Event Camera Meets Resource-Aware Mobile Computing: Abstraction, Algorithm, Acceleration, Application

HAOYANG WANG, RUISHAN GUO, PENGTAO MA, CIYU RUAN, XINYU LUO, WENHUA DING, TIANYANG ZHONG, Shenzhen International Graduate School, Tsinghua University, China
JINGAO XU, Carnegie Mellon University, USA
YUNHAO LIU, Global Innovation Exchange, Tsinghua University, China
XINLEI CHEN*, Shenzhen International Graduate School, Tsinghua University, China

With the increasing complexity of mobile device applications, these devices are evolving toward high agility. This shift imposes new demands on mobile sensing, particularly in achieving high-accuracy and low-latency. Event-based vision has emerged as a disruptive paradigm, offering high temporal resolution and low latency, making it well-suited for high-accuracy and low-latency sensing tasks on high-agility platforms. However, the presence of substantial noisy events, lack of stable, persistent semantic information, and large data volume pose challenges for event-based data processing on resource-constrained mobile devices. This paper surveys the literature from 2014 to 2025 and presents a comprehensive overview of event-based mobile sensing, encompassing its fundamental principles, event *abstraction* methods, *algorithm* advancements, and both hardware and software *acceleration* strategies. We discuss key *applications* of event cameras in mobile sensing, including visual odometry, object tracking, optical flow, and 3D reconstruction, while highlighting challenges associated with event data processing, sensor fusion, and real-time deployment. Furthermore, we outline future research directions, such as improving the event camera with advanced optics, leveraging neuromorphic computing for efficient processing, and integrating bio-inspired algorithms. To support ongoing research, we provide an open-source *Online Sheet* with recent developments. We hope this survey serves as a reference, facilitating the adoption of event-based vision across diverse applications.

Additional Key Words and Phrases: Mobile Sensing; Event Camera; Event-based Vision

1 INTRODUCTION

Mobile sensing. With ongoing advancements in sensor technology and the proliferation of sophisticated computing capabilities within embedded systems, mobile devices (e.g., drones and autonomous vehicles) have emerged as the most groundbreaking innovations in recent years [1–5]. As illustrated in Fig.1, these devices are increasingly deployed in a variety of novel applications, including last-mile delivery [5], industrial inspection [6–8], rapid relief-and-rescue [9], aerial imaging [10] and sky networking [11], particularly within smart city scenarios. To perform these tasks, which require extensive interaction with the external environment, mobile devices must possess the ability to: (*i*) *awareness their own state*, including location and orientation [12–14], (*ii*) *comprehend their surroundings*, (e.g., environmental structure and map) [15], and (*iii*) *understand their relationship with the environment*, such as the spatio-temporal relationships between mobile devices and objects within their surroundings [16]. Achieving these capabilities has become a focal point of interest within the mobile computing community.

High agility trend of mobile devices. In context of smart city scenarios, the demand for mobile devices is progressively increasing, with their mission profiles evolving toward execution of 4D tasks, which are characterized as deep, dull, dangerous, and dirty [17]. The expanding scale of cities also necessitates that mobile devices complete

*Xinlei Chen is the corresponding author

Authors' addresses: Haoyang Wang, Ruishan Guo, Pengtao Ma, Ciyu Ruan, Xinyu Luo, Wenhua Ding, Tianyang Zhong, Shenzhen International Graduate School, Tsinghua University, Shenzhen, China; Jingao Xu, Carnegie Mellon University, Pittsburgh, USA; Yunhao Liu, Global Innovation Exchange, Tsinghua University, Beijing, China; Xinlei Chen, chen.xinlei@sz.tsinghua.edu.cn, Shenzhen International Graduate School, Tsinghua University, Shenzhen, China.

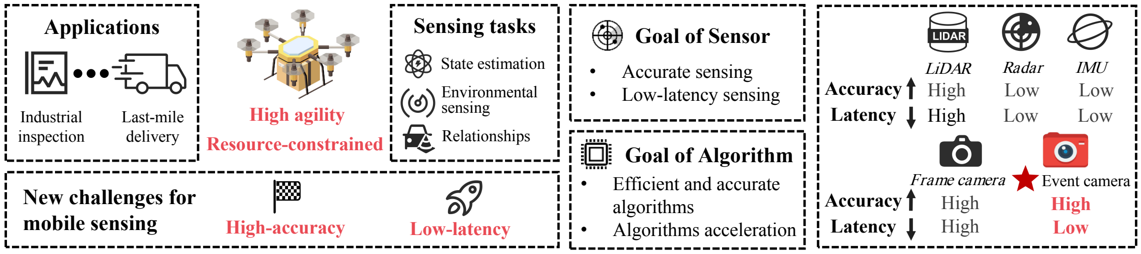


Fig. 1. Mobile devices are used in various applications, with key tasks including state estimation, environment perception, and understanding device-environment interactions. As devices become more agile, mobile sensing faces higher demands for accuracy and latency. This requires tight coordination between sensors and algorithms: *(i)* sensors must capture high-precision data with minimal delay; *(ii)* algorithms must efficiently process data within resource constraints. Traditional sensors fall short of these needs, whereas event cameras—capable of asynchronously capturing pixel-level intensity changes with microsecond latency—offer transformative potential. This survey presents a comprehensive review of event cameras and the development of efficient, accurate algorithms.

various tasks within shorter time frames, driving their evolution toward high-speed operation [1, 18]. Consequently, the development of mobile devices is exhibiting new trends toward high agility. For instance, DJI's industrial inspection drones can cruise at speeds of 21 m/s [19], while Wing's delivery drones fly at 30 m/s to deliver packages [20].

New challenges for mobile sensing. As mobile devices evolve toward high agility design, mobile sensing is required to advance toward *high accuracy* and *low latency*, enabling these mobile devices to perceive their state and surroundings in *millimeter*-level accuracy with *millisecond*-level latency, thereby facilitating faster responses and more precise adjustments. This evolution establishes new objectives for the sensors and data processing algorithms involved in mobile sensing tasks: (*i*) *On the sensor input front*, it is essential to acquire higher accuracy raw data with lower latency. (*ii*) *On data processing algorithms front*, efficient processing of sensor measurements is essential to enhance accuracy in mobile sensing tasks while optimizing performance on resource-constrained platforms.

Existing sensors for mobile sensing. However, existing sensors increasingly struggle to meet the high accuracy and low latency demands of mobile sensing, especially for environmental perception and interaction in high-agility devices. *(i) Radar-based solutions* employ sensors such as LiDAR [1, 21], mmWave radar [22], and ultrasound radar, which emit signals and estimate distances based on reflections. These methods track distance changes to update device positions and infer spatial relationships. However, they suffer from either high latency or limited accuracy: LiDAR achieves millimeter-level accuracy but requires point accumulation into frames at low frequencies (e.g., 10 Hz), introducing delays up to 100 ms; mmWave radar offers millisecond latency but lacks sufficient spatial resolution for millimeter accuracy. *(ii) Camera-based methods* use monocular [23, 24] and stereo cameras [25] for self-localization and environment mapping via SLAM [26]. Yet, these approaches are computationally demanding and limited by low temporal resolution (<30 Hz), relatively high latency (>30 ms), and standard dynamic range (60 dB), making them inadequate for high-agility mobile platforms.

New sensor: Event camera. The event cameras are novel bio-inspired sensors that outputs pixel-wise intensity changes in an asynchronous manner [27, 32]. Unlike frame cameras, it generates output based on scene dynamics rather than a global clock that is independent of the scene [33]. The event cameras offer four key advantages that align well with the requirements of mobile sensing tasks for high-agility mobile devices: (*i*) The *μs-level temporal resolution* refers to the time interval between two consecutive samples. A higher temporal resolution implies a smaller interval, enabling event cameras to capture high-speed motions without motion blur and thereby supporting accurate perception during fast operations [34]. (*ii*) The *μs-level sensing latency* denotes the time required for the sensor to respond to a change in illumination by producing an output. A lower latency allows environmental changes to be reported to mobile

Table 1. Summary of topics covered in various studies

Topic	Design of event camera	Adv. of event camera	Gen. model of event	HW design	Products	Datasets	Repr.	Denoising	Filtering and feature ext.	Matching	Mapping	Accel.	App.	Adv. in mobile comp.	Challenge in mobile comp.
[27] (2020)	✓	✓	✓	✓	✓		✓		✓	✓	✓				
[28] (2023)		✓					✓		✓	✓	✓		✓		
[29] (2024)		✓	✓		✓	✓							✓		
[30] (2024)		✓				✓	✓		✓	✓	✓				
[31] (2024)		✓	✓	✓	✓		✓	✓					✓		
This survey	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

***Note**: Abbreviations used: Avg. (Advantage), Gen. model (Generation model), HW (Hardware), Repr. (Representations), feature ext. (Feature extraction), Accel. (Acceleration), App. (Application), mobile comp. (Mobile computing).*

platforms almost instantaneously [35]. (*iii*) The *high dynamic range (HDR)*, which is 140 *dB* compared to 60 *dB* for standard cameras, making it effective in diverse lighting conditions [33, 36]. (*iv*) The *low power consumption* (e.g., 0.5 *W*) makes it particularly suitable for efficient designed mobile devices [37]. These advantages position event cameras as a promising technology to empower mobile devices designed for high-speed operation and efficiency.

