Visual Navigation Datasets for Event-based Vision: 2014-2021

Andrejs Zujevs and Agris Nikitenko
Faculty of Computer Science and Information Technology, Riga Technical University, Latvia

Keywords: Datasets, Event-based Vision, Neuromorphic Vision, Visual Navigation, Concise Review.

Abstract: Visual navigation is becoming the primary approach to the way unmanned vehicles such as mobile robots and drones navigate in their operational environment. A novel type of visual sensor named dynamic visual sensor or event-based camera has significant advantages over conventional digital colour or grey-scale cameras. It is an asynchronous sensor with high temporal resolution and high dynamic range. Thus, it is particularly promising for the visual navigation of mobile robots and drones. Due to the novelty of this sensor, publicly available datasets are scarce. In this paper, a total of nine datasets aimed at event-based visual navigation are reviewed and their most important properties and features are pointed out. Major aspects for choosing an appropriate dataset for visual navigation tasks are also discussed.

1 INTRODUCTION

The essential functionality of mobile robots is their ability to navigate in an operating environment. The classic way to navigate a mobile robot in its environment is to count wheel turns and then estimate a motion path. It is called wheel odometry. An alternative to wheel odometry is inertial odometry, which utilises an inertial measurement unit (IMU), which can measure angular velocity, linear acceleration, and the magnetic field.

Visual navigation, which started to be used in mobile robots and drones quite recently, uses visual sensors (digital cameras) as its main source of data and is a more accurate approach. Visual navigation can be divided into two areas of research (Scaramuzza and Fraundorfer, 2011): visual odometry and visual simultaneous localization and mapping (visual SLAM). The former provides only relative pose estimation - that is, only the local position of a vehicle on a map - whereas the latter deals with the global position of a vehicle on a map. Visual SLAM uses loop-closures (previously seen parts of operational environments) that allow to fully re-estimate the actual position of the vehicle by using all the previously seen data. Therefore, visual SLAM is a computationally expensive approach and has important limitations when operating on real-time systems and mobile robots or micro-drones. On the other hand, visual odometry is more efficient and requires significantly fewer computational resources. However, visual navigation systems equipped with conventional digital cameras also have limitations such as motion blur effect, data redundancy, relatively low dynamic range, power-consuming and computationally expensive devices.

Event-based vision is a new generation of computer vision. It involves a dynamic visual sensor (DVS), also called an event-based camera (EBC) or 'silicon retina' (Brandli et al., 2014), as the primary sensor. The DVS is a biologically inspired alternative to conventional digital cameras designed to overcome their limitations. The DVS imitates the operating principle of the retina. Instead of transmitting all the pixels of a frame (as in the case of conventional digital cameras) from the image sensor, the DVS asynchronously transmits only the pixels that undergo some threshold brightness intensity changes. DVS cameras are power-efficient, have a high dynamic range and high temporal resolution. Thus, the DVS is a particularly promising sensor for use in mobile robots and drones as the main component of event-based visual navigation.

Developing methods of visual navigation requires a source of repeatable data. Thus, datasets are the most exploited resource for different kinds of benchmarks, evaluation of algorithms, models training, and performance measurement. Publicly available datasets are useful when real sensors are not physically available or when research is mainly concerned with methods rather than data preparation. A dataset is a volume of specific data stored in a structured way

Zujevs, A. and Nikitenko, A.
Visual Navigation Datasets for Event-based Vision: 2014-2021.
DOI: 10.5220/0010607105070513
In *Proceedings of the 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2021)*, pages 507-513
ISBN: 978-989-758-522-7
Copyright © 2021 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

and documented for other users. Datasets are useful when either real sensors are not available or when it is necessary to use data prepared in a specific way. This paper focuses on the datasets aimed at visual navigation tasks (e.g. structure from motion and particularly for Visual Odometry (VO), reconstruction, segmentation, and visual SLAM) using a DVS camera. The rest of the paper is organized as follows. Section 2 provides a brief description of the event-based vision. Section 3 offers a concise survey of publicly available event-based visual navigation datasets. Finally, Section 4 discusses the reviewed datasets and provides general conclusions.

