APEIRON: a Multimodal Drone Dataset Bridging Perception and Network Data in Outdoor Environments

Nunzio Barone
nunzio.barone@poliba.it
Politecnico di Bari
Bari, Italy

Walter Brescia
walter.brescia@poliba.it
Politecnico di Bari
Bari, Italy

Saverio Mascolo
saverio.mascolo@poliba.it
Politecnico di Bari
Bari, Italy

Luca De Cicco
luca.decicco@poliba.it
Politecnico di Bari
Bari, Italy

ABSTRACT

Unmanned Aerial Vehicles (UAVs), commonly denoted as drones, are being increasingly adopted as platforms to enable applications such as surveillance, disaster response, environmental monitoring, live drone broadcasting, and Internet-of-Drones (IoD). In this context, drone systems are required to carry out tasks autonomously in potentially unknown and challenging environments. As such, deep learning algorithms are widely adopted to implement efficient perception from sensors, making the availability of comprehensive datasets capturing real-world environments important. In this work, we introduce APEIRON, a rich multimodal aerial dataset that simultaneously collects *perception data* from a stereocamera and an event based camera sensor, along with measurements of *wireless network links* obtained using an LTE module. The assembled dataset consists of both perception and network data, making it suitable for typical perception or communication applications, as well as cross-disciplinary applications that require both types of data. We believe that this dataset will help promoting multidisciplinary research at the intersection of multimedia systems, computer networks, and robotics fields. APEIRON is available at https://c3lab.github.io/Apeiron/.

CCS CONCEPTS

• **Computer systems organization** → *Robotic autonomy*; *Sensors and actuators*; • **Information systems** → Multimedia information systems; • **General and reference** → **Measurement**; • **Networks** → **Network measurement**.

ACM Reference Format:
Nunzio Barone, Walter Brescia, Saverio Mascolo, and Luca De Cicco. 2024. APEIRON: a Multimodal Drone Dataset Bridging Perception and Network Data in Outdoor Environments. In *ACM Multimedia Systems Conference 2024 (MMSys '24), April 15–18, 2024, Bari, Italy*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3625468.3652186

MMSys '24, April 15–18, 2024, Bari, Italy

ACM ISBN 979-8-4007-0412-3/24/04
https://doi.org/10.1145/3625468.3652186

1 INTRODUCTION

Unmanned Aerial Vehicles (UAVs), commonly denoted as drones, are being increasingly adopted in various applications, including surveillance [5], disaster response [16], environmental monitoring [1, 9], live drone broadcasting [11, 17, 21], and Internet-of-Drones (IoD) [7]. Achieving full autonomy for drone systems is a key requirement for carrying out complex missions in Beyond Visual Line-Of-Sight (BVLOS) scenarios. To this end, drones require an accurate perception of the environment to effectively navigate the surroundings. Several sensing modalities can be leveraged, such as Time-of-Flight (ToF) sensors (f.i., LiDARs), radars, RGB and depth cameras, etc. At the same time, drones often require long-range wireless communication systems, such as cellular networks, to deliver real-time information, such as telemetry, the output of given sensors, or live videos, to ground operators or live audiences [12]. The communication layer is important not only for improving situational awareness in scenarios such as disaster response and surveillance, but also for drone swarms [10, 15], and various entertainment scenarios as well, f.i., to stream live sports events and for news coverage [20].

In this context, as learning based algorithms are being increasingly adopted to implement efficient end-to-end autonomous drone systems, the availability of comprehensive public datasets capturing real-world environments becomes important. The literature provides several datasets collected using UAVs that can be grouped into two main categories: 1) *perception datasets* [4, 8, 13], which collect data typically used for computer vision tasks, visual SLAM, etc, and 2) *network-related* datasets focusing on scenarios such as IoD [12, 14]. Nevertheless, at the best of authors' knowledge, there is no public dataset that combines data coming from perception sensors with network measurements.

In this work, we introduce APEIRON, a rich multimodal aerial dataset that collects perception data from a variety of sensors and measurements of network bandwidth obtained using an LTE module. In particular, concerning perception, we equip a self-built hexarotor drone with 1) a stereocamera, namely the ZED2i from Stereolabs[1] and 2) an Event-Based camera, the Prophesee EVK4 HD[2]. Along with such data, the Flight Control Unit (FCU) used, the PixHawk, enables capturing GPS, gyroscope, accelerometer, magnetometer, barometer and, environmental temperature data, as well.