Event-based data processing algorithms. Event cameras offer high-accuracy, low-latency data acquisition but face three main processing challenges: *(i)* Event cameras' sensitivity to illumination causing significant noise, *(ii)* Event data lacks stable, persistent texture and cannot provide consistent information, as events are generated only at image edges depending on motion and scene texture, complicating feature extraction and long-term data association. *(iii)* Large data volume leading to high computational load on mobile devices. Efficient and accurate event data processing is thus crucial for resource-limited, agile mobile platforms to perceive their state and environment quickly. This survey reviews event processing algorithms across six key stages: event representation, denoising, filtering and feature extraction, matching, mapping, and hardware/software acceleration.

Difference between existing surveys. As shown in Fig.2, this survey extends previous surveys by focusing on *how event cameras enable high-agility, resource-constrained mobile devices to achieve high-accuracy and low-latency self-state estimation and environmental understanding* and extending prior surveys by incorporating literature from 2023–2025. It outlines the processing workflow of event data and reviews the advancements at each stage of this workflow. Using the key metrics of accuracy and efficiency in mobile computing, we summarize various methods in each stage to provide a deeper analysis of cutting-edge research, and assess each stage under latency, accuracy, and power constraints. It also covers work on event-based hardware and software acceleration, offering insights for deploying event cameras on resource-constrained mobile devices. Finally, we discuss applications of event cameras on mobile platforms.

Contribution. The main contribution of the survey paper is summarized as follows.

(1) We extend prior surveys by presenting a comprehensive review of how event cameras empower high-agility, resource-constrained mobile devices to achieve accurate, low-latency self-state estimation and environmental sensing.

(2) We present an in-depth introduction to event generation models, the design of event camera hardware, existing commercial products, and benchmark datasets for event-based vision.

(3) We highlight unique advantages of event cameras in mobile sensing, as well as the specific challenges they encounter.

(4) We categorize event stream processing methods into distinct stages and present a comprehensive review of each, covering event stream representation, data processing algorithms, acceleration strategies, and applications on event-based mobile platforms. Furthermore, we evaluate these stages with respect to latency, accuracy, and power constraints.

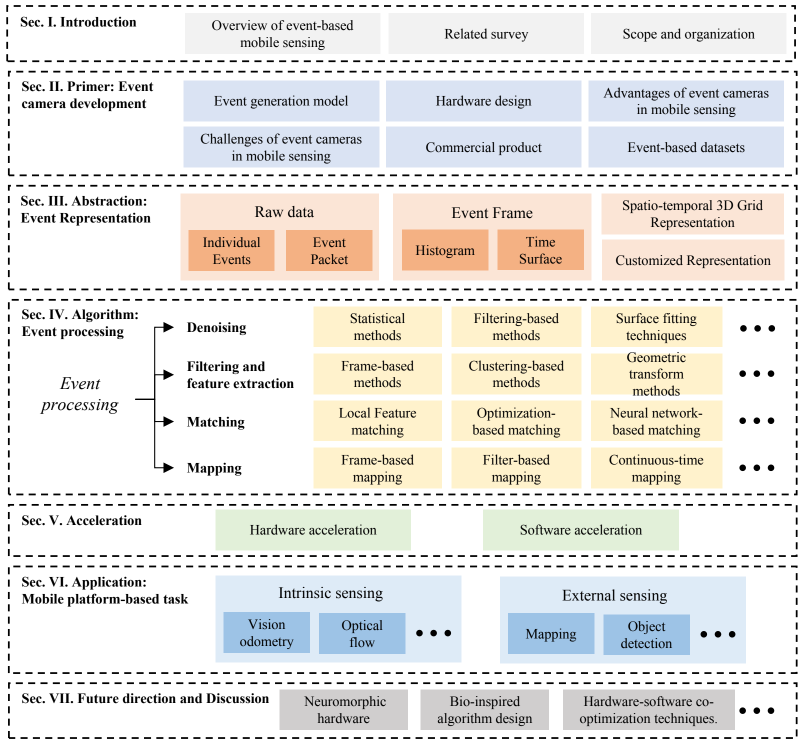


Fig. 2. Structure of this survey.

(5) We present our insights and solutions for future trends, with a particular focus on bio-inspired event camera hardware design, algorithm development, and hardware-software co-optimization techniques.

Online resource. This survey presents a comprehensive review of event-based sensing systems, focusing on key technological advancements and practical applications. To further support the research community, we have established an open-source *Online Sheet* [1], which is adapted from [38]. This online sheet will be regularly updated, ensuring access to the latest developments and fostering continued innovation in event-based sensing systems.

Organization. Fig.2 illustrates the survey structure. Sec. 2 introduces event camera fundamentals, including principles, hardware, products, datasets, and their pros and cons on mobile devices. Sec. 3 covers event stream **abstraction** methods, while Sec. 4 reviews event processing **algorithms** such as denoising, filtering, matching, and mapping. Sec. 5 discusses hardware/software **acceleration** for resource-constrained devices, and Sec. 6 highlights mobile **applications** of event cameras. Sec. 7 outlines future directions, with conclusions in Sec. 8.

2 PRIMER: EVENT CAMERA DEVELOPMENT

2.1 Event generation model

Unlike conventional cameras that capture images at fixed intervals, event cameras operate asynchronously by detecting changes in log intensity at individual pixels, generating events only when significant changes occur. This characteristic enables event cameras to achieve exceptionally high temporal resolution and effectively mitigate motion blur, particularly in scenarios involving fast-moving objects.

[1] Event-based mobile sensing resource

Table 2. Classification of Event Camera Types

Indicator	Original event camera[42]	DVS [40]	ATIS [41]	DAVIS [43]
Development origin	Mahowald & Mead, Caltech	Inspired by silicon retina	Evolved from DVS	Advanced DVS/ATIS
Output data type	Log brightness	Change in brightness	Change + absolute brightness	Change + absolute brightness
Sensor type	Large pixels, CMOS	CMOS	CMOS	CMOS
Pixel structure	Single-pixel design	Smaller, simpler pixels	Dual-subpixel	Shared pixel/subpixel
Brightness measurement	Continuous-time	Change only	Absolute + change	Absolute + change
Dynamic range	Limited	Narrow	High	Moderate
Event synchronization	Basic	Fast reset	Potential mismatch	Slow sampling
Noise filtering	Minimal	Simple	Complex	Advanced

Each event is defined by the pixel location where the change occurs, the timestamp, and the polarity. Formally, an event can be represented as $e_k = (x_k, t_k, p_k)$, where e_k denotes the event, x_k specifies the pixel location, t_k represents the timestamp, and p_k indicates the polarity of the change. The change in log intensity is given by $\Delta L(x_k, t_k) = L(x_k, t_k) - L(x_k, t_k - \Delta t_k)$, where $L(x_k, t_k)$ represents the log intensity at pixel x_k and time t_k. As shown in Fig.3, an event is triggered only when $|\Delta L(x_k, t_k)|$ exceeds a predefined threshold C. The polarity p_k is assigned as +1 if $\Delta L(x_k, t_k) > 0$, and -1 otherwise. This event-driven paradigm substantially reduces redundant data, enhancing processing efficiency, conserving computational resources, and enabling deployment in resource-constrained systems such as embedded devices [39]. In practice, threshold C can be adjusted to meet specific application needs and impacts event camera performance. A high C reduces sensitivity, may missing subtle changes, while a low C increases noise-triggered events, causing redundancy.

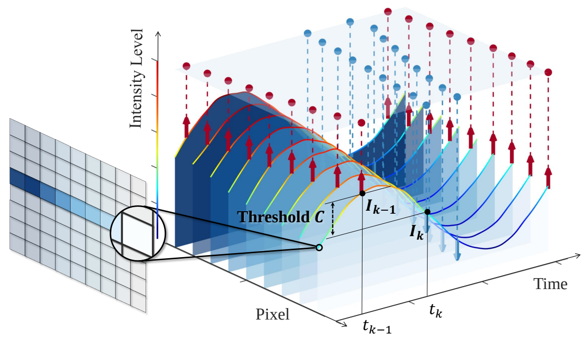


Fig. 3. Principle of the Event Cameras: Events are generated based on changes in logarithmic light intensity over time.

