IBISCape: A Simulated Benchmark for multi-modal SLAM Systems Evaluation in Large-scale Dynamic Environments

Abanob Soliman1 · Fabien Bonardi1 · Désiré Sidibé1 · Samia Bouchafa1

Received: 30 March 2022 / Accepted: 5 October 2022 / Published online: 19 October 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
The development process of high fidelity SLAM systems depends on their validation upon reliable datasets. Towards this goal, we propose IBISCape, a simulated benchmark that includes data synchronization and acquisition APIs for telemetry from heterogeneous sensors: stereo-RGB/DVS, LiDAR, IMU, and GPS, along with the ground truth scene segmentation, depth maps and vehicle ego-motion. Our benchmark is built upon the CARLA simulator, whose back-end is the Unreal Engine rendering a high dynamic scenery simulating the real world. Moreover, we offer 43 datasets for Autonomous Ground Vehicles (AGVs) reliability assessment, including scenarios for scene understanding evaluation like accidents, along with a wide range of frame quality based on a dynamic weather simulation class integrated with our APIs. We also introduce the first calibration targets to CARLA maps to solve the unknown distortion parameters problem of CARLA simulated DVS and RGB cameras. Furthermore, we propose a novel pre-processing layer that eases the integration of DVS sensor events in any frame-based Visual-SLAM system. Finally, extensive qualitative and quantitative evaluations of the latest state-of-the-art Visual/Visual-Inertial/LiDAR SLAM systems are performed on various IBISCape sequences collected in simulated large-scale dynamic environments.

Keywords Benchmark · Multi-modal · Datasets · LiDAR · Event cameras · Calibration · SLAM

1 Introduction

Autonomous vehicles navigating in unknown and dynamic environments need to rely on accurate perception systems for real-time 3D mapping. These perception systems must function optimally in all weather conditions and situations. That enables the vehicle to make decisions for its passengers or the surrounding pedestrians and cars. To this objective, many novel technologies have been developed over the last decade. Some use vision sensors such as monocular Visual Odometry (VO) [1], which can suffer from estimations up to a scale factor. Innovative solutions to estimate this scale factor by fusion with another sensor like mono/stereo Visual-Inertial Odometry (VIO) [2–4] and RGB-D SLAM [5] to add depth information have been proposed. Other works use LiDAR [6] sensor that provides high precision point clouds mapping of the scene, or use the GPS [7] for localization using satellite signal triangulation. Multi-modal datasets can enrich and broaden the research in the Simultaneous Localization and Mapping (SLAM) field, mainly applied to Autonomous Ground Vehicles (AGVs) navigation in large-scale and dynamic environments. These environments have specific characteristics, such as the dynamic range of the objects’ intensities in the scene. For example: mapping an indoor small room with proper lighting can be of higher quality than mapping a road in a city (large-scale) at night with high intensity fog, rain, and wind (outdoors dynamic environment). The advantages of system multi-modality appear when depending on cameras with high dynamic range, such as the DAVIS sensor and regular low-cost cameras and sensors (IMU/GPS). This multi-modality leads to completing the data shortages during the scene mapping and AGV’s localization.
Nowadays, multi-modal frameworks of sensors have proven to be attracting the attention of many researchers in robotics perception for different tasks such as calibration [8, 9] and odometry [10, 11]. That is due to the fact that heterogeneous sensors that perceive the environment allow the acquisition of complementary information data about the scene. Moreover, sensors multi-modality can also include redundancy such as stereo-DVS or stereo-RGB cameras configurations. Having redundancy in the system setup can improve both the precision and the quality of the collected scene landmarks. Furthermore, some sensors have a high temporal resolution and are sensitive to the scene intensity changes, such as the DAVIS sensor (Event Camera) [12]. While other sensors can efficiently detect and track landmarks and scene features in the 3D spatial domain, such as RGB-D cameras [13] and LiDAR [14].

Simulated datasets [11, 15–18] provide the possibility to have sequences in various complex scenarios. Moreover, setting a hardware data acquisition framework with a specific configuration can be costly and time-consuming and is prone to multiple limitations such as the carrier (car, handheld, drone), weather conditions, sensors configuration, and synchronization. Furthermore, open sourcing the data acquisition APIs with configurable calibration targets can widen the research horizon in multi-modal calibration and sensors synchronization to reach reliable and easy algorithms to implement.

**IBISCape main contributions** to mitigate all these hardware configuration constraints and to facilitate the multi-modal data synchronization and acquisition process are:

- A benchmark of 43 sequences for multi-modal LiDAR/VI-SLAM applications, besides open-sourcing our multi-modal data acquisition APIs.
- A simulated core sensor suite of most visual-inertial sensors used in assessing visual SLAM systems, along with providing high resolution frames of variable quality depending on the dynamic level of the scene. The full sensor setup is represented in Fig. 1.
- A solution to calibrate CARLA [19] RGB and DVS cameras with unknown distortion values.
- An advanced high quality 3-channel events pre-processing layer for frame-based Visual-SLAM systems based on the Event Spike Tensor (EST) representation method [20], that can outperform the latest state-of-the-art methods, especially in dynamic environments with adverse conditions.
- A comprehensive and extensive evaluation of state-of-the-art VI systems using IBISCape sequences collected in dynamically simulated large-scale environments, along with a fair comparison with the publicly available real world SLAM systems evaluation benchmarks.

This article is organized as follows: in Section 2, we discuss the advantages and novelty of our benchmark compared to the related datasets in the field of multi-modal visual localization, including the state-of-the-art V/VI/LiDAR SLAM algorithms. Section 3 explains the data acquisition APIs methodologies and the system calibration in details. Then, an extensive evaluation of the most recent Odometry/SLAM systems using 31 IBISCape SLAM sequences with multiple modalities is represented in Section 4. Finally, in Section 5, we provide concluding remarks about our work including evaluation observations that motivate and push the development process of new multi-modal SLAM techniques forward, especially in dynamic and large-scale environments based on new findings.

2 Related Works

2.1 Existing Datasets

The main goal of our benchmark’s data acquisition APIs is to collect multi-modal sequences suitable for most robotics perception evaluation, including scene understanding, calibration, and complete SLAM systems. IBISCape APIs are highly configurable concerning the intrinsic and extrinsic setup of the sensors and include all CARLA sensors till the version (0.9.11).
Table 1 compares the recent SLAM systems evaluation benchmarks from the sensors types and configuration point of view along with the carrier and ground truth information. Compared to the most recently publicly available benchmarks, IBISCape includes all the sensors needed to evaluate all the state-of-the-art VIO algorithms in any desired configuration including data rates and mono/stereo setups.

Since IBISCape is a simulated benchmark, the GT data for the poses, vehicle controls, scene segmentation, and depth maps are rendered in high precision. This high precision GT data can significantly improve fitting the models of novel data driven VIO architectures that lacks this high quality training data with the real world datasets and hence improving the prediction accuracy.

Table 1: Core sensor suite comparison of latest VIO evaluation benchmarks

| Benchmark Name | RGB          | Depth        | DVS*         | LiDAR*       | IMU          | GT             | Carrier |
|----------------|--------------|--------------|--------------|--------------|--------------|----------------|---------|
| Real World Platform |              |              |              |              |              |                |         |
| TUM-RGBD [21] | Mono@30Hz    | Mono@30Hz    | -            | -            | Accel@500Hz  | MoCap@300Hz    | Handheldc |
| KITTI [22]    | Stereo@15Hz  | -            | -            | 1@10Hz,100m  | 1@100Hz     | GPS            | Car     |
| Malaga Urban [23] | Stereo@20Hz | -            | -            | 5@75Hz,30m   | 1@100Hz     | GPS            | Car     |
| UMic NCLT [24] | Omni@5Hz     | -            | -            | 1@10Hz,100m  | 1@100Hz     | GPS/IMU/LiDAR  | Segway   |
| EuRoC [25]    | Stereo@20Hz  | -            | -            | -            | 1@200Hz     | Laser/Vicon    | MAV     |
| Zurich [26]   | Mono@30Hz    | -            | -            | -            | 1@100Hz     | GPS            | MAV     |
| PennCOSYVIO [27] | Stereo@20Hz | -            | -            | 3@200Hz      | Markers     | Handheld       |         |
| TUM-VI [28]   | Stereo@20Hz  | -            | -            | 1@200Hz      | MoCap       | Handheld       |         |
| Oxford [29]   | Stereo@16Hz  | Mono@30Hz    | -            | -            | 1@500Hz     | Vicon          | Handheld |
| KAIST [30]    | Stereo@10Hz  | -            | -            | -            | 1@200Hz     | GPS            | Car     |
| OIVIO [31]    | Stereo@30Hz  | -            | -            | -            | 1@100Hz     | MoCap          | Handheld |
| UZH-FPV [32]  | Stereo@30Hz  | -            | APS/DVS/IMU  | -            | 1@500Hz     | Laser          | UAV     |
| UMA-VI [33]   | Stereo@25Hz  | -            | -            | -            | 1@250Hz     | Camera         | Handheld |
| Blackbird [34] | Stereo@120Hz | Mono@60Hz    | -            | -            | 1@100Hz     | MoCap          | UAV     |
| VCU-RVI [35]  | Mono@30Hz    | Mono@30Hz    | -            | -            | 1@100Hz     | MoCap          | Handheld |
| TUM-VIE [36]  | Stereo@20Hz  | -            | Stereo ≤10^6e/s | -          | 1@200Hz     | MoCap          | Helmet   |
| Simulated Platform |              |              |              |              |              |                |         |
| VIODE [15]    | Stereo@20Hz  | -            | -            | 1@200Hz     | Simulation   | UAV            |         |
| EVENTSCAPE [11] | Mono@25Hz   | Mono@25Hz    | Mono ≤10^6e/s| -            | -            | Simulation     | Car     |
| Paris-CARLA-3D [16] | 6@2Hz        | -            | 1@10Hz,80m   | -            | -            | Simulation     | Car     |
| KITTI-CARLA [17] | Stereo@10Hz | -            | 1@10Hz,80m   | -            | Sim@1000Hz  | Car            |         |
| SynWoodScape [18] | 5@10Hz       | 5@10Hz       | 5@10Hz       | 1@10Hz      | 1@10Hz      | Sim./GPS@10Hz  | Car     |
| IBISCape b,e  | Stereo@20Hz  | Mono@20Hz    | Stereo ≤10^6e/s| 1@20Hz,100m | 3@200Hz     | Sim./GPS@200Hz | Car     |

* e/s is DVS events per second
b Segmentation frames classify any visible object by displaying it in a different color according to its label (for example, pedestrians in a different color than cars). At the beginning of the simulation, each scene element is created with a tag. In the CARLA simulator, there are 23 segmentation tags with the possibility of adding new tags https://carla.readthedocs.io/en/latest/tuto_D_create_semantic_tags/
c Some sequences where collected using a Robot for SLAM systems evaluation
d Annotations for the dynamic objects in the scene are generated using scripts
e No available DVS and LiDAR sensors detailed specifications
f LiDAR sensor rotation rate [Hz] and range [m]
jection of the pixels acquired from a setup of multiple pinhole cameras without addressing the effect of the actual radial-tangential distortions of CARLA RGB/DVS pinhole cameras.

- A more in-depth performance analysis of the DAVIS sensor V-SLAM algorithms in the dynamic environments with adverse weather conditions is needed.
- Since CARLA is an outdoor environment simulator, the acquired data should imitate that of the real world platforms, as a result the SLAM systems evaluation results on these simulated outdoors datasets should be as close as possible to the evaluations performed using real world datasets.

As an overview of the capabilities of the IBISCape benchmark, we collect simulated sequences on a car equipped with most of the low-cost sensors that can be used in the field of robotics perception. This simulation is thoroughly controlled by an autopilot that navigates the car on traffic-aligned roads. Furthermore, weather and scene constituents, including other cars and pedestrians, can be autonomously controlled within our APIs, resulting in datasets that can contend with the real world benchmarks in the literature.

### 2.2 Dynamic Environment Simulation

Minoda et al. [15] introduce the concept of dynamic scene simulation with moving vehicles. In the works [16–18], sequences are collected with some limited pre-defined weather conditions in CARLA. In our benchmark, we extrapolate the concept of dynamic scene simulation to an entire dynamic environment simulation. This simulation includes moving vehicles and pedestrians, as well as a weather class. The weather ticking function updates the weather states every CARLA world tick with a specific speed factor and update frequency. The weather states that can be controlled are clouds, rain, wind, fog, humidity intensity, and sun angles.

A particular observation from sample IBISCape sequences in Fig. 2 is that our weather update algorithm generates dynamic weather with high intensity fog, rain, and wind with average percentages of 70%, 45%, and 70%, respectively. These dynamic weather conditions result in high trajectory estimation errors due to map loss using existing VIO algorithms. This observation is further verified in Section 4 where we compare the trajectory estimation accuracy in diverse weather conditions.

These weather challenges motivate the development of new VIO techniques based on the hybridization of heterogeneous multi-modal sensors to complete the shortages in the map lost during navigation. In Table 2, a brief comparison regarding the scene dynamic class and the amount of information being processed is represented in the camera’s frame resolution for all benchmarks represented in Table 1. The dynamic level indicators ([C]lear/[M]oderate/[D]ynamic) in Table 2, represent the severity of the [W]eather constituents such as: rain, fog, wind and lack of luminosity besides indicating the amount and speed of moving objects in the [S]cene such as other vehicles and walking pedestrians.

### 2.3 Visual Odometry Techniques

The novel VI systems are divided into two prominent techniques: loosely and tightly coupled fusion methodologies [37]. In loosely coupled fusion [38], the camera is used as a black-box pose estimator [1], and an Extended Kalman Filter or an optimizer is applied to fuse the visual pose estimate with the pre-integrated noisy pose from IMU [39]. Whereas in tightly coupled fusion, the scene descriptors (feature points) from the camera are directly inserted to the filter or optimizer to be fused with the IMU readings of the accelerometer and gyroscope using a model that estimates the pose, visual scale, IMU biases, and also re-project the optimized features to build a precise map of the scene.
In our article we focus on evaluating the most recent VI systems: BASALT and ORB-SLAM3 that share the same mapping layer concept based on ORB descriptors. However, their tracking architectures, IMU pre-integration methodologies, and loop-closing constraints are different. In Section 4.1, a qualitative performance analysis of BASALT and ORB-SLAM3 on multiple IBISCape SLAM sequences is performed.

Since the DAVIS camera is a visual sensor with the highest dynamic range and temporal resolution (up to 1MHz), it can be deemed one of the efficient sensors to deal with high speed robotics scenarios [47] where conventional cameras may fail. Event cameras work on an unconventional caption technology based on the asynchronous detection of image intensity changes through all pixels on the retina. Novel open-source event-based VO algorithms have been developed in the last few years, including: monocular tracking (EVO) [48], mapping (EMVS) [49], and stereo mapping and tracking (ESVO) [47] methods.

However, the current approaches have a computational complexity limitations based on the number of events and the frame resolution. Another DAVIS sensor limitation is the navigation in high rain, dense fog and dark outdoor environments. This limitation is recently studied in [50] by fusing RGB frames with DVS events in an object detection application. In this work, we propose a novel low complexity events-only pre-processing layer that outputs a high quality 3-channel event tensors that can outperform the data driven approach (E2VID) [51], especially in outdoors environments with adverse weather conditions.

The LiDAR sensor operates on an efficient ranging technology that measures the distance to target objects based on the time lapse between the emitted and received laser rays. LiDAR has a sensing range up to 200 meters and a Field Of View (FOV) up to 360°. Due to its operational technology and technical capabilities, the LiDAR can be deemed as the most reliable sensor for Odometry (LOAM) [52] and SLAM (MULLS) [53] tasks in large-scale dynamic environments.

### 3 Core Sensor Suite

Sensors in IBISCape APIs are highly configurable according to the intended mission, we have set an initial sensor configuration for our experiments that can be easily changed. This initial configuration of the IBISCape core sensor suite is given in Table 1.

Table 4 shows the distribution of IBISCape sequences with different sensor modalities and configurations in all dynamic environmental conditions. All datasets in every sensor suite are synchronized during acquisition and time-stamped in nano-seconds for high precision. Moreover, during the sequence collection, the vehicle control forces are saved as normalized vectors within the range [0, 1] and the steering angle in the range [-1, 1].

Simulated LiDAR intrinsics are given in Table 3, where atmosphere_attenuation_rate is a factor that defines the sensor wave length and atmospheric conditions. To ensure a better realistic LiDAR measurements, CARLA defines a random drop proportion of points with a general drop rate factor and a drop rate factor based on the point intensity.

These control commands are normalized with respect to their maximum attained value based on the chosen vehicle dynamics. One of the advantages of CARLA simulator is that we can tune the physical properties of the vehicle and its wheels.

### Table 2 Benchmarks dynamic scene information

| Benchmark       | RGB Resolution [px] | Level\(^a,b\) |
|-----------------|---------------------|--------------|
| TUM-RGBD        | 1x640 × 480         | C            |
| KITTI           | 2x1384 × 1032       | C            |
| Malaga Urban    | 2x1024 × 768        | M            |
| Umich NCLT      | 1x1600 × 1200       | D (W/S)      |
| EuRoC           | 2x752 × 480         | D (S)        |
| Zurich          | 1x1024 × 768        | M            |
| PennCOSYVIO     | 2x752 × 480         | C            |
| TUM-VI          | 2x1024 × 1024       | C            |
| Oxford          | 2x1280 × 960        | M            |
| KAIST           | 2x1600 × 1200       | M            |
| OIVIO           | 2x1280 × 720        | C            |
| UZH-FPV         | 2x640 × 480         | C            |
| UMA-VI          | 2x1024 × 768        | C            |
| Blackbird       | 2x1024 × 768        | C            |
| VCU-RVI         | 1x640 × 480         | D (S)        |
| TUM-VIE         | 2x1024 × 1024       | D (S)        |
| VIODE           | 2x752 × 480         | D (S)        |
| EVENTSCAPE      | 1x512 × 256         | C            |
| Paris-CARLA-3D  | 6x2048 × 2048       | M (W/S)      |
| KITTI-CARLA     | 2x1392 × 1024       | M (W/S)      |
| SynWoodScape    | 1x1024 × 1024       | M (W/S)      |
|                | 2x1280 × 966        | M (W/S)      |
|                | 2x3264 × 2448       | M (W/S)      |
| IBISCape (Ours) | ×1024 × 1024        | D (W/S)      |

\(^a\) C: Clear, M: Moderate, D: Dynamic
\(^b\) W: Weather, S: Scene

The tightly coupled VI systems can be approached using two architectures: filter-based like MSCKF [40] and ROVIO [41], and optimization-based such as VINS-Mono [42], OKVIS [43], and recently ORB-SLAM3 [44] and BASALT [45]. In the work of Delmerico et al. [46], they compare all these VIO algorithms (except the recent works: ORB-SLAM3 and BASALT) in moderately constrained environments with respect to the dynamic level of the scene. They conclude that ROVIO and VINS-Mono are the best performing techniques concerning system latency, robustness, and accuracy.

In our article we focus on evaluating the most recent VI systems: BASALT and ORB-SLAM3 that share the same mapping layer concept based on ORB descriptors. However, their tracking architectures, IMU pre-integration methodologies, and loop-closing constraints are different. In Section 4.1, a qualitative performance analysis of BASALT and ORB-SLAM3 on multiple IBISCape SLAM sequences is performed.
Simulated GPS data is collected with all setups and synchronized with the GT pose. A text file with every frame explains its dataset files contents in detail. The data access manual for the 43 sequences and the acquisition APIs is given in details in the link in the extended data section in Appendix A.

3.1 Cameras Intrinsic & Extrinsic Calibration

One of the advantages of IBISCape benchmark is providing calibration targets for evaluating multi-modal calibration algorithms as well as SLAM systems performance analysis. The more erroneous the calibration parameters, the more incorrect pose is estimated. Although the intrinsic calibration parameters of CARLA cameras can be configured directly in the APIs, there is no direct way to set the lens distortion coefficients till version (0.9.11). Consequently, we propose introducing the first calibration targets (Checkerboard \((7 \times 7)\) and AprilGrid \((6 \times 6)\)) to one CARLA map (Town 3). Moreover, to excite all angles, especially the pitch and roll angles which are not easy to be simulated in a car, we introduce artificial bumps in the form of bubbles and waves, as shown in Fig. 3.

Instead of simulating blinking LED lights that cannot be used in a multi-modal calibration framework that includes RGB frames, we use Kalibr [54] to calibrate the stereo RGB cameras and the stereo DVS sensors after performing a frame reconstruction from events using the generic framework E2CALIB [55] (sample in Fig. 4). Since active illumination cannot be used to calibrate conventional cameras such as mono/stereo-RGB cameras, E2CALIB with the traditional calibration targets makes it possible to calibrate DVS sensors as any conventional camera. Hence, all cameras’ intrinsic and extrinsic parameters in a multi-modal framework can be calibrated irrespective of their caption technology, i.e., frames or events.

All IBISCape cameras operate on a global shutter mode with a FOV of \(90^\circ\). Table 5 shows the specific DVS sensors parameters set during our simulations, including the positive/negative thresholds associated with an increment in brightness change along with their white noise standard deviation for positive/negative events.

The known camera model is a pinhole model with unknown distortion parameters for RGB and DVS cameras. We calibrate our cameras using Kalibr pinhole-radial-tangential and

---

**Table 3** Simulated LiDAR characteristics

| Parameter                              | Value set in CARLA |
|----------------------------------------|--------------------|
| channels                               | 64                 |
| range                                  | 100.0 \([\text{m}]\) |
| points_per_second                      | 5e6                |
| rotation_frequency                     | 20 \([\text{Hz}]\)  |
| upper_fov                              | 15.0°              |
| lower_fov                              | \(-25.0°\)         |
| horizontal_fov                         | 360.0°             |
| atmosphere_attenuation_rate            | 0.004              |
| dropoff_general_rate                   | 0.45               |
| dropoff_intensity_limit                | 0.8                |
| dropoff_zero_intensity                 | 0.4                |
| noise_stddev                           | 0.0                |

**Table 4** IBISCape sequences & sensor setup

| Acquisition Sensor Suite | Clear | Mod. | Dyn. |
|--------------------------|-------|------|------|
| Calibration              |       |      |      |
| IMU                      | 2     | –    | –    |
| 2xRGB+IMU (SVI)          | 2     | –    | –    |
| 2xDVS+2xIMU (ESVI)       | 2     | –    | –    |
| RGB-D                    | 2     | –    | –    |
| LiDAR+2xRGB              | 2     | –    | –    |
| Full Sensor Setup        | 2     | –    | –    |
| SLAM                     |       |      |      |
| 2xRGB+IMU (SVI)          | 2     | 2    | 3    |
| 2xDVS+2xIMU (ESVI)       | 2     | 2    | 2    |
| RGB-D                    | 2     | 2    | 2    |
| LiDAR+2xRGB              | 2     | 2    | 2    |
| Full Sensor Setup        | 2     | 2    | 2    |
| Total = 43               | 22    | 10   | 11   |

All calibration evaluation sequences are collected in Clear weather only to achieve the best frame quality for robust and precise calibration results.

Simulated GPS data is collected with all setups and synchronized with the GT pose. A text file with every framework explains its dataset files contents in detail. The data

---

**Fig. 3** Excitation of the vehicle pitch and roll angles using bubble bumps for reliable calibration results using Kalibr
pinhole-equidistant distortion models. The calibration process is validated based on two criteria:

- The estimated stereo baselines (extrinsics) compared to the GT values set in our acquisition APIs (see Tables 6 and 7).
- The quality of the optimization process that can be determined from the pixels re-projection errors and the number of optimization constraints (see Table 8).

Based on these two criteria and the obtained results, we can conclude that the pinhole-radtan camera-distortion model best fits both RGB and DVS cameras simulation in CARLA. This conclusion is due to its lowest re-projection errors and highest stereo baseline estimation accuracy.

We provide all the calibration configuration files and various ROS scripts to convert the raw dataset files to rosbag and .h5 file formats for Kalibr and E2CALIB frameworks.

### 3.2 Simulated IMU Calibration

IBISCape novel calibration methodology is based on fixing the high quality calibration target in the center of the frame and moving the vehicle towards it with a complete manual control. Furthermore, adding bumps in its way in the form of a big wave and spherical bubbles, can ensure the sufficient excitation of the inertial sensor for precise system (IMU+cameras) calibration.

In CARLA, IMU measurements are modeled as most low-cost real world IMUs containing a particular bias $b$ and white gaussian noise $n$. Thus, the GT angular velocities $\omega$ and linear accelerations $a$ in the IMU frame are modeled as

$$\omega_{GT} = \omega_{gyro} - b_{g} - n_{g}, \quad a_{GT} = a_{accel} - b_{a} - n_{a}.$$  

The standard deviation $\sigma_{\omega g}, \sigma_{b a}, \sigma_{n g}, \sigma_{b g}$ values are given in Table 9, and Allan Deviation plots are given in Fig. 5 calibrated by the IMU Still Calibration Tool in [33] using a 300 [hrs] of IMU simulated sequence.

In Table 9, IMU still calibration shows a remarkable difference between the GT values we set in CARLA and the estimated ones. This is an expected observation, since in a simulation environment the standard deviation and bias values set as GT are the dynamic IMU covariance values which can’t be estimated by the static Allan deviation method [56].

Till CARLA version 0.9.11, the acceleration bias standard deviation value cannot be manually set within the simulation. As a result, an accurate and reliable IMU still

### Table 5 Simulated DVS characteristics

| Parameter                  | Value set in CARLA |
|---------------------------|--------------------|
| +ve/-ve_threshold         | 0.3                |
| sigma_+ve/-ve_threshold   | 0.0                |
| refractory_period_ns      | 0.0                |
| log_eps                  | 0.05               |

### Table 6 Stereo DVS sensors and RGB Cameras intrinsic parameters estimation using Kalibr

| Camera Model | $f_x$ | $f_y$ | $c_x$ | $c_y$ | $k_1$ | $k_2$ | $k_3$ | $k_4$ |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| DVS          |       |       |       |       |       |       |       |       |
| cam0-radtan  | 517.07| 517.59| 506.41| 513.27| -2.32e-3| 7.12e-4| 1.97e-4| -8.87e-4|
| cam1-radtan  | 517.45| 517.79| 504.48| 512.89| -8.34e-4| -1.08e-3| 9.11e-5| -1.28e-3|
| cam0-ear        | 375.85| 373.29| 573.79| 513.44| -0.0122 | 1.9684 | -3.8539 | 2.82   |
| cam1-ear        | 370.65| 368.16| 572.65| 513.1 | 0.2912  | 0.2954 | -0.2626 | 0.2344 |
| GT cam0/cam1  | 512.0 | 512.0 | 512.0 | 512.0 | -       | -       | -       | -      |
| RGB          |       |       |       |       |       |       |       |       |
| cam0-radtan  | 513.55| 513.07| 511.0 | 510.26| 1.92e-3 | -1.83e-3| -8.5e-4 | 2.1e-4 |
| cam1-radtan  | 512.51| 512.87| 512.0 | 512.1 | -2.75e-3| 3.16e-3 | 3.7e-4  | -3.8e-4|
| cam0-ear        | 511.11| 511.18| 511.24| 511.0 | 0.3533  | 0.065  | 0.181  | -0.058 |
| cam1-ear        | 512.41| 512.31| 512.0 | 512.34| 0.3269  | 0.1084 | 0.1495 | -0.0505|
| GT cam0/cam1  | 512.0 | 512.0 | 512.0 | 512.0 | -       | -       | -       | -      |

$f_x$ and $f_y$, $c_x$ and $c_y$ are the focal lengths and principal point coordinates, respectively. $k_1, k_2, k_3, k_4$ are the radial and tangential distortion coefficients, respectively. Calibration is performed using the Checkerboard target.
3.3 Inter-sensor Extrinsic Parameters

The CAD model of the GT extrinsic relation between all the sensors in a full sensor setup is represented in Fig. 6. Calibration is essential to obtain simulated datasets with usable IMU measurements. We evaluate the IBIScape Stereo-Visual Inertial (SVI) sequences using the still calibration values for the IMU noises.

Table 7 Estimation quality is further validated by comparison to the stereo baselines set in CARLA

| Camera Model     | Stereo Baseline (t [m]) |
|------------------|-------------------------|
| DVS              |                         |
| cam0-radtan      | q=[3.18e-4 -1.77e-3 3.17e-5 1] |
| cam1-radtan      | t=[-0.1986 0.0009 0.0131] |
| cam0-equid       | q=[3.69e-4 -1.16e-3 -1.14e-4 1] |
| cam1-equid       | t=[-0.1902 0.003 0.0115] |
| GT cam0/cam1     | q=[0 0 0 1], t=[-0.2 0 0] |
| RGB              |                         |
| cam0-radtan      | q=[-0.0021 1.45e-4 4e-5 1] |
| cam1-radtan      | t=[0.0103 -0.004] |
| cam0-equid       | q=[-0.0014 -3.5e-4 -2e-5 1] |
| cam1-equid       | t=[-0.413 0.005 0.01] |
| GT cam0/cam1     | q=[0 0 0 1], t=[-0.4 0 0] |

cam0 and cam1 are the left and right cameras, respectively. Bold denotes an accurately estimated stereo baseline.

Table 8 Re-projection errors & optimization constraints

| Camera Model     | Re-projection errors [px.] | Edges |
|------------------|----------------------------|-------|
| DVS              |                            |       |
| cam0-radtan      | [0.000132, -0.000016]      | 61397 |
| cam1-radtan      | [0.000163, -0.000009]      | 61397 |
| cam0-equid       | [-0.000740, 0.000008]      | 61397 |
| cam1-equid       | [-0.000703, 0.002294]      | 61397 |
| RGB              |                            |       |
| cam0-radtan      | [-0.000034, -0.000007]     | 29008 |
| cam1-radtan      | [0.0000034, 0.0000007]     | 29008 |
| cam0-equid       | [-0.000067, 0.000001]      | 29008 |
| cam1-equid       | [0.000064, 0.0000007]      | 29008 |

Table 9 Simulated IMU still calibration results

| Parameter     | CARLA               | Calibrated |
|---------------|---------------------|------------|
| $\sigma_{ba}$ | [m/s$^2$/$\sqrt{Hz}$] | -          | 4.983e-3  |
| $\sigma_{va}$ | [m/s$^3$/$\sqrt{Hz}$] | 7e-2       | 3.167e-6  |
| $\sigma_{bg}$ | [rad/s/$\sqrt{Hz}$]  | -          | 2.839e-4  |
| $\sigma_{vg}$ | [rad/s$^2$/$\sqrt{Hz}$] | 4e-3       | 1.916e-7  |

The axes shown on the camera’s center-line are given for the visual sensors only: RGB, DVS, Depth, Segmentation. All IMUs axes conventions are similar to that shown on the IMU0 center-line. Axes color and direction conventions coincide precisely with the Top view CAD model in Fig. 1.

There is no orientation change between cameras i.e. $\bar{\theta} = [0, 0, 0]$ and all cameras have the relative rotation $q_{cam_i} = [0, 0, 0, 1]$. In Table 10, we give the exact GT values for each sensor location with respect to the IMU0 (body) axes.

In the RGB-D sensor setup, the simulated RGB and Depth cameras have a concentric configuration where both the focal centers are coincided. Moreover, IBIScape data acquisition APIs are written to be highly configurable with respect to the inter-sensor extrinsic parameters with the ease of adding and removing sensors.

4 Evaluation

4.1 Efficient VI Systems

We use our IBIScape sequences to evaluate state-of-the-art monocular and stereo VI-SLAM algorithms which are ORB-SLAM3 [44] and BASALT VIO [45]. Their choice is because they are the latest state-of-the-art SLAM (ORB-SLAM3) and VIO (BASALT) algorithms. Accordingly, their extensive evaluation on new large-scale and dynamic environment (scene and weather) IBIScape sequences can facilitate detecting their limitations and performance regarding their accuracy and robustness.

BASALT uses a sparse set of FAST keypoints, tracks them between consecutive frames based on optical flow (KLT) [57], and uses a pyramidal resolution method to ensure reliable and robust tracking in large-scale displacements tracking. Two layers for local bundle adjustment and global pose graph optimization are implemented for precise localization, mapping, and loop-closing. Furthermore, partial marginalization non-linear factors are applied to remove the IMU and feature outlier measurements for constant latency localization.

ORB-SLAM3 is developed to withstand a prolonged duration of low visual information. When a map is disturbed, it initiates a new map that will be smoothly merged with previous maps when revisiting similarly mapped areas. That results in a robust system that operates in dynamic environments and is much more accurate and robust than previous approaches.

Both ORB-SLAM3 and BASALT relate to the optimization-based tightly-coupled fusion stereo VI systems. In Section 4.2.1, a detailed evaluation of their performance in large-scale dynamic environments is performed. In Section 4.2.3, our stereo event cameras configuration is used to evaluate the latest open-source stereo DVS.
mapping and tracking method ESVO [47], along with the novel event-based mapping method EMVS [49]. In the work of E2VID [51], the authors evaluate their event-based frame reconstruction method in the application of monocular VIO, and their method has shown superior performance compared to the other frame-based and event/frame-based methods in comparison. However, these experiments are carried out on indoor sequences with ideal environmental conditions.

In our evaluations on IBISCape, inspired by the work of E2VID, we extrapolate these experiments to include stereo V-SLAM systems in outdoors dynamic environments. Then, we propose an alternative 3-channel event-based frame reconstruction layer that can outperform the quality of E2VID visually as shown in Figs. 11 and 12 and numerically as given in Table 12.

Table 10  IBISCape full sensor setup extrinsics

| Sensor              | X,Y,Z Translation to IMU0 [m] |
|---------------------|-------------------------------|
| Left RGB0           | [0.0, -0.2, -2.8]             |
| Right RGB1          | [0.0, 0.2, -2.8]              |
| left DVS0 + IMU1    | [0.0, 0.1, -2.8]              |
| Right DVS1 + IMU2   | [0.0, -0.1, -2.8]             |
| RGB-D cameras       | [0.0, 0.0, -2.8]              |
| GPS                 | [0.2, 0.0, -2.8]              |
| LiDAR               | [0.0, 0.0, -3.0]              |
| GT Segmentation     | [0.0, 0.0, -2.8]              |
| GT Pose             | [0.0, 0.0, 0.0]               |

Fig. 5  IMU log-log scaled plot of Allan-variances over the cluster time. We calculate the IMU noises $a_i$ and $a_n$ at cluster times 1 sec. and 3 sec. with slopes $+1/2$ respectively.

Fig. 6  Full sensor setup CAD model (Front view)
In Section 4.2.5, an extensive in-depth evaluations of the latest LiDAR based Odometry/SLAM algorithms MULLS [53] and an advanced version of LOAM [52] in dynamic environments with adverse weather conditions is provided. All LiDAR SLAM sequences simulate multiple loop closure detection situations. Section 4.2.6 compares the evaluation process which is run on 31 IBISCape sequences given in Table 11 simulated in various large-scale dynamic environments with the real world evaluations on the state-of-the-art benchmarks in literature.

### 4.2 Performance Analysis

To ease the comparison with the previous and future SLAM system benchmarks, the performance analysis is done using the two known SLAM systems evaluation metrics defined in [58]:

(i) The RMS of Absolute Trajectory Error (ATE) for all (n) estimated poses, and defined as:

$$\text{ATE}(\hat{T}^{(1:n)}, T_{gt}^{(1:n)}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} ||t_i||^2 [m]},$$

where $\hat{T}^{(1:n)}, T_{gt}^{(1:n)} \in SE(3)$ are the estimated and ground truth trajectories, respectively. $t_i \in R(3)$ is the translation vector of the absolute trajectory error $E_i$ at time step $i$ where $E_i(R, t_i) = T_{gt}^{-1} T_{rel} \hat{T}_i \in SE(3)$, and $T_{rel}$ is rigid-body transformation corresponding to the least-squares solution that maps the $\hat{T}$ trajectory onto the $T_{gt}$ trajectory calculated by optimization.

(ii) Relative Pose Error (RPE) at every i-th frame, and defined as:

$$\text{RPE}(\hat{T}^{(1:n)}, T_{gt}^{(1:n)}) = ||\delta t_i|| [m].$$

where $\delta t_i$ is the translation vector of the relative pose error $e_i(\delta \theta_i, \delta t_i) = (T_{gt(i)}^{-1} T_{rel(i)} \hat{T}_i) \in \mathfrak{se}(3)$ at time step $i$ with a fixed time interval $\Delta$ for our local trajectory increments. For the orientations RPE values are given in degrees, we use the same formula after replacing the translation vector $\delta t_i$ with the rotation part $\delta \theta_i$ in $e_i$ by applying the vee operator to the skew-symmetric error matrix:

$$\text{RPE}(\hat{T}^{(1:n)}, T_{gt}^{(1:n)}) = ||[\delta \theta_i]_\vee|| [rad]$$

We discuss a thorough descriptive and analytical evaluation for the latest state-of-the-art SLAM systems in the following sub-sections. The descriptive and analytical studies for every sensor setup raise the confidence in the novelty and usability of the IBISCape benchmark, using the calibrated RGB and DVS cameras distortion parameters along with the IMU still calibration.

| Table 11 IBISCape sequences specifications |
|--------------------------------------------|
| Sequence | Specifications |
| | Length [m] | Duration [sec] | Size | Loop | Closure |
| Full Setup | Clear-1 | 214.6313 | 60.52 | 1211 | – |
| | Clear-2 | 251.0401 | 70.55 | 1412 | – |
| | Moderate-1 | 368.9815 | 71.08 | 1422 | – |
| | Moderate-2 | 104.5391 | 29.92 | 599 | – |
| | Dynamic-1 | 217.9678 | 70.24 | 1405 | – |
| | Dynamic-2 | 61.2707 | 23.38 | 468 | – |
| SVI Setup | Clear-1 | 140.2081 | 70.16 | 1404 | – |
| | Clear-2 | 141.1631 | 71.45 | 1429 | ✓ |
| | Moderate-1 | 253.8933 | 64.40 | 1288 | – |
| | Moderate-2 | 330.6167 | 85.98 | 1719 | – |
| | Dynamic-1 | 248.6546 | 72.35 | 1448 | – |
| | Dynamic-2 | 289.0983 | 74.01 | 1480 | – |
| Accident | Clear-1 | 23.6777 | 6.13 | 123 | – |
| | | | | | |
| RGB-D Setup | Clear-1 | 223.1038 | 74.95 | 1500 | – |
| | Clear-2 | 360.5324 | 89.55 | 1792 | – |
| | Moderate-1 | 209.1469 | 72.65 | 1454 | – |
| | Moderate-2 | 233.6294 | 70.00 | 1401 | – |
| | Dynamic-1 | 208.0217 | 65.50 | 1311 | – |
| | Dynamic-2 | 406.3022 | 75.65 | 1514 | – |
| ESVI Setup | Clear-1 | 116.3213 | 23.98 | 4672 | – |
| | Clear-2 | 251.5679 | 60.61 | 12123 | – |
| | Moderate-1 | 264.6535 | 72.91 | 13980 | – |
| | Moderate-2 | 274.8627 | 61.13 | 4390 | – |
| | Dynamic-1 | 333.2455 | 71.54 | 13997 | ✓ |
| | Dynamic-2 | 15.0866 | 23.87 | 11771 | – |
| LiDAR Setup | Clear-1 | 38.2770 | 13.85 | 278 | ✓ |
| | Clear-2 | 64.8373 | 22.75 | 456 | ✓ |
| | Moderate-1 | 111.0011 | 37.50 | 751 | ✓ |
| | Moderate-2 | 235.0245 | 76.90 | 1539 | ✓ |
| | Dynamic-1 | 81.3070 | 28.90 | 579 | ✓ |
| | Dynamic-2 | 146.7856 | 52.25 | 1046 | ✓ |

* Number of camera frames, events packets or laser scans
To ensure a fair evaluation process, all the data acquisition APIs and benchmarking experiments are executed on a 16 GB RAM laptop computer running 64-bit Ubuntu 20.04.3 LTS with AMD(R) Ryzen 7 4800h ×16 cores 2.9 GHz processor and a Radeon RTX NV166 Renoir graphics card.

4.2.1 SVI Setup Evaluation

IBISCape Stereo Visual Inertial (SVI) sequences push one of the limits of the ORB-SLAM3 system as mentioned in [44], which is the IMU initialization of planar motion of vehicles like cars. In Fig. 7(A), this limitation constraint was further tested using the Dynamic 1 sequence with significantly dimmed light and rapid scene motions. The ORB-SLAM3 IMU initialization failed to start with the mapping layer. This failure has led to a significant trajectory drift due to the map loss. This IMU initialization failure is also observed in the Dynamic 2 sequence with the BASALT system.

In Table 12, the other sequences, Clear 1,2, Moderate 1,2, show superior performance for the trajectory estimation using the ORB-SLAM3 system over BASALT based on both overall ATE and incremental RPE values. IBISCape SVI sequences are provided in raw and rosbag formats, along with the evaluation configuration files .json and .yaml for BASALT and ORB-SLAM3.

Although sharing ORB keypoints for loop-closing in BASALT and scene descriptors in ORB-SLAM3, BASALT has shown superior accuracy and robustness regarding the visual-inertial sub-system than an early version of ORB-SLAM [45]. This better performance is due to the inertial layer of BASALT that utilizes recovered non-linear factors summarizing IMU and visual tracking on the higher layer of VIO.

However, the latest version of ORB-SLAM3 proved to be much more accurate than BASALT during evaluation on most of the IBISCape sequences, as shown in the performance analysis results in Table 12. Despite the superior performance of ORB-SLAM3 over BASALT, we note that the trajectory estimation is much faster in BASALT than in ORB-SLAM3. This evaluation observation validates the proposed comparison in Tab. (I) in [44].

4.2.2 RGB-D Setup Evaluation

One of the advantages of IBISCape sequences is the variety of its sensors’ multi-modality. While SVI sequences can provide the scene depth information by stereo RGB cameras and augment the scale factor using the inertial measurements, IBISCape RGB-D sequences offer another sensor modality to measure the scene depth: the depth camera. After alignment with the GT and scale factor recovery using the GPS measurements, we evaluate two ORB-SLAM3 algorithms: the monocular RGB and the RGB-D SLAM systems. In Fig. 7, it is evident that adding the depth information results in more accurate trajectory estimation with a minor map loss in dynamic weather.

We notice this map loss clearly with the mono-RGB using Dynamic 1,2, Moderate 1,2 sequences. However, in clear weather sequences Clear 1,2, the monocular RGB SLAM can outperform the RGB-D SLAM as seen in Table 12 with respect to the ATE values. IBISCape RGB-D sequences are provided in raw format with the RGB and Depth cameras association.txt file for every sequence, along with the evaluation configuration .yaml files for ORB-SLAM3 RGB-D and mono-RGB systems.

4.2.3 ESVI Setup Evaluation

IBISCape event-based sequences are designed to address two corner case scenarios introduced to the event-based monocular/stereo VO algorithms. The first scenario is the planar motion in large-scale environments; this scenario leads to the generation of millions of events fired at locations in the scene that can be tens of meters away from each other. These environments consume much time to reconstruct a map, leading to significant processing time gaps between the tracking and mapping layers of the odometry algorithm.

As a result, the ESVO [47] experiments on IBISCape sequences fail during the trajectory estimation giving an error indicating inconsistency between the tracking and mapping layers, although maps initialize successfully. Accordingly, in order to assess ESVO, the evaluation is run with down sampling the rosbags playing time by a factor of 0.0005. This leads to high system latency during evaluations, for example: a 23 seconds sequence needs (23/0.0005) seconds to be evaluated i.e. nearly 12 hours.

Despite this extremely high system latency during evaluations, Fig. 8 reports noticeable low ATE values when compared to the other frame-based SLAM system. During EMVS [49] event-based mapping evaluations, we notice a significant map loss due to high fog as seen in sequence Dynamic 1 of the FULL setup, or rapid motions as seen in sequence Clear 2 of the ESVI setup.

The second scenario is the dynamic weather, including fog and rain droplets that can cause random asynchronous events. Hot pixels in real world DVS can be hardware defects, but simulated DVS can indicate random rain/fog firings in CARLA. Applying a hot pixel filter can detect and remove these unexpected events. Figure 9 shows a sample of the hot pixels removed due to fog and rain. Removing hot pixels in DVS sensor is based on two criteria: the highest N pixels firing most events, or the pixels firing greater than \( n_e \times \sigma \) events. Where \( n_e \) is event...
[A] SVI Setup Sequences

[B] RGB-D Setup Sequences
occurrence standard deviation multiplier and $\sigma$ is the event occurrence standard deviation.

The second corner case effect is witnessed during evaluating EMVS [49], where black dense blocks of point cloud points are accumulated on the trajectory during navigation in heavy rain and fog.

In order to evaluate IBISCape stereo-DVS calibration parameters, we construct stereo-RGB frames from the events using the E2VID pre-trained model [51]. We assess the ORB-SLAM3 stereo RGB SLAM system on these reconstructed frames. Despite filtering the scene from noisy events resulting from fog and rain droplets, Table 12 shows a complete failure in trajectory estimation in the case of Dynamic 1 sequence denoted by italic font. Due to the dynamic weather conditions and rapid system dynamics E2VID frame reconstruction fails with most of IBISCape sequences. Consequently, another event-based frame reconstruction method is needed to consider these two corner cases without losing the high dynamic range that DVS sensors can provide.

To this extent, we propose a novel DVS sensor pre-processing layer called the Event 3-Channel Tensor (E3CT) based on the Event Spike Tensor (EST) representation method [20]. The E3CT is a pre-processing layer that combines the benefits of both EST and Histograms of Averaged Time Surfaces (HATS) [59] concepts, this allows its modeling to be simple with reliable time information. The E3CT concept is illustrated in Algorithm 1, where in line (19) an event volume is created in the form of a 4D tensor $(n,2,c,h,w)$, then in line (20) we update every event’s $e(x,y,p,t)$ linearly weighted histogram. Where $t_i = \frac{i}{\delta}$ is the sequential number of the current event time, $c_i = \left(\frac{i - t_0}{\delta}\right) \mod c$ the sequential number of the current event nano-time bin, $[\_]_c$ is the closest nano-time bin number to the right and left, $r_i = c \times \frac{i - t_0}{\delta} - 0.5$ is the relative temporal distance to the centre of the corresponding nano-time bin, $t_i = \frac{i}{\delta}$ is the sequential number of the current nano-time bin, $t_0$ is the initial event timestamp [nsec], $\delta = \frac{2\pi}{\Delta}$ is the nano-time bin interval [nsec].

For a visualized illustration for the nano-time bins concept see Fig. 10, we show an example of two positive events fall with the only two possibilities. The first falls exactly on the nano-time bin edge between the two nano-tbin $n$, $n+1$. This event will contribute equally to both first and middle channels of this E3CT. The second falls totally in the n-1 nano-tbin, so it will have a high contribution weight to the last channel and a low contribution weight to the middle and no contribution to the first channel. As, the number of nanobins increase the number of contribution slices will increase and the quality of the E3CT will improve.

In Fig. 11, we represent six cases to compare DVS sensor event arrays preprocessed using the E3CT and E2VID methods without any post processing. These corner cases show the effectiveness of the E3CT to construct frames that preserve the scene artifacts even in the harshest weather conditions. This is due to the high sampling duration (1 sec) set during the E3CT construction. Selecting high sampling duration along with a hot pixel filter can efficiently suppress the rain and fog events contributions to the E3CT as seen with Cases 3-6. Figure 12 gives an in-depth illustration to the quality of the constructed E3CTs on sample artifacts from IBISCape sequences.

The only case where the E2VID can evaluate event-based SLAM systems performance in trajectory estimations is clear weather with a low dynamic scene level as in the Clear 1 sequence. A rapid change in the scene can cause an instantaneous map loss that affects the whole trajectory estimation even if the weather is clear, as in the Clear 2 sequence. IBISCape ESVI sequences are provided in raw .npz (NumPy arrays) and bag formats for the stereo-DVS events, along with timestamps.csv file includes the start and end timestamps for every time surface. Besides, the E2VID grayscale frames reconstruction results and the locations of the hot pixels for the stereo-DVS are also provided.

4.2.4 FULL Setup Evaluation

The most significant contribution of the IBISCape benchmark is its FULL sensor setup sequences, where all sequences contain a combination of all the available sensors simulated in clear/moderate/dynamic weather environments. As a result, a complete comprehensive quantitative evaluation of all the SLAM systems mentioned in the previous sub-sections can be compared on the same sequence for every specific weather condition, as seen in Fig. 13. We represent in Table 12 an extensive qualitative assessment of the state-of-the-art SLAM systems based on the six FULL setup sequences. Regarding Clear 1,2, the trajectory estimation is aligned with the ground truth profile until a rapid motion occurs and the events map is disturbed. Each IBISCape FULL setup sequence is equipped with all the data formats as given with the specialized setup sequences.

Based on all the evaluation observations, we can conclude that the current pre-trained models to reconstruct frames can be unreliable specially in dynamic weather and large-scale environments as represented in Fig. 11. This gives the most important advantage of IBISCape benchmark providing thousands of
Table 12 ORB-SLAM3 (SVI, RGB-D, mono-RGB, stereo-RGB(E2VID, E3CT)), BASALT, MULLS and A-LOAM performance analysis based on both ATE and RPE evaluation metrics using IBIScape sequences in all simulated dynamic environments

| Sequence | Method 1 | Method 2 |
|----------|----------|----------|
| FULL Setup - I ORB-SLAM3 - SVI | | BASALT |
| Clear-1 | 13.7184 | 18.7082 |
| Clear-2 | 12.3043 | 12.0170 |
| Moderate-1 | 32.8159 | 49.9634 |
| Moderate-2 | 3.7829 | 11.8746 |
| Dynamic-1 | 17.2807 | 0.2190±0.1318 |
| Dynamic-2 | 4.9187 | 0.1645±0.4232 |
| FULL Setup - II ORB-SLAM3 - RGB-D | | ORB-SLAM3 - mono-RGB |
| Clear-1 | 20.2653 | 3.5142 |
| Clear-2 | 14.8820 | 18.4484 |
| Moderate-1 | 40.1021 | 67.1074 |
| Moderate-2 | 3.5969 | 0.4995±0.0667 |
| Dynamic-1 | 11.5730 | 22.2793 |
| Dynamic-2 | 15.5917 | 17.2632 |
| FULL Setup - III E2VID-SVO | | E3CT-SVO (Ours) |
| Clear-1 | 84.7657 | 70.9616 |
| Clear-2 | 156.8587 | 103.4646 |
| Moderate-1 | 157.9537 | 203.4386 |
| Moderate-2 | 29.1791 | 37.1249 |
| Dynamic-1 | 235.7885 | 91.2599 |
| Dynamic-2 | 52.1609 | 0.4045±0.0208 |
| SVI Setup ORB-SLAM3 - SVI | | BASALT |
| Clear-1 | 3.1262 | 12.2769 |
| Clear-2 | 1.6666 | 0.1859±0.0234 |
| Moderate-1 | 11.5160 | 4.0626 |
| Moderate-2 | 8.8561 | 0.0514±0.0845 |
| Dynamic-1 | 50.7355 | 0.1290±0.0712 |
| Dynamic-2 | 9.5503 | 0.1087±0.5707 |
| Accident | 16.2158 | 27.4657 |
| RGB-D Setup ORB-SLAM3 - RGB-D | | ORB-SLAM3 - mono-RGB |
| Clear-1 | 20.9667 | 0.3587±0.1527 |
| Clear-2 | 5.9339 | 0.1519±0.3228 |
| Moderate-1 | 2.8882 | 12.4161 |
| Moderate-2 | 13.5358 | 0.1853±0.1512 |
| Dynamic-1 | 8.7264 | 0.2324±0.4938 |
| Dynamic-2 | 12.0050 | 0.2414±0.1277 |
| ESVI Setup E2VID-SVO | | E3CT-SVO (Ours) |
| Clear-1 | 62.2875 | 0.4169±0.1158 |
| Clear-2 | 121.0946 | 0.4808±0.9041 |
| Moderate-1 | 79.8216 | 2.5671 |
| Moderate-2 | 9.4286 | 0.1005 |
| Dynamic-1 | 65.8318 | 15.1911 |
| Dynamic-2 | 106.0616 | 121.4955 |
| LiDAR Setup MULLS (with/without Loop Closure) | | A-LOAM |
| Clear-1 | 0.5593 | 0.1393±0.2929 |
| Clear-2 | 5.5411 | 0.1431±0.0161 |
| 5.5662 | 0.1431±0.0161 |
Table 12 (continued)

| Sequence  | Method 1                | Method 2                |
|-----------|-------------------------|-------------------------|
|           | ATE [m] | RPE [m] | RPE [deg] | ATE [m] | RPE [m] | RPE [deg] |
| Moderate-1 | 0.9192 | 0.2061±0.0791 | 0.9073±0.9341 | 3.6931 | 0.1501±0.0201 | 0.6811±0.6512 |
|           | 0.9431 | 0.2061±0.0782 | 0.9063±0.9341 |          |          |          |
| Moderate-2 | 0.5391 | 0.1861±0.1122 | 0.3613±0.7033 | 5.9511 | 0.1531±0.0141 | 0.2761±0.5062 |
|           | 0.5371 | 0.1871±0.1121 | 0.3611±0.7041 |          |          |          |
| Dynamic-1 | 0.5711 | 0.1882±0.0693 | 1.1321±0.9351 | 2.3042 | 0.1391±0.0171 | 0.8871±0.6581 |
|           | 0.5492 | 0.1882±0.0693 | 1.1321±0.9351 |          |          |          |
| Dynamic-2 | 1.6193 | 0.1821±0.0941 | 1.2051±0.9251 | 3.0251 | 0.1401±0.0151 | 0.8712±0.6642 |
|           | 1.6483 | 0.1821±0.0941 | 1.2051±0.9244 |          |          |          |

Relative Pose Error (RPE) is formulated in terms of the mean ± standard deviation. Bold denotes the lowest error metric for the current sequence.

Fig. 8  Pose estimation by ESVO and point cloud reconstruction by EMVS algorithms on ESVI and FULL sensor setups.
event arrays collected in a way to ease the retraining of the current models and motivates the development of new approaches to process events in such scenarios and corner cases.

The most prominent conclusion from evaluations on the FULL setup is that in outdoors dynamic weather where the dynamic range of the scene is considerably high, DVS sensor cannot be reliable to estimate the pose of the AGV with the current event-based SLAM systems. This conclusion is since events are fired asynchronously with high frequency, causing the visual sensor to be susceptible to weather constituents like rain or fog, which can degrade the estimation performance. Accordingly, our multi-modal datasets with the simulated corner cases can be the building block of choose-case scenarios for selecting the most efficient combination of multi-modal VI sensors for AGVs navigating in adverse conditions.

### 4.2.5 LiDAR Setup Evaluation

During the LiDAR based SLAM systems quantitative evaluation, we can observe significantly low RMS ATE values with MULLS systems for all the sequences compared to A-LOAM system as given in Table 12, and the lowest RMS ATE corresponds to MULLS with the loop closure option.

![Algorithm 1](https://example.com/algorithm1.png)

**Algorithm 1** Event 3-channel tensor (E3CT) pre-processing layer for frame-based systems

| Input: Packets of Events Arrays @f ep Hz |
| Output: E3CT (RGB Frame) @f ec Hz |

1. \( hot\_pixels \leftarrow \text{Hot Pixel Array} \)  \( \triangleright \) Fig. 9
2. \( n \leftarrow 24 \)  \( \triangleright \) #Temporal bins
3. \( c \leftarrow 3 \)  \( \triangleright \) #Channels
4. \( h \times w \leftarrow 1024 \times 1024 \)  \( \triangleright \) Frame dimensions
5. \( \Delta t \leftarrow 1e9 \) (1 sec)  \( \triangleright \) Sampling duration [nsec]
6. \( e_l \leftarrow [t, x, y, p] \)  \( \triangleright \) Events List
7. **for each** \( \text{packet} \in \text{event\_packets} \) **do**
8. \( \text{if} \) \( \text{packets} < f ep / f ec \) **then**
9. \( \text{Load events in current packet} \)
10. \( \text{if} \) \( \text{length} \) \( \text{packet} > 2 \times h \times w \) **then**
11. \( \text{Remove} \) \( hot\_pixels \) **from** \( \text{packet} \)
12. Add packet to \( e_l \)
13. **else**
14. Add packet to \( e_l \)
15. \( t_f \leftarrow e_l[t][-1] \)
16. packet += 1
17. **else**
18. \( C \leftarrow \mathcal{V}(e_l, n, 2 \times c, \Delta t) \)  \( \triangleright \) Event Volume
19. \( C(t_i, c_i, y, x) \leftarrow \max (0, 1 - \|t^* - t^*\|) \)
20. \( C_{\text{Total}} = \sum_{n=0}^{24} C(t_i, c_i, y, x) \)
21. visualize\( (C_{\text{Total}}) \)  \( \triangleright \) RGB Frame
22. **end if**
23. **end for**

![Fig. 10](https://example.com/figure10.png)

**Fig. 10** Graphical illustration of E3CT nano-time bins and events contribution to each channel. Nano-tbins are replaced with micro-tbins for real world DVS sensors

![Fig. 11](https://example.com/figure11.png)

**Fig. 11** The effect of adverse weather conditions on DVS events and ORB feature extraction. E3CT preserves both the high dynamic range property with the pixel temporal information of the DVS sensor, and the high quality with rich information (3-channels) of the RGB frames in all weather conditions in both static and dynamic scenes. Six cases are tested with an ascending difficulty form clear to adverse weather and from static to dynamic scene.
enabled. However, the RPE translation and rotation components slightly show a relatively lower values for A-LOAM compared to MULLS systems.

Figure 14 shows a more detailed qualitative evaluations, the efficient LiDAR point cloud registration method (TEASER) [60] results are shown in red rectangles presenting the feature matching between two consecutive scans on the left and the global registration results on the right using the Neighborhood Category Context (NCC) encoding. In the blue colored rectangles, we shows NCC encoding results that provide an independent description of every feature extracted from the source and target scans without any additional computational operations that can increase the system latency.

4.2.6 Comparative Evaluation

To sum up all the latest state-of-the-art evaluations of nearly 80 experiments using the IBISCape benchmark, we provide a quantitative analysis of the mean value of all errors in Table 13. The average of experiments with the E3CT-SVO show error values that are considerably less than that of the E2VID-SVO. This gives an indication that future developments of event-based SLAM systems using the E3CT event representation method that can benefit from all the 3-channels information will result in low latency and high accuracy system. Then, in order to have a thorough quantitative comparison of all the methods, a weighted normalized accuracy parameter (ACC) of all the SLAM systems evaluation parameters is proposed:

\[
ACC = 0.5 \times \frac{ATE_{\text{min}}}{ATE_{\text{method}}} + 0.25 \times \frac{RPE_{\text{Trans}}}{RPE_{\text{method}}} + 0.25 \times \frac{RPE_{\text{Rot}}}{RPE_{\text{method}}} [\mu \text{rad}].
\]  

Weights are distributed with 50% for the RMS ATE values and 50% for the RPE values divided equally between translation and rotation error values. The SLAM system that provides the lowest ATE and RPE values will give an Accuracy \( \geq 1 \) which is the highest Accuracy value. This qualitative analysis is represented in Fig. 15, where the SLAM system accuracy is compared to its system latency.

Since IBISCape benchmark targets a realistic simulation for the state-of-the-art SLAM systems evaluation, we compare the evaluation results on all the 31 IBIScape SLAM sequences with the real world publicly available datasets based on the RMS ATE values as given in their original papers in Fig. 16. To ensure a fair comparison, E2VID results reported in our work can’t be compared to that in [51], because the back-end VIO estimation method [43] using E2VID as a pre-processing layer is different than our evaluation back-end method (ORB-SLAM3/stereo-RGB).

The main outcome of this IBISCape versus real world benchmarks comparison, is that the IBISCape dataset as well as the data acquisition APIs are designed to realistically simulate outdoor environments that researchers can be confident to use in their novel AGVs SLAM systems reliability evaluation in adverse weather dynamic environments. Furthermore, towards reliable semantic SLAM systems, transfer-learning models from simulators to real world is
indispensable, specially in scenarios where real world data is not easy to collect or at dangerous situations. IBISCape APIs can also provide high-end training and testing data for transfer-learning applications.

5 Conclusion

This article proposes the IBISCape simulated heterogeneous sensors benchmark in large-scale dynamic environments along with 43 sequences suitable for multi-modal calibration & LiDAR/VI-SLAM evaluation. We also demonstrated new efficient algorithms for data synchronization during the acquisition process and a new iterative solution to estimate the unknown distortion coefficients of CARLA simulated cameras. Using multiple adverse weather conditions, we have shown their impact on the latest state-of-the-art SLAM systems trajectory estimations.

A novel event-based pre-processing layer is presented based on the Event Spike Tensor representation called the Event 3-Channel Tensor (E3CT). This efficient model-based layer produces high dynamic range 3-channel event frames and is validated on multiple adverse conditions where it is witnessed to outperform other learning-based pre-trained models. Accordingly, E3CTs will open new paths for working on model-based multiple channel event-based representations for more robust event-based SLAM systems.

The performance analysis includes a description of the sequence upon which the evaluation is done and the special conditions and corner cases simulated within every sequence to push the limits of the SLAM systems under assessment. The analytical study includes a comprehensive evaluation of the SLAM system performance and a quantitative comparison based on the ATE and RPE values. We hope that this new dataset will help in advancing the research in the field of multi-modal heterogeneous sensors fusion applied to Autonomous Ground Vehicles (AGVs) navigation in large-scale and dynamic environments.

As a future research trend, it will be indispensable to develop new efficient multi-modal: calibration and SLAM algorithms based on the fusion of heterogeneous sensors with different caption and spectral technologies. That allows the SLAM system to better estimate the trajectory

Fig. 13 Trajectories estimated by ORB-SLAM3 and BASALT SLAM systems using IBISCape sequences with FULL sensor setup and comparing to their ground truth and GPS paths. ORB-SLAM3 algorithms involved are: Monocular RGB, Stereo-RGB (S-RGB) with E2VID and E3CT (Ours), Stereo Visual Inertial (SVI), and RGB-D SLAM systems. While for BASALT, the SVI algorithm is assessed
Fig. 14 LiDAR sensor setup sequences qualitative detailed evaluation. From right to left: in red rectangle the point cloud features matching and global registration of two consecutive scans, in blue rectangle geometric feature points extraction, MULLS loop closure detection by Pose Graph Optimization (PGO), and trajectories estimated by MULLS and A-LOAM LiDAR Odometry/SLAM systems and comparison to their ground truth and GPS paths.
based on a reliable continuous-time 3D scene mapping. Finally, an in-depth investigation is needed concerning the effect of map loss on SLAM systems estimations during long-term navigation in large-scale and dynamic weather environments.

Appendix A: Extended Data

We generate data by eight acquisition APIs with four sensor setups mentioned in Table 4 in two groups: 1. calibration and 2. SLAM. SLAM data acquisition APIs run on all CARLA maps with an autopilot for traffic-aligned navigation. On the other hand, calibration APIs run on our modified CARLA-map with manual vehicle control to apply desired motions to collect sequences with basic or complex motions. Both AprilGrid and Checkerboard targets are introduced during acquisition. Half of the calibration sequences are collected using the AprilGrid $6 \times 6$ and the other half using the Checkerboard $7 \times 7$.

In order to operate all sensors in the same acquisition API on multiple frequencies, we develop the following procedure: the core data acquisition concept is that the CARLA world clock ticks with the highest frequency sensor in the setup. After that, the system waits to listen to all sensors sending data at this tick, updates the weather conditions, and waits for a new world tick. This allows the acquisition of all sensors data with its occurrence timestamps. Then, one can apply any synchronization/calibration algorithms on the collected datasets as in [8, 10]. We apply this methodology (see Program 1) to all sensor setups except the RGB-D setup, which requires time-synchronized and registered frames.

On the contrary, the CARLA world ticks with the lowest frequency sensor in the LiDAR/RGB-D setup with CARLA synchronous_mode acquisition (see Program 2). All the spawned sensors in the setup are stacked in a queue waiting for the world’s tick to start listening to the data. Although all sensors operate with their frequencies, the API reads the measurements of all sensors simultaneously at the timestamp of that CARLA world tick.

The open source data acquisition APIs and all sequences can be accessed using the Github repository: https://github.com/AbanobSoliman/IBISCape.git.
In the repository there is a complete manual on how to execute the APIs in all setups and options, including a library developed for IBISCape dataset files format to be processed using Robotic Operating System (ROS) based algorithms. Besides the Python based ROS tools, we attach the configuration files for all the assessed algorithms along with the Kalibr calibration results. A more in-detail insights of the IBISCape benchmark is available in the supplementary multimedia file.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s10846-022-01753-7.

**Author Contributions** All authors contributed to the study conception and design. The first draft of the manuscript was written by A.S., and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This work is supported by the French Ministry of Higher Education, Research and Innovation (MESRI). Author A.S. has received a Ph. D. grant from MESRI covering this research.

**Declarations**

**Ethics approval** The submitted work is original and not have been published elsewhere in any form or language.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Ethics Declarations** This research work is based on computer simulation open source software and did not involve human participants or animals. Hence, Consent to participate and Consent for publication are not applicable.

**Competing Interests** The authors have no relevant financial or non-financial interests to disclose.

**References**

1. Forster, C., Zhang, Z., Gassner, M., Weltherger, M., Scaramuzza, D.: Svo: Semidirect visual odometry for monocular and multicamera systems. IEEE Trans. Robot. 33(2), 249–265 (2017). https://doi.org/10.1109/TRO.2016.262335
2. Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P.: Keyframe-based visual-inertial odometry using nonlinear optimization. Int. J. Robot. Res. 34 (2014). https://doi.org/10.1177/0278364914554813
3. Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D.: On-manifold pre-integration for real-time visual-inertial odometry. IEEE Trans. Robot. 33(1), 1–21 (2017). https://doi.org/10.1109/TRO.2016.2597321
4. Qin, T., Li, P., Shen, S.: Vins-mono: A robust and versatile monocular visual-inertial state estimator. IEEE Trans. Robot. 34(4), 1004–1020 (2018). https://doi.org/10.1109/TRO.2018.2853729
5. Kerl, C., Sturm, J., Cremers, D.: Dense visual slam for rgb-d cameras. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (2013)
6. Alliez, P. et al.: Real-time multi-slam system for agent localization and 3d mapping in dynamic scenarios. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4894–4900 (2020)
28. Schubert, D. et al.: The tum vi benchmark for evaluating visual-inertial odometry. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1680–1687 (2018)

29. Judd, K.M., Gammell, J.D.: The Oxford multimotion dataset: Multiple se (3) motions with ground truth. IEEE Robotics and Automation Letters 4(2), 800–807 (2019)

30. Jeong, J., Cho, Y., Shin, Y.-S., Roh, H., Kim, A.: Complex urban dataset with multi-level sensors from highly diverse urban environments, Int. J. Rob. Res. 38(6), 642–657 (2019)

31. Kasper, M., McGuire, S., Beckman, C.: A benchmark for visual-inertial odometry systems employing onboard illumination. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2019)

32. Delmerico, J., Cieslewski, T., Rebecq, H., Faessler, M., Scaramuzza, D.: Are we ready for autonomous drone racing? the uzh-fpv drone racing dataset. 2019 International Conference on Robotics and Automation (ICRA) (2019)

33. Zuñiga-Noël, D., Jaenal, A., Gomez-Ojeda, R., Gonzalez-Jime
nez, J.: The uma-vi dataset: Visual-inertial odometry in low-textured and dynamic illumination environments. Int. J. Rob. Res. 39(9), 1052–1060 (2020)

34. Antonini, A., Guerra, W., Murali, V., Sayre-McCord, T., Karaman, S.: The blackbird uav dataset. Int. J. Rob. Res. 39(10–11), 1346–1364 (2020)

35. Zhang, H., Jin, L., Ye, C.: The vcu-rvi benchmark: Evaluating visual inertial odometry for indoor navigation applications with an rgd-8 camera. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 6209–6214 (2020). https://doi.org/10.1109/IROS45743.2020.9341713

36. Klenk, S., Chui, J., Demmel, N., Clemers, D.: Tum-vie: The tum stereo visual-inertial event dataset. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8601–8608 (2021). https://doi.org/10.1109/IROS51168.2021.9636728

37. Yuan, C., et al.: A novel fault-tolerant navigation and positioning method with stereo-camera/micro electro mechanical systems inertial measurement unit (mems-imu) in hostile environment. Micromachines 9, 626 (2018). https://doi.org/10.3390/mi9120626

38. Faessler, M. et al.: Autonomous, vision-based flight and live dense 3d mapping with a quadrotor micro aerial vehicle. J. Field Robot. 33(4), 431–450 (2016). https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21581

39. Lynen, S., Achtelik, M.W., Weiss, S., Chli, M., Siegwart, R.: A robust and modular multi-sensor fusion approach applied to multi navigation. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (2013)

40. Mourikis, A.L., Roumeliotis, S.I.: A multi-state constraint kalman filter for vision-aided inertial navigation. Proceedings 2007 IEEE International Conference on Robotics and Automation (2007)

41. Bloesch, M., Omari, S., Hutter, M., Siegwart, R.: Robust visual inertial odometry using a direct ekf-based approach. 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS) (2015)

42. Qin, T., Li, P., Shen, S.: Vins-mono: A robust and versatile monocular visual-inertial state estimator. IEEE Trans. Robot. 34(4), 1004–1020 (2018)

43. Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P.: Keyframe-based visual-inertial odometry using nonlinear optimization. Int. J. Rob. Res. 34(3), 314–334 (2015)

44. Campos, C., Elvira, R., Rodriguez, J.J.G., M.Montiel, J.M., Tardos, J.: Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. IEEE Trans. Robot. 1–17 (2021). http://dx.doi.org/10.1109/TRO.2021.3075644. https://doi.org/10.1109/tro.2021.3075644

45. Usenko, V., Demmel, N., Schubert, D., Stueckler, J., Clemers, D.: Visual-inertial mapping with non-linear factor recovery. IEEE Robotics and Automation Letters (RA-L) & Int. Conference on
Intelligent Robotics and Automation (ICRA) 5(2), 422–429 (2020). https://doi.org/10.1109/LRA.2019.2961227

46. Delmerico, J., Scaramuzza, D.: A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots. 2018 IEEE International Conference on Robotics and Automation (ICRA) (2018)

47. Zhou, Y., Gallego, G., Shen, S.: Event-based stereo visual odometry. IEEE Trans. Robot. 37(5), 1433–1450 (2021). https://doi.org/10.1109/TRO.2021.3062252

48. Gehrig, D., Gehrig, M., Hidalgo-Carrió, J., Scaramuzza, D.: Video to events: Recycling video datasets for event cameras. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), pp. 3583–3592 (2020). https://doi.org/10.1109/CVPR42600.2020.00364

49. Rebecq, H., Gallego, G., Mueggler, E., Scaramuzza, D.: EMVS: Event-based multi-view stereo—3D reconstruction with an event camera in real-time. Int. J. Comput. Vis. 126, 1394–1414 (2018). https://doi.org/10.1007/s11263-017-1050-6

50. Tomy, A., Paigwar, A., Mann, K.S., Renzaglia, A., Laugier, C.: Fusing Event-based and RGB camera for Robust Object Detection in Adverse Conditions. ICRA 2022 - IEEE International Conference on Robotics and Automation (2022). https://hal.archives-ouvertes.fr/hal-03591717

51. Rebecq, H., Ranfil, R., Kolton, V., Scaramuzza, D.: Events-to-video: Bringing modern computer vision to event cameras. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR) (2019)

52. Zhang, J., Singh, S.: Loam: Lidar odometry and mapping in real-time. Robotics: Science and Systems (2014)

53. Pan, Y., Xiao, P., He, Y., Shao, Z., Li, Z.: Mulls: Versatile lidar slam via multi-metric linear least square. 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 11633–11640 (2021). https://doi.org/10.1109/ICRA48506.2021.9561364

54. Rehder, J., Nikolic, J., Schneider, T., Hinzmann, T., Siegwart, R.: Extending Kalibr: Calibrating the extrinsics of multiple imus and of individual axes. 2016 IEEE International Conference on Robotics and Automation (ICRA) (2016)

55. Muglikar, M., Gehrig, M., Gehrig, D., Scaramuzza, D.: How to calibrate your event camera. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1403–1409 (2021)

56. Galleani, L., Tavella, P.: The dynamic allan variance. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 56(3), 450–464 (2009). https://doi.org/10.1109/TUFFC.2009.1064

57. Tomasi, C., Kanade, T.: Detection and tracking of point. Int. J. Comput. Vis. 9, 137–154 (1991)

58. Chen, W. et al.: An overview on visual slam: From tradition to semantic. Remote Sensing 14(13) (2022). https://www.mdpi.com/2072-4292/14/13/3010. https://doi.org/10.3390/rs14133010

59. Sironi, A., Brambilla, M., Bourdis, N., Lagorce, X., Benosman, R.: Hats: Histograms of averaged time surfaces for robust event-based object classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1731–1740 (2018)

60. Yang, H., Shi, J., Carlone, L.: Teaser: Fast and certifiable point cloud registration. IEEE Trans. Robot. 37(2), 314–333 (2020)

61. Zhou, Y. et al. Semi-dense 3d reconstruction with a stereo event camera. Proceedings of the European conference on computer vision (ECCV), pp. 235–251 (2018)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Abanob Soliman received his B.Sc. degree (Hons.) in Aerospace Engineering from Cairo University, in 2018. In 2019, he was working as a Teaching Assistant in the Electronics and Computer Engineering department at Nile University. Then, he received his M.Sc. degree in Smart Aerospace and Autonomous Systems from Université d’Evry/Université Paris-Saclay, in 2020. Since 2020, he is working towards his PhD in signal and image processing from Université d’Evry/Université Paris-Saclay in the IBISC laboratory under the supervision of Prof. Samia Bouchafa, Prof. Désiré Sidié, and Dr. Fabien Bonardi. His research interests include; visual-inertial odometry, sensor fusion, depth estimation, optimization, and neuromorphic vision. He is an IEEE graduate student member.

Fabien Bonardi received in 2017 the Ph.D. degree in electrical and computer engineering from Normandie University, UNIROUEN (LITIS Lab). He is assistant professor at Université d’Evry/Université Paris-Saclay since 2018. His research concerns computer vision, visual perception in robotics and sensors fusion for non-conventional cameras.

Désiré Sidié is a full professor at Université Evry - Paris Saclay, and a member of the IBISC laboratory. His research interests include computer vision for autonomous vehicles, and medical image analysis. He has authored more than 90 papers in international journals and conferences. He has received many award with his students including, ICRA Best Vision paper award finalist, and SPIE Medical Imaging Best Student paper award. He is an IEEE senior member.

Samia Bouchafa received a Ph.D. degree in electrical and computer engineering from University Paris VI and the French Research Institute in Transportation (IFSTTAR, then University Gustave Eiffel) in 1998. In 1999, she was an assistant professor at University Paris XI (IEF Lab), then an associate professor in 2011 (HDR). Since 2012, she has a full professor position at Université d’Evry/Université Paris-Saclay. Since 2015, she has been coordinating the second year of a Master in Electrical Engineering, specializing in “Automatic mobile systems” at the University of Paris-Saclay. Since 2019, she has been the director of the IBISC laboratory. Her research concerns computer vision, visual perception, multimodal vision for autonomous systems, motion analysis, stereovision, visual odometry, and localization.