2 EVENT-BASED VISION

Event-based vision is a new technology of visual data generation by a visual sensor, as well as of the way this new type of visual data is processed. Instead of generating a sequence of image frames, a DVS sensor produces a stream of events. Each event represents a particular pixel's intensity level change above a certain threshold value. An event is a tuple of x,y coordinates of the pixel, with a timestamp measured in microseconds and polarity, which represents the direction of the intensity level change. The DVS produces data only for scenes - views of operational environments from the sensor's perspective - with movement caused either by the sensor's ego-motion or movements in the scenes themselves, for example, see Figure 1.

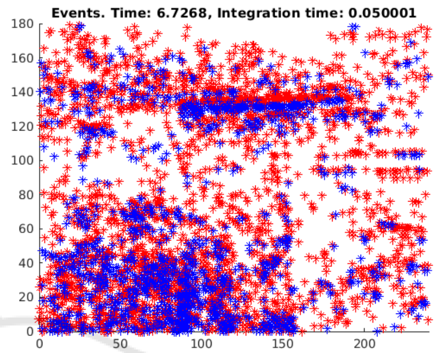
Address-Event Representation (AER) (Conradt et al., 2009) is a standard for communication, processing, and storage of event data, which was first introduced in (Mahowald, 1992). Subsequently, the jAER project was introduced by the event-based vision community in 2006. It provides API for work with various versions of DVS, as well as many different methods for event data processing [1]. Within the jAER project, many groups of researchers provide their own methods of implementation (Mueggler, 2017), (Brandli et al., 2016), (Katz et al., 2012), (Rueckauer and Delbruck, 2016), (Liu and Delbruck, 2018), (Benosman et al., 2014). Another resource related to jAER is the C library[2] named cAER, which is an optimized jAER project for embedded computers and is distributed as a standalone library. Since 2019, Inivation AG has been developing a new software development library[3] for DVSs, with interfaces for C++, Python and ROS.

[1]https://github.com/SensorsINI/jaer/

[2]https://github.com/inivation/libcaer

[3]https://inivation.com/dvp/

(a)



(b)

Figure 1: Example form the dataset (Zujevs et al., 2021): (a) colour frame acquired by an RGB-D camera, (b) events represented according to the colour frame scene and produced by the camera's ego-motion, and then accumulated over a short period of time; red markers are positive events (pixels whose intensity increases) and blue markers are negative events (pixels whose intensity decreases). The events are timestamped with microsecond resolution.

The field of event-based vision is growing fast. In (Gallego et al., 2020), a survey of event-based vision is presented.

3 RECENT DATASETS

The dataset[4] presented in (Barranco et al., 2016) contains data sequences from a DAVIS240 (events and APS frames) and a Microsoft Kinect Sensor (RGB-D sensor). The sensors are mounted on a Pan-Tilt Unit (PTU-46-17P70T) on board a Pioneer 3DX Mobile Robot. The PTU provides the pan and tilt angles and angular velocities while the mobile robot provides the direction of translation and speed. The dataset contains real and synthetic data in a total of 40 static sequences for the indoor environment - an office with or without people. The data sequences contain ob-

[4]https://github.com/fbarranco/eventVision-evbench

jects of different sizes, textures and shapes, and the sensors are rotated or translated to some degree. The events are provided both in AERdat2.0 data format and in matlab files while depth is provided in pgm and matlab files. Also, the author includes the synthetic data generated from conventional CV datasets. The ground truth is provided by 3-D motion parameters in textual data format for the 3-D translation and 3-D pose of the camera (respecting the DAVIS coordinates). In addition, the dataset provides the ground truth as 2-D image motion fields generated from the depth and the 3D motion. Calibration of the DAVIS and the RGB-D sensor is also provided in the dataset.