2.2 Hardware design

In this part, we will introduce hardware design of modern event cameras, as illustrated in Tab. 2.

General hardware architecture. As shown in Fig.4, event cameras employ CMOS sensors for low-latency operation, which involves three steps: (i) incident light generates electron–hole pairs, (ii) electrons are collected under an electric field, and (iii) readout circuits convert them into voltage signals for logarithmic brightness computation. The event circuit filters noise, applies a threshold, and triggers an event when exceeded, transmitting it to the processor.

DVS event camera [40]. The Dynamic Vision Sensor (DVS), inspired by the silicon retina, detects brightness changes via capacitance coupling and resets after each measurement. It outputs only changes, enabling smaller pixel sizes but restricting output to event data, which limits information extraction in static scenes.

ATIS event camera [41]. The Active Time-Image Sensor (ATIS) uses subpixels to measure absolute brightness, doubling pixel area compared to DVS but enabling wide dynamic and static ranges for robust imaging under extreme lighting. Its limitation lies in potential misalignment between absolute brightness (averaged across pixels) and event data (triggered per pixel), especially during high-speed motion.

DAVIS event camera [43]. The Dynamic and Active Vision Sensor (DAVIS) is capable of outputting both absolute brightness and event-based data. In DAVIS, pixels and subpixels share the same sensor, enabling a more compact design. As a result, the pixel area is smaller than that of the ATIS, with only a modest 5% increase in size compared to the DVS. However, the sampling speed of the DAVIS circuit is slower than that of the DVS circuit.

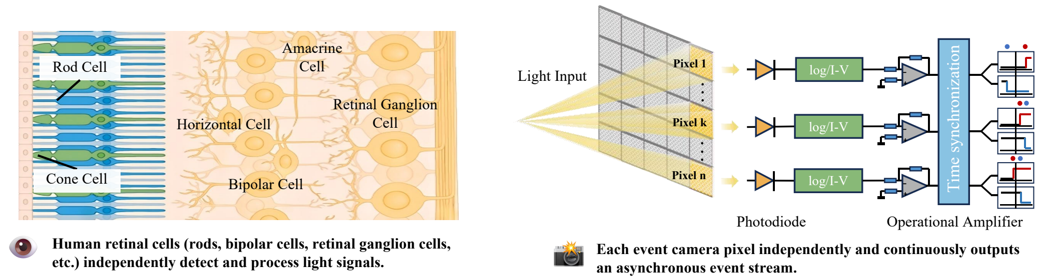


Fig. 4. Working Mechanism of Event Cameras: Inspired by the rod cells in the human eye, event camera operates at the pixel level, independently transforming light into voltage signals to capture intensity variations.

Beyond foundational architectures, modern event cameras are optimized for mobile platforms. Key trends include more versatile readouts (e.g., hardware-generated event-accumulation views in the CeleX-V [44]), mobile-friendly interfaces like MIPI CSI-2 (on the CeleX5-MIPI [45]), and ultra-low power optimization specifically targeting battery-powered mobile devices (e.g., Prophesee's GENX320-class sensors [46]). A detailed comparison of these commercial products is presented in Section 2.5.

2.3 Advantages of event cameras in mobile sensing

High temporal resolution. The high agility of mobile devices induces rapid environmental changes, often causing motion blur in frame cameras and limiting the responsiveness of radar- or camera-based solutions due to their low spatio-temporal resolution. Temporal resolution, defined as the interval between consecutive samples, improves with shorter intervals. Event cameras, with microsecond-level resolution and motion-blur-free sensing, enable timely detection of both environmental dynamics and device state.

Low sensing latency. The high agility of mobile devices demands rapid awareness of both environmental and self-state changes. Sensing latency, defined as the time required for a sensor to respond to environmental variations, is relatively long in frame cameras due to their global exposure time (20 ms), which delays reactions and increases collision risk. In contrast, event cameras employ independent pixels that trigger events immediately upon brightness changes, achieving sub-millisecond latency and enabling mobile devices to detect and respond to changes instantaneously.

High dynamic range (HDR). Mobile devices are increasingly used in challenging environments, such as low-light nighttime and bright daytime settings, requiring reliable sensing across varying lighting conditions. Standard frame cameras have a dynamic range of about 60 dB, making them less effective in extreme lighting. Event cameras operate on a logarithmic scale with independent pixels, offering a very high dynamic range with >120 dB. This enables them to adapt to both extremely dark and bright conditions, making mobile devices suitable for a wider range of scenarios.

Low power consumption. Mobile devices in complex urban environments often prioritize efficiency but face limited computation and power. Frame cameras require heavy processing and energy, whereas event cameras transmit only brightness changes, reducing redundant data and easing computational and power demands to improve efficiency.

2.4 Challenges of event cameras in mobile sensing

Since event cameras operate fundamentally differently from frame-based cameras by capturing per-pixel brightness changes asynchronously as events, their integration into mobile devices poses several challenges:

(i) How to mitigate event bursts and accurately extract features from event data, given the high-speed operation of mobile devices and event cameras' lack of stable, persistent semantic information? Event cameras are highly sensitive to illumination changes, with even minor variations triggering numerous events. On high-agility mobile devices, rapid scene changes captured by onboard event cameras can trigger event bursts, generating thousands of events in a short

time. Taking the DAVIS 346 event camera (346 × 260 resolution) as an example, let v denote the translational velocity of the drone and w its angular velocity. Under normal flight conditions (2m/s $\leq v \leq$ 4m/s, $5^\circ/s \leq w \leq 15^\circ/s$), the event generation rate is 298 e/ms. During rapid translation (18m/s $\leq v \leq$ 25m/s, $5^\circ/s \leq w \leq 15^\circ/s$), the rate increases to 945 e/ms, while in rapid rotation (2m/s $\leq v \leq$ 4m/s, $75^\circ/s \leq w \leq 95^\circ/s$), it further rises to 1437 e/ms [39]. In contrast, the data rate of a frame-based camera is independent of motion dynamics and scene content. Regardless of whether the scene is static or highly dynamic, frames are sampled at a fixed rate, e.g., 1MP × 30fps = 30M pixels/s, resulting in a stable and uniform data stream. This comparison highlights the non-uniform nature of event data: static scenes generate almost no events, while high-speed motion can trigger explosive outputs far exceeding frame rates. This sparsity–burst duality, coupled with the lack of stable texture or semantic cues, makes efficient processing difficult, as meaningful signals are often buried in motion-induced noise.

(ii) How to efficiently process a large volume of event data given on-board constrained resources? Mobile devices typically rely on low-power embedded systems for efficiency, inherently limiting their computational capacity. For example, a smartphone SoC such as the Qualcomm Snapdragon 8Gen3 has a typical power budget of 5–7W, with its integrated NPU delivering around 30$TOPS$ at INT8 precision. By contrast, a laptop equipped with an Intel Core i7-13700H can consume up to 45W per CPU package and provide 1–2 $TFLOPS$ of FP32 compute power. For high-performance drones, the NVIDIA Jetson Orin NX module consumes 15–25W and delivers 100 $TOPS$ of AI compute, optimized for onboard vision tasks. In comparison, a NVIDIA GeForce RTX 4060Ti (160W) offers up to 22 $TFLOPS$ FP32 and 288 $TOPS$ INT8, while the RTX 4090 (450W) reaches 83 $TFLOPS$ FP32 and over 1.3 $POPS$ INT8. Clearly, mobile devices trade computational resources for energy efficiency. Meanwhile, onboard event cameras capture rapid scene changes, generating large volumes of events that demand efficient processing. For instance, the iniVation DVXplorer can reach peak output rates above 1$Meps$, the Sony IMX636 sensor used in Prophesee Gen4.1 exceeds 10$Meps$, and the high-speed CelePixel Taurus supports up to 240$Meps$. Such workloads heavily strain limited computational resources. In a typical scenario, an algorithm processing a 5$Meps$ event stream with 100 GOP/s demand is trivial for an RTX 4060Ti (288 $TOPS$), leaving ample headroom for other tasks. For a smartphone NPU (30 $TOPS$), however, achieving low latency and high energy efficiency is challenging due to memory sharing constraints and thermal throttling.

2.5 Commercial product & comparation

Event cameras are increasingly entering the commercial market, with several companies offering distinct products. Representative suppliers include iniVation, Prophesee, Lucid Vision Labs and CelePixel. Their commercial offerings emphasize features relevant to mobile and embedded contexts, such as low-power operation, compact modules, mobile-friendly interfaces, or robust industrial connectivity.