[1]https://store.stereolabs.com/en-it/products/zed-2i
[2]https://www.prophesee.ai/event-camera-evk4/

It is important to note that, while RGB-D cameras provide valuable insights on the scene such as point clouds and depth images, due to their limited dynamic range, their performance can be impaired in conditions of sudden change of light (e.g. moving from a dark room to a bright one). This may result in gaze effects when the camera is looking at a light source (e.g. the sun), which usually leads to incorrect or inaccurate perception of the environment.

Event-based cameras [6] are light-based sensors that address these challenges. These devices are composed of a matrix of pixels which, instead of detecting light levels, as classical RGB cameras do, detect changes in light intensity for each pixel independently. The detection of this intensity variation is denoted as an *Event*. Each event is produced in output asynchronously. This enables such sensors to capture data at a frame rate that is higher than several orders of magnitude than that of classical RGB cameras.

Concerning network data, we employ an LTE module to carry out active TCP measurements of the downlink and uplink. The metrics collected in the network traces comprise estimated available bandwidth and low level metrics such as RTT, congestion window, packet losses, etc.

The resulting dataset is a collection of perception and network data, enabling both typical perception or communication applications, as well as cross-domain applications that require both data types. We argue that the APEIRON dataset will help fostering multidisciplinary research at the intersection of multimedia systems, computer networks, and robotics fields.

2 RELATED WORK

In this section we discuss the datasets that are more closely related to the one proposed in this work. In particular, we group the datasets available publicly into two categories: 1) datasets collecting network-related data from drones, discussed in Section 2.1; 2) perception datasets obtained with drones equipped with sensors, presented in Section 2.2.

2.1 Network-related datasets

In [18], a study is conducted measuring Long Term Evolution (LTE) network performance using drones in a rural setting. The study focuses on interference caused by drones and the related effects on the utilization of network resources.

The authors of [12] provide raw LTE data captured using a drone in rural and urban-like environments. The collected dataset aims at improving UAV communication, navigation, and study cellular coverage at different altitudes. The paper also offers post-processing and experiment code for further research.

In [14], authors present an analysis of data from numerous drone flights at varying altitudes and locations, using multiple cellular carriers. This data was collected in a medium-sized urban environment. The data, along with the operating system, chipset information, drone instrumentation, and server-side packet captures, is offered as an open-source repository for the community.

All the datasets described above collect network-related data, but do not provide any perception data.

2.2 Perception datasets

In [3], authors present a method for tracking power lines using a drone equipped with an event camera similar to the one used in this work. It identifies lines in the event stream by detecting planes in the spatio-temporal signal and tracks them over time. The method is evaluated in real-world flights along a powerline, showing significant improvements over existing approaches. While this work highlights the potential of event cameras for industrial applications, the associated dataset is not multi-modal as it only contains perception data from the event based camera and does not contain networking data.

Authors of [2] introduce a dataset for high-speed drone racing. The dataset includes over 27 sequences in an indoor and an outdoor environment, with more than 10 km of flight distance, captured on a first-person-view (FPV) racing quadrotor flown by an expert pilot. The dataset features camera images, inertial measurements, event-camera data, and precise ground truth poses. While the dataset provides insightful perspectives on FPV quadrotors, it lacks networking information, while focusing on a very specific application.

In [8] authors present DroneFace, a dataset consisting of facial RGB images taken from a camera mounted on a long stick as if they were taken from a drone at various distances and heights. The dataset contains both frontal and side portrait images of subjects, for facial recognition training purposes. In [13] an aerial dataset was captured at different altitudes (50 to 500 meters) collecting monocular RBG images of one roundabout in Cyprus at 3840×2160 resolution. The dataset was then annotated to test how the detection of cars is affected by the altitude of the drone.

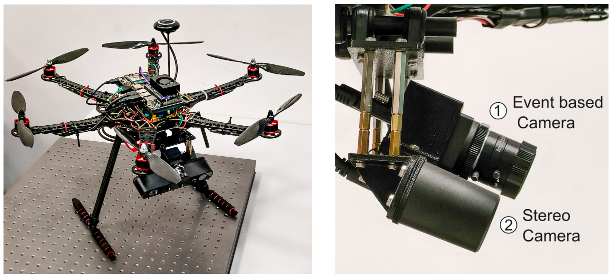
The dataset most closely related to ours is discussed in [4]. In particular, the dataset incorporates data from i) a mmWave radar sensor, ii) a full HD RGB camera (1080p), iii) a 240×180 resolution event-based camera, and iv) a gyroscope. Nevertheless, the dataset presented in this paper differs from [4] in several ways. Firstly, [4] focuses on collision avoidance tasks in indoor scenarios, whereas our dataset is collected in outdoor scenarios with a wider range of possible applications. Moreover, our work collects data using a significantly higher resolution event-based camera (i.e. 1080×720 instead of 240×180) and features RGB-D images instead of RGB images. The inclusion of depth information in the RGB-D data can serve as ground truth, facilitating applications such as monocular depth estimation from the event-based camera. Finally, compared to [4], the APEIRON dataset proposed in this work also introduces network measurements.

3 METHODOLOGY

This section presents the methodology employed to collect the APEIRON dataset. Section 3.1 describes the data collection platform, including information on the drone and onboard equipment employed. Next, Section 3.2 presents the software stack implemented to collect data.

3.1 Data collection platform

Figure 1(a) shows the self-built hexarotor drone employed to collect the APEIRON dataset. The sensors and hardware related to the features of the dataset are reported in Table 1.



(a) Hexarotor drone (b) Sensors placement

Figure 1: Hexarotor drone and sensors placement

Table 1: Sensors and hardware employed for the collection of perception and networking data

Device	Description
Prophesee EVK4 HD	*Data type*: Events *Sensor*: Sony IMX636 HD 1280×720 pixels, 1/2.5" CMOS *Pixel latency*: < 100 μs *Dynamic range*: ≥ 120 dB
Stereolabs ZED2i	*Data types*:RGB-D (L/R monocular and depth), point cloud, acceleration, angular rate, magnetometer, barometric pressure, temperature, pose estimate *Sensor*: 2688×1520 pixels, 1/3" CMOS, 2.1 mm focal length *FoV*: Horizontal 110°, Vertical 70° *Baseline*: 12 cm *Depth range*: 20 m *Capture res.*: 720p@60 fps *Dynamic range*: 64.6 dB *Positional tracking accuracy*: ±1 mm *Orientation accuracy*: ±0.1°
PixHawk 2.4.8 (FCU)	*Data types*: IMU data, GPS data, barometric pressure, temperature, flight logs
Quectel EG25-G LTE module	*Max Bandwidth*: 150Mbps DL/ 50Mbps UL

In the following, we describe in more details the employed drone, sensors, and their placement on the drone.

UAV frame and propulsion.
The hexacopter is based on the DJI S550 frame type, a structure chosen because it provides enough space for customization and payload capacity. The propulsion system is built using six 2212 920 KV brushless motors coupled with 1045 propellers controlled by 30 A ESCs powered by a 4S battery. This setup allows to lift a total weight of 3 kg (drone, battery, sensors, computing).

Flight controller unit.
The FCU employed is an off-the-shelf *Pixhawk 2.4.8* which was configured with the latest stable PX4 version[3]. This unit is responsible for the stabilization and position control of the UAV, allowing the execution of autonomous missions defined by waypoints lists. In this work it will be employed in two main GPS aided flight modes: 1) *Position flight mode*: the UAV is controlled by an human operator to allow the fine grained control of the path when inspecting close targets and for test purposes; 2) *Mission flight mode*: a list of waypoints and procedures is loaded onto the FCU, allowing repeatability.

[3]At the time of writing 1.14.0

Perception sensors.
Stereocamera: Stereolabs' ZED2i is an RGB-D camera equipped with internal accelerometer, gyroscope, magnetometer, barometer, and temperature sensor. This device comes with a Software Development Kit (SDK) which enables the computation of depth images, self-pose tracking, environment reconstruction (in the form of a point cloud), etc.
Event-Based camera: the EVK4 HD is the latest evaluation kit produced by Prophesee, featuring the IMX636, an HD CMOS sensor with a pixel latency of 100 μs and a dynamic range greater than 120 dB. The event processing is embedded on the device, allowing very fast event processing. This device has an equivalent temporal resolution greater than 10000 FPS. The data produced from this sensors can be stored in a raw format and consist in the events encoded in the EVT3 format. Each detected event is described by the following data: 1) X and Y *position*, 2) *event polarity*, a boolean value that defines whether the brightness level has increased or decreased, and 3) event *timestamp*[4]. These data can be exported in post production to any format (e.g., CSV) in order to be used more easily.
Sensors displacement: the FCU, along with its onboard sensors, are mounted to be joint with the main plate of the drone and at its center (zero displacement). Sensors are mounted to be joint together and to share the field of view as shown in Figure 1(b). In particular, custom mounting plates have been designed and 3D printed to mount the EB camera to the stereo camera, which already provided mounting points. Then, another mounting plate has been designed to attach the stereo camera to the drone. This mounting plate has a 30° offset with respect to the plane on which the FCU lays. This allows the sensors to have a field of view which can frame both the ground and the front view, in order to capture features which can be useful for many applications, such as environment reconstruction, self localization, etc.