The dataset[5] published in (Weikersdorfer et al., 2014) contains 26 data sequences (each 20-60 seconds long) from an eDVS and a PrimeSense sensor (a colour camera equipped with a depth sensor). Ground truth data are provided in the *bvh* data format from an OptiTrack V100 motion capture system. Data from other sensors are provided in the text data format. Events are provided in both the eDVS's pixel coordinates and in the PriveSensor's pixel coordinates with depth values. The data sequences are mostly provided in 640x480 resolution at 30Hz. Each data sequence is accompanied by the estimated path of the proposed SLAM method. The dataset contains data of hand-held 6-DOF motion in static and dynamic office scenes with and without people. Along with the dataset, the authors propose a novel event-based 3D V-SLAM (EB-SLAM-3D) and eDVS calibration method, which uses a checkerboard calibration target and a blinking LED in the centre to estimate the pixel-to-pixel correspondence between the eDVS and the RGB-D sensors.

In (Mueggler et al., 2017b) proposed dataset is aimed at comparing event-based SLAM methods. The dataset[6] containing a total of 27 data sequences from DAVIS240 and synthetic data sequences (each 2-133 seconds long) is presented. The sequences provide hand-held and slider motions. The dataset includes the following objects and scences: patterns, wall poster, boxes, outdoors, dynamic, calibration, office, urban scenes, scenes with objects captured by a motorized linear slider, 3 synthetic planes, and 3 synthetic walls. The ground truth data are provided by a motion capture system and by the DAVIS's IMU, and, for some data sequences, by the slider's position. For the data captured in outdoor environments, no ground truth data are provided. Events and IMU data are provided in text files while images are available in png files. The data sequences are also available in rosbag data containers. The authors provide the first version of a DVS simulator based on the BLENDER tool.

Paper (Binas et al., 2017) offers a dataset[7] intended to investigate event camera applications in automatic driver assistance systems (ADAS). A new update of the dataset is presented in (Hu et al., 2020). It was used for training a neural network to predict the instantaneous steering angle using data from a DAVIS346. For all the recordings, the camera was mounted in a fixed position behind a windshield. A polarisation filter was used in some recordings to reduce the windshield and hood glare. The dataset consists of a total of over 12 hours of a car driving under various weather, driving, road, and lighting conditions for seven consecutive days with a total mileage of 1000km, comprising different types of roads. The data were stored in the HDF5 data format. A lot of car parameters were read at 10Hz rate (e.g. steering wheel angle, accelerator pedal position, engine speed etc.). The typical duration of the data sequences is 1-60 min. The data in the dataset tend to be unbalanced. The authors also provide Python-based tools[8] for data visualization and export.

In (Zhu et al., 2018), the authors present the first work[9] where a synchronized stereo pair of DAVIS346B was installed on a sensor rig and then mounted on a hexacopter, on the roof of a car and a motorcycle. Data were gathered in different environments and at different illumination levels. From each DAVIS camera, the following streams of data were recorded: grey-scale images, events and IMU data. Additionally, a stereo camera (VI sensor from Skybotix) and a LIDAR (Velodyne VLP-16 PUCK LITE) were used, and data were recorded from the LIDAR, an indoor and outdoor motion capture system, and a GPS sensor. A total of 14 data sequences are available. Ground truth is provided by the motion capture system for indoor and outdoor scenes. For other data sequences where the motion capture system was not available, LIDAR odometry was used. GPS data accompany the ground truth data. The data sequences are provided in rosbag and hdf5 data containers.

In (Scheerlinck et al., 2019), the first color-event dataset[10] recorded by the color version of DAVIS346 is provided. This is a general-purpose dataset without ground truth. Also, the updated version of ESIM (Event-based Camera Simulator) (Rebecq et al., 2018) for color events generation is presented. The dataset contains the following types of scenes: simple objects, indoor/outdoor, people, and

[5] http://ebvds.neurocomputing.systems/EBSLAM3D/index.html

[6] http://rpg.ifi.uzh.ch/davis_data.html

[7] http://sensors.ini.uzh.ch/databases.html

[8] https://code.ini.uzh.ch/jbinas/ddd17-utils

[9] https://daniilidis-group.github.io/mvsec/

[10] http://rpg.ifi.uzh.ch/CED.html

various lighting conditions (daylight, indoor light, low light), as well as camera motions (linear, 6-DOF motion) and dynamic motions.