Inivation is a prominent leader in the event camera industry, specializes in developing high-resolution, energy-efficient event cameras [47]. Inivation's flagship sensors, like the DAVIS240 and DAVIS346, combine event-based and frame-based data, enabling real-time analysis of both modalities. With resolutions up to 1 megapixel, they rank among the highest-resolution event cameras available. Designed for excellent low-light performance and outdoor durability, these sensors are ideal for applications in robotics, autonomous systems, and more.

Prophesee is a leading innovator in event-based vision technology, renowned for its cameras' ultra-high dynamic range exceeding 120 dB [48]. Their flagship Metavision sensors, including the EVK4 HD and EVK5 HD cameras, offer microsecond-level temporal resolution for applications requiring rapid response and high accuracy, such as autonomous driving and industrial automation. Recently, they have introduced the GENX320, a new sensor designed with features particularly beneficial for mobile sensing applications. With its compact 320×320 resolution, small 6.3μm pixels, and

Table 3. Specifications of Event Camera Models for Mobile/Embedded Sensing Applications

Supplier	Model	Resolution	Dynamic Range (dB)	Pixel Size (µm)	Power Consumption	Mobile/Embedded Features
iniVation	DVXplorer	640×480	90-120	9.0	Max 12W	Multi-camera synchronization; Robust aluminum casing
	DVXplorer Micro	640×480	110	9.0	<140mA @ 5VDC	Compact and lightweight design
	DAVIS346	346×260	120	18.5	Typical 180mA @ 5VDC	Hybrid output (provides both event and frame data)
	DAVIS346 AER	346×260	120	18.5	Typical 180mA @ 5VDC	Integrated IMU for self-contained visual-inertial sensing
Prophesee	EVK4 HD	1280×720	>86	4.86×4.86	Typical 0.5W	High-resolution sensor with external trigger support
	EVK5 HD	1280×720	>110	4.86×4.86	Typical 0.5W	On-board advanced processing; Hardware trigger support
	GENX320	320×320	>120	6.3	3mW	Ultra-low power consumption; Compact, mobile-optimized design
CelePixel	CeleX-V	1280×800	-	9.8	400mW	Multi-mode output (Event, Grayscale, Accumulated frames)
	CeleX5-MIPI	1280×800	-	9.8	-	Designed for direct, low-level System-on-Chip (SoC) integration
Lucid Vision Labs	TRT009S-EC	1280×720	120	4.86	-	Industrial-grade robustness and reliable data streaming

ultra-low power consumption, the GENX320 addresses key constraints in battery-powered mobile devices, making it well-suited for applications such as AR/VR headsets, drones, and other embedded platforms [46].

Lucid Vision Labs, a manufacturer of industrial cameras [49], offers the TRT009S-EC Triton camera built on the Sony IMX636 event sensor. It utilizes a GigE Vision interface with Power over Ethernet (PoE) to ensure robust and reliable data streaming. This design prioritizes operational stability for mobile robotic platforms over the ultra-low power consumption typical of other mobile-optimized sensors.

CelePixel developed smart sensory platforms with a unique on-chip processing architecture. While its official website is inactive, technical resources remain available via software repositories. Their CeleX-V sensor [44] is notable for a multi-mode capability, outputting events, full frames, and hardware-generated event-accumulation views. The CeleX5-MIPI [45] variant targets mobile integration, providing a MIPI CSI-2 interface for low-power connection.

Comparation of products. Table 3 compares commercial event cameras, revealing significant diversity in their specifications. Resolutions range from 320x320 to high-definition 1280x800 pixels, while a high dynamic range (often >110 dB) is a common strength for challenging lighting. A critical differentiator is power consumption, which spans from just 3 mW for mobile-optimized models to 12 W for high-performance ones. This reflects divergent design goals, from ultra-low-power, compact devices to performance-focused industrial models with advanced features like integrated IMUs, hybrid outputs, and multi-camera synchronization for complex robotic systems.

2.6 Event-based datasets

As shown in Tab. 4, event-based datasets serve as key benchmarks for robotics and visual perception.

MVSEC [50]. The MVSEC (Multi Vehicle Stereo Event Camera) dataset integrates data from event camera, stereo frame-based camera, LiDAR, IMU, motion capture, and GPS, making it a comprehensive resource for tasks such as stereo vision, SLAM, and autonomous driving. It provides synchronized stereo event streams and frame-based stereo images captured from vehicles navigating diverse driving environments, including urban roads and highways.

DVS-Pedestrian [51]. The DVS-Pedestrian dataset is a benchmark dataset designed specifically for pedestrian detection and tracking, captured using a DVS in various urban environments. It features sequences of pedestrians performing various actions (walking, running, standing) under different lighting conditions and backgrounds.

DDD20 [52]. The DDD20 (Dynamic Driving Dataset 2020) features synchronized frame camera, event camera, and IMU data from real-world urban and highway environments. It includes vehicles, pedestrians, cyclists and supports

Table 4. Event-based Datasets

Dataset name	Year	Data Volume	Perspective	Participants	Lighting Conditions	Annotation Count	Application Scenario
MVSEC [50]	2018	-	Dynamic	Pedestrians, vehicles	Daytime	-	Driving, handheld scenes
DVS-Pedestrian [51]	2019	0.1 hours, 4.6K annotations	Dynamic	Pedestrians	Daytime	4.6K	Walking street
DDD20 [52]	2020	51 hours	Dynamic	Pedestrians, vehicles	Daytime, night	-	Driving
1 Mpx Automotive [56]	2020	15 hours	Dynamic	Cars, pedestrians, two-wheelers	Daytime, night	25M bboxes	Object Detection
DSEC [53]	2021	1 hour, 390K annotations	Dynamic	Pedestrians, vehicles, scenes	Daytime, night	390K	Driving
TUM-VIE [54]	2021	21 video clips	Dynamic	Static objects	Standard light, low-light	-	3D perception and navigation
FE108 [57]	2021	1.5 hours	Dynamic	21 object types	LL, HDR, fast motion	208K frames	Object Tracking
VECtor [55]	2022	12 video clips	Dynamic	Static objects	Standard light, low-light, HDR	-	SLAM
eTraM [58]	2024	10 hours, 2M annotations	Static	Vehicles, pedestrians, micro-mobility	Daytime, night, twilight	2M	Intersections, roadways, streets
LLE-VOS[59]	2024	70 video clips	Dynamic	Pedestrians, other targets	Normal, low-light	5600	Gym, classroom, zoo
REVD [60]	2024	21 sequences	Dynamic	Various scenes	Varied	21 paired seq.	Video Deblurring
SDE [61]	2024	91 sequences	Dynamic	Indoor, outdoor scenes	Low-light, normal	30K+ pairs	Low-Light Enhancement

key tasks like object detection, tracking, and motion prediction. Additionally, DDD20 provides vehicle control signals (steering, throttle, braking) and is ideal for evaluating frame-event fusion in driving assistance systems.

DSEC [53]. The DSEC (Dynamic and Static Environment for Cars) dataset is a large-scale dataset for autonomous driving research, designed to test algorithms in both dynamic and static environments. It contains data from multiple sensors, including event cameras, LiDAR, and RGB cameras, captured from vehicles moving through urban streets and highways. The DSEC dataset is particularly useful for tasks like visual odometry, 3D reconstruction, and object tracking in dynamic environments. Its multimodal nature allows for the development of algorithms that can handle challenges of autonomous driving, such as dealing with fast-moving objects and varying light conditions.

TUM-VIE [54]. The TUM-VIE (TUM Visual-Inertial Evaluation) dataset is designed for evaluating VIO and SLAM algorithms. It contains synchronized data from event cameras, monocular cameras and IMUs, recorded during various motion scenarios, including both indoor and outdoor environments. This dataset is particularly useful for testing algorithms that combine visual and inertial data to estimate the camera's position and orientation in real time.

VECtor [55]. The VECtor Event Dataset is the first SLAM benchmark dataset captured using a fully synchronized multi-sensor setup, including event-based and regular stereo cameras, RGB-D sensors, LiDAR, and IMU, with complete 6-DoF ground truth for diverse scenarios. It captures the full spectrum of motion dynamics and environmental conditions while providing precise calibration, specifically designed to address challenges unique to dynamic vision sensors.

eTraM [58]. The eTraM (Event-based Traffic Monitoring) dataset a fully event-based traffic perception benchmark captured using a high-resolution Prophesee EVK4 HD event camera. It features annotated traffic data with diverse vehicles, pedestrians, and micro-mobility objects under challenging lighting and weather conditions, including high glare, overexposure, underexposure, nighttime, twilight, and rainy days.

LLE-VOS [59]. The LLE-VOS (Low-Light Event-based Video Object Segmentation) dataset is designed for event-based video object segmentation in low-light conditions, providing synchronized event and frame data with ground-truth masks. It captures challenging scenarios including night-time and indoor scenes.

FE108 [57]. The FE108 dataset is a large-scale, frame-event-based resource for object tracking. It comprises 108 sequences with 21 object types under challenging conditions like low light, HDR, and fast motion. With high-frequency ground truth for both domains, it is highly suitable for evaluating multi-modal tracking algorithms.

1 Mpx Automotive Detection Dataset [56]. This dataset is the first large-scale, high-resolution benchmark for automotive object detection. It provides over 14 hours of recordings with more than 25 million bounding box annotations for cars, pedestrians, and two-wheelers, making it ideal for training robust detectors for autonomous driving.

REVD [60]. The Real-world Event Video Deblurring (REVD) dataset is the first real-world benchmark for its task. It offers synchronized high-resolution blurred videos, sharp ground-truth counterparts, and event streams captured in scenes with extreme motion blur, serving as a critical resource for video restoration algorithms.

SDE dataset [61]. The SDE dataset is a large-scale, real-world benchmark for low-light image enhancement. It consists of over 30,000 spatially and temporally aligned image-event pairs captured in varied lighting conditions. This precise alignment enables the development of robust enhancement techniques for real-world scenarios.

Event-based datasets provide essential benchmarks for robotics and autonomous driving (e.g., MVSEC, DSEC, DDD20), SLAM/VIO (e.g., TUM-VIE, VECtor), and traffic monitoring (e.g., eTraM). They also support research in object detection/tracking (e.g., 1 Mpx Automotive, FE108, DVS-Pedestrian) and image restoration under motion blur or low light (e.g., REVD, SDE, LLE-VOS). By offering rich multimodal data, these datasets highlight the strengths of event cameras in high-speed, low-light, and HDR scenarios, advancing beyond traditional vision systems.

2.7 Synthetic data generation: simulators and approaches

While deep learning has advanced event-based vision, progress depends on large, diverse datasets, yet collecting dense ground truth for tasks like optical flow is costly [27]. Simulators address this by generating synthetic event data with pixel-accurate annotations in controllable environments, enabling systematic training and evaluation. This section reviews major simulation approaches, categorized as physics-based rendering, video-to-event conversion, neural rendering, and AI-driven generation, as well as quality assessment methods (Tab. 5).

Physics-based rendering and simulation This approach typically involves rendering high-framerate video from a 3D environment and then converting these frames to events based on a pixel model. A pioneering tool, ESIM [62], synthesizes events by interpolating logarithmic intensity changes to provide “perfect” data with microsecond resolution. For large-scale generation, frameworks like InteriorNet [63] create photorealistic indoor scenes with event streams and corresponding multi-modal data (RGB-D, IMU, semantics). In contrast, some simpler methods just threshold frame differences, a faster but less precise technique [64].

Video-to-event conversion To leverage existing video archives, video-to-event (V2E) converters transform standard video into event streams. The v2e toolbox [65], for instance, implements a sophisticated pixel model that accounts for non-ideal behaviors like bandwidth limits, threshold mismatch, and noise, yielding highly realistic DVS events. Other robust tools, including the V2CE Toolbox [66] and the Prophesee Video to Event Simulator [67], also provide robust methods for converting frame-based video into the event domain.

Neural rendering and AI-driven event generation More recently, neural rendering and generative AI have emerged. Neural Radiance Fields (NeRFs) offer a new direction; EvDNeRF [68], the first dynamic NeRF trained on events, can synthesize event streams of dynamic scenes from novel viewpoints. Pushing this further, generative AI models like Text-to-Events [69] bypass intermediate frames entirely, using a latent diffusion model to generate event data directly from text prompts, though currently limited to specific domains like gestures.

Differentiable simulators and quality assessment Recent research thrusts are differentiable simulation and quality assessment. SENPI [70] exemplifies the former, offering a fully differentiable framework for end-to-end co-optimization of sensor parameters and a network’s architecture. To address sim-to-real gap, Event Quality Score (EQS) [71] provides a metric by comparing latent features from a network processing real versus synthetic data. The resulting score correlates with real-world performance and can serve as a loss function to improve simulators.

Table 5. Summary of Event Camera Simulators and Generation Methods

Simulator/Method	Type	Core Mechanism	Key Features/Output	Noteworthy Aspects/Limitations
ESIM	Physics-based	Interpolates log-intensity from rendered video to trigger threshold-based events.	Event streams, intensity frames, depth maps.	Foundational open-source tool. Generates 'perfect' events but lacks realistic noise models.
InteriorNet	Physics-based (Integrated)	Renders large, photo-realistic indoor scenes and converts the high-framerate video to events.	Event streams, RGB-D, IMU, semantics, optical flow.	Mega-scale dataset with rich ground truth (RGB-D, IMU) and realistic physics.
v2e Toolbox	Video-to-Events	Converts video to events using a detailed DVS pixel model with various noise sources.	Realistic DVS event streams.	High realism, especially for low-light, by modeling sensor non-idealities (noise, bandwidth).
V2CE	Video-to-Events (Learning-based)	Uses a 3D UNet to predict event voxels, then a statistical model (LDATI) for continuous timestamps.	High-fidelity, continuous event streams.	Claims SOTA. Solves 'temporal layering' problem by generating continuous, non-discrete timestamps.
EvDNeRF	Neural Rendering	Trains a dynamic NeRF directly on event data to synthesize novel event streams.	Predicted event streams, intensity, depth.	First dynamic NeRF for events; can render from novel viewpoints. Training may be unstable.
Text-to-Events	Generative AI	Generates events directly from text prompts using a latent diffusion model.	Synthetic event frames.	Bypasses intermediate video generation. Currently limited to specific domains (e.g., gestures).
SENPI	Differentiable Simulator	Fully differentiable PyTorch library modeling the entire event generation pipeline.	High-fidelity pseudo-event tensors.	Allows co-optimization of sensor parameters and network architectures.
EQS	Quality Assessment Metric	Measures realism by comparing latent features from a network processing real vs. synthetic data.	A differentiable realism score.	Differentiable metric for raw events that correlates with sim-to-real performance. Can be used as a loss.

3 ABSTRACTION: EVENT REPRESENTATION

Event data is often processed and transformed into various representations to extract meaningful information (features) for solving specific tasks. Here, we review popular representations of event data, which range from simple, hand-crafted transformations to more elaborate methods, as shown inFig.5.

3.1 Raw events

Raw events offer high fidelity, retaining complete temporal and spatial information, making them ideal for event-driven processing, especially in applications using SNNs. While raw events provide detailed and precise information, they come with challenge of managing substantial data loads and ensuring proper alignment of events across time.

Individual Events. Raw events $e_k \doteq (\mathbf{x}_k, t_k, p_k)$ are utilized by event-by-event processing methods such as probabilistic filters and Spiking Neural Networks (SNNs). These methods build additional information from past events or external knowledge and fuse it with incoming events asynchronously to produce an output [72, 73].

Event Packet. The event set $E = \{e_k\}^{N_e}$ retains precise timestamp and polarity information for each event. Selecting the appropriate packet size N_e is crucial to meet the assumptions of the algorithm (e.g., constant motion speed throughout packet's duration), which varies depending on the task [74, 75]

3.2 Event frame (2D Grid)

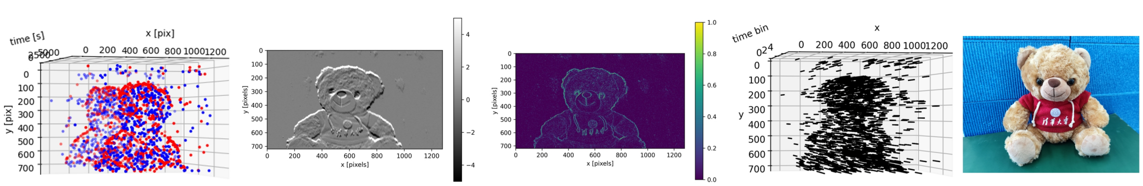
Events within a spatio-temporal neighborhood are converted into a 2D grid—often by counting events or accumulating polarity per pixel—forming an Event Frame compatible with standard image-based algorithms. While easy to implement, this representation can lose temporal information, suffer from motion blur in dynamic scenes, struggle under HDR conditions, and fail to fully exploit the sparsity of event data, reducing efficiency.

Histogram. This representation converts events into histograms, offering an activity-driven sample rate. Although not fully aligned with the event-based paradigm, it has proven effective [76, 77]. Traditional 2D histograms discretize events into bins, while the Activity-Aware Event Integration Module extends them with spatiotemporal operations, capturing finer details and improving dynamic-scene performance, particularly in semantic segmentation [78].

Time Surface. A Time Surface (TS) is a 2D map where each pixel stores a single time value (e.g., the timestamp of the last event at that pixel). Events are converted into an image whose "intensity" is a function of the motion history at that location, with larger values corresponding to more recent motion. TSs explicitly expose the rich temporal information of the events and can be updated asynchronously [79–81].

3.3 Spatio-temporal 3D grid representation

A Voxel Grid is a 3D space-time histogram where each voxel corresponds to a pixel and time interval, preserving temporal information better than 2D projections. With polarity, it forms a discretized scalar field on the image plane, using



(a) (b) (c) (d) (e)

Fig. 5. Event representations method (a) Raw event (b) 2D histogram (c) Time surface (d) Voxel grid (e) RGB picture.

accumulation or kernel distribution for sub-voxel accuracy [82–84]. This representation retains rich spatiotemporal detail for precise event tracking but requires more memory and computation, posing challenges for real-time use.

3.4 Customized representation

Customized event representations address task-specific limitations of traditional methods by combining spatial, temporal, or domain-specific features to boost efficiency and accuracy. For example, [85] employs adaptive filtering for motion deblurring to preserve high-frequency details, while [86] introduces the 2D-1T Event Cloud Sequence (ECS), which separates spatial and temporal components to retain sparsity and capture both geometry and motion, aiding recognition.

Task-specific designs improve performance by exploiting relevant features and domain knowledge, but their generalizability across tasks remains limited. For example, the adaptive filtering representation in [85] is tightly coupled with a generative model and an Extended Kalman Filter, making it powerful for dynamic tracking but hard to adapt to recognition or reconstruction tasks that require spatially structured outputs. Similarly, the ECS representation [86] yields compact embeddings suitable for classification but lacks the explicit geometric and kinematic detail needed for precise tracking or mapping. Moreover, both are deeply embedded in their respective architectures, so transferring them to other paradigms demands significant redesign and retraining. Consequently, while customized representations demonstrate how tailoring to specific assumptions can achieve state-of-the-art performance, they also reveal the trade-off between specialization and generalization. Future research toward more unified or transferable event representations, along with standardized benchmarks for cross-task robustness, would be valuable for broadening applicability.

4 ALGORITHM: EVENT PROCESSING

4.1 Event-based denoising

Motivation. As event cameras see increasing use in high-speed, low-latency applications, effective denoising becomes crucial for reliable vision tasks. These bio-inspired sensors detect brightness changes with high temporal precision but are highly sensitive to noise, degrading event stream quality and affecting tasks like reconstruction and detection. Noise arises from both external (e.g., ambient light changes) and internal sources (e.g., leakage currents), causing spurious events known as background activity (BA), which wastes bandwidth and reduces accuracy. Over time, hot pixels continuously emit false events, introducing hot noise. Unlike conventional cameras that suppress noise via image integration, event cameras amplify it due to logarithmic encoding and differential sampling. This results in more BA, missed events, timing noise, and redundant trailing events after edge arrivals, complicating robust denoising efforts.

Challenge. Event cameras face significant challenges in noise mitigation due to diverse noise sources, annotation limitations, computational constraints, and non-uniform noise distributions. One major challenge lies in the diversity of noise types intrinsic to event cameras: *(i) BA noise*, triggered by junction leakage currents or low-light conditions, generates spurious events. *(ii) Hot pixels*, common during prolonged usage or high-speed scenarios, produce persistent erroneous signals. *(iii) Temporal noise* introduces stochastic timing variations, while *(iv) structural noise* arises from edge inconsistencies or redundant trailing events. Their distinct spatiotemporal and statistical characteristics complicate simultaneous mitigation. The lack of high-quality annotated datasets further limits supervised denoising methods, as creating paired data is labor-intensive, and synthetic datasets suffer from domain gaps. Multi-modal fusion with frame or IMU data helps, but struggles with extreme motion blur or lighting changes.

Computational efficiency is another bottleneck, many algorithms prioritize accuracy over latency, making them unsuitable for power-constrained systems, especially in low-light conditions where noise worsens. Moreover, the spatially and temporally non-stationary nature of noise challenges traditional methods that assume uniform noise. For example, BA noise often appears in localized bursts, while other noise types vary over time. Robust solutions must handle these irregularities for consistent performance.

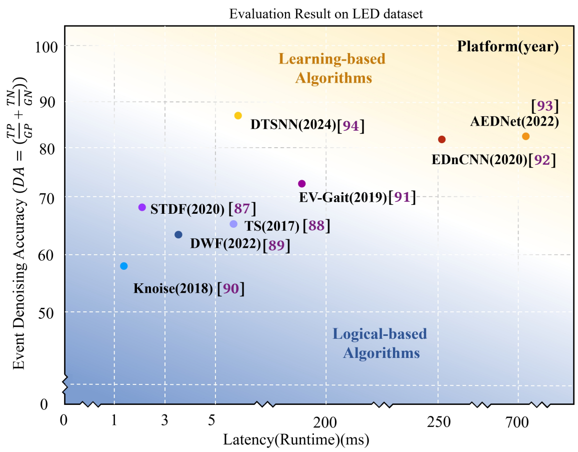


Fig. 6. Comparison of event-based denoising algorithms in computational cost vs. performance, all evaluated on the LED dataset.

Literature review. Event denoising has achieved notable advancements through diverse methodologies, including statistical approaches, filtering-based techniques, surface fitting and deep learning, significantly improving noise suppression in event-based vision (Fig. 6). To ensure fair comparison among these methods despite their diverse assumptions, we adopt the large-scale real-world LED dataset [94], which provides paired noisy/clean events under controllable illumination and noise levels. All methods are evaluated using the same metric denoising accuracy $DA = \frac{1}{2}(\frac{TP}{GP} + \frac{TN}{GN})$, which decomposes into signal retention (SR) and noise removal (NR), thus jointly capturing the two key aspects of denoising quality. As illustrated in Fig. 6, this unified setting enables an objective and reproducible comparison of statistical, filtering-based, and deep learning strategies.

Statistical methods. Early techniques relied on statistical analysis to identify outliers by evaluating event density in local spatio-temporal neighborhoods [95]. Delbruck *et al.* [96] pioneered density-based filtering, leveraging local context to suppress noisy events. Subsequent efforts [89, 90] enhanced these techniques with optimized event storage strategies to reduce computational complexity and improve processing efficiency. However, these approaches often require manual parameter tuning to adapt to varying noise conditions, limiting their scalability and generalizability.

Filtering-based methods. To better exploit the temporal and asynchronous nature of event data, several filtering algorithms have been developed. *(i) Temporal filters* [97] remove redundant events by leveraging temporal correlations. *(ii) Spatial filters* [98] isolate motion-related events by analyzing pixel intensity changes. *(iii) Spatio-temporal filters* [95, 99] combine both strategies to suppress background activity (BA) noise. For example, Liu *et al.* [100] showed that integrating spatial and temporal filtering reduces BA noise while preserving critical motion events.

Surface fitting techniques. Surface fitting methods offer an alternative approach, particularly effective for smoothing event data in continuous motion scenarios. Methods like EV-Gait [91] and the Guided Event Filter (GEF) [101] employ local plane fitting , optical flow estimation, and image gradients for noise smoothing. The time surface (TS) method [88, 97] transforming event streams into monotonically decreasing representations, effectively mitigates sparsity but exhibits limited performance in low light or highly dynamic scenes.

Deep learning-based methods. Recent advances in deep learning have revolutionized event denoising, enabling automated solutions by training on noisy-clean event pairs [102, 103]. Sparse feature learning (K-SVD [104], MLPF [90])

Table 6. Comprehensive Comparison of event filtering and feature extraction methods

Method	Advantages	Disadvantages	Applications
Frame-based methods [108, 109]	High computational efficiency; Low memory consumption; Simple implementation	Poor accuracy in high-speed scenarios; Limited temporal precision; Loss of fine event details	Basic feature detection; Initial event processing
Surface of Active Events-based methods [110, 111]	High temporal accuracy; Precise event timing preservation; Good feature localization	High memory overhead; Reduced processing efficiency; Complex computational requirements	Temporal-spatial feature detection; Optimization tasks.
Clustering-based methods [112, 113]	Linear time complexity; Efficient memory usage; Fast stream processing	Accuracy depends on parameters; Poor precision in complex scenes; Unstable performance	Dense event processing; Dynamic scene analysis
Geometric transform-based methods [114–116]	High tracking accuracy; Precise motion estimation; Robust feature detection	Heavy computational load; Low processing efficiency; High resource consumption	Fast motion tracking; Extreme illumination scenarios
Temporal filtering-based methods [117, 118]	Fast processing speed; Efficient memory utilization; Good real-time performance	Accuracy affected by noise; Precision loss in filtering; Detail preservation issues	High temporal precision tasks; Real-time processing; Resource-constrained systems
Asynchronous methods [119–121]	Low latency processing; High temporal accuracy; Efficient event handling	Complex implementation; Resource intensive	Low-latency applications; High-speed corner detection; Noise-heavy environments
Hybrid methods [122–124]	High detection accuracy; Robust feature extraction; Multi-layer filtering	Computationally heavy; Complex parameter tuning	High-speed tasks; Multi-feature scenarios
Neural networks-based methods [117, 125–127]	High feature accuracy; Tracks complex dynamics; Biologically inspired	Training-intensive; High computational demands; Accuracy-speed tradeoffs	Complex dynamic environments; Biological vision systems
Frame-event hybrid methods [128–130]	High spatial-temporal accuracy; Precise feature matching; Robust performance	Complex design; High computational cost; Complex resource management	Robotics; Precision tasks; High-speed tracking

methods focus on sparse feature extraction and event probability estimation. EDnCNN [92] integrates frame and IMU data to classify events as signal or noise. EventZoom [105] employs a U-Net architecture for efficient noise-to-noise denoising, achieving superior performance in handling noisy event streams. AEDNet [93] processes raw DVS data while preserving inherent spatio-temporal correlations. EDformer [106] introduces an event-by-event transformer model, while EDmamba [107] leverages state space modeling to achieve efficient noise-aware spatiotemporal denoising.

