



HAL
open science

Simulating Aerial Event-based Environment: Application to Car Detection

Ismail Amessegher, Hajer Fradi, Clémence Liard, Jean-Philippe Diguët,
Panagiotis Papadakis, Matthieu Arzel

► **To cite this version:**

Ismail Amessegher, Hajer Fradi, Clémence Liard, Jean-Philippe Diguët, Panagiotis Papadakis, et al.. Simulating Aerial Event-based Environment: Application to Car Detection. European Robotics Forum, Mar 2024, Rimini, Italy. 10.1007/978-3-031-76424-0_26 . hal-04497648v2

HAL Id: hal-04497648

<https://hal.science/hal-04497648v2>

Submitted on 8 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Simulating Aerial Event-based Environment: Application to Car Detection

Ismail Amessegher¹, Hajer Fradi¹, Clmence Liard^{1,2}, Jean-Philippe Diguët²,
Panagiotis Papadakis^{1,2}, Matthieu Arzel¹

¹IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France

²CNRS, IRL CROSSING, Adelaide, Australia

Abstract. With the primary goal of enhancing the efficiency of drones for research and rescue missions through the exploitation of neuromorphic sensors and event-based vision, our focus in this work lies in setting up a simulated environment that can be used for synthetic data generation. In particular, we employ Unreal Engine to generate scenes suitable for the case of vehicle perception, followed by a dynamic event-based simulation environment in interaction with the AirSim and v2e software tools. The synthetic event data acquired in this simulated environment is shown to provide a valuable resource for training Artificial Intelligence (AI) systems and more particularly for the task of car detection using YOLOv7.

Keywords: Drones, Simulation, Events, Synthetic Data, Car Detection

1 Introduction

The increasing frequency of natural disasters calls for new solutions to improve the effectiveness and efficiency of disaster response services. According to a United Nations review, in 2019 only, around 400 climate-related and geographical events resulted in tens of thousands of human losses. In this context, first-response by Search-And-Rescue (SAR) robots has shown its potential to operate in unstructured and hostile environments. A SAR team typically composed of aerial and/or ground vehicles operates remotely via a control post [7]. To accomplish such missions, efforts must be directed towards multiple aspects, including:

- Safe movement: enhancing safety with onboard AI assistance that ensures reliable obstacle detection.
- Fast movement: achieving low-latency environment capture for higher drone reactivity.
- Energy preservation: prioritizing artificial vision solutions that are conducive to low energy consumption.

SAR missions would therefore significantly benefit from technologies enabling higher energy efficiency combined with improved navigation autonomy of drones.

To fulfill this objective, our focus in this study is to investigate newly emerging bio-inspired sensors known as event or neuromorphic cameras [5]. Thanks to their unique properties including lightweight, energy-efficient, and high-frequency data acquisition, an increasing interest has been shown in exploiting these new sensors for perception as well as the control of drones [3].

It is crucial to note that event-based vision requires additional dataset acquisition and subsequent training, which are more readily accessible for conventional RGB data. Due to various constraints in the extensive deployment and usage of real robots for training and testing, the use of a simulated environment can reduce the risk of damage and compensate for the lack of real data via abundant simulated trajectories following previous works [1]. On this basis, in this work, we integrate complementary tools that jointly serve the goal of generating synthetic event aerial data that are finally used to train a proof-of-concept, custom car detector.

2 Simulated Event-based Environment

In the literature, object detection or tracking by drones has been predominantly performed using conventional frame-based cameras [8, 2, 4]. Event cameras are newly emerging bio-inspired sensors that asynchronously measure per-pixel brightness changes. Known also as neuromorphic sensors, they are different from conventional frame-based sensors where images are captured at a fixed frame rate. Their resulting output is a stream of events encoding the time, position, and sign of brightness changes. Event sensors have several advantages compared to conventional RGB cameras including high microsecond temporal resolution, very high dynamic range (140 dB vs. 60 dB), low power consumption, and high pixel bandwidth (in the order of kHz) [5]. Given all these advantages, these cameras have witnessed an increasing interest in drone applications.

At present, event-based ground-truth data acquired from airborne drones are scarce and limited in volume and diversity. To compensate for the risk of drone crashes when acquiring data from real images and allow for more accurate annotations in nearly infinite environmental scenarios, we opt for using a simulated pipeline. In particular, inspired by research works that use AirSim simulator [10] and Unreal Engine (UE), we build a simulated event-based environment coupled with the v2e toolbox [6] for synthetic event data creation.

In detail, AirSim (Aerial Informatics and Robotics Simulation) is an open-source simulation platform developed by Microsoft for unmanned robots. It is built using Unreal Engine as its graphics engine [9], which is particularly powerful and versatile. UE boasts impressive graphics capabilities, stunning visual effects, and strong support for virtual and augmented reality. AirSim comes with photo-realistic camera, depth and thermal sensors. For event-based data, the v2e toolbox allows to convert conventional image data into event-based data [6] while adjusting various parameters related to the image conversion, such as timing, output video dimensions and noise levels to simulate real-world event data collection conditions.

3 Data Generation

Since aerial images are expensive to be collected, and hard to be manually annotated, synthetic data becomes the most feasible solution to provide large amounts of automatically annotated data. For the task of car detection, the training data is created using the Modular Neighborhood pack ¹. The scene depicts a suburban neighborhood and contains vehicles, which are suitable for our application of car detection. For RGB images, the expected size is 640x480 and images need to be captured at a frequency of 200Hz. Since at this high acquisition frequency, data could be significantly blurred, it is worth making efforts in Unreal Engine and AirSim to slow down the simulation. Initially, we could capture images at a frequency of 20 Hz. By making time flow 10 times slower within the simulation, data capture would reach 200 Hz. AirSim allows us also to select the date at which the simulation takes place. This parameter enables data captures with four types of sunlight conditions: day, night, day followed by night, and night followed by day. This helps to test the recognition capabilities under varying lighting conditions.

To construct an image from raw events, various representations could be investigated [13]. In our approach, we opt to aggregate events within a temporal window of $\Delta t = 5$ ms, aligning with the initial frequency of the simulated environment. The resulting collected dataset comprises a total number of 10982 frames, with 7191 for training and 3791 for testing. The training set consists of 3216 samples acquired under daytime conditions and 3975 during the night. Similarly, the test set includes 2301 samples under daytime conditions and 1490 under nighttime conditions. For the annotations, bounding boxes with a size of less than 240 pixels are excluded. Figure 1 shows two sample images (daytime and nighttime), with their corresponding event frames and annotations.

4 Experiments

We choose to evaluate our prototype for the task of car detection, conducting experiments on the YOLOv7 detector [11], motivated by its effectiveness for real-time applications as well as its more accurate results compared to previous YOLO versions [12]. Reportedly, it outperforms other real-time object detection models in both speed and accuracy. In the training phase, we use a mini-batch size equal to 64 and 30 epochs. The model is trained on an NVIDIA RTX A6000 GPU. The performance of car detector in the test set of our created synthetic event data is evaluated in terms of mean Average Precision (mAP) of detections at Intersection Over Union (IOU) equal to 0.5 regarding the ground truth boxes. The achieved mAP attains 0.63, with precision and recall both reaching 0.93 and 0.63, respectively. These outcomes affirm that we succeed in precisely detecting cars within the simulated data, though there is room for improvement in terms

¹ <https://www.unrealengine.com/marketplace/en-US/product/modular-neighborhood-pac>

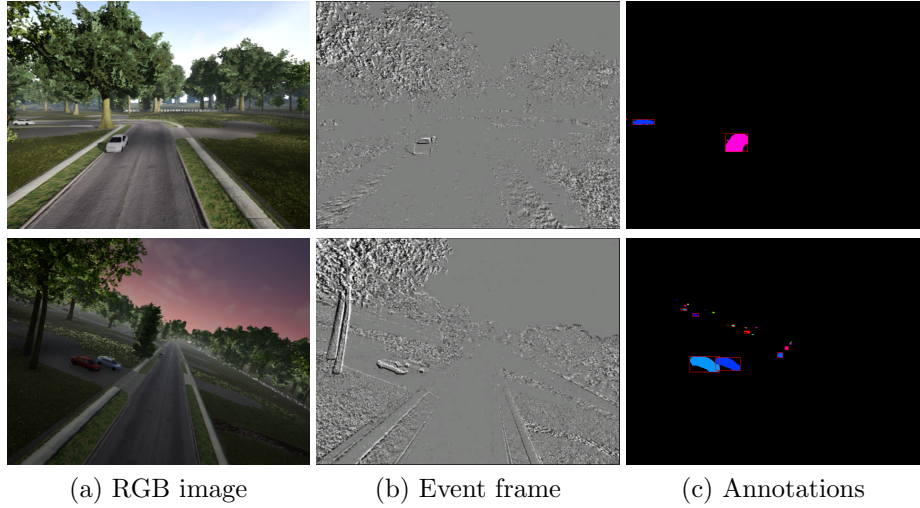


Fig. 1. Two sample images of the simulated dataset with their corresponding generated event frames and annotation boxes in the segmentation map. In the first row, a frame during the day, and in the second row, another sample under nighttime conditions.

of recall. Increasing true positives by detecting more objects could be achieved by expanding the size of the training data.

5 Conclusion

In this work, we build a simulation environment based on Unreal Engine, AirSim and v2e toolbox to simulate dynamic aerial events. These tools are assessed in generating event-based data, which is used to train a custom car detection model. By means of experiments using the YOLOv7 detector, the utility of the simulated event data for training a custom model under different lighting conditions. Having used the Modular Neighborhood environment scene driven by its relevance to our use-case, we could subsequently improve the variety of objects and scene conditions via the addition of further scenes.

6 Acknowledgement

This work has received a French government support granted to the Labex CominLabs excellence laboratory and managed by the National Research Agency in the “Investing for the Future” program under reference ANR-10-LABX-07-01 from September 2022 to December 2024. In this program, the project associated to this work is *LEASARD* (Low-Energy deep neural networks for Autonomous Search-And-Rescue Drones).

References

1. Mitriakov Andrei, Papadakis Panagiotis, and Garlatti Serge. An open-source software framework for reinforcement learning-based control of tracked robots in simulated indoor environments. *Advanced Robotics*, 36(11):519–532, 2022.
2. Yaru Cao, Zhijian He, Lujia Wang, Wenguan Wang, Yixuan Yuan, Dingwen Zhang, Jinglin Zhang, Pengfei Zhu, Luc Van Gool, Junwei Han, et al. Visdrone-det2021: The vision meets drone object detection challenge results. In *Proceedings of the IEEE/CVF International conference on computer vision*, pages 2847–2854, 2021.
3. Rika Sugimoto Dimitrova, Mathias Gehrig, Dario Brescianini, and Davide Scaramuzza. Towards low-latency high-bandwidth control of quadrotors using event cameras. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4294–4300. IEEE, 2020.
4. Hajer Fradi, Lorenzo Bracco, Flavia Canino, and Jean-Luc Dugelay. Autonomous person detection and tracking framework using unmanned aerial vehicles (uavs). In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1047–1051. IEEE, 2018.
5. Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020.
6. Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic dvs events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1312–1321, 2021.
7. Geert-Jan M. Kruijff, Fiora Pirri, Mario Gianni, Panagiotis Papadakis, Matia Pizzoli, Arnab Sinha, Viatcheslav Tretyakov, Thorsten Linder, Emanuele Pianese, Salvatore Corrao, Fabrizio Priori, Sergio Febrini, and Sandro Angeletti. Rescue robots at earthquake-hit mirandola, italy: A field report. In *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2012.
8. Anitha Ramachandran and Arun Kumar Sangaiah. A review on object detection in unmanned aerial vehicle surveillance. *International Journal of Cognitive Computing in Engineering*, 2:215–228, 2021.
9. Andrew Sanders. *An introduction to Unreal engine 4*. CRC Press, 2016.
10. Shital Shah, Debadepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics: Results of the 11th International Conference*, pages 621–635. Springer, 2018.
11. Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7464–7475, 2023.
12. Hangyue Zhao, Hongpu Zhang, and Yanyun Zhao. Yolov7-sea: Object detection of maritime uav images based on improved yolov7. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 233–238, 2023.
13. Nikola Zubić, Daniel Gehrig, Mathias Gehrig, and Davide Scaramuzza. From chaos comes order: Ordering event representations for object recognition and detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12846–12856, 2023.