

Drone Detection with Event Cameras

Gabriele Magrini
University of Florence
Florence

gabriele.magrini@unifi.it

Lorenzo Berlincioni
University of Florence
Florence

lorenzo.berlincioni@unifi.it

Luca Cultrera
University of Florence
Florence

luca.cultrera@unifi.com

Federico Becattini
University of Siena
Siena

federico.becattini@unisi.it

Pietro Pala
University of Florence
Florence

pietro.pala@unifi.it

Abstract

The diffusion of drones presents significant security and safety challenges. Traditional surveillance systems, particularly conventional frame-based cameras, struggle to reliably detect these targets due to their small size, high agility, and the resulting motion blur and poor performance in challenging lighting conditions. This paper surveys the emerging field of event-based vision as a robust solution to these problems. Event cameras virtually eliminate motion blur and enable consistent detection in extreme lighting. Their sparse, asynchronous output suppresses static backgrounds, enabling low-latency focus on motion cues. We review the state-of-the-art in event-based drone detection, from data representation methods to advanced processing pipelines using spiking neural networks. The discussion extends beyond simple detection to cover more sophisticated tasks such as real-time tracking, trajectory forecasting, and unique identification through propeller signature analysis. By examining current methodologies, available datasets, and the distinct advantages of the technology, this work demonstrates that event-based vision provides a powerful foundation for the next generation of reliable, low-latency, and efficient counter-UAV systems.

1. Introduction

The proliferation of small unmanned aerial vehicles (UAVs), commonly known as drones, has marked a significant technological turning point, unlocking a vast range of applications from commercial logistics and precision agriculture to infrastructure inspection and cinematography. However, this increasing accessibility has concurrently given rise to pressing challenges in security and public safety [67]. The potential for malicious use, including unau-

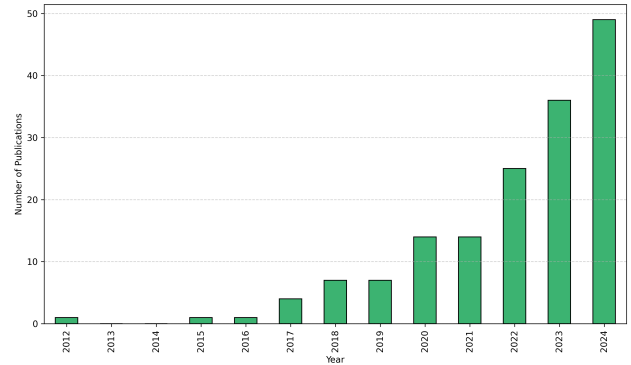


Figure 1. **Event camera based drone detection works trend.** We show the number of papers trough the years 2012-2024 as results of the query on Google Scholar (event_camera OR neuromorphic_vision) AND (drone OR UAV) AND (detection).

thorized surveillance, the disruption of controlled airspace, smuggling of contraband, and threats to critical infrastructure, has established the development of robust and reliable counter-UAV systems as a paramount research priority. The intrinsic characteristics of these aircraft, namely their small size, high agility, low acoustic and thermal signatures, and often non-metallic composition, present a serious detection challenge for traditional surveillance technologies.

Among the various sensing modalities employed for drone detection, vision-based systems are particularly compelling due to their passive nature and ability to provide rich information for target classification and identification [29, 46]. Nevertheless, conventional frame-based cameras, which acquire images at a fixed rate, suffer from fundamental limitations when tasked with detecting drone targets. The high angular velocity of a drone, especially in

close-range scenarios, frequently leads to significant motion blur, which degrades feature representation and complicates detection and tracking. Furthermore, the limited dynamic range of standard CMOS sensors makes them highly susceptible to challenging lighting conditions. A drone silhouetted against a bright sky can be lost in sensor saturation, while one operating in low-light or shadowed regions may fail to generate sufficient contrast for detection [38, 44]. These limitations necessitate a paradigm shift in visual sensing to overcome the specific challenges posed by drone detection.

In response to these shortcomings, a new class of bio-inspired sensors, known as event cameras, has emerged as a technology for high-speed and high-dynamic-range applications [19]. Unlike traditional cameras that capture a sequence of intensity frames, event cameras operate asynchronously. Each pixel independently monitors for changes in log-scale brightness. When a change exceeds a programmable threshold, the pixel generates an *event* containing its spatial coordinates, a microsecond-resolution timestamp, and the polarity of the brightness change. This event-driven data acquisition results in a sparse and continuous stream of information that encodes the dynamic aspects of a scene. The key advantages of this paradigm include a temporal resolution on the order of microseconds, which virtually eliminates motion blur; a very high dynamic range of over 120 dB, enabling robust performance in extreme lighting conditions; and low power consumption, as static parts of the scene generate no data.

The unique properties of event cameras are well-suited to the drone detection problem [35]. The high temporal resolution ensures that the fine details and motion trajectory of a fast-moving drone are captured without blur, providing a rich signal for detection and tracking algorithms [34, 42]. The high dynamic range allows for consistent detection in scenarios with severe backlighting or poor illumination, where conventional cameras would fail. Furthermore, the inherent data sparsity, which naturally filters out static backgrounds and highlights moving objects, serves as a powerful attentional mechanism. This significantly reduces data redundancy and allows computational resources to be focused on the target of interest, making event cameras an ideal sensor for real-time, low-latency, and power-efficient counter-UAV systems [5, 27]. In this work, we study these distinct advantages, analyzing the rapid diffusion of event cameras for drone detection (see Fig. 1).

2. Drone Recognition Across Domains

Recognition of flying objects generally falls into two main categories: natural entities like birds and insects, and artificial ones such as drones, flying vehicles and, generally speaking, Unmanned Aerial Vehicles (UAVs). Among these, drone recognition has attracted significant attention

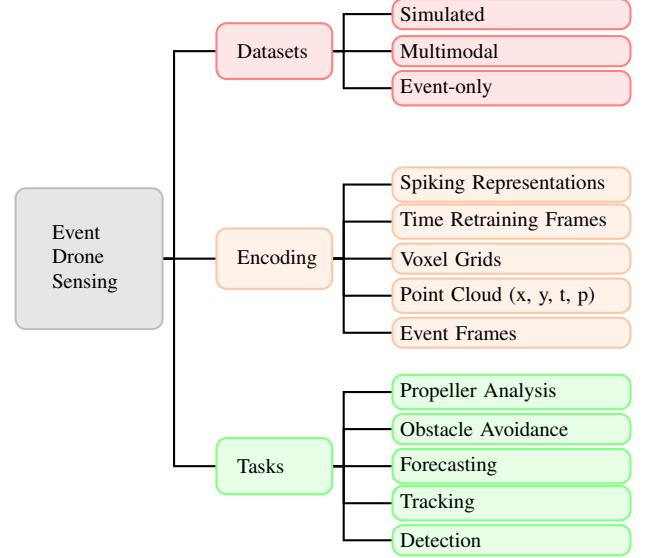


Figure 2. **Drone Sensing Taxonomy.** We group the related works taken into account and extrapolate a precise taxonomy.

due to growing security concerns, including risks related to unauthorized surveillance, airspace violations, and breaches of restricted zones.

Significant effort has been dedicated to improving the detection of unidentified flying objects across multiple sensor domains. The most common approach relies on the RGB domain, leveraging the maturity, low cost, and high quality of standard cameras [4, 11, 13, 20, 26, 41, 51, 65, 77]. Research in this area includes methods for detecting flying objects from a single moving camera [4] or tracking small unmanned aerial systems (UAS) [51]. Others use visible camera systems [26], develop real-time aerial localization and tracking systems [65], and enable collision avoidance through visual detection [77]. Beyond standard visual light, other optical sensors are also employed. Thermal cameras are used for night-time UAV detection [1], while Short-Wave Infrared (SWIR) cameras provide robust detection capabilities in both day and night scenarios [4]. Some studies adopt alternative optical bands, such as SWIR [10, 44] or combine multiple bands, such as VIS, SWIR, MWIR, and LWIR for a more comprehensive evaluation [20]. Further diversifying optical methods, some works employ depth maps [7] or LiDAR sensors for detect-and-avoid applications [15]. Non-optical methods are also common. Radar systems, such as a 35 GHz FMCW system, are used for drone detection [16], and cost-effective Radio Frequency (RF) based detection methods have been investigated [45]. Finally, acoustic sensors can detect drones by identifying their sound signatures [10, 40], a technique that is sometimes paired with optical detection methods [10].

3. Event-based Drone Sensing

Event-based drone sensing can be declined according to several subtasks and can be addressed in different ways depending on how the event stream is represented. We summarize the key distinctions in Fig. 2. In the following, we provide a formal definition of event streams and drone sensing from two different viewpoints: drone detection and drone fencing. Let \mathcal{E} denote a stream of events, defined as

$$\mathcal{E} = \{e_i = (x_i, y_i, p_i, t_i)\}_{i \in \mathbb{N}},$$

where $x_i \in [0, W - 1]$ and $y_i \in [0, H - 1]$ represent the spatial coordinates of the event e_i , $p_i \in \{0, 1\}$ denotes its polarity, and $t_i \in [0, \infty)$ is its timestamp.

Drone Detection The objective of the drone detection task is to train a detector $\mathcal{D}(\mathcal{I}_t)$ that produces a set of predicted bounding boxes

$$\hat{B} = \{\hat{b}_k = (\hat{x}_k, \hat{y}_k, \hat{w}_k, \hat{h}_k, \hat{t}_k)\}_{k \in \mathbb{N}},$$

where each bounding box is defined by its top-left corner (\hat{x}_k, \hat{y}_k) , its width \hat{w}_k , height \hat{h}_k , and associated timestamp \hat{t}_k . The detector operates on the input \mathcal{I}_t , which can be either the event stream, but also a combination of the event stream \mathcal{E} and a corresponding RGB stream \mathcal{V} , collected over a temporal window of duration Δ ending at detection time t . All detections at time t must be based solely on data available up to time t , without access to any future information.

Drone Fencing Differently from drone detection, the goal of drone fencing is not the precise localization of a drone within a scene, but rather to determine its passage or mere presence within a predefined surveilled area. In the drone fencing task, the objective is to train an identifier $\mathcal{F}(\mathcal{I})$ that produces a binary output \mathcal{P}_t at any given time t , indicating the drone presence in the scene.

Challenges of Drone Detection The steep increment in usage and diffusion of drones in the last years has dramatically increased the related challenges for their detection. This can be linked to 4 main causes, that affect event-based drone detection:

- *Variety*: modern drones are characterized by a notable variety of shape, dimensions and goals, making detection based solely on appearance hard to generalize in the wild;
- *High speed*: the speed range in drones is notable and is not trivial to retrieve information at high fps;
- *Extreme scenarios*: the conditions in which drone detection is necessary covers also extreme weather scenarios, low light environments, and harsh meteorological conditions (eg. rain, snow) which are challenging for event cameras too [9, 38];

- *Distractors and concurrent objects*: The operational environments often contain numerous other flying objects such as birds, other drones, airplanes, and even insects [11, 36], which can act as distractors and confuse detection systems.

4. Event-based Drone Detection

A significant variety of event-based drone detectors is present in the literature. Their characteristics mainly depend on the nature of the data and the way in which the data is represented: on the one hand, the availability of other modalities in addition to the event stream requires the model to handle multiple parallel streams; on the other hand, different ways of representing events may require different types of architecture to process them. A comparison of common event representations is shown in Fig. 3, where we depict a moving drone as an event point cloud, a voxel grid, or as two frame-based representations.

Frame Based In contrast to the native, asynchronous nature of event data, frame-based approaches transform streams of events into dense, temporally integrated representations that mimic traditional image formats. This conversion enables the direct application of established computer vision algorithms and architectures, which require structured inputs and are difficult to adapt to irregular, sparse data.

Several recent works on drone detection with event cameras [17, 35, 38, 39, 78] rely on frame-like representations obtained by accumulating events over a fixed temporal window Δt . For instance, Mandula et al. [39] accumulate events within constant intervals to produce two-channel representations: one for positive and one for negative brightness changes, which are then processed by a CNN-based detector. The authors also exploit this representation to spatially align event and RGB frames for annotation transfer. Similarly, Magrini et al. [35] propose a multimodal detection pipeline in which event frames are fused with RGB inputs in a network based on Detr [6]. In this setup, events are grouped in intervals of $\Delta t = 1/F$, where F is the RGB frame rate. In Magrini et al. [38], synchronized RGB and event frames are used for detection, tracking, and trajectory forecasting of UAVs in challenging real-world conditions.

In Eldeborg et al. [17], a Spiking Neural Network (SNN) and an Artificial Neural Network (ANN) are trained using artificial event frames. The input to both models is a 128×128 frame with two polarity channels, obtained either by aggregating events within 50ms (ANN) or by feeding them sequentially in 1ms slices (SNN). In Zundel et al. [78], event frames are fed into a YOLO-based detector [50], and DBSCAN [18] is used for post-processing to identify and track drone clusters.

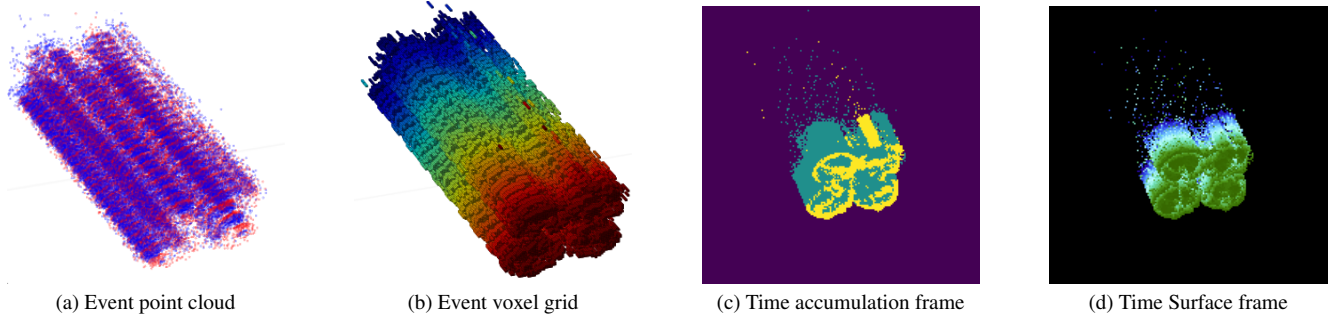


Figure 3. **Event data representations.** Comparison of different views of the event data for a drone. Samples from the FRED dataset [38].

These studies demonstrate that frame-based event representations can be effectively integrated into conventional vision pipelines. They enable low-latency inference, efficient implementation on existing hardware, and compatibility with widely used neural architectures such as ResNet [23], MobileNet [25], and YOLO. Moreover, frame-based inputs simplify data augmentation, training routines, and multimodal fusion. However, this approach sacrifices fine-grained temporal details by aggregating events over time, which can blur fast motion and miss brief dynamics critical for detecting small, fast drones. It also discards the original sparsity and temporal resolution from the event stream.

Point Cloud Based Event camera streams can be seen as time-continuous point clouds, thanks to the extreme temporal granularity offered by the neuromorphic sensor. At the same time, all the information is retained. Yet the different nature of the point cloud topology and semantics, when compared to typical 3D point clouds, makes the usage of standard point-based methods non-trivial. In the context of event-based tiny object detection, and specifically small or distant drones, Chen et al. [9] propose the Event-based Sparse Segmentation Network (EV-SpSegNet), a U-shaped architecture designed for direct processing of sparse event point clouds. A key component is the Grouped Dilated Sparse Convolution Attention (GDSCA) module, which leverages sparse convolutions and patch attention to efficiently capture multi-scale local and long-range spatiotemporal contextual features, critical for identifying the continuous curve-like trajectories of small moving targets. The key idea is that the movement described by an object in the 3D event point cloud is distinctive of its nature and can therefore be used for object detection by first segmenting the point cloud. Complementing this, the novel Spatiotemporal Correlation (STC) Loss is introduced to specifically promote event continuity and suppress isolated noise, further enhancing the network’s ability to segment tiny objects in event streams. Another approach, presented in Magrini et al. [36], is based on the classification of algorithmically detected objects in the event stream. In this case, the task is

to correctly classify multiple classes of unidentified flying objects, specifically drones, birds and insects, using a point cloud-based network. Specifically, the usage of PointNet [47] and PointNet++ [48] demonstrates a strong capability in distinguishing these classes by the implicit nature of their movements in an event point cloud.

Voxels Voxel grid representations [75] discretize the spatio-temporal volume of events into a regular 3D structure. The 2D sensor plane is extended along the temporal axis, which is partitioned into time bins. Each voxel encodes the number of events occurring at a given pixel location during a specific temporal interval. This representation captures temporal dynamics more explicitly than accumulated frames, while preserving a dense and regular structure compatible with standard deep learning architectures. Liu et al. [34] exploit voxel representations to compute event-driven optical flow. While not specific to drone detection, the voxel-based representation enables real-time motion analysis [74], a key component for detecting fast-moving objects. Wang et al. [62] present a more direct application, where EventVOT introduces a high-resolution benchmark and voxel-based baseline for object tracking, including UAVs. In this work, the voxel grid captures fine temporal motion information, and its regular shape facilitates the fusion with RGB modalities.

Time-retaining Frames While conventional frame representations aggregate events into static histograms over a fixed temporal window Δt , time-aware frame representations aim to retain the temporal structure of the event stream within a frame-like format. These encodings not only capture spatial event density but also preserve the relative timing of individual events, enabling networks to exploit dynamic motion cues. A classic example of time-retaining approach is Time Surface [54], where each pixel value encodes the timestamp of the most recent event. Similarly, Temporal Binary Representation (TBR) [28] and its spiking variant SpikeTBR [37] encode temporal information into the pixel values of frame representations. Mitrokhin et

al.[42] leverage a motion-compensated time image to segment moving objects from background motion. The representation encodes per-pixel timing information, allowing the system to disentangle drone trajectories from camera-induced motion.

End-to-end Neuromorphic approaches This scenario involves a complete pipeline of neuromorphic sensing, processing, and computing. Unlike hybrid approaches, where event data is converted into intermediate representations to interface with conventional deep learning models, end-to-end neuromorphic systems maintain the spike-based nature of the signal throughout the entire processing chain. This includes both Spiking Neural Network (SNN) architectures and dedicated neuromorphic hardware, such as Intel’s Loihi [14] or SynSense’s Speck [66], which are designed to exploit the sparse, event-driven characteristics of the input. These systems offer unmatched energy efficiency and latency, making them especially suitable for edge deployments in scenarios like drone detection, where power and bandwidth constraints are critical. In Kirkland et al. [31], the authors propose a low-power UAV detection system based on a retina-inspired neuromorphic vision sensor, leveraging a deep convolutional Spiking Neural Network trained using Spike-Timing Dependent Plasticity (STDP). The system is designed to detect drones by processing asynchronous event streams in real time, demonstrating the potential of STDP-trained SNNs for efficient aerial threat detection in edge scenarios. Another significant advantage of end-to-end neuromorphic pipelines is the extremely low power consumption; in Eldeborg et al. [17] the authors present a low-power drone detection system utilizing the Speck processor and embedded event-based camera to establish a virtual tripwire, with a Spiking neural network tasked with classifying the presence of the drone in the scene (drone fencing task, see Sec. 3). In Zhang et al.[68], the authors propose a low-power UAV object detection system based on event cameras, leveraging a Spiking Swin Transformer architecture that combines spiking neurons with hierarchical transformer features. In this case, no deployment on a real neuromorphic chip has been used.

5. Beyond Neuromorphic Drone Detection

Drone Tracking A natural extension of the drone detection task is drone tracking. In the drone tracking task, drone instances must be spatio-temporally identified across the entirety of a stream of events. Several event-based general-purpose tracking approaches have been proposed in the literature. Early works adapted traditional computer vision techniques, such as combining Discriminative Correlation Filters (DCF) with Convolutional Neural Network (CNN) features by first converting event streams into frame-like representations [32]. SiamEvent [8] shifted the focus to

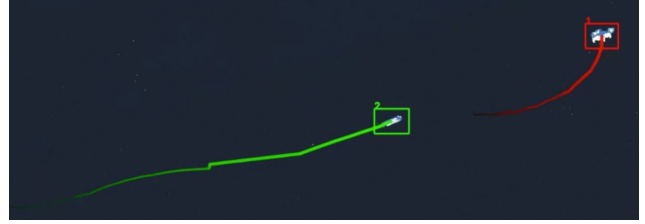


Figure 4. **Drone tracking task.** Drone tracking annotations taken from the FRED dataset [38]. Different colors correspond to different drones.

learning an edge-aware similarity metric more aligned with the information content of events. More recently, spiking neural networks have also been used for object tracking [69].

The first work to address drone tracking [42] proposed the EED dataset, which contained small drones among other objects. The task was addressed by developing an unconstrained motion segmentation model that estimated the effects of 3D camera motion from the event stream. A larger benchmark for object tracking including also UAVs, EventVOT [62], was recently proposed. The authors also present HDETrack, a tracking framework that leverages knowledge distillation, training a lightweight, event-only student tracker to mimic a powerful, multimodal RGB-Event teacher network, achieving high performance with low latency. Recently, some works have focused solely on tracking drones. Iaboni et al. [27] proposed a low-cost motion capture system that uses an event camera to track multiple quadrotors in real-time. The method combines YOLOv5 with KD-tree tracking, trained on synthetic event frames. End-to-end models for real-time neuromorphic detection and tracking of UAVs have also been proposed [55]: an asynchronous, event-by-event method, combines corner tracking with clustering to identify targets, a design choice aimed at minimizing computational load. A more traditional model is also proposed in the same work, where the event stream is first accumulated into event frames.

The authors of the FRED dataset [38] instead introduced a larger-scale benchmark for drone detection and tracking, better formalizing the task and proposing several tracking-by-detection baselines, combining object detections such as YOLO [30], RT-DETR [73] and FasterRCNN [22] with the state-of-the-art tracker ByteTrack [70]. An RGB-Event multimodal tracker is also proposed, based on ER-DETR [35]. MOTA (Multi-Object Tracking Accuracy), IDF1 (Identity F1 score), ID Switch, Precision and Recall are used to evaluate the trackers. An example is shown in Fig. 4.

Drone Forecasting Forecasting the position of fast-moving objects is an extremely challenging task. How-

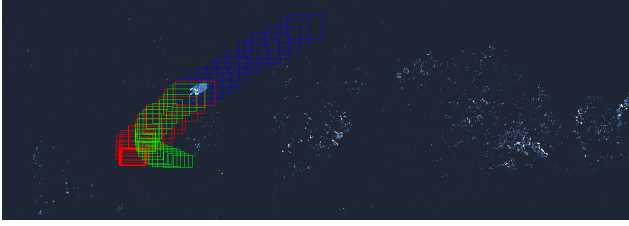


Figure 5. **Drone forecasting task.** Annotation for the forecasting task from FRED [38]. Blue: past; Green: GT; Red: prediction.

ever, the benefits of event cameras have been demonstrated in the literature [43]. Despite this, only a few works exist on forecasting the trajectories of UAVs with event cameras, which will likely become a fundamental application in the near future. Liang et al. [33] proposes an unsupervised approach that utilizes raw LiDAR point clouds to extract drone trajectories and aligns them with event camera images through motion consistency to generate pseudo-labels. By combining kinematic estimation with a visual Mamba neural network in a self-supervised manner, the method predicts future drone trajectories, outperforming supervised image-only and audio-visual baselines in long-horizon predictions. Magrini et al. [38] instead propose the only publicly available benchmark for event-based drone forecasting. The authors present blind references based on LSTMs or transformers that observe only drone coordinates and show that event-based models can largely improve the performance. Interestingly, event-based models consistently outperform RGB-based ones and the proposed multimodal approach outperforms unimodal models for short-term predictions. An example is shown in Fig. 5.

Propeller Blade Analysis A highly effective method for drone identification using event cameras is the analysis of the visual signature generated by propellers. As can be seen in Fig. 6, the rapid rotation of blades, which causes motion blur for standard cameras, creates a rich, high-frequency spatio-temporal pattern perfectly suited for the microsecond temporal resolution of event sensors [2, 52, 56]. These sensors leverage their high temporal resolution, high dynamic range (HDR), and sparse, event-driven output to capture the periodic brightness changes from the blades, effectively filtering out the static background [24, 52, 56]. This unique signature serves as a robust feature to distinguish drones from other objects and effectively acts as an uncooperative, natural fiducial marker for detection and tracking [52].

Several methodologies have been developed to analyze this propeller signature. Frequency-domain techniques treat the event stream as a periodic signal to extract its core frequency components. This analysis is conceptually analogous to the use of micro-Doppler signatures in radar systems for drone identification [16, 17]. More advanced prob-

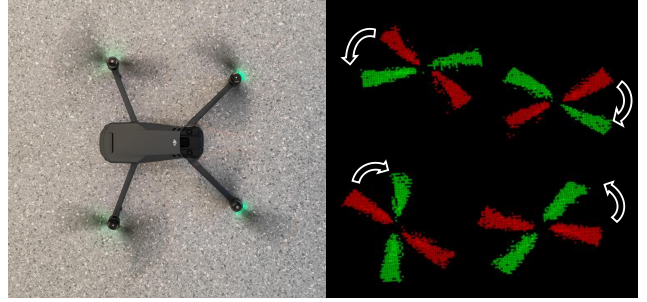


Figure 6. **Propeller Blade Analysis.** Drone propeller blades observed with an event camera. Image taken from [71].

abilistic methods offer greater robustness. One of these approaches models local event arrivals as Poisson processes to statistically distinguish the periodic bursts from propeller blades against background noise in real-time [56]. Finally, learning-based approaches have leveraged deep neural networks. The seminal work, EVPropNet, trained a CNN exclusively on a large-scale synthetic dataset of propeller events [52]. This network transfers directly to real-world scenarios without retraining, enabling complex on-drone applications like autonomous following and mid-air landing on unmarked drones, shifting the paradigm from passive surveillance to active robotic interaction.

These analysis methods enable precise quantitative measurements of a drone’s operational state. The field of event-based tachometry focuses on estimating propeller rotational speed with high fidelity. Systems like EV-Tach have demonstrated performance comparable to commercial laser tachometers, with relative errors as low as 0.03%, while functioning in difficult lighting conditions [68, 72]. Other methods, like EB-ASM, also achieve high accuracy and have demonstrated the crucial capability of measuring the speeds of multiple propellers simultaneously with a single camera, which is essential for monitoring multi-rotor drones [2]. Beyond speed, the analysis can be extended to full state estimation. For instance, by fitting an ellipse to the aggregated event cloud generated by a propeller, it is possible to derive the drone’s pitch and roll angles, providing valuable attitude information [56]. The evolution of these applications shows a clear progression from simple detection to sophisticated tachometry and state estimation, establishing propeller signature analysis as a key technology for both remote surveillance and advanced autonomous robotics.

In the work by Stewart et al [57, 58], the authors present a virtual fence system for drone detection based on event cameras. The approach leverages the unique rotational signature of drone propellers, captured as periodic patterns in the event stream. By computing temporal histograms and applying spectral analysis, the system detects drones by identifying characteristic frequency peaks associated with

Dataset	Resolution	RGB / Event	Duration	Drone-Centric	Drone Types	Tasks
Anti-UAV [29]	1920 × 1080	✓ / ×	4h:25m	✓	≥ 6	Det, Track
Pawelczyk et al. [46]	640 × 480	✓ / ×	51446 frames	✓	-	Det
EED [42]	240 × 180	× / ✓	≤ 0h:30m	×	-	Det, Track
VisEvent [61]	346 × 260	✓ / ✓	≤ 5h	×	-	Det
EventVOT [62]	1280 × 720	× / ✓	≤ 5h	×	-	Det
F-UAV-D [39]	1280 × 720	✓ / ✓	0h:30m	✓	2	Det
Ev-UAV [9]	346 × 240	× / ✓	0h:15m	✓	-	Det, Track
Ev-Flying [36]	1280 × 720	× / ✓	1h:07m	×	1	Det, Track
NeRDD [35]	1280 × 720	✓ / ✓	3h:30m	✓	2	Det
CRSOT [76]	1280 × 720	✓ / ✓	2h:50m	✓	-	Det, Track
FRED [38]	1280 × 720	✓ / ✓	7h:07m	✓	5	Det, Track, Forecast

Table 1. **Comparison of event-based drone datasets.** RGB/Event: availability of modalities. Drone-centric: whether the drone is the primary subject. Task: which tasks the dataset is aimed at (Detection, Tracking, Forecasting)

propeller motion. A substantial improvement was then presented [59], in which the propeller analysis is used directly for the drone spatial detection instead of only presence detection. These methods are purely algorithmic and need no neural network or training data.

6. Datasets

Before delving deep into the existing datasets and their characteristics, we need to make a further distinction. Since event cameras alone provide for many advantages not present in other sensors, many works in the literature focus on data gathered by this sensor only. On the other side, the usage of multiple sensors concurrently presents the advantage of better reliability, especially in the presence of challenging scenarios and a more complete coverage of the scene by using complementary domains (e.g., the RGB domain can retain color information that may be crucial for distinction between drones and other kinds of flying objects). Tab. 1 provides an overview of event-based drone sensing datasets.

Event Only Datasets Many methods rely solely on event data due to its rich temporal and spatial content. Single-modality data is easier to collect, requiring only labeling and no sensor synchronization. We can make a further distinction: drone-centered datasets and drone-containing datasets. This difference pertains to the simple inclusion of drone instances in a broader dataset (drone-containing) versus datasets that focus on drone detection as their primary objective (drone-centered). A collection of 10,000 event frames, among the first drone-centered event based dataset, was first presented by Iaboni et al. [27]. Here, the dataset focuses on multi-quadrotor localization and tracking. The system uses a ceiling-mounted event camera to monitor quadrotors flying indoors and outdoors. The dataset cap-

tures multiple quadrotors (up to 6) in motion across various lighting conditions, flight speeds, and altitudes. In [9], the authors present Ev-UAV, a dataset comprising 147 sequences and over 2.3 million event-level annotations, featuring extremely small targets. The dataset captures diverse scenarios, including urban clutter and extreme lighting conditions. An alternative approach is presented in Da et al. [12]: here the authors present SFERA, a novel stereo fisheye event camera system developed for fast, omnidirectional drone detection and tracking. Two synchronized fisheye Prophesee cameras provide 360° spherical coverage. The dataset includes over 35 minutes of annotated recordings of drones in outdoor environments under various flight trajectories. Annotations provide per-frame drone positions projected in spherical coordinates. Other works focus on generic object detection, with a broader focus on large quantities of classes, but also including recordings of drones [42, 62, 63].

Multi-modal Datasets Beyond the event camera-only datasets, another subset of datasets comprises complementary domains to be used with the neuromorphic data. These kinds of approaches have a number of advantages as well as some practical disadvantages. On one hand, they provide for more resilient data to domain-specific limitations: one common event camera problem is in fact related to scenarios in which objects may remain stationary for a certain period of time, leading to a total loss of information about said objects. In this scenario, a complementary RGB camera can improve the chance of retaining information [60, 64]. RGB data also helps to distinguish drones from birds, insects, and other flying agents. Some of the major downsides of these approaches are related to the difficulty in setting up the recording station when multiple sources are present. A general-purpose dataset, comprising also annotation of

drones, is presented in Wang et al. [61], in which the authors gather 820 video pairs (RGB + event) collected across real-world scenarios featuring fast motion, low illumination, and cluttered backgrounds. The dataset focuses on tracking and adopts a low-resolution event camera (346×240). Similarly, Zhu et al. [76] tackle this approach with a more diverse data scenario, with the main goal of tracking different objects at high resolution. In Mandula et al. [39], the authors introduce F-UAV-D, a dedicated RGB-neuromorphic dataset for event-based UAV detection. The dataset was captured using the Prophesee EVK4 (1280×720) for the event stream and Sony’s IMX219 for the RGB frames. The dataset comprises approximately 30 minutes of event-frame videos featuring fast-moving drones in both controlled indoor and outdoor scenarios. In Magrini et al. [35], the authors introduce NeRDD (Neuromorphic-RGB Drone Detection), a multimodal benchmark dataset with spatio-temporally synchronized event (DVS) and RGB data, with over 3.5 hours of annotated drone footage collected from stationary ground cameras. Another similar dataset, FRED [38], is composed of a total of more than 7 hours of annotations for drone detection, tracking and forecasting, which includes high-resolution recordings of five drone models and challenging settings such as rain, low light, insect distractors, and indoor/outdoor scenarios.

Simulators and Synthetic Data Since data gathering is a complex and time-consuming process, especially regarding the labeling phase, another frequently common method of creating drone-related datasets is through the simulation of drone-populated environments or the generation of synthetic data. Labeling is potentially free, and scene parameters (e.g., lighting, weather, motion) are fully controllable.

Simulators are also frequently paired with event simulation networks, which create synthetic events from temporally upsampled RGB or grayscale images [21, 49]. This can be applied to both On-Drone [3] and Off-Drone approaches. The main disadvantage of this approach is the limited environment realism, with risks of domain shift when compared to real recordings and scenarios. Simulations may be slow and computationally heavy, with limited lighting and temporal realism, both critical for event data.

Another direction for synthetic data generation is the usage of 3D environments with embedded physical properties, which can result in more realistic data and light simulation. Specialized tools for drone flight simulation have also been developed: Shah et al. [53] proposed an event-based camera simulation module that takes the output from a regular simulated camera, applies a non-linear operation, subtracts successive frames, and then thresholds the outcome. For accurately capturing a rotating propeller’s spectral signature, this approach would demand an exceptionally high frame rate from the simulated camera. In Eldeborg et al. [17], the

authors propose a mix of a small portion of real neuromorphic data combined with a substantial amount of synthetic data, simulated using Blender: in this case, the goal is to simulate the rotating propellers of a UAV.

7. Conclusions

In this paper, we addressed the critical challenge of drone detection, where traditional cameras are often hindered by motion blur and limited dynamic range. Event cameras offer a robust alternative, using their high temporal resolution, high dynamic range, and sparse data output to effectively capture small, fast-moving drones in challenging lighting conditions. The field is advancing rapidly, with methods evolving from converting event data into frames for use in conventional neural networks to processing it in its raw point cloud form or with spiking neural networks. Research has also progressed beyond simple detection to tackle more complex tasks such as drone tracking, trajectory forecasting, and unique identification through the analysis of propeller blade signatures. Event-based vision addresses the limits of conventional sensors, enabling efficient, low-latency counter-UAV solutions.

Acknowledgements

This paper was partially funded by Leonardo S.p.A. and by the project ”Collaborative Explainable neuro-symbolic AI for Decision Support Assistant”, CAI4DSA, CUP B13C23005640006. This work was partially supported by the Piano per lo Sviluppo della Ricerca (PSR 2023) of the University of Siena, project FEATHER: Forecasting and Estimation of Actions and Trajectories for Human–robot interactions.

References

- [1] Petar Andrašić, Tomislav Radišić, Mario Muštra, and Jurica Ivošević. Night-time detection of uavs using thermal infrared camera. *Transportation research procedia*, 28:183–190, 2017. 2
- [2] George O de A Azevedo, Leandro H de S Silva, Agostinho Freire, Rogério P De Araújo, and Bruno JT Fernandes. Multiple simultaneous rotation event-based angular speed measurement. *IEEE Sensors Journal*, 2025. 6
- [3] Anish Bhattacharya, Marco Cannici, Nishanth Rao, Yuezhan Tao, Vijay Kumar, Nikolai Matni, and Davide Scaramuzza. Monocular event-based vision for obstacle avoidance with a quadrotor. In *Conference on Robot Learning*, pages 4826–4843. PMLR, 2025. 8
- [4] Gabriel Carlisle Birch and Bryana Lynn Woo. Counter unmanned aerial systems testing: Evaluation of vis swir mwir and lwir passive imagers. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2017. 2
- [5] Hao Cao, Jingao Xu, Danyang Li, Zheng Yang, and Yunhao Liu. Eventboost: Event-based acceleration platform

- for real-time drone localization and tracking. In *IEEE IN-FOCOM 2024-IEEE Conference on Computer Communications*, pages 1851–1859. IEEE, 2024. 2
- [6] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 3
- [7] Adrian Carrio, Sai Vemprala, Andres Ripoll, Srikanth Saripalli, and Pascual Campoy. Drone detection using depth maps. In *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 1034–1037. IEEE, 2018. 2
- [8] Yujeong Chae, Lin Wang, and Kuk-Jin Yoon. Siamevent: Event-based object tracking via edge-aware similarity learning with siamese networks. *arXiv preprint arXiv:2109.13456*, 2021. 5
- [9] Nuo Chen, Chao Xiao, Yimian Dai, Shiman He, Miao Li, and Wei An. Event-based tiny object detection: A benchmark dataset and baseline. *arXiv preprint arXiv:2506.23575*, 2025. 3, 4, 7
- [10] Frank Christnacher, Sébastien Hengy, Martin Laurenzis, Alexis Matwyschuk, Pierre Naz, Stéphane Schertzer, and Gwenael Schmitt. Optical and acoustical uav detection. In *Electro-Optical Remote Sensing X*, pages 83–95. SPIE, 2016. 2
- [11] Angelo Coluccia, Alessio Fascista, Arne Schumann, Lars Sommer, Anastasios Dimou, Dimitrios Zarpalas, Miguel Méndez, David De la Iglesia, Iago González, Jean-Philippe Mercier, et al. Drone vs. bird detection: Deep learning algorithms and results from a grand challenge. *Sensors*, 21(8): 2824, 2021. 2, 3
- [12] Daniel Rodrigues Da Costa, Maxime Robic, Pascal Vasseur, and Fabio Morbidi. A new stereo fisheye event camera for fast drone detection and tracking. In *IEEE International Conference on Robotics and Automation*, 2025. 7
- [13] Fardad Dadboud, Hamid Azad, Varun Mehta, Miodrag Bolic, and Iraj Mantegh. Drift: Autonomous drone dataset with integrated real and synthetic data, flexible views, and transformed domains. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6900–6910. IEEE, 2025. 2
- [14] Mike Davies, Narayan Srinivasa, Tsung-Han Lin, Gautham Chinya, Yongqiang Cao, Sri Harsha Choday, Georgios Dimou, Prasad Joshi, Nabil Imam, Shweta Jain, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *Ieee Micro*, 38(1):82–99, 2018. 5
- [15] Maarten Uijt De Haag, Chris G Bartone, and Michael S Braasch. Flight-test evaluation of small form-factor lidar and radar sensors for suas detect-and-avoid applications. In *2016 IEEE/AIAA 35th digital avionics systems conference (DASC)*, pages 1–11. IEEE, 2016. 2
- [16] Jędrzej Drozdowicz, Maciej Wielgo, Piotr Samczynski, Krzysztof Kulpa, Jarosław Krzonkalla, Maj Mordzonek, Marcin Bryl, and Zbigniew Jakielaszek. 35 ghz fmcw drone detection system. In *2016 17th International Radar Symposium (IRS)*, pages 1–4. IEEE, 2016. 2, 6
- [17] Anton Eldeborg Lundin, Rasmus Winzell, Hanna Hamrell, David Gustafsson, and Hannes Övrén. Drone detection using a low-power neuromorphic virtual tripwire. In *European Conference on Computer Vision*, pages 105–120. Springer, 2024. 3, 5, 6, 8
- [18] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, pages 226–231, 1996. 3
- [19] 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. 2
- [20] Sai Ram Ganti and Yoohwan Kim. Implementation of detection and tracking mechanism for small uas. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 1254–1260. IEEE, 2016. 2
- [21] Daniel Gehrig, Mathias Gehrig, Javier Hidalgo-Carrió, and Davide Scaramuzza. Video to events: Recycling video datasets for event cameras. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2020. 8
- [22] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 5
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4
- [24] S. Hoseini, G. Orchard, A. Yousefzadeh, B. Deverakonda, T. Serrano-Gotarredona, and B. Linares-Barranco. Passive localization and detection of quadcopter uavs by using dynamic vision sensor. In *2017 5th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, pages 81–85, 2017. 6
- [25] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 4
- [26] Shuowen Hu, Geoffrey H Goldman, and Christoph C Borel-Donohue. Detection of unmanned aerial vehicles using a visible camera system. *Applied optics*, 56(3):B214–B221, 2017. 2
- [27] Craig Iaboni, Deepan Lobo, Ji-Won Choi, and Pramod Abichandani. Event-based motion capture system for on-line multi-quadrotor localization and tracking. *Sensors*, 22(9):3240, 2022. 2, 5, 7
- [28] Simone Undri Innocenti, Federico Becattini, Federico Pernici, and Alberto Del Bimbo. Temporal binary representation for event-based action recognition. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 10426–10432. IEEE, 2021. 4
- [29] Nan Jiang, Kuiran Wang, Xiaoke Peng, Xuehui Yu, Qiang Wang, Junliang Xing, Guorong Li, Qixiang Ye, Jianbin Jiao, Zhenjun Han, et al. Anti-uav: A large-scale benchmark for vision-based uav tracking. *T-MM*, 2021. 1, 7
- [30] Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024. 5

- [31] Paul Kirkland, Gaetano Di Caterina, John Soraghan, Yiannis Andreopoulos, and George Matich. Uav detection: a stdp trained deep convolutional spiking neural network retinaneuromorphic approach. In *International Conference on Artificial Neural Networks*, pages 724–736. Springer, 2019. 5
- [32] Hongmin Li and Luping Shi. Robust event-based object tracking combining correlation filter and cnn representation. *Frontiers in neurorobotics*, 13:82, 2019. 5
- [33] Hanfang Liang, Shenghai Yuan, Fen Liu, Yizhuo Yang, Bing Wang, Zhuyu Huang, Chenyang Shi, and Jing Jin. Label-free long-horizon 3d uav trajectory prediction via motion-aligned rgb and event cues. *arXiv preprint arXiv:2507.03365*, 2025. 6
- [34] Min Liu and Tobi Delbruck. Edflow: Event driven optical flow camera with keypoint detection and adaptive block matching. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9):5776–5789, 2022. 2, 4
- [35] Gabriele Magrini, Federico Becattini, Pietro Pala, Alberto Del Bimbo, and Antonio Porta. Neuromorphic drone detection: an event-rgb multimodal approach. In *European Conference on Computer Vision*, pages 259–275. Springer, 2024. 2, 3, 5, 7, 8
- [36] Gabriele Magrini, Federico Becattini, Giovanni Colombo, and Pietro Pala. Ev-flying: an event-based dataset for in-the-wild recognition of flying objects. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4947–4955, 2025. 3, 4, 7
- [37] Gabriele Magrini, Federico Becattini, Luca Cultrera, Lorenzo Berlincioni, Pietro Pala, and Alberto Del Bimbo. Spike-tbr: a noise resilient neuromorphic event representation. *arXiv preprint arXiv:2506.04817*, 2025. 4
- [38] Gabriele Magrini, Niccolò Marini, Federico Becattini, Lorenzo Berlincioni, Niccolò Biondi, Pietro Pala, and Alberto Del Bimbo. Fred: The florence rgb-event drone dataset. In *Proceedings of the 33rd ACM International conference on multimedia*, 2025. 2, 3, 4, 5, 6, 7, 8
- [39] Jakub Mandula, Jonas Kühne, Luca Pascarella, and Michele Magno. Towards real-time fast unmanned aerial vehicle detection using dynamic vision sensors. In *Proceedings of the 2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6, 2024. 3, 7, 8
- [40] József Mezei, Viktor Fiaska, and András Molnár. Drone sound detection. In *2015 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 333–338. IEEE, 2015. 2
- [41] Sahaj K Mistry, Shreyas Chatterjee, Ajeet K Verma, Vinit Jakhetiya, Badri N Subudhi, and Sunil Jaiswal. Drone-vs-bird: Drone detection using yolov7 with csrt tracker. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–2. IEEE, 2023. 2
- [42] Anton Mitrokhin, Cornelia Fermüller, Chethan Parameshwara, and Yiannis Aloimonos. Event-based moving object detection and tracking. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–9. IEEE, 2018. 2, 5, 7
- [43] Marco Monforte, Luna Gava, Massimiliano Iacono, Arren Glover, and Chiara Bartolozzi. Fast trajectory end-point prediction with event cameras for reactive robot control. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4036–4044, 2023. 6
- [44] Thomas Müller. Robust drone detection for day/night counter-uav with static vis and swir cameras. In *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VIII*, pages 302–313. SPIE, 2017. 2
- [45] Phuc Nguyen, Mahesh Ravindranatha, Anh Nguyen, Richard Han, and Tam Vu. Investigating cost-effective rf-based detection of drones. In *Proceedings of the 2nd workshop on micro aerial vehicle networks, systems, and applications for civilian use*, pages 17–22, 2016. 2
- [46] Maciej Pawełczyk and Marek Wojtyra. Real world object detection dataset for quadcopter unmanned aerial vehicle detection. *IEEE Access*, 8:174394–174409, 2020. 1, 7
- [47] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 4
- [48] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 4
- [49] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In *Conference on robot learning*, pages 969–982. PMLR, 2018. 8
- [50] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 3
- [51] Artem Rozantsev, Vincent Lepetit, and Pascal Fua. Flying objects detection from a single moving camera. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4128–4136, 2015. 2
- [52] Nitin J Sanket, Chahat Deep Singh, Chethan M Parameshwara, Cornelia Fermüller, Guido CHE de Croon, and Yiannis Aloimonos. Evpropnet: Detecting drones by finding propellers for mid-air landing and following. *arXiv preprint arXiv:2106.15045*, 2021. 6
- [53] 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, 2017. 8
- [54] Amos Sironi, Manuele Brambilla, Nicolas Bourdis, Xavier Lagorce, and Ryad Benosman. Hats: Histograms of averaged time surfaces for robust event-based object classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1731–1740, 2018. 4
- [55] Rikard Fridsén Skogsberg, Jens Grundmark, and Matts Björck. Event-based real-time detection and tracking of uavs. In *Emerging Imaging and Sensing Technologies for Security and Defence IX*, pages 30–40. SPIE, 2024. 5
- [56] Radim Spetlik, Tereza Uhrová, and Jiří Matas. Efficient real-time quadcopter propeller detection and attribute estimation with high-resolution event camera. In *Scandinavian Conference on Image Analysis*, pages 217–230. Springer, 2025. 6

- [57] Terrence Stewart, Marc-Antoine Drouin, Guillaume Gagne, and Guy Godin. Drone virtual fence using a neuromorphic camera. In *International Conference on Neuromorphic Systems 2021*, pages 1–9, 2021. 6
- [58] Terrence Stewart, Marc-Antoine Drouin, Michel Picard, Frank Billy Djupkep Dizeu, Anthony Orth, and Guillaume Gagné. A virtual fence for drones: Efficiently detecting propeller blades with a dvxplorer event camera. In *Proceedings of the International Conference on Neuromorphic Systems 2022*, pages 1–7, 2022. 6
- [59] Terrence Stewart, Marc-Antoine Drouin, Michel Picard, Frank Billy Djupkep, Anthony Orth, and Guillaume Gagné. Using neuromorphic cameras to track quadcopters. In *Proceedings of the 2023 International Conference on Neuromorphic Systems*, pages 1–5, 2023. 7
- [60] Abhishek Tomy, Anshul Paigwar, Khushdeep S Mann, Alessandro Renzaglia, and Christian Laugier. Fusing event-based and rgb camera for robust object detection in adverse conditions. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 933–939. IEEE, 2022. 7
- [61] Xiao Wang, Jianing Li, Lin Zhu, Zhipeng Zhang, Zhe Chen, Xin Li, Yaowei Wang, Yonghong Tian, and Feng Wu. Visevent: Reliable object tracking via collaboration of frame and event flows. *IEEE Transactions on Cybernetics*, 2023. 7, 8
- [62] Xiao Wang, Shiao Wang, Chuanming Tang, Lin Zhu, Bo Jiang, Yonghong Tian, and Jin Tang. Event stream-based visual object tracking: A high-resolution benchmark dataset and a novel baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19248–19257, 2024. 4, 5, 7
- [63] Xiao Wang, Yu Jin, Wentao Wu, Wei Zhang, Lin Zhu, Bo Jiang, and Yonghong Tian. Object detection using event camera: A moe heat conduction based detector and a new benchmark dataset. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 29321–29330, 2025. 7
- [64] Zihao W Wang, Peiqi Duan, Oliver Cossairt, Aggelos Katsaggelos, Tiejun Huang, and Boxin Shi. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1609–1619, 2020. 7
- [65] Yuanwei Wu, Yao Sui, and Guanghui Wang. Vision-based real-time aerial object localization and tracking for uav sensing system. *IEEE Access*, 5:23969–23978, 2017. 2
- [66] Man Yao, Ole Richter, Guangshe Zhao, Ning Qiao, Yunnan Xing, Dingheng Wang, Tianxiang Hu, Wei Fang, Tugba Demirci, Michele De Marchi, Lei Deng, Tianyi Yan, Carsten Nielsen, Sadique Sheik, Chenxi Wu, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-based dynamic computing with asynchronous sensing-computing neuromorphic chip. *Nature Communications*, 15(1), 2024. 5
- [67] Jawad Yousaf, Huma Zia, Marah Alhalabi, Maha Yaghi, Tasnim Basmaji, Eiman Al Shehhi, Abdalla Gad, Mohammad Alkhedher, and Mohammed Ghazal. Drone and controller detection and localization: Trends and challenges. *Applied Sciences*, 12(24):12612, 2022. 1
- [68] Heng Zhang, Nuo Chen, Miao Li, and Wei An. Spiking swin transformer for uav object detection based on event cameras. In *2024 12th International Conference on Information Systems and Computing Technology (ISCTech)*, pages 1–6. IEEE, 2024. 5, 6
- [69] Jiqing Zhang, Bo Dong, Haiwei Zhang, Jianchuan Ding, Felix Heide, Baocai Yin, and Xin Yang. Spiking transformers for event-based single object tracking. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 8801–8810, 2022. 5
- [70] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. In *European conference on computer vision*, pages 1–21. Springer, 2022. 5
- [71] Guangrong Zhao, Yiran Shen, Ning Chen, Pengfei Hu, Lei Liu, and Hongkai Wen. High speed rotation estimation with dynamic vision sensors. *arXiv preprint arXiv:2209.02205*, 2022. 6
- [72] Guangrong Zhao, Yiran Shen, Ning Chen, Pengfei Hu, Lei Liu, and Hongkai Wen. Ev-tach: A handheld rotational speed estimation system with event camera. *IEEE Transactions on Mobile Computing*, 23(6):7483–7498, 2023. 6
- [73] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. Detsr beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16965–16974, 2024. 5
- [74] Alex Zihao Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multi-vehicle stereo event camera dataset: An event camera dataset for 3d perception. *IEEE Robotics and Automation Letters*, 3(3):2032–2039, 2018. 4
- [75] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Unsupervised event-based learning of optical flow, depth, and egomotion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 989–997, 2019. 4
- [76] Yabin Zhu, Xiao Wang, Chenglong Li, Bo Jiang, Lin Zhu, Zhixiang Huang, Yonghong Tian, and Jin Tang. Crsot: Cross-resolution object tracking using unaligned frame and event cameras. *arXiv preprint arXiv:2401.02826*, 2024. 7, 8
- [77] Tamás Zsedrovits, Ákos Zarándy, Bálint Vanek, Tamás Péni, József Bokor, and Tamás Roska. Collision avoidance for uav using visual detection. In *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, pages 2173–2176. IEEE, 2011. 2
- [78] Antoine Zundel, Cédric Demonceaux, Nicolas Hueber, Damien Spittler, Guillaume Strub, and Sébastien Changey. Bimodal vision system combining standard camera and dynamic vision sensor for detecting and tracking fast uavs. In *Emerging Imaging and Sensing Technologies for Security and Defence IX*, page 1320402. SPIE, 2024. 3