Paper (Bryner et al., 2019) presented a method that tracks the 6-DOF pose of an event-based camera in an initially known environment described by a photometric 3D map (intensity + Depth) created using the classical approach of dense 3D reconstruction. The method uses direct event data without employing features, and it was successfully evaluated on real and synthetic data. The dataset[11] was released for public use. It includes the acquired images and the ground truth of the camera's trajectory. In this paper, the authors are more focused on the localization on a given map. Ground truth data for real data were provided by a motion capture system. A total of 23 data sequences are provided within rosbag data containers.

The first dataset (Zujevs et al., 2021) aimed at visual navigation tasks in different types of agricultural environment for the autumn season is publicly available[12]. It provides a total of 21 data sequences in 12 scenarios. The data sequences were gathered by a sensor bundle with the following elements onboard: a DVS240, a Lidar (OS-1, 16 channel), an RGB-D (Intel RealSense i435) and environmental sensors. The dataset is accompanied by sensors calibration results and raw data used during the sensors calibration procedure. For each sequence, a video demonstrating its content is provided. Ground truth is provided by three LIDAR SLAM methods, where a Cartographer (Hess et al., 2016) estimated the loop closure more accurately than the other two methods.

The first dataset (Gehrig et al., 2021) aimed at driving scenarios in challenging illumination conditions, where data are recorded from two monochrome high-resolution event-based cameras[13], two RGB cameras (FLIR Blackfly S USB3), a LIDAR (Velodyne VLP-16), and GPS (GNSS receiver), is available in[14]. In total, it provides 53 sequences 12-2255 seconds long. All the involved sensors were intrinsically and extrinsically calibrated. Ground truth data are provided by GPS and estimated depth from fusing the LIDAR data with event and frame-based camera data. The data are provided in the text, png and hdf5 data formats. All the aforementioned datasets are summarized in Table 1.

[11] http://rpg.ifi.uzh.ch/direct_event_camera_tracking/

[12] https://ieee-dataport.org/open-access/agri-ebv-autumn

[13] Prophesee PPS3MVCD, 640x480 pixels.

[14] http://rpg.ifi.uzh.ch/dsec.html

4 DISCUSSIONS

In this section, the aspects of the dataset usage are discussed. Obviously, the most important factor in choosing an appropriate dataset is the visual task(s) that has(ve) to be performed. Some of the common visual navigation tasks are 2-D/3-D motion estimation (Gallego et al., 2016), scene reconstruction(Kim et al., 2016), visual SLAM(Vidal et al., 2018) and image motion estimation (also called optical flow) (Benosman et al., 2014). All the reviewed datasets are appropriate for 2-D/3-D motion estimation, at least when using only the data from a DVS sensor. Other sensors can improve estimation results if a motion estimation method or a framework uses sensor fusion. For the motion estimation task, good results are obtained by fusing DVS data with IMU and colour or grey-scale image, as proposed in (Weikersdorfer et al., 2014), and (Zhu et al., 2017). However, an additional requirement arises - the need for the ground truth motion path. All the datasets, except for (Scheerlinck et al., 2019), provide ground truth data (either via a motion capture system or estimated using data from the other used sensors, for example, LIDAR data).

The image motion estimation task requires depth data, such data are available in (Barranco et al., 2016), (Weikersdorfer et al., 2014), (Zhu et al., 2018), (Gehrig et al., 2021), and (Zujevs et al., 2021) dataset.

Scene reconstruction allows to reconstruct a scene - by using an event stream - as grey-scale images, all the datasets are appropriate for this task.

The visual SLAM task allows estimating the global position of a mobile robot or a drone on a map. Visual SLAM requires loop closures in data sequences, and the following datasets contain loop closures: (Mueggler et al., 2017b), (Weikersdorfer et al., 2014), (Zhu et al., 2018) and (Zujevs et al., 2021). Ground truth is also an additional requirement for SLAM method evaluation and comparison purposes.

The data format used in data sequences of a dataset is the second important aspect that should be taken into account. There are three common data formats used in datasets: textual (data are stored in text files), native binary (data are stored in native binary files associated with the appropriate sensor), rosbag (data containers used by ROS[15]). Usually, datasets use mixed data types of data sequences. For example, in the dataset (Barranco et al., 2016), textual, binary and Matlab files are used to store data, and, in (Mueggler et al., 2017b), textual, binary and rosbag files are used.

[15] Robot Operating System

Table 1: Summary of visual navigation datasets: 2014-2021.

Year	2014	2016	2016	2017	2018	2019	2019	2021	2021
Paper	(Weikersdorfer et al., 2014)	(Barranco et al., 2016)	(Mueggler et al., 2017b)	(Binas et al., 2017)	(Zhu et al., 2018)	(Scheerlinck et al., 2019)	(Bryner et al., 2019)	(Zujevs et al., 2021)	(Gehrig et al., 2021)
Visual task									
3-D motion estim.	•	•	•	•	•	•	•	•	•
Visual-SLAM	•		•		•			•	•
Sensors used									
DAVIS		•	•	•	•	•	•		
DVS	•							•	•
RGB-D	•	•					•	•	•
LIDAR					•			•	•
Other	•	•		•	•				•
Sensors mounting point									
Hand-held	•		•		•		•		•
Car				•	•				
Mobile platform		•				•		•	
Drone					•				
Other			•						
Environment: outdoor									
City and country			•	•	•	•			•
Tunnels				•		•			•
Highways				•	•	•			•
Agricultural env.								•	
Environment: indoor									
Office with/without people	•	•	•		•	•	•		
Simple objects		•	•		•	•			
Posters and HDR			•						
Agricultural env.								•	
Ground truth is provided by									
Motion Capt.Syst.	•		•				•		
IMU		•	•		•			•	
Other odometry	RGB-D		Slider pos.	GPS	GPS LIDAR		3-D map	LIDAR	GPS LIDAR Depth
Data and data format used									
Number of seq.	26	40	27	-	14	84	23	21	53
Sequence length	20-60sec		2-133 sec	12hours		25ms-28min	10-45 sec	111-337sec	12-2255sec
Data format	text bvh	AER matlab pgm	text png rosbag	hdf5	rosbag hdf5	rosgab	rosbag	rosbag text AER png pcd	hdf5 text png
Dataset location									
URL link	Link	Link	Link	Link	Link	Link	Link	Link	Link

Another two aspects that should be considered are the availability of ground truth and the sensor coordinate systems used. Ground truth allows to compare methods in a quantitative way by applying different kinds of metrics. Each visual navigation task has its own type of ground truth. Another aspect is sensor coordinate system used as primary within data sequences. There are two common approaches: (1) the calibration parameters are provided to make your own transformation between sensors coordinate systems (body frames), and (2) all the data are already transformed into the main coordinate system of one of the visual sensors.

In addition, another important factors that influence the choice of a particular dataset are the environment and the motion type of the camera. As shown in Table 1, datasets are dedicated to indoor and outdoor environments. The differences between these environments include illumination conditions, the type of a scene - static or dynamic (where objects in a scene are moving)- types of objects and their shape and pattern, camera mounting place (on a car, on a mobile robot, on a hexacopter, and hand-held mounted). Depending on the visual navigation task and the requirements for the used methods, an appropriate dataset should be used. In many situations, the availability of ground truth is also a major requirement, which allows to do a quantitative analysis, however, the accuracy of the ground truth might be different. If a motion capture system is used to generate the ground truth, then the accuracy is high. Unfortunately, a motion capture system is not always available, especially for outdoor scenes. Hence, the ground truth is estimated from the data of the sensor used, for example, LIDAR data.