4.2 Event-based filtering and feature extraction

Motivation. Unlike traditional frame-based cameras capturing images at fixed intervals, event cameras operate asynchronously, producing sparse data only when significant brightness changes occur. This dynamic output demands specialized filtering to isolate meaningful events for effective feature extraction. Event filtering highlights important scene dynamics—like motion and edges—while removing irrelevant or redundant data. This enhances efficiency and accuracy in tasks such as object detection, motion tracking, and scene reconstruction, especially in fast-moving or rapidly changing lighting conditions. By reducing noise and preserving critical features, filtering supports robust, low-latency processing essential for real-time applications.

Challenge. The asynchronous nature of event data poses key challenges for filtering, as there's no universal definition of a "significant" event. Variations in event frequency and timing complicate maintaining temporal coherence, which is vital for accurate motion representation during feature extraction. Balancing filtering complexity with real-time processing is also difficult—advanced methods may boost accuracy but risk adding latency. Parameter tuning is another challenge, since optimal settings vary with environmental conditions and must adapt to dynamic, unpredictable scenes. This need for adaptability makes robust filtering crucial for reliable event-based vision. Overcoming these issues is essential to enhance feature extraction and ensure consistent real-world performance.

Literature review. Event-based feature extraction has progressed from adapting frame-based methods to developing specialized event-driven approaches (Tab. 6). Early work modified traditional techniques, such as applying Harris corner detection [108] to event accumulation frames. A major advance was the Surface of Active Events (SAE) [110], which records the timestamp of the latest event per pixel, preserving temporal precision while improving efficiency and feature extraction performance.

Feature extraction methods have advanced in both accuracy and efficiency. eFAST [119] shifted from gradient-based to faster comparison-based operations tailored for event data, while Arc* [111] improved detection speed and repeatability through refined SAE filtering. Hybrid methods like FA-Harris [122] and TLF-Harris [123] balanced efficiency and robustness via candidate selection and multi-layer filtering. FEAST [125] introduced unsupervised extraction with spiking neuron-like units, and ROT-Harris [124] used tree-based processing to surpass traditional 2D approaches. Together, these works highlight a shift toward practical, scalable implementations.

Parallel to advancements in feature extraction, event filtering has emerged as a critical preprocessing step for robust feature extraction, leveraging the asynchronous and high-resolution temporal data of event cameras. Temporal filtering-based methods have significantly improved data quality and computational efficiency by prioritizing meaningful scene changes [117, 118]. Clustering-based methods, such as eCDT [112], dynamically group events to represent dense streams compactly while minimizing complexity. However, such methods often face challenges in maintaining robustness under varying conditions due to sensitivity to parameter configurations [113].

Parametric filtering has further advanced event-based vision by employing geometric transformations to filter irrelevant data. For instance, EKLT [114] aligns events to improve feature extraction, while curve-fitting techniques construct smooth spatio-temporal trajectories [115], excelling in scenarios with rapid motion or extreme illumination. Asynchronous methods, such as HASTE [120], process individual events in real time using hypothesis-driven transformations, and proximity-based trackers continuously refine feature locations, effectively suppressing noise and supporting low-latency applications [121].

Integrating filtering with feature extraction has greatly improved the robustness and efficiency of event-based vision systems. For instance, combining filters with classical algorithms like Harris or FAST enables reliable corner detection, while shape detection methods using ICP and Hough transforms enhance performance in high-speed, high-dynamic-range settings [128]. Modern tracking approaches also exploit the low latency and high temporal resolution of event cameras by incorporating probabilistic associations and spatio-temporal constraints, resulting in more stable features and improved tracking accuracy [129, 130].

Neural network-based filtering integrates asynchronous event processing with biologically inspired architectures. Pulse-based neural networks [126] exploit the high temporal resolution of event data for precise feature tracking, showing strong potential in dynamic environments. Yet, current models still fall short of the detection accuracy achieved by convolutional networks [117, 127]. Ongoing efforts to unify filtering, feature extraction, and tracking are pushing event-based vision toward broader applicability while preserving efficiency and robustness in real-world scenarios.

4.3 Event-based matching

Motivation. Matching involves identifying corresponding features between event streams captured at different times or viewpoints, forming the basis for tasks like visual odometry, video interpolation, and mobile sensing. Event cameras, unlike RGB cameras that operate at fixed intervals, trigger only on brightness changes, making them ideal for dynamic scenes. They capture edges and textures efficiently while avoiding redundant data. The precise temporal information in event data also enables accurate motion inference. By mitigating issues like motion blur and low frame rates common in traditional cameras, event cameras offer distinct advantages for matching tasks—even with single-modal data—demonstrating strong performance in high-frequency, fast-changing environments.

Challenge. Event-based matching faces major challenges due to the sparsity of events in pixel space and their uneven temporal distribution, which hinder robust feature extraction—especially in low-light or static scenes. This calls for specialized methods adapted to event camera properties. Moreover, the ultra-high temporal resolution produces massive data volumes, raising computational costs. Achieving real-time performance in resource-limited settings thus requires algorithms that balance efficiency and accuracy, minimizing latency while ensuring reliability in dynamic env..

Table 7. Event-based Matching Algorithm Comparison

Algorithm	Sample	Input	Advantages	Disadvantages
Local feature matching [131–140]	Optical Flow Feature descriptor: SIFT/SURF/ORB	Batch	Low latency, noise resistance Low computational resource	Difficult to handle global motion Difficult to handle repeat texture Need further extraction method
Optimization-based matching [141–143]	Contrast maximization Levenberg-Marquardt Graph optimization	Batch/Asynchronous	Global consistency Higher accuracy after iterations Detailed design of optimizer	Higher computational complexity Higher time latency Hard to converge Sensitive to initial values
Neural network-based matching [144–148]	CNN SNN GNN	Batch/Asynchronous	Adaptive learning Less manual intervention Efficient inference	Large amounts of training data Need parameter tuning Poor generalization ability

Literature review. In visual data processing, matching algorithms are essential for motion estimation, visual odometry, and feature tracking. They fall into three main categories—local feature, optimization-based, and deep learning-based matching—each with distinct strengths and limitations (Tab. 7).

Local feature matching methods. These methods focus on detecting and associating local features within visual data [149]. Techniques such as optical flow [131, 132] and descriptors like SIFT [135], SURF [136], and ORB [137] exemplify this approach. They offer low latency, robustness to noise, and low computational demands, making them well-suited for mobile, autonomous, and robotic applications [139, 140]. However, local feature matching faces challenges with event streams, especially under large global motion, repetitive textures, or heavy reliance on specific feature extractors [150]. Its performance also declines in scenes with sparse features or dramatic changes, underscoring the need for more advanced and robust techniques.

Optimization-based matching methods. These methods aim to ensure global consistency by solving optimization problems to align event data or trajectories. Techniques such as contrast maximization [141–143], the Levenberg-Marquardt algorithm, and graph-based optimization are widely used. They offer high flexibility and accuracy, supporting both batch and asynchronous processing, which makes them valuable in event camera applications. However, these approaches are computationally demanding and sensitive to initial conditions—poor initialization can lead to local optima or non-convergence. Their high memory and processing requirements also limit real-time usability, especially for large-scale data. Achieving a balance between efficiency and accuracy remains a core challenge.

Deep learning-based matching methods. These methods mark a transformative shift by employing CNNs [146], SNNs [144, 145], and GNNs [148] to automatically learn representations and matching strategies, surpassing handcrafted descriptors in accuracy. GNNs are well-suited to the asynchronous nature of event streams and excel in high-speed, dynamic environments. However, these models are computationally demanding, rely on large annotated datasets, and struggle to generalize, while training remains time-consuming and requires careful hyperparameter tuning.

Selecting the most suitable matching algorithm depends on the specific task requirements, including computational constraints, real-time processing needs, data characteristics, and scene complexity. As technology advances, these algorithms are likely to evolve and converge, enabling the development of more versatile and robust visual matching techniques capable of addressing a broader spectrum of applications.

4.4 Event-based mapping

Motivation Building on event matching, asynchronous event streams from different viewpoints can be fused across poses to incrementally reconstruct a dense 3D map. Event-based sensing leverages spatial sparsity and high temporal resolution, reducing computation by focusing on regions of change and minimizing motion blur to preserve edges during fast motion. These properties enable real-time, efficient, and robust mapping in dynamic environments.

Challenge. Using cameras—including event cameras—for mapping and depth estimation is feasible but challenging. Monocular cameras lack direct depth information, requiring techniques like multi-frame perspective changes or fusion

Table 8. Comparison of different mapping methods.

Types of Mapping Methods	Advantages	Disadvantages or Challenges
Frame-based mapping[151–156]	✓ Improves the accuracy of depth estimation ✓ Enhances the quality of map construction ✓ Lower latency and computation cost	✗ Requires measuring and updating depth information between current and initial detection frames, potentially increasing computational load
Filter-based mapping [157–160]	✓ High robustness, suitable for rapidly changing environments ✓ Dynamically adjusts the system's map representation	✗ Requires a large number of parameters to represent camera poses, potentially increasing computational complexity
Continuous-time mapping [161–163]	✓ Reduces parameter complexity ✓ Improves mapping accuracy and efficiency ✓ Simultaneously updates camera poses and 3D landmarks	✗ Requires handling continuous curve interpolation and optimization problems ✗ Increasing computational difficulty ✗ High latency due to frequent states update
Spatio-temporal Consistency[164]	✓ Improves the accuracy of map construction ✓ Optimizes motion parameters	✗ Requires iteratively searching for the closest points and applying a pruned ICP algorithm ✗ High computational cost and latency

with auxiliary sensors (e.g., IMUs) to infer depth. Stereo event camera setups can estimate depth more directly through disparity. However, event cameras' high temporal resolution leads to significant computational demands, as processing their continuous asynchronous data in real time is resource-intensive. Additionally, aligning event data with other sensor modalities complicates feature extraction and synchronization calibration. These challenges intensify in dynamic environments, where maintaining sensor consistency is vital for accurate depth estimation.

Literature review. Mapping plays a fundamental role in constructing a 3D representation of the environment based on visual features captured by cameras. With advancements in event-based vision, event cameras have become an increasingly valuable source of information in visual simultaneous localization and mapping, offering high-frequency, asynchronous data well-suited for real-time processing in dynamic scenes. Various event-based mapping approaches have been proposed, each exhibiting distinct characteristics and advantages Tab. 8.

Frame-based mapping methods. Frame-based mapping methods often use event-derived 2D representations with depth filters to iteratively refine scene depth via feature triangulation. For example, [153] estimates poses relative to planar structures by minimizing reprojection errors, while [155] adopts the SVO algorithm [156] for pose estimation from event feature correspondences. These methods typically model depth with Gaussian–uniform filters, updating estimates from feature comparisons guided by poses. Additionally, [152] improves mapping by solving pose estimation as a least-squares problem on 2D–3D line constraints.

Filter-based mapping methods. Filter-based mapping techniques address the asynchronous nature of event data by continuously updating maps during camera tracking. Line-based vSLAM refines maps by measuring distances between incoming events and reprojected 3D lines, generating point cloud reconstructions. These methods often use the Hough transform to extract 3D line features, leveraging spatial correlations for robustness in dynamic scenes. For example, [157] estimates poses via distances between back-projected event rays and planar points, while [160] improves precision with a probabilistic measurement function on planar surfaces. More recent strategies, such as [158], update filter states by evaluating event-to-line distances, often combining EMVS [159] with Hough-based line extraction to strengthen event-line associations and enhance spatial accuracy.

Continuous-Time mapping methods. To reduce the high parameter count in filter-based discrete pose representations, continuous-time mapping replaces discrete poses with smooth trajectory models such as B-splines or Gaussian processes, enabling interpolation from local control states. This lowers complexity and supports joint optimization of poses and landmarks, improving accuracy and efficiency. For example, [161] employs B-splines, while [162] use Gaussian process motion models to interpolate poses at arbitrary timestamps. [162] further integrates incremental SfM for consistent refinement of control states and landmarks, and [164] introduces a spatio-temporal constraint based on equal-time event pairs to enhance rotational accuracy. In practice, these methods iteratively establish correspondences and apply pruned ICP for spatial consistency, yielding more precise event-based maps.

5 ACCELERATION

Event-based sensing generates data only when pixel intensity changes, drastically reducing redundancy and offering efficiency, low power, and minimal latency, which is well-suited for high-speed, energy-constrained applications. With growing demand on mobile and edge platforms such as drones, vehicles, and wearables, real-time and efficient processing is critical but remains challenging under limited resources. Balancing accuracy and efficiency thus becomes central for deployment. To address this, optimized hardware accelerators and specialized software must exploit event sparsity to cut computation while preserving accuracy (Fig.7, Fig.8), enabling real-time event-based vision on mobile devices for advanced autonomous applications.

5.1 Hardware acceleration

Motivation. Event-based vision transforms visual processing by generating data only from scene changes, achieving sparsity, high temporal resolution, and low latency—ideal for real-time, power-limited applications. Yet, conventional processors built for dense synchronous data struggle with sparse, asynchronous streams, making specialized hardware essential. Hardware accelerators bridge raw events to high-level tasks like detection, tracking, and reconstruction by exploiting sparsity and parallelism to cut redundancy, streamline data flow, and support massive parallel processing. This enables precise, low-latency, and energy-efficient event-based systems ready for practical deployment.

Challenge. Traditional CPUs, constrained by sequential execution, struggle to process sparse, asynchronous event streams in real time, often failing to meet the low-latency, high-throughput demands of tasks like detection and tracking on resource-limited platforms. GPUs, though powerful for dense parallel workloads, are optimized for structured image data; when applied to irregular event streams, they suffer from poor resource utilization, higher latency, and increased power draw. In contrast, FPGAs offer a reconfigurable, massively parallel architecture that can be tailored to the unique characteristics of event data. Through custom dataflow designs, sparse convolutions, and asynchronous pipelines, they deliver low-latency, energy-efficient performance without relying on dense matrix operations. Their fine-grained hardware control and inherent parallelism make them especially well-suited for mobile and embedded platforms, where efficiency and performance must coexist.

Literature review. Recent research on hardware acceleration for event-based vision can be broadly categorized into three main approaches: neuromorphic computing, event-driven deep neural network (DNN) acceleration, and hardware optimization techniques aimed at enhancing efficiency and reducing power consumption.

Neuromorphic computing for event-based vision. Neuromorphic computing, inspired by the brain's dynamic processing, mimics biological neurons and synapses to enable efficient event-driven computation. SNNs typically run on custom neuromorphic chips for low-power, high-efficiency processing. Notable devices include BrainScales [165], Spinnaker [166], Neurogrid [167], TrueNorth [168], Darwin [169], and more recently, Loihi [170], Tianjic [171], and Speck [172]. A key challenge is improving energy efficiency via high-level brain-inspired mechanisms. Among these, the asynchronous chip Speck [172] stands out as a sensing-computing SoC that fully leverages sparse, event-driven processing. Operating at ultra-low power (0.70 mW in real-time), Speck demonstrates neuromorphic computing's promise for power-constrained mobile and edge systems.

Event-driven DNN acceleration. Neuromorphic computing provides a bio-inspired paradigm for event-driven processing, while another crucial research direction focuses on hardware acceleration tailored to event-based deep learning models. Unlike conventional GPUs, which fail to fully exploit the sparsity of event data, specialized architectures have been developed to achieve higher efficiency. (i) Sparse dataflow architectures: The combinable dynamic sparse dataflow architecture (ESDA) [173] realizes a configurable sparse dataflow model on FPGAs. By employing a unified sparse token–feature interface to interconnect parameterizable network modules, ESDA reduces both latency and power consumption during event-based DNN inference, making it well-suited for edge deployment. (*ii*) Optimized fusion for event-based vision: EventBoost [174] accelerates event–image fusion through a dedicated hardware accelerator on the Zynq SoC platform. By optimizing fusion workloads in real time, EventBoost mitigates inefficiencies typical of CPU- and GPU-based processing, significantly boosting performance for event-driven visual tasks.

Hardware optimization for efficient event processing. Beyond neuromorphