorithms, we provide the tools[2] to transform the ground-truth data from the WGS84 coordinates to the local frame/ENU frame based on the original points.

2) LiDAR Point Cloud Maps Generation: Utilizing the ground-truth pose for each frame in conjunction with their corresponding LiDAR point clouds, we accumulate these point clouds to construct a highly accurate LiDAR point cloud map to depict the TsingMa Bridge in Fig.6. The map encompasses rich spatial information, providing a detailed 3D reconstruction of the bridge and its surrounding areas.

V. DATASET APPLICATIONS

A. Visual SLAM Evaluations

As shown in Table IV., we evaluate the performance of VINS-MONO [35], ORB-SLAM3 [36], and ESVIO [37]

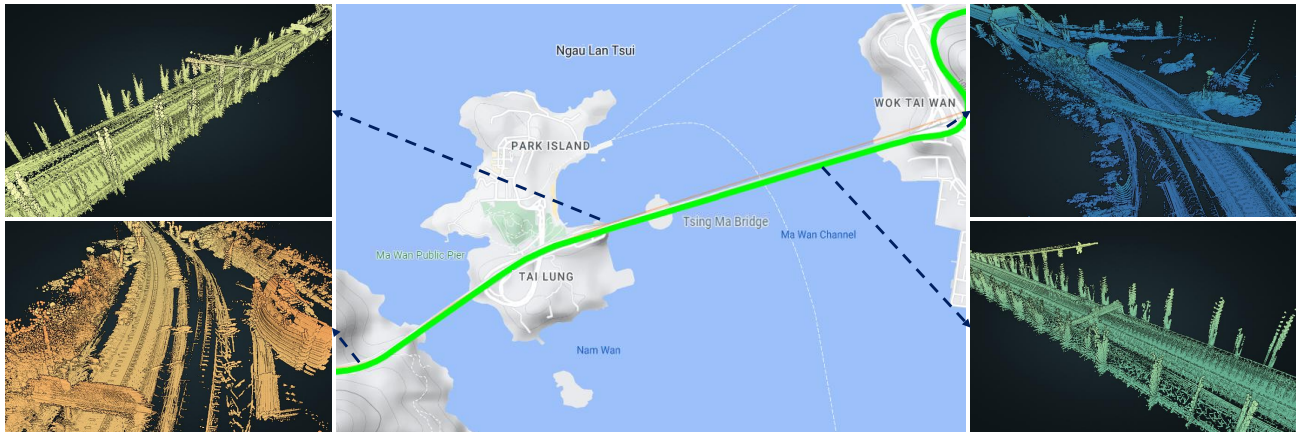[2]https://github.com/IPNL-POLYU/UrbanNavDataset/tree/master/tools/gt_vis

Fig. 6. The vehicle poses ground-truth on Google map with the LiDAR point cloud maps of Tsing Ma bridge.

across various scenes and lighting conditions on our dataset. The accuracy is quantified using mean position error (MPE, %), which aligns the estimated trajectory with ground-truth through 6-DOF transformation (in SE3) computed by the tool [38]. For the VINS-Mono, we evaluate it separately using RGB images and infrared images. Due to the resolution provided by industrial cameras being higher in contrast to the infrared camera, we achieve superior performance when utilizing RGB images. The ORB-SLAM3 often fails to robustly track features during high-speed vehicle movements, potentially resulting in the tracking thread restarts. The ESVIO leverages the complementary advantages of event streams and RGB images, allowing it to handle the lack of texture in RGB images under broad illumination conditions to achieve higher accuracy. Fig.7 compares event, RGB images, and infrared images under different lighting conditions. The RGB images offer rich texture under regular luminance scenes in contrast to events and infrared images offer comparatively limited information, e.g., the infrared image struggles to accurately discern traffic left-turn symbol on the ground. Conversely, the RGB image may lose numerous environmental features under the conditions of low light or over-exposure. The infrared camera can capture infrared radiation beyond the visible spectrum and event cameras can detect pixel-level intensity changes at low latency. Both the event camera and the infrared camera are more resilient in external varying lighting conditions, providing effective visibility compared to the industrial camera, e.g., in nighttime scenes, event cameras can capture road signs, and the infrared camera can clearly capture the surrounding bushes.
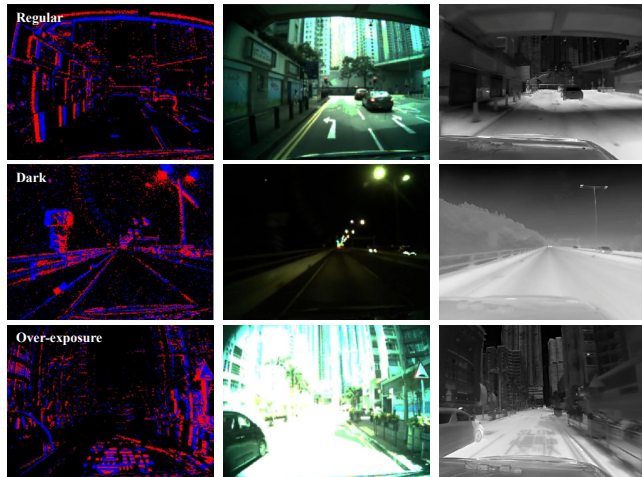

Fig. 7. Comparison of event streams, RGB images, and infrared images under different lighting conditions.

B. LiDAR SLAM Evaluations

Table V. demonstrates the performance of LIO-SAM [39], LVI-SAM [40], Fast-LIO2 [41], Point-LIO [42] across various scenes on our dataset. We use the same criteria introduced in Section V-A to evaluate the localization accuracy.

TABLE V. The MPE(%) of different LiDAR localization algorithms on ECMD datasets

Sequence	LIO-SAM [39] LIO	LVI-SAM [40] LVIO	Fast-LIO2 [41] LIO	Point-LIO [42] LIO
Dense street day easy a	0.25	0.25	**0.23**	0.25
Urban road day medium	**0.67**	*failed*	6.23	6.34
Tunnel easy a	**1.00**	*failed*	1.36	20.84
Highway day medium b	**0.51**	*failed*	0.63	*failed*
Bridge day difficult a	**0.78**	*failed*	1.33	15.52
Suburban road night medium b	**0.61**	*failed*	1.73	0.72

Due to the tilt-mounted LiDAR setups (see Fig.1), we are able to acquire point clouds of towering buildings situated on both sides of the street. This installation approach compensated for the lack of vertical point clouds compared to the horizontally mounted LiDAR. In Fig.8, red point clouds are generated from a horizontally mounted LiDAR while white and green point clouds are generated from tilt-

TABLE IV. The MPE(%) of different visual localization algorithms on ECMD datasets

Sequence	VINS-Mono [35] Mono VIO 2.30[MP] image	VINS-Mono [35] Infrared VIO 0.33[MP] infrared	ORB-SLAM3 [36] Stereo VO 2.30[MP] image	ESVIO [37] Stereo EVIO 2.30[MP] image 0.31[MP] event
Dense street night easy a	**0.34**	2.82	10.01	0.48
Urban road day easy b	14.26	**7.78**	*failed*	9.63
Highway day easy a	1.48	2.65	*failed*	**1.04**
Suburban road day easy b	2.36	3.48	9.04	**1.35**

mounted LiDARs. We evaluate the performance of LOAM [43] using three different LiDARs (center, left, and right)in *Dense_street_day_medium_circle_a* sequences.

The MPE of LOAM using center LiDAR is 1.02%, compared to 8.67% using the left LiDAR and 2.00% using the right LiDAR. LOAM using the left LiDAR exhibits significant drift since it initially captures minimal point cloud information. Although the LOAM merely using tilt-mounted LiDAR produces less accurate results compared to the center LiDAR, multi-LiDAR fusion can integrate complementary information, thereby improving localization accuracy and constructing more precise point cloud maps. Meanwhile, tunnel scenes present challenges for LiDAR SLAM. We capture three consecutive frames of LiDAR point clouds at two-second intervals in Fig.9. It is evident that these LiDAR point clouds exhibit high similarity in the tunnel environment, potentially resulting in degradation phenomena and inaccurate state estimation.
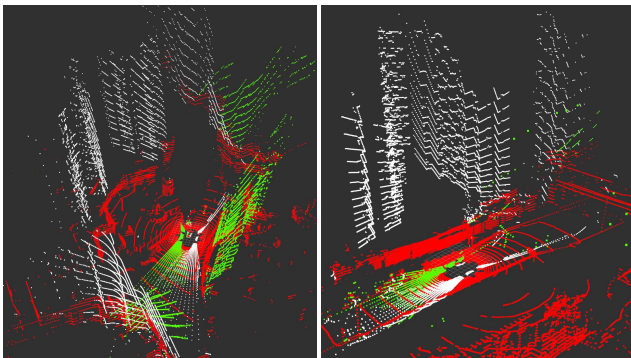

Fig. 8. Point clouds of dense street with high-rising buildings captured from three LiDARs, including two with angled installations.
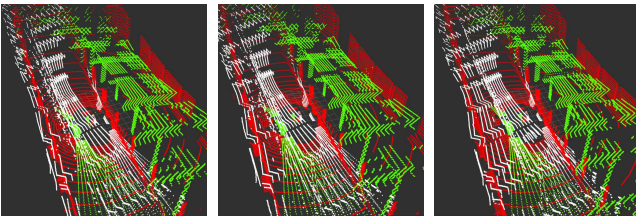

Fig. 9. LiDAR point clouds are sequentially captured in tunnel scenes. The repetitive patterns lead to the degradation challenges for localization.

VI. ISSUES

A. LiDAR-Camera Interference

Due to space limitations, we positioned LiDAR closer to the event cameras. As a consequence, the infrared wavelengths emitted by LiDAR directly impinge on the photoreceptor of event cameras, resulting in continuous disturbances and flickering in the captured images and event streams. To address this issue, we implement infrared filters on event cameras to counteract the effect. However, this intervention led to a compromise, resulting in a degradation of the quality of the recorded event data.

B. Event Artifacts

During the night or low illumination scenarios, we observed that when event cameras were directly toward a glowing light source, such as street lights or store lighting, event streams would exhibit persistent flickering and produce artifacts around the light source. This could potentially lead to a distorted view of the observed object. We postulate this phenomenon is related to the inherent principle of event cameras, and presently, there is no known solution to address this issue.

VII. CONCLUSIONS

In this paper, we propose an event-centric autonomous driving dataset generated with multiple sensors across various scenarios for developing SLAM algorithms. All sensors undergo meticulous calibration and are temporally synchronized at the hardware level. We employ the GNSS-RTK/INS navigation system, which provides centimeter-level accuracy, to acquire precise ground-truth of the vehicle. Furthermore, we conduct the evaluation of various state-of-the-art visual and LiDAR SLAM algorithms while identifying their constraints. We hope this dataset could contribute to the development of visual and LiDAR SLAM. In future work, we intend to expand the dataset to encompass additional tasks, including semantics, optical flow, depth estimation, etc.

REFERENCES

[1] X. Liu, W. Wen, and L.-T. Hsu, "Glio: Tightly-coupled gnss/lidar/imu integration for continuous and drift-free state estimation of intelligent vehicles in urban areas," *IEEE Transactions on Intelligent Vehicles*, 2023.

[2] J. Huang, S. Zhao, T. Zhang, and L. Zhang, "Mc-veo: A visual-event odometry with accurate 6-dof motion compensation," *IEEE Transactions on Intelligent Vehicles*, 2023.

[3] P. Yan, W. Wen, and L.-T. Hsu, "Integration of vehicle dynamic model and system identification model for extending the navigation service under sensor failures," *IEEE Transactions on Intelligent Vehicles*, 2023.

[4] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, *et al.*, "Event-based vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 1, pp. 154–180, 2020.

[5] W. Guan, P. Chen, Y. Xie, and P. Lu, "Pl-evio: Robust monocular event-based visual inertial odometry with point and line features," *IEEE Transactions on Automation Science and Engineering*, 2023.

[6] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, 2018.

[7] M. Gehrig, W. Aarents, D. Gehrig, and D. Scaramuzza, "Dsec: A stereo event camera dataset for driving scenarios," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4947–4954, 2021.

[8] G. Mollica, S. Felicioni, M. Legittimo, L. Meli, G. Costante, and P. Valigi, "Ma-vied: A multisensor automotive visual inertial event dataset," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[9] A. Hadviger, V.-J. Štironja, I. Cvišić, I. Marković, S. Vražić, and I. Petrović, "Stereo visual localization dataset featuring event cameras," in *2023 European Conference on Mobile Robots (ECMR)*. IEEE, 2023, pp. 1–6.

[10] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam," *The International Journal of Robotics Research*, vol. 36, no. 2, pp. 142–149, 2017.

[11] S. Klenk, J. Chui, N. Demmel, and D. Cremers, "Tum-vie: The tum stereo visual-inertial event dataset," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 8601–8608.

[12] L. Gao, Y. Liang, J. Yang, S. Wu, C. Wang, J. Chen, and L. Kneip, "Vector: A versatile event-centric benchmark for multi-sensor slam," *IEEE Robotics and Automation Letters*, 2022.

[13] J. Yin, A. Li, T. Li, W. Yu, and D. Zou, "M2dgr: A multi-sensor and multi-scenario slam dataset for ground robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2266–2273, 2021.

[14] J. Jiao, H. Wei, T. Hu, X. Hu, Y. Zhu, Z. He, J. Wu, J. Yu, X. Xie, H. Huang, *et al.*, "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, 2022, pp. 3851–3856.

[15] J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, "Are we ready for autonomous drone racing? the uzh-fpv drone racing dataset," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6713–6719.

[16] J. P. Rodríguez-Gómez, R. Tapia, J. L. Paneque, P. Grau, A. G. Eguíluz, J. R. Martínez-de Dios, and A. Ollero, "The griffin perception dataset: Bridging the gap between flapping-wing flight and robotic perception," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1066–1073, 2021.

[17] J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, "Ddd17: End-to-end davis driving dataset," *arXiv preprint arXiv:1711.01458*, 2017.

[18] Y. Hu, J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, "Ddd20 end-to-end event camera driving dataset: Fusing frames and events with deep learning for improved steering prediction," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.

[19] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "Hats: Histograms of averaged time surfaces for robust event-based object classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1731–1740.

[20] C. Scheerlinck, H. Rebecq, T. Stoffregen, N. Barnes, R. Mahony, and D. Scaramuzza, "Ced: Color event camera dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[21] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 6, pp. 1964–1980, 2019.

[22] T. Fischer and M. Milford, "Event-based visual place recognition with ensembles of temporal windows," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6924–6931, 2020.

[23] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[24] K. Chaney, F. Cladera, Z. Wang, A. Bisulco, M. A. Hsieh, C. Korpela, V. Kumar, C. J. Taylor, and K. Daniilidis, "M3ed: Multi-robot, multi-sensor, multi-environment event dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4015–4022.

[25] A. J. Lee, Y. Cho, Y.-s. Shin, A. Kim, and H. Myung, "Vivid++: Vision for visibility dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6282–6289, 2022.

[26] L.-T. Hsu, F. Huang, H.-F. Ng, G. Zhang, Y. Zhong, X. Bai, and W. Wen, "Hong kong urbannav: An open-source multisensory dataset for benchmarking urban navigation algorithms," vol. 70, no. 4. Institute of Navigation, 2023.

[27] P. De Tournemire, D. Nitti, E. Perot, D. Migliore, and A. Sironi, "A large scale event-based detection dataset for automotive," *arXiv preprint arXiv:2001.08499*, 2020.

[28] S. Kennedy, J. Hamilton, and H. Martell, "Architecture and system performance of span-novatel's gps/ins solution," in *Proceedings of IEEE/ION PLANS 2006*, 2006, pp. 266–274.

[29] J. Kannisto, T. Vanhatupa, M. Hannikainen, and T. Hamalainen, "Software and hardware prototypes of the ieee 1588 precision time protocol on wireless lan," in *2005 14th IEEE Workshop on Local & Metropolitan Area Networks*. IEEE, 2005, pp. 6–pp.

[30] R. Langley, "Nmea 0183: A gps receiver," *GPS world*, vol. 6, no. 7, pp. 54–57, 1995.

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

[32] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza, "How to calibrate your event camera," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1403–1409.

[33] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 6, pp. 1964–1980, 2019.

[34] F. Zhu, Y. Ren, and F. Zhang, "Robust real-time lidar-inertial initialization," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3948–3955.

[35] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[36] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, 2021.

[37] P. Chen, W. Guan, and P. Lu, "Esvio: Event-based stereo visual inertial odometry," *IEEE Robotics and Automation Letters*, vol. 8, no. 6, pp. 3661–3668, 2023.

[38] M. Grupp, "evo: Python package for the evaluation of odometry and slam." https://github.com/MichaelGrupp/evo, 2017.

[39] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2020, pp. 5135–5142.

[40] T. Shan, B. Englot, C. Ratti, and D. Rus, "Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 5692–5698.

[41] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "Fast-lio2: Fast direct lidar-inertial odometry," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2053–2073, 2022.

[42] D. He, W. Xu, N. Chen, F. Kong, C. Yuan, and F. Zhang, "Point-lio: Robust high-bandwidth light detection and ranging inertial odometry," *Advanced Intelligent Systems*, p. 2200459, 2023.

[43] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and systems*, vol. 2, no. 9. Berkeley, CA, 2014, pp. 1–9.

Chen Peiyu received the BSc degrees in automation from the Nanjing University of Science and Technology, China, and MSc degrees in computer control & automation from the Nanyang Technological University, Singapore, in 2020 and 2022, respectively. He is currently working forward the Ph.D. degree at the University of Hong Kong. His research interests include robotics, visual-inertial simultaneous localization and mapping, nonlinear control, and so on.

Guan Weipeng obtained his Bachelor degree in Electronic Science & Technology, as well as Master degree in Control Theory and Control Engineering from the South China University of Technology. He is currently pursuing his PhD degree in Robotics at the University of Hong Kong. He has worked with several reputable organizations, including: Samsung Electronics, Huawei Technologies, The Chinese Academy of Sciences, The Chinese University of Hong Kong, The Hong Kong University of Science and Technology, etc. He has also served as a technical consultant for multiple companies, such as TCL. Moreover, he has published over 60 research articles in prestigious international journals and conferences, as well as holds more than 40 authorized patents. His research interests primarily focus on robotics, event-based VO/VIO/SLAM, visible light positioning, etc.

Huang Feng received his bachelor's degree from Shenzhen University in Automation in 2014 and MSc in Electronic Engineering at Hong Kong University of Science and Technology in 2016. He is a Ph.D. student in the Department of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University. His research interests including localization and sensor fusion for autonomous driving.

Zhong Yihan obtained his bachelor's degree in process equipment and control engineering from Guangxi University in 2020 and a Master's degree from The Hong Kong Polytechnic University (PolyU). He is currently a Ph.D. student at the Department of Aeronautical and Aviation Engineering (AAE) of PolyU. His research interests include factor graph optimization-based collaborative positioning and low-cost localization.

Wen Weisong (Member, IEEE) received a BEng degree in Mechanical Engineering from Beijing Information Science and Technology University (BISTU), Beijing, China, in 2015, and an MEng degree in Mechanical Engineering from the China Agricultural University, in 2017. After that, he received a PhD degree in Mechanical Engineering from The Hong Kong Polytechnic University (PolyU), in 2020. He was also a visiting PhD student with the Faculty of Engineering, University of California, Berkeley (UC Berkeley) in 2018. Before joining PolyU as an Assistant Professor in 2023, he was a Research Assistant Professor at AAE of PolyU since 2021. He has published 30 SCI papers and 40 conference papers in the field of GNSS (ION GNSS+) and navigation for Robotic systems (IEEE ICRA, IEEE ITSC), such as autonomous driving vehicles. He won the innovation award from TechConnect 2021, the Best Presentation Award from the Institute of Navigation (ION) in 2020, and the First Prize in Hong Kong Section in Qianhai-Guangdong-Macao Youth Innovation and Entrepreneurship Competition in 2019 based on his research achievements in 3D LiDAR aided GNSS positioning for robotics navigation in urban canyons. The developed 3D LiDAR-aided GNSS positioning method has been reported by top magazines such as Inside GNSS and has attracted industry recognition with remarkable knowledge transfer.

Hsu Li-Ta received B.S. and Ph.D. degrees in aeronautics and astronautics from National Cheng Kung University, Taiwan, in 2007 and 2013, respectively. He is currently an associate professor with the Department of Aeronautical and Aviation Engineering. The Hong Kong Polytechnic University, before he served as a post-doctoral researcher at the Institute of Industrial Science at the University of Tokyo, Japan. In 2012, he was a visiting scholar at University College London, the U.K. His research interests include GNSS positioning in challenging environments and localization for pedestrians, autonomous driving vehicle, and unmanned aerial vehicle.

Lu Peng obtained his BSc degree in automatic control and MSc degree in nonlinear flight control both from Northwestern Polytechnical University (NPU). He continued his journey on flight control at Delft University of Technology (TU Delft) where he received his PhD degree in 2016. After that, he shifted a bit from flight control and started to explore control for ground/construction robotics at ETH Zurich (ADRL lab) as a Postdoc researcher in 2016. He also had a short but nice journey at University of Zurich & ETH Zurich (RPG group) where he was working on vision-based control for UAVs as a Postdoc researcher. He was an assistant professor in autonomous UAVs and robotics at Hong Kong Polytechnic University prior to joining the University of Hong Kong in 2020.

Prof. Lu has received several awards such as 3rd place in 2019 IROS autonomous drone racing competition and best graduate student paper finalist in AIAA GNC (top conference in aerospace). He serves as an associate editor for 2020 IROS (top conference in robotics) and session chair/co-chair for conferences like IROS and AIAA GNC for several times. He also gave a number of invited/keynote speeches at multiple conferences, universities and research institutes.