Computing. The companion computer is an *NVIDIA Jetson Xavier NX 16 GB* with a 512 GB SSD storage device. The board is employed to collect sensors data and perform network measurements. The software stack designed to collect sensors' data is described in Section 3.2.

Connectivity. The communication with the FCU is guaranteed by a remote radio controller and a Holybro SiK telemetry module, while the connectivity of the companion computer is provided by the *Quectel EG25-G* LTE module[5].

3.2 Software stack

Figure 2 shows the software architecture used for the collection of perception and network data.

Data collection requires synchronization among each data source to allow the usage of multi-modal data sources. To this end, we

[4]Expressed in μs relative to camera activation.
[5]https://www.quectel.com/product/lte-eg25-g

store a timestamp for each data source. This allows to synchronize different data sources in post processing. A Docker image has been developed for each data source collection. Each running image (*container*) writes data in a dedicated directory, mounted as a volume. The data collection is launched using the Docker *compose* tool.

Perception.

Stereo camera. Regarding ZED2i, we used a Docker image[6] provided by StereoLabs to record the live feed information in a proprietary format denoted as SVO, using the *ZED_SVO_Recording*[7] tool from ZED SDK 4.0. This tool allows to simulate a real camera by reading the recorded SVO files. This way, it is possible to playback the data flow captured by the camera, including video feed and sensors' data. This functionality supports the offline usage of the standard SDK, allowing the generation of data such as depth maps, dense point clouds, pose estimates, and maps' meshes. Computation of such data in offline mode allows to use a workstation with higher computation capabilities to leverage the more demanding SDK configurations, reaching the highest quality available, otherwise inaccessible with an embedded computer such as the Jetson Xavier. The SVO format also stores the UNIX timestamp for each frame and sensor data, enabling synchronization with other data sources in post production.

Event camera. Concerning the EVK4 HD, the OpenEB[8] SDK, an open version of the Metavision SDK provided by Prophesee, was used to record a "raw" file storing EVT3 data and a "bias" file representing the calibration parameters used for the camera. This allows to playback and manipulate the entire event camera's data flow. The events are stored along with a timestamp that is the number of microseconds elapsed since the capture was started. To allow a unique representation in time, we also provide a text file containing the UNIX timestamp at which the event-based camera was started. In post-production events can be exported to CSV files along with their timestamp. This eases the use of events without the need to interface with the metavision SDK.

Network.

Network traces are obtained by starting a TCP flow either in the downlink or uplink direction using the `iperf3` tool. In particular, an `iperf3` server instance is run on the remote server, while the client is executed on the on-board computer. Two data types are stored: 1) TCP traffic packets in PCAP format obtained using `tcpdump`, and 2) timestamped socket statistics such as RTT, Goodput, cwnd, ssthresh, packet losses, by adapting `ss-pretty`[9] to our needs. The process is similar for both uplink and downlink measurements, the only difference is on where `ss-pretty` is executed. Indeed, since the congestion control algorithm runs at the sender, `ss-pretty` is executed either on the on-board computer to measure the uplink channel, or on the remote server for the downlink channel case.

4 THE APEIRON DATASET

In this Section, we present the APEIRON dataset. We start describing the considered scenarios (Section 4.1), then we present the dataset format (Section 4.2) and, finally, we provide a quantitative analysis of the dataset (Section 4.3).

4.1 Scenarios

At the time of writing, the APEIRON dataset collects two different scenarios: an industrial scenario (IND) and an open field scenario (OF). Multiple trajectories at different heights have been executed for the same scenario, in order to capture different points of view. Each scenario is selected to highlight possible sensors' limitations and strengths.

In particular, the first scenario, shown in Figure 3(a), is close to an *industrial building*: the drone executes several trajectories capturing the sides and the surroundings of the building. This scenario can be used, f.i., to study Visual Inertial Odometry (VIO) localization techniques [19] or for 3D reconstruction applications. The scenario alternates frames with plenty of features with frames with very few visual features.

Figure 3(b) shows the *open field* scenario: the drone captures data of an open field at different heights. In this case, due to the scenario's nature, few visual features are available. Therefore, it represents a very challenging environment for VIO applications.

To implement the trajectories two different approaches were used: i) manually controlled flights in position mode; ii) autonomous flights. The first approach was used to perform close inspections to buildings to ensure safety of the operations. The second one employs the *QGroundControl* tool to define missions by giving a waypoint list to the FCU, this allows smoother and repeatable trajectories.

4.2 Dataset format

Data collected in APEIRON are organized in *runs*: trajectories executed by the drone in a given scenario and identified by the timestamp, expressed as "run-*day-month-year-hours-minutes-seconds*". Each run is associated to a directory, containing:

- `event.bias`: a file describing the biases employed to capture data from the Event-Based camera;
- `event.raw.timestamp`: a file containing the initial UNIX timestamp used as offset for the raw events, relative to the initial boot of the camera;
- `event.raw`: the file collecting the events, produced by the *Metavision* SDK;
- `zed2i.svo`: the recordings of the *ZED2i* camera, obtained from the *Stereolabs*' SDK;
- `log-day-month-year-hours-minutes-seconds.ulg`: it contains the flight logs, containing all the data captured by PixHawk's sensors, such as GPS data, accelerometer, magnetometer, control inputs for the motors, etc;
- `tcpdump-down(up).pcap`: the TCP packets capture for the down(up)-link communication to a remote server, collected using *tcpdump*;
- `tcp-internals-down(up).log`: output of the *ss-pretty* tool, leveraged to capture the socket related statistics;
- `run.json`: it contains metadata of the run.

In the dataset Github repository we also provide preprocessing tools and Jupyter notebooks which implement the starting functionalities to access data.

[6] stereolabs/zed:4.0-tools-devel-l4t-r35.2
[7] https://github.com/stereolabs/zed-sdk/tree/master/recording
[8] https://github.com/prophesee-ai/openeb
[9] https://github.com/gaddman/ss-pretty

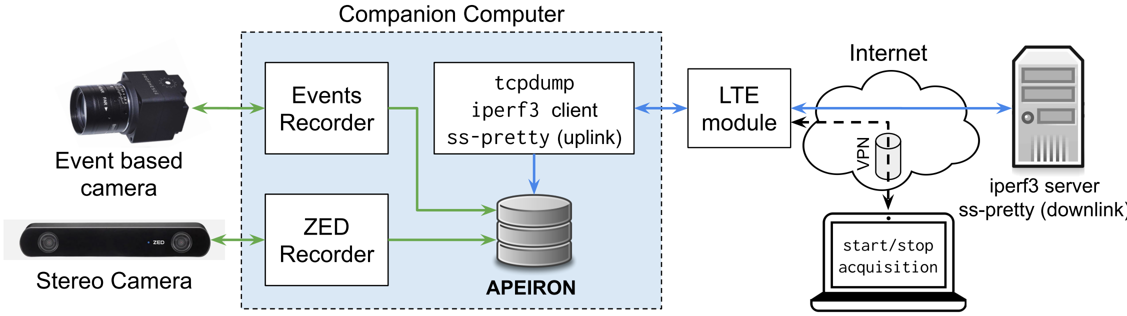
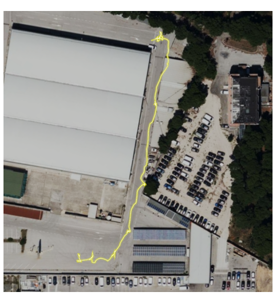


Figure 2: Software and connection architecture composed by: the Jetson Xavier on-board of the UAV, the remote server used as target for the uplink/downlink network traces, the user SSH interface used to start/stop and monitor the collection process via a remote shell.

Table 2: APEIRON dataset main features

Run	Flight duration [min]	RGB frames	Avg events rate [Mev/s]	EVT3 vs SVO File size [GB]	GoodPut [Mbps]	Average RTT [ms]	Direction	PX4 Online Flight Preview
IND-1	03:20	10324	3.6	2.5 vs 7.1	2.389	184.271	Upstream	http://bit.ly/49kb3LE
IND-2	04:02	12031	3.6	3.0 vs 8.3	N/A	N/A	Downstream	https://bit.ly/48XWCNt
IND-3	04:11	11299	5.5	4.4 vs 7.9	4.372	190.886	Downstream	https://bit.ly/490x1DE
IND-4	03:58	11079	4.7	3.7 vs 7.7	N/A	N/A	Upstream	https://bit.ly/48YQB3a
OF-1	04:14	11142	6.0	4.6 vs 9.9	9.721	412.106	Upstream	https://bit.ly/42y38YY
OF-2	02:15	5520	7.5	2.9 vs 5.0	3.873	144.713	Downstream	https://bit.ly/3HMtMni
OF-3	03:07	9220	9.2	4.8 vs 7.7	7.003	493.119	Downstream	https://bit.ly/3SItrs6
OF-4	03:35	9513	11.8	7.6 vs 9.7	6.477	104.132	Upstream	N/A



(a) Industrial Scenario

(b) Open Field Scenario

Figure 3: Example scenarios and trajectories

4.3 Analysis of the dataset

The main features of APEIRON are summarized in Table 2. In particular, the dataset collects a total of 100 GB of data for an approximate 30-minute flight time.

To ease flight review, Table 2 also reports the link to the *PX4 Flight review* online tool for each run. The dataset also includes an uncontrolled crash (OF-4). We decided to keep such kind of data to provide researchers with information of such a kind of event in order to enable and spur the development of techniques, f.i., for early detection of failures.

4.3.1 Network traces. To give insights on the collected network traces, Figure 4(a) and (b) show respectively the CDF of goodput and RTT for each of the runs. Concerning the goodput (Figure 4(a)), the highest median value is obtained in an open field scenario (OF-2) and, as expected, on the downlink direction. Nonetheless, it is also associated with a higher standard deviation compared to the other runs. The measured RTTs shown in Figure 4(b) exhibit in one case a heavy tailed distribution (f.i., OF-1). Upon inspection of the logs, this is due to a (low) number of outlier values (RTT≈7 s) measured towards the end of the run. The other runs exhibit RTT distributions that are typical in mobile scenarios. On the other hand, in industrial scenarios, while the median RTT is lower compared to the former case, the standard deviation is reduced.

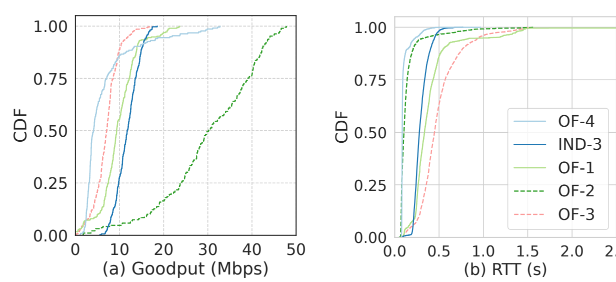
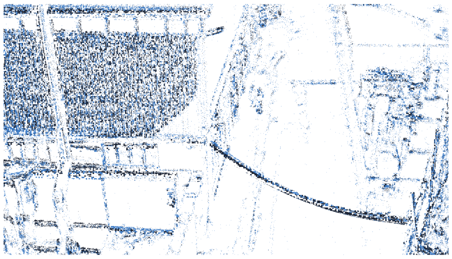


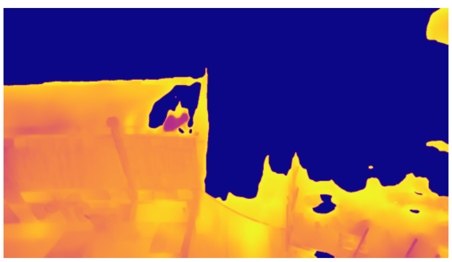
Figure 4: Network traces: continuous (dashed) lines refer to uplink (downlink) measurements.

4.3.2 Perception. To provide a better understanding of the data representation of these sensors, Figure 5 shows a frame sample from the run IND-4 for several sensing modalities. In particular, the figure presents a frame from, respectively: the EB camera, the RGB image, the depth image and the related point cloud from run IND-4. In order to keep RGB and EB images comparable, the RGB image has been properly cropped to match the FOV of the EB Camera.

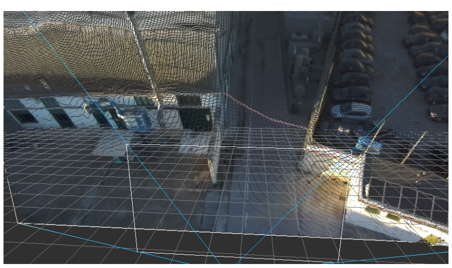


(a) EB Camera

(b) RGB Camera (Left)



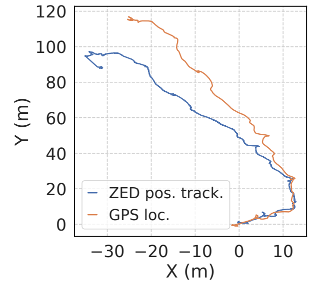
(c) Depth Camera



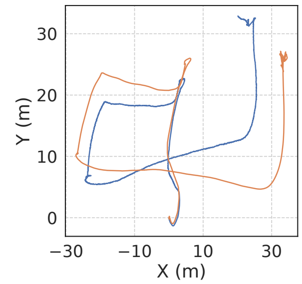
(d) ZED Local Point Cloud

Figure 5: Perception data (run IND-4)