Finally, there are no event-based datasets aimed at event-based feature detection and tracking for visual navigation tasks. That is essential aspect for the 2-D/3-D motion estimation by using feature detecting tracking methods, for example, Arc*(Alzugaray and Chli, 2018), eHarris(Vasco et al., 2016), and eFast(Mueggler et al., 2017a).

5 CONCLUSIONS

Datasets aimed at event-based visual navigation are currently scarce because of the novel type of the dynamic vision sensor used. Event-based methods for all the mentioned visual navigation tasks are also scarce. Another difficulty is the rare availability of event-based methods implementations in open source resources. This fact complicates the evaluation and application of the proposed methods in real robotic systems.

The reviewed datasets are an important contribution to the development of event-based visual navigation methods. These datasets provide data sequences for different types of environment from DVSs, depth sensors, RGB-D, LIDAR, and their IMUs. In total, nine datasets were summarized, in different groups of features. Each dataset is accompanied by a data location link. All the mentioned datasets have ground truth data, except for one dataset, which provides data from a new colour version of the DVS camera. Another, currently unique, dataset is aimed at agricultural environments, where data are recorded in such settings as a forest, a meadow, a cattle farm, etc.

Choosing an appropriate dataset is an essential task for successful evaluation and development of new methods as well as for their quantitative and qualitative analysis. The type of environment and the type of camera motion used (fast, slow, rotational, and translational) within n-DOF are two major factors. While there is a sparse availability of event-based visual navigation datasets, there are no datasets that provide data for event-based feature detection and tracking. This direction of event-based visual navigation is based on the classical approach to how motion is estimated from frame-based data. Based on all of the above, the design of new datasets is highly necessary since it will lead to the development and better availability of new methods.

ACKNOWLEDGEMENTS

A.Zujevs is supported by the European Regional Development Fund within the Activity 1.1.1.2 “Post-doctoral Research Aid” of the Specific Aid Objective 1.1.1 (No.1.1.1.2/VIAA/2/18/334), while A.Nikitenko is supported by the Latvian Council of Science (lzp-2018/1-0482).

REFERENCES

Alzugaray, I. and Chli, M. (2018). Asynchronous Corner Detection and Tracking for Event Cameras in Real Time. *IEEE Robotics and Automation Letters*, 3(4):3177–3184.

Barranco, F., Fermuller, C., Aloimonos, Y., and Delbruck, T. (2016). A Dataset for Visual Navigation with Neuromorphic Methods. *Frontiers in Neuroscience*, 10(FEB):1–9.

Benosman, R., Clercq, C., Lagorce, X., Sio-Hoi Ieng, and Bartolozzi, C. (2014). Event-Based Visual Flow. *IEEE Transactions on Neural Networks and Learning Systems*, 25(2):407–417.

Binas, J., Neil, D., Liu, S.-C., and Delbruck, T. (2017). DDD17: End-To-End DAVIS Driving Dataset. pages 1–9.

Brandli, C., Berner, R., Minhao Yang, Shih-Chii Liu, and Delbruck, T. (2014). A 240x180 130 dB 3 us Latency Global Shutter Spatiotemporal Vision Sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341.

Brandli, C., Strubel, J., Keller, S., Scaramuzza, D., and Delbruck, T. (2016). ELiSeD — An event-based line segment detector. In *2016 Second International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, pages 1–7. IEEE.

Bryner, S., Gallego, G., Rebecq, H., and Scaramuzza, D. (2019). Event-based, Direct Camera Tracking from a Photometric 3D Map using Nonlinear Optimization. In *2019 International Conference on Robotics and Automation (ICRA)*, volume 2019-May, pages 325–331. IEEE.

Conradt, J., Berner, R., Cook, M., and Delbruck, T. (2009). An embedded AER dynamic vision sensor for low-latency pole balancing. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 780–785. IEEE.

Gallego, G., Delbruck, T., Orchard, G. M., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A., Conradt, J., Daniilidis, K., and Scaramuzza, D. (2020). Event-based Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1.

Gallego, G., Lund, J. E. A., Mueggler, E., Rebecq, H., Delbrück, T., and Scaramuzza, D. (2016). Event-based, 6-dof camera tracking for high-speed applications. *ArXiv*, abs/1607.03468.

Gehrig, M., Aarents, W., Gehrig, D., and Scaramuzza, D. (2021). Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, 6(3):4947–4954.

Hess, W., Kohler, D., Rapp, H., and Andor, D. (2016). Real-time loop closure in 2D LIDAR SLAM. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, volume 2016-June, pages 1271–1278. IEEE.

Hu, Y., Binas, J., Neil, D., Liu, S.-C., and Delbruck, T. (2020). DDD20 End-to-End Event Camera Driving Dataset: Fusing Frames and Events with Deep Learning for Improved Steering Prediction. *arXiv*.

Katz, M. L., Nikolic, K., and Delbruck, T. (2012). Live demonstration: Behavioural emulation of event-based vision sensors. In *2012 IEEE International Symposium on Circuits and Systems*, pages 736–740. IEEE.

Kim, H., Leutenegger, S., and Davison, A. J. (2016). Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 349–364. Springer, Cham, eccv 2016. edition.

Liu, M. and Delbruck, T. (2018). Adaptive Time-Slice Block-Matching Optical Flow Algorithm for Dynamic Vision Sensors. In *British Machine Vision Conference 2018*.

Mahowald, M. (1992). *VLSI Analogs of Neuronal Visual Processing: A Synthesis of Form and Function*. PhD thesis, California Institute of Technology Pasadena, California.

Mueggler, E. (2017). *Event-based Vision for High-Speed Robotics*. PhD thesis, University of Zurich.

Mueggler, E., Bartolozzi, C., and Scaramuzza, D. (2017a). Fast Event-based Corner Detection. In *Procedings* *of the British Machine Vision Conference 2017*, volume 1, pages 1–11. British Machine Vision Association.

Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., and Scaramuzza, D. (2017b). The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *The International Journal of Robotics Research*, 36(2):142–149.

Rebecq, H., Gehrig, D., and Scaramuzza, D. (2018). ESIM: an Open Event Camera Simulator. In Billard, A., Dragan, A., Peters, J., and Morimoto, J., editors, *Proceedings of The 2nd Conference on Robot Learning*, pages 969–982. PMLR.

Rueckauer, B. and Delbruck, T. (2016). Evaluation of Event-Based Algorithms for Optical Flow with Ground-Truth from Inertial Measurement Sensor. *Frontiers in Neuroscience*, 10(APR).

Scaramuzza, D. and Fraundorfer, F. (2011). Tutorial: Visual odometry. *IEEE Robotics and Automation Magazine*, 18(4):80–92.

Scheerlinck, C., Rebecq, H., Stoffregen, T., Barnes, N., Mahony, R., and Scaramuzza, D. (2019). CED: Color Event Camera Dataset. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, volume 2019-June, pages 1684–1693. IEEE.

Vasco, V., Glover, A., and Bartolozzi, C. (2016). Fast event-based Harris corner detection exploiting the advantages of event-driven cameras. *IEEE International Conference on Intelligent Robots and Systems*, 2016-Novem:4144–4149.

Vidal, A. R., Rebecq, H., Horstschaefer, T., and Scaramuzza, D. (2018). Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High-Speed Scenarios. *IEEE Robotics and Automation Letters*, 3(2):994–1001.

Weikersdorfer, D., Adrian, D. B., Cremers, D., and Conradt, J. (2014). Event-based 3D SLAM with a depth-augmented dynamic vision sensor. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 359–364. IEEE.

Zhu, A. Z., Atanasov, N., and Daniilidis, K. (2017). Event-based visual inertial odometry. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:5816–5824.

Zhu, A. Z., Thakur, D., Ozaslan, T., Pfrommer, B., Kumar, V., and Daniilidis, K. (2018). The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception. *IEEE Robotics and Automation Letters*, 3(3):2032–2039.

Zujevs, A., Pudzs, M., Osadcuks, V., Ardavs, A., Galauskis, M., and Grundspenkis, J. (2021). Agri-EBV-Autumn.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS