Stereo Visual Localization Dataset Featuring Event Cameras

Antea Hadviger[1], Vlaho-Josip Štironja[1], Igor Cvišić[1], Ivan Marković[1], Sacha Vražić[2], Ivan Petrović[1]

***Abstract*—Visual odometry and SLAM methods are facing increasingly complex scenarios and novel solutions are needed to offer more accurate and reliable results in challenging environments. Standard cameras are challenged under low light conditions or very high-speed motion, as they suffer from motion blur and operate at a limited frame rate. These problems can be alleviated by using event cameras – asynchronous visual sensors that offer complementary advantages compared to standard cameras, as they do not suffer from motion blur and support high dynamic range. Although there is a number of existing datasets intended for visual odometry and SLAM that contain event data, most of them are collected using monocular sensors and limited either in terms of camera resolution or ground truth availability. Our work aims to complement this by further supporting development of robust stereo visual odometry and SLAM algorithms, allowing to exploit both event data and intensity images. We provide both indoor sequences with 6-DoF motion and outdoor vehicle driving sequences that additionally contain 3D lidar data. All sequences contain data from a synchronized high-resolution stereo event and standard cameras, whereas ground truth trajectories are provided by either a motion capture system or an RTK GNSS receiver.**

***Index Terms*—event cameras, stereo cameras, visual odometry and SLAM, sensor fusion**

The dataset is available at http://www.bitbucket.com/unizg-fer-lamor/event-dataset.

I. Introduction

Event cameras, also known as dynamic vision sensors (DVS), are asynchronous biologically inspired sensors that detect changes in brightness intensity on a pixel-by-pixel basis. Intensity changes are reported as events as they occur, rather than capturing whole scenes with a fixed frame-rate like traditional cameras. Event cameras have numerous advantages over traditional cameras, including high temporal resolution in microseconds, low latency, high dynamic range up to 120 dB, and low power consumption. These sensors have a great potential to be particularly useful in challenging scenarios, such as scenes with dynamic illumination or high-speed motion. However, given that the output of event cameras is fundamentally different from traditional cameras, the full potential of these sensors can only be realized by developing new asynchronous processing methods. Nevertheless, event cameras have already proven to be valuable for various robotic perception tasks such as depth estimation, visual odometry, and simultaneous localization and mapping (SLAM). More details on event cameras, event-based methods, and results can be found in a detailed survey [1].

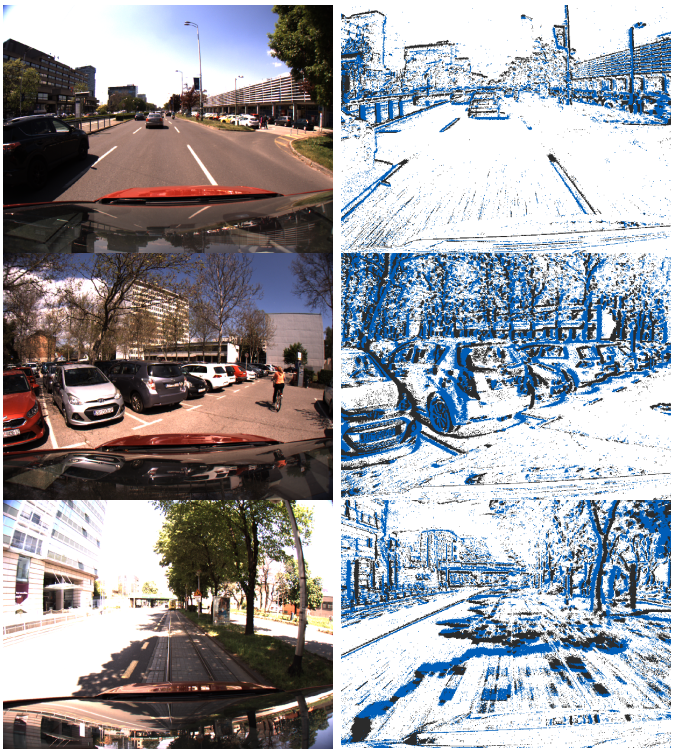

Fig. 1: Sample intensity images (left) and accumulated events (right; color indicates event polarity) from the outdoor sequences. Third sample shows a situation where the left part of the intensity image is overexposed due to bright sunlight as the car is passing through the shaded area, while the event cameras can handle this scenario well, given their high dynamic range.

Even though state-of-the-art visual SLAM algorithms based on intensity images achieve high accuracy in favorable conditions, there still remains a challenge to achieve robust visual SLAM under low light or high dynamic range conditions, as well as very high-speed motions, making it imperative to investigate alternative sensors. Using the complementary advantages of event and standard cameras for SLAM, visual

This work has been supported by the Croatian Science Foundation under contract No. I-3137-2019. The research has been carried out within the activities of the *Centre of Research Excellence for Data Science and Cooperative Systems* supported by the Ministry of Science and Education of the Republic of Croatia.

[1] Antea Hadviger, Vlaho-Josip Štironja, Igor Cvišić, Ivan Marković, and Ivan Petrović are with the University of Zagreb Faculty of Electrical Engineering and Computing, Laboratory for Autonomous Systems and Mobile Robotics, Croatia. {antea.hadviger, vlaho-josip.stironja, igor.cvisic, ivan.markovic, ivan.petrovic}@fer.hr

[2] Sacha Vražić is with Rimac Automobili, Croatia. sacha.vrazic@rimac-automobili.com

odometry, and depth estimation has already been proven to achieve interesting results, as reviewed in a recent comprehensive survey on event-based SLAM [2]. First attempt to combine events with intensity images and inertial measurements for robust visual SLAM was presented in [3]. More recently, [4] presents a visual-inertial odometry method that tightly-coupled the events, intensity images, and IMU by leveraging point and line features. This method based on monocular cameras was subsequently extended to support stereo cameras in [5]. Similarly, the work in [6] explores the possibility of using event-based visual-inertial odometry for navigation of planetary robots.

As the complexity of problems addressed by event-based odometry and SLAM increases in terms of the type of scene, illumination, or speed of motion, there is a rising interest for datasets featuring event cameras coupled with other sensors to enable researchers to explore novel methods with proper evaluation benchmarks at their disposal. Visual SLAM datasets such as KITTI [7], EuRoC [8], and TUM-RGBD [9] have pushed the field forward by providing high-quality data and ground truth with clear quantitative evaluation guidelines. Even though there is a number of existing datasets intended for visual odometry and SLAM that contain event-based data, most of them are monocular and limited either in terms of camera resolution or ground truth availability. Our work aims to provide a dataset to support development of robust stereo visual odometry and SLAM algorithms, allowing to exploit both event data and intensity images. We offer both indoor sequences with 6-DoF motion and outdoor sequences collected with a car driving through urban areas. All sequences contain data from a stereo high-resolution event and standard camera, while ground truth trajectories are provided from either a motion capture system or an RTK GNSS receiver. Outdoor sequences also feature 3D lidar data to support depth estimation development.

The rest of the paper is organized as follows. Section II presents related event-based datasets intended for visual odometry and SLAM. In Section III we present utilized sensors along with describing synchronization and calibration methods used for collecting the data. Finally, we conclude the paper in Section V.

II. Related Work

The first event camera simulator was presented in [10], as well as a collection of datasets including data from a DAVIS event camera (240×180 resolution), which also provides intensity images and IMU measurements. Both synthetic and real-world sequences are available. For indoor sequences, ground truth for 6-DoF motion is recorded using a motion capture system. DDD17 [11] and DDD20 [12] are both large scale datasets collected with a monocular DAVIS346 camera (346×240) mounted on a driving car. Since these datasets are intended for automated driving applications, additional data from the vehicle, such as steering angle and speed, are also provided. However, 6-DoF pose ground truth is not available, as only 2D translation can be inferred from the provided GPS latitude and longitude. The UZH-FPV Drone Racing Dataset [13] is a specialized visual-inertial odometry dataset intended for 6-DoF flying drone localization in high-speed scenarios. It features event data from the miniDAVIS346, intensity images, and IMU measurements, along with precise ground truth poses.

MVSEC is the first dataset with synchronized stereo event cameras [14]. It features a variety of data captured in different illumination levels and environments. The indoor sequences were collected with a handheld rig and a hexacopter, while sensors were mounted on top of a car and a motorcycle for outdoor driving sequences. Each DAVIS346 camera provides event streams, grayscale intensity images, and IMU readings. Additionally, 3D lidar, motion capture, and GPS data are utilized to provide ground truth trajectories and depth. However, low resolution of the DAVIS346 camera (346×240), as well as smaller stereo baseline, are limiting factors for odometry and SLAM accuracy, especially for outdoor driving scenarios. Similarly, the DSEC dataset [15] contains data from a stereo event camera, a stereo standard camera, and 3D lidar, but with the added benefit of a higher event camera resolution. DSEC addresses the problem of low event camera resolution and small stereo baseline as the main drawbacks of MVSEC by featuring Prophesee Gen3.1 event cameras with 640×480 pixels. Even though DSEC provides a very rich amount of data collected by driving in a variety of illumination conditions, it does not contain ground truth trajectories. Furthermore, it does not cover very high speed scenarios where standard cameras would suffer from motion blur. The TUM-VIE dataset [16] consists of a large variety of head-mounted sequences, targeting VR applications, and handheld sequences, in both indoor and outdoor environments. The dataset contains stereo event data, stereo grayscale frames, as well as IMU data. However, ground truth poses are only available at the beginning and the end of each sequence, therefore limiting trajectory evaluation. The Prophesee Gen4 event cameras contain a larger sensor of 1280×720 pixels, which is significantly larger than the ones featured in the DSEC and MVSEC datasets.[im:Treba li ova rečenica? Poslije se nigdje ne koristi ova kamera...] The VECtor dataset [17] is captured by a full hardware-synchronized sensor suite that includes a Prophesee Gen3 event stereo camera, a regular stereo camera, an RGB-D sensor, a lidar, and an IMU, while ground truth trajectories are provided by a motion capture system, or by matching motion-compensated lidar scans with a dense point cloud of the environment captured by the laser scanner. Sequences are collected in large and small-scale environments. Each small-scale sequence comes in two variants (normal or fast), depending on the speed of exerted motion. The ViViD++ dataset [18] provides data from a sensor system including RGB, thermal, event, depth, and inertial measurements, along with ground truth RTK-GPS trajectory in outdoor driving scenarios. However, only a single event and standard camera is deployed.

Even though HD event cameras with a resolution of 1280×720 are available on the market and provide more

TABLE I: Properties of the sensors used for data collection

Sensor	Description
2× DVXplorer event camera	640 × 480 pixels 200μs temporal resolution up to 110 dB 165 Mega events/s
2× FLIR Blackfly S camera BFS-PGE-31S4C-C	2048 × 1536 pixels up to 60 dB global shutter 10 Hz
3× IMU integrated in DVXplorer and Ouster	3D accelerometer 3D gyroscope
MoCap OptiTrack Flex13 indoor only	accurate 6-DoF pose 850nm IR light 100 Hz
Spatial FOG Dual GNSS outdoor only	3D pose and orientation up to 8mm accuracy (with RTK) 40 Hz
Ouster OS1-128 Lidar outdoor only	128 channels Vertical FoV: 45° Angular resolution: 0.35° 10 Hz

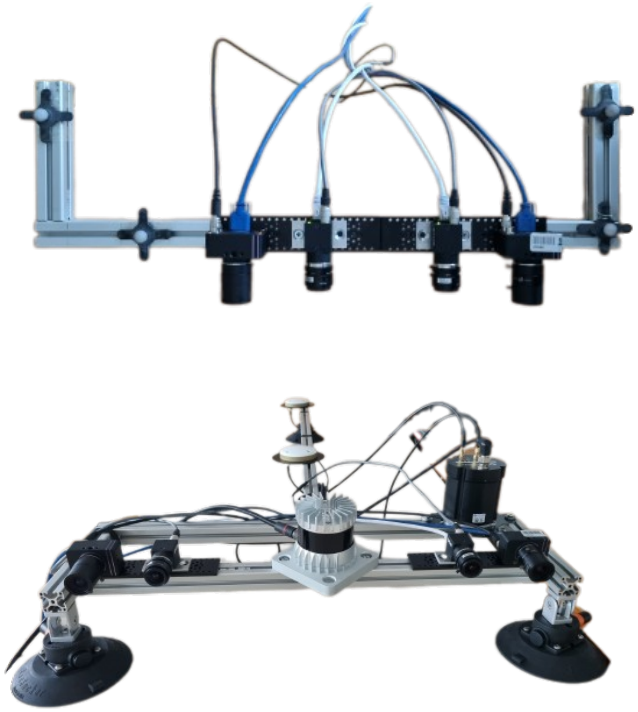

Fig. 2: Sensor setups for collecting indoor (upper image) and outdoor sequences (lower image). Event and standard stereo cameras are mounted on an aluminum rig. Indoor setup includes motion capture markers, while the GNSS dual antenna and lidar are mounted on the outdoor rig.

detailed information about the scene, we found that the event rate exceeds the bandwidth limit during very rapid motions in scenes with rich texture, leading to significant loss of events during the recording. Bandwidth limitation is the main reason why we opted for using event cameras with 640×480 resolution for recording our dataset, as we wanted to prevent data loss and preserve all the events generated during very high speed motions, thus enabling more accurate performance of the event-based algorithms. Furthermore, the main motivation for collecting this dataset was the lack of ground truth trajectories in stereo datasets that feature high-resolution event cameras, e.g., DSEC and TUM-VIE, which we believe can complement and serve greatly visual odometry and SLAM development and evaluation.

III. The Proposed Dataset

Both our indoor and outdoor setup (Figure 2) feature a stereo event camera (with an integrated IMU) and a stereo standard camera, but with different baselines and intrinsic parameters. Additionally, the indoor rig contains motion capture markers, while the outdoor rig equips a 3D lidar (with an integrated IMU) and a dual GNSS antenna. [im:A vrlo napredni AHRS sustav s fiber gyrom? :-)] This section provides details about the individual sensors (summarized in Table I) and describes how we approached synchronization and calibration. All sequences are collected and published in the ROS bag format.

A. Sensor Setup

For both indoor and outdoor data acquisition we used the DVXplorer stereo event camera kit. The cameras have a spatial resolution of 640×480 pixels (VGA) and offer asynchronous output as a stream of events with temporal resolution of 200 μs, while intensity frames are not available. Their dynamic range is between 90 and 110 dB (3-100k lux with 99.9% of pixels respond to 27.5% contrast, 0.3-100k lux with 50% of pixels respond to 80% contrast). The two event cameras are enclosed in anodized aluminum cases and mounted on the rig in a horizontal stereo setup, with a baseline of 30 cm for the indoor and 60 cm for outdoor sequences. The cameras are connected to the computer using USB 3.0. The maximum throughput of events that they can handle is 165 Mega-events per second, but in practice it is also limited by the USB 3.0 bandwidth. Both event cameras are equipped with 4-12 mm varifocal CS-mounted lenses. To suppress irratic events generated by infrared (IR) flashes from the motion capture system, we additionally placed IR cut filters, in the form of thin rectangular pieces of glass directly on top of the sensor, under the lens. To capture and visualize event data, we used the ROS event camera driver developed by iniVation [1]. The DVXplorer sensor also encapsulates a 6-axis IMU (gyro and accelerometer) with a sampling rate of 800 Hz. The event cameras parameters, usually referred to as biases, regulate the signal-to-noise ratio and the event rate of the generated stream. For outdoor and indoor sequences with brighter lighting, we set the DVXplorer biases to default values as suggested by

[1] https://gitlab.com/inivation/dv/dv-ros

the manufacturer (sensitivity level 3 out of 5). For indoor sequences with less lighting, the biases are set to promote slightly higher sensor sensitivity level (4 out of 5), since there is less contrast in the scene. Intensity color images are obtained using a stereo pair of FLIR Blackfly S GigE global shutter cameras, with a spatial resolution of 2048×1536 pixels with auto-exposure enabled. Even though the cameras can operate at a higher frequency, we had to lower their frame rate to 10 Hz due to bandwidth limitations. The cameras are powered through a Power-Over-Ethernet (PoE) switch and connected to the computer via Ethernet. They are rigidly mounted on the rig in a horizontal stereo setup along the same line as the event cameras. The baselines for the indoor and outdoor setup are 10 cm and 40 cm, respectively. We used an inhouse ROS camera driver that extended the existing camera drivers and wrapped them in a ROS node. Unfortunately, even with the frequency lowered to 10 Hz, frames are occasionally dropped at random from either left or right camera. In case of a dropped frame, we still retain the corresponding stereo frame in the ROS bag. Lidar data was collected using the Ouster OS1-128 high-resolution scanning lidar sensor. It has a 128 channel vertical resolution that gives 2.62 million points per second, with the vertical field of view of $45°$. Angular resolution is $0.35°$ with the output rate of 10 Hz. Additionally, Ouster OS1-128 is equipped with an IMU MPU 9250 that has an output rate of 100 Hz.

To provide ground truth poses for the indoor sequences, we used the OptiTrack motion capture system with 12 Flex13 cameras installed in our laboratory. Four reflective infrared markers were placed on the rig. The motion capture system reports both position and orientation of the rigid body defined by the markers with millimeter level precision. In our setup, the motion capture software is ran on a different computer than the one used for recording the data. The poses are transmitted to the destination computer via WiFi and captured using the dedicated ROS driver at a frequency of 100 Hz.

For outdoor sequences, Spatial FOG Dual GNSS (with RTK) poses represents the ground truth proxy. We used a TopCon base station with the quad-constellation RTK Net G5 receiver and a G5-A1 antenna mounted on the top of a 13-story building. The used GNSS has factory claimed horizontal and vertical position accuracy of 0.008 m and 0.015 m, respectively, with an output frequency of 40 Hz. On the vehicle we placed the GNSS antennas away from the other sensors to ensure that there is interference due to multiple running data cables.

B. Indoor Sequences

Indoor sequences (Figure 3) are acquired in our laboratory with rich scene setup generating many events. The motion capture OptiTrack system is used to record the ground truth trajectories. The rig with stereo event and standard cameras was handheld in all the indoor sequences with 6-DoF motion being present in all of the sequences. One part of the dataset was recorded in bright daylight with all the blinds open (event camera settings set to default). Given that the standard cameras

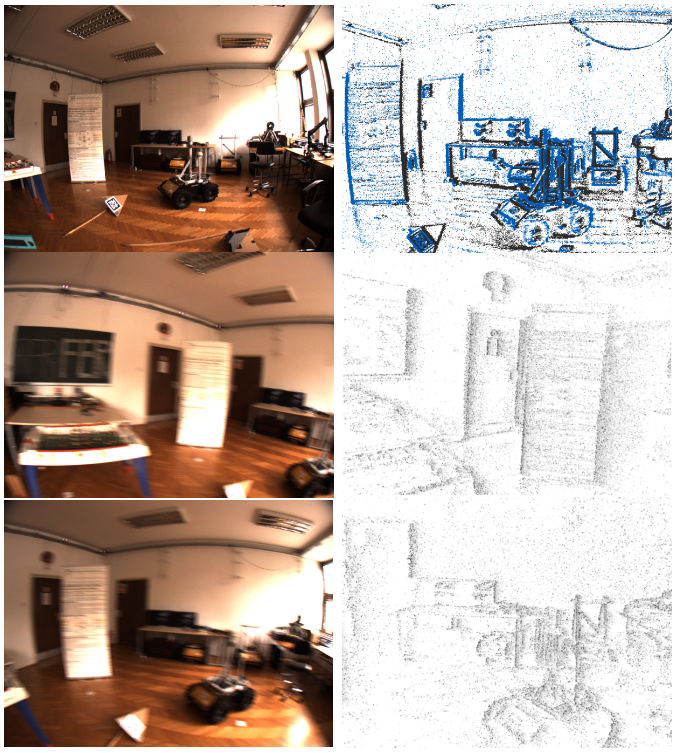

Fig. 3: Sample intensity images (left) and accumulated events/event time surfaces (right) of indoor sequences. In the second and third sample, motion blur is present in the intensity images, whereas event time surfaces remain sharp.

operate with auto-exposure settings, the exposure changes according to the amount of light coming into the sensor. Thus, there might be some overexposed or underexposed frames as the cameras are rapidly changing poses, from pointing towards to windows, where a lot of light is present, and away from the windows. The other part of the dataset is collected with the blinds closed and with less natural light (event camera sensitivity set to *high*). We provide several sequences named with the prefix *slow*, where speed of motion is limited so that intensity images do not suffer from motion blur. In contrast, other sequences contain rapid motions, inducing motion blur in the intensity images, thus making them more challenging for standard visual algorithms.

C. Outdoor Sequences

To collect the data for outdoor driving sequences (Figure 1), we mounted the multi-sensor rig on the car roof. The car was driven in the central area of the city of Zagreb, Croatia. Given the high traffic in the urban settings, all sequences contain moving objects such as pedestrians and other vehicles. We cut out the parts of the recordings where the car is still at the traffic lights for a longer period of time. Note that the traffic lights, some illuminated billboards, and the lights of the other vehicles generate bursts of events due to flickering. However, during daytime, this does not cause any additional noise or bandwidth issues. Due to auto-exposure of



Fig. 4: Ground truth trajectory for a part of our outdoor driving sequences, obtained with a dual antenna GNSS system mounted on the sensor rig placed on the top of the car.

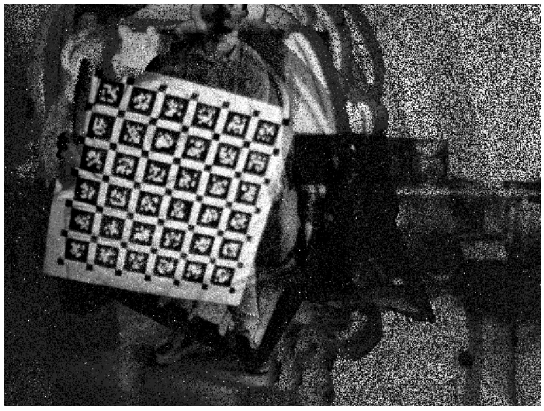

Fig. 5: Sample image used for event camera calibration, obtained by accumulating events based on their polarities within a predefined interval, and using an exponential decay kernel to determine each event's contribution to the image intensity. The image is an output of the *dv-ros-accumulation* node of the iniVation ROS toolbox.

the standard cameras, parts of some frames can be overexposed or underexposed, given the bright daylight conditions (third sample in Fig. 1). Due to the lens and camera positioning, a part of the car hood is visible in the images. Sample ground truth trajectory obtained with the GNSS RTK system can be seen in Fig. 4.

IV. Synchronization and Calibration

A. Camera Synchronization and Temporal Calibration

Two standard FLIR Blackfly S cameras are synchronized to capture images at the same time by physically connecting their GPIO pins with a synchronization cable. Similarly, two DVXplorer event cameras are also synchronized from the hardware side using a dedicated synchronization cable. Since they do not capture intensity images, we do not need to worry about triggering their exposure at the same time, but their synchronization allows us to keep the timestamps of the generated events consistent across both sensors, on a microsecond precision level.

Since the clocks between the standard and event cameras are not synchronized on the hardware level, we performed temporal calibration to ensure timestamp consistency across all visual sensors. To estimate the time delay between sensors, we used the open-source ROS toolbox called Calirad [19], which implements a method for multisensor calibration based on Gaussian processes estimated moving object trajectories, resulting in reliably estimated temporal parameters. Specifically, for visual sensor calibration we used a square AprilTag [20] marker with a side length of 14.35 cm printed on a planar rigid board. Using the AprilTag detector and intrinsic camera parameters, which need to be estimated prior to this procedure, we obtain continuous 3D positions of the target in the coordinate systems of the cameras, which Calirad uses to estimate the time delay between sensors.[im:Je li ovo rađeno prije svakog snimanja?]

B. Camera Calibration

To perform intrinsic and extrinsic calibration of all four visual sensors, we use a grid of 6x6 AprilTags as a calibration target that allows for robust data association. We move the AprilTag grid in front of the static sensor rig to capture enough data that is subsequently used as the input for the Kalibr toolbox [21]. The intensity images from the standard cameras can be directly used as the input, but DVXplorer event cameras only capture events, which need to be processed in a way that would allow for the tool to detect AprilTags. Therefore, we make use of the intensity image reconstruction tool available in the iniVation event camera driver. Namely, the events are accumulated during the time interval set using the configuration parameters and reported in the form of intensity frames at the configured frame rate. Instead of naïvely accumulating events based on their polarity within a certain interval, the contribution of events to the intensity of the image is calculated by an exponential decay kernel that accounts for the events' timestamps. The resulting images, as seen in Fig. 5, can be used as an input for Kalibr to perform intrinsic and extrinsic calibration. Even though frame acquisition between the standard and event cameras is not synchronized, being able to choose an arbitrarily high event frame rate enables us to have all four intensity images acquired at approximately the same time to allow correct extrinsic calibration. We ensure that the AprilTag detections cover the entire field of view of the cameras and that they are captured at various scales and orientations.

C. IMU-Camera Calibration

The iniVation ROS event camera driver provides a tool which takes as input the stream of the IMU data coming from the camera capture node and estimates the IMU biases. The camera must be placed on a stable and leveled surface with the gravity vector in the same direction of the Y axis of the camera frame, e.g., a flat table or a floor. After performing intrinsic camera calibration, we collected images of a static AprilTag calibration pattern while moving and rotating the camera

around all the axes to obtain extrinsic and temporal parameters between the IMU and the event camera. [todo:@vjs, note the used tool]

D. Lidar Extrinsics Calibration

The lidar and camera are calibrated using the open-source package [22]. The package evaluates calibration parameters using Perspective-n-Point RANSAC with Levenberg-Marquardt refinement. In order to perform the calibration, various target-based sequences with a checkerboard moving and rotating around all axes were recorded. The extrinsic calibration between lidar and its integrated IMU was assumed to be equal to the manufacturer specifications.

E. Motion Capture to Camera Calibration

OptiTrack motion capture system provides 3D position and orientation of the center of the rigid body defined by the markers placed on our rig. However, the center of the markers is not fully aligned to any camera frame, thus requiring extrinsic calibration to obtain the transformation from the camera coordinate system to the coordinate system of the motion capture markers. For this purpose, we use the same AprilTag as for the camera temporal calibration described in Section IV-A. The AprilTag detector gives us 3D positions and orientations of the target in the camera frame ${}^{\text{cam}}\mathbf{T}_{\text{tag}}$ while the motion capture system provides the position of the rig set in the motion capture frame ${}^{\text{mocap}}\mathbf{T}_{\text{rig}}$. We used these transforms, simultaneously obtained by moving the rig in front of the AprilTag, to finally estimate the position of the rig marker with respect to the camera ${}^{\text{rig}}\mathbf{T}_{\text{cam}}$ by solving the hand-eye calibration problem defined in [23] by the following equations:

$$ {}^{\text{mocap}}\mathbf{T}_{\text{rig}}^{(i)}\,{}^{\text{rig}}\mathbf{T}_{\text{cam}}\,{}^{\text{cam}}\mathbf{T}_{\text{tag}}^{(i)} = {}^{\text{mocap}}\mathbf{T}_{\text{rig}}^{(0)}\,{}^{\text{rig}}\mathbf{T}_{\text{cam}}\,{}^{\text{cam}}\mathbf{T}_{\text{tag}}^{(0)} \quad (1) $$

$$ \left({}^{\text{mocap}}\mathbf{T}_{\text{rig}}^{(0)}\right)^{-1}\,{}^{\text{mocap}}\mathbf{T}_{\text{rig}}^{(i)}\,{}^{\text{rig}}\mathbf{T}_{\text{cam}} = {}^{\text{rig}}\mathbf{T}_{\text{cam}}\,{}^{\text{cam}}\mathbf{T}_{\text{tag}}^{(0)}\left({}^{\text{cam}}\mathbf{T}_{\text{tag}}^{(i)}\right)^{-1} \quad (2) $$

We used the OpenCV *calibrateHandEye* function to solve this problem.

V. Conclusion

In this paper we have proposed a dataset aimed for developing robust stereo visual odometry and SLAM algorithms using synchronized stereo event and standard cameras of high resolution and sufficient baselines. We included 6-DoF sequences collected indoors with a handheld rig, as well as outdoor driving sequences obtained with a sensor rig mounted on the car roof. In the former case, a motion capture system provided ground truth trajectories, whereas for the latter an RTK GNSS receiver is deployed for outdoor data collection.[im:Bitno je naglasiti da nije samo RTK GNSS, to je samo dio jednadžbe. Već imamo i AHRS sustav s fiber gyrom koji je u fuziji s RTK GNSS-om. MOžda mi je negdje u članku još promaklo to naglasiti pa vidite treba li takav komentar još negdje staviti.]

References

[1] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis *et al.*, "Event-based vision: A survey," *arXiv preprint arXiv:1904.08405*, 2019.

[2] K. Huang, S. Zhang, J. Zhang, and D. Tao, "Event-based simultaneous localization and mapping: A comprehensive survey," *arXiv preprint arXiv:2304.09793*, 2023.

[3] A. R. Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high speed scenarios," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 994–1001, 2018.

[4] W. Guan, P. Chen, Y. Xie, and P. Lu, "Pl-evio: Robust monocular event-based visual inertial odometry with point and line features," *arXiv preprint arXiv:2209.12160*, 2022.

[5] P. Chen, W. Guan, and P. Lu, "Esvio: Event-based stereo visual inertial odometry," *IEEE Robotics and Automation Letters*, 2023.

[6] F. Mahlknecht, D. Gehrig, J. Nash, F. M. Rockenbauer, B. Morrell, J. Delaune, and D. Scaramuzza, "Exploring event camera-based odometry for planetary robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8651–8658, 2022.

[7] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.

[8] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.

[9] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 573–580.

[10] H. Rebecq, D. Gehrig, and D. Scaramuzza, "Esim: an open event camera simulator," in *Conference on robot learning*. PMLR, 2018, pp. 969–982.

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

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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*, vol. 7, no. 3, pp. 8217–8224, 2022.

[18] 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.

[19] J. Peršić, L. Petrović, I. Marković, and I. Petrović, "Spatiotemporal multisensor calibration via gaussian processes moving target tracking," *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1401–1415, 2021.

[20] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 3400–3407.

[21] 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.

[22] V. Heethesh, "Ros camera lidar calibration package, available: https://github.com/heethesh/lidar_camera_calibration," 2019.

[23] K. Daniilidis, "Hand-eye calibration using dual quaternions," *The International Journal of Robotics Research*, vol. 18, no. 3, pp. 286–298, 1999.