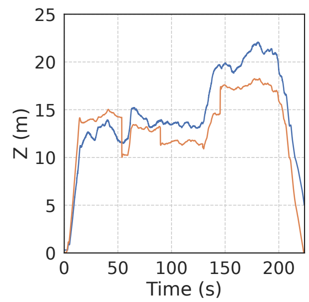
An interesting insight is provided by comparing the measured GPS position with the pose estimate obtained from the ZED SDK's Positional Tracking tool that employs the data stored in the SVO file to produce such estimate. Figure 6 compares the two trajectories obtained by the two localization systems for two different scenarios. The first one is an industrial scenario (IND4, Figures 6(a) and 6(c)), with many visual features. The second is an open field scenario (OF2, Figures 6(b) and 6(d)), with much less visual features. The figure analyzes the trajectories in the X and Y axes (Figures 6(a) and 6(b)) and on the Z axis by the flight duration (Figures 6(c) and 6(d)). As expected, the pose estimation from the ZED SDK's Positional Tracking shows a drift in both scenarios, exacerbated by the lack of useful visual features in the open field environment.



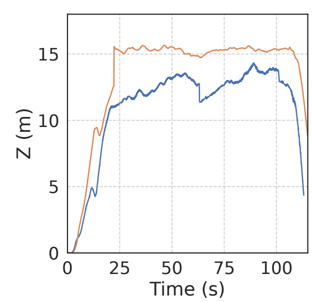
(a) IND4 - XY



(b) OF2 - XY



(c) IND4 - Z



(d) OF2 - Z

Figure 6: Localization using GPS or ZED Positional Tracking

5 CONCLUSIONS

In this paper, we have presented APEIRON, a multi-modal public dataset that combines perception and network data from drones in outdoor scenarios. The dataset includes multiple types of sensors: a high-resolution event-based camera, an RGB-D camera, gyroscope, accelerometer, and GPS data, providing rich and diverse information for various applications. APEIRON is unique in its combination of networking and perception data in outdoor settings, making it a valuable resource for multidisciplinary research at the intersection of multimedia systems, computer networks, and robotics fields.

The dataset presented in this work has several potential applications in areas such as object detection, tracking, and recognition, as well as for multimedia systems applications involving streaming of sensors data to a GCS. We hope that this dataset will inspire further research and development in the fields of perception and networking for drones.

In future works, we plan to expand the APEIRON dataset by adding more scenarios, sensors, and trajectories. We also plan to develop new algorithms and applications that leverage the unique features of the dataset, such as event-based vision and depth estimation.

ACKNOWLEDGMENTS

This work was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART", CUP: D93C22000910001) and the AGREED project (ARS01_0025) funded by the Italian Ministry of University and Research.

REFERENCES

[1] Saeid Asadzadeh, Wilson Jose de Oliveira, and Carlos Roberto de Souza Filho. 2022. UAV-based remote sensing for the petroleum industry and environmental monitoring: State-of-the-art and perspectives. *Journal of Petroleum Science and Engineering* 208 (2022), 109633.

[2] Jeffrey Delmerico, Titus Cieslewski, Henri Rebecq, Matthias Faessler, and Davide Scaramuzza. 2019. Are We Ready for Autonomous Drone Racing? The UZH-FPV Drone Racing Dataset. In *2019 International Conference on Robotics and Automation (ICRA)*. 6713–6719. https://doi.org/10.1109/ICRA.2019.8793887

[3] Alexander Dietsche, Giovanni Cioffi, Javier Hidalgo-Carrio, and Davide Scaramuzza. 2021. PowerLine Tracking with Event Cameras. In *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*.

[4] Julien Dupeyroux, Raoul Dinaux, Nikhil Wessendorp, and Guido De Croon. 2022. A Novel Obstacle Detection and Avoidance Dataset for Drones. In *System Engineering for Constrained Embedded Systems* (Budapest, Hungary) *(DroneSE and RAPIDO)*. Association for Computing Machinery, New York, NY, USA, 8–13. https://doi.org/10.1145/3522784.3522786

[5] Jianrui Fan, Lei Lei, Shengsuo Cai, Gaoqing Shen, Pan Cao, and Lijuan Zhang. 2024. Area Surveillance with Low Detection Probability using UAV swarms. *IEEE Transactions on Vehicular Technology* 73, 2 (2024), 1736–1752.

[6] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. 2022. Event-Based Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 1 (2022), 154–180. https://doi.org/10.1109/TPAMI.2020.3008413

[7] Mirmojtaba Gharibi, Raouf Boutaba, and Steven L Waslander. 2016. Internet of drones. *IEEE Access* 4 (2016), 1148–1162.

[8] Hwai-Jung Hsu and Kuan-Ta Chen. 2017. DroneFace: An Open Dataset for Drone Research. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (Taipei, Taiwan) *(MMSys'17)*. 187–192. https://doi.org/10.1145/3083187.3083214

[9] Katharin R. Jensen-Nau, Tucker Hermans, and Kam K. Leang. 2021. Near-Optimal Area-Coverage Path Planning of Energy-Constrained Aerial Robots With Application in Autonomous Environmental Monitoring. *IEEE Transactions on Automation Science and Engineering* 18, 3 (2021), 1453–1468. https://doi.org/10.1109/TASE.2020.3016276

[10] Han Jiang, Yanchun Chang, Liying Yang, Xu Liu, and Yuqing He. 2023. Cooperative Exploration of Heterogeneous UAVs in Mountainous Environments by Constructing Steady Communication. *IEEE Robotics and Automation Letters* 8, 11 (2023), 7249–7256.

[11] Muhammad Asif Khan, Emna Baccour, Zina Chkirbene, Aiman Erbad, Ridha Hamila, Mounir Hamdi, and Moncef Gabbouj. 2022. A Survey on Mobile Edge Computing for Video Streaming: Opportunities and Challenges. *IEEE Access* 10 (2022), 120514–120550. https://doi.org/10.1109/ACCESS.2022.3220694

[12] Sung Joon Maeng, Ozgur Ozdemir, Ismail Guvenc, Mihail L Sichitiu, Magreth Mushi, and Rudra Dutta. 2023. LTE I/Q data set for UAV propagation modeling, communication, and navigation research. *IEEE Communications Magazine* 61, 9 (2023), 90–96.

[13] Rafael Makrigiorgis, Christos Kyrkou, and Panayiotis Kolios. 2023. How High can you Detect? Improved accuracy and efficiency at varying altitudes for Aerial Vehicle Detection. In *Proc. of the International Conference on Unmanned Aircraft Systems (ICUAS)*. 167–174.

[14] Gonzalo J Martinez, Grigoriy Dubrovskiy, Shangyue Zhu, Alamin Mohammed, Hai Lin, J Nicholas Laneman, Aaron D Striegel, Ravikumar V Pragada, and Douglas R Castor. 2021. An Open, Real-World Dataset of Cellular UAV Communication Properties. In *Proc. of the International Conference on Computer Communications and Networks (ICCCN)*. 1–6.

[15] Iago Medeiros, Azzedine Boukerche, and Eduardo Cerqueira. 2021. Swarm-based and energy-aware unmanned aerial vehicle system for video delivery of mobile objects. *IEEE Transactions on Vehicular Technology* 71, 1 (2021), 766–779.

[16] Balmukund Mishra, Deepak Garg, Pratik Narang, and Vipul Mishra. 2020. Drone-surveillance for search and rescue in natural disaster. *Computer Communications* 156 (2020), 1–10.

[17] Javad Sabzehali, Vijay K. Shah, Qiang Fan, Biplav Choudhury, Lingjia Liu, and Jeffrey H. Reed. 2022. Optimizing Number, Placement, and Backhaul Connectivity of Multi-UAV Networks. *IEEE Internet of Things Journal* 9, 21 (2022), 21548–21560. https://doi.org/10.1109/JIOT.2022.3184323

[18] Joonas Sae, Richard Wirén, Juhani Kauppi, Helka-Liina Maattanen, Johan Torsner, and Mikko Valkama. 2019. Public LTE network measurements with drones in rural environment. In *Proc. of the IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*. IEEE, 1–5.

[19] Davide Scaramuzza and Friedrich Fraundorfer. 2011. Visual Odometry [Tutorial]. *IEEE Robotics and Automation Magazine* 18, 4 (2011), 80–92. https://doi.org/10.1109/MRA.2011.943233

[20] Congxi Song, Biao Han, Xiaolan Ji, Yahui Li, and Jinshu Su. 2023. AI-driven Multipath Transmission: Empowering UAV-based Live Streaming. *IEEE Network* in press (2023).

[21] Xiaoli Wang, Aakanksha Chowdhery, and Mung Chiang. 2017. Networked drone cameras for sports streaming. In *Proc. of the IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 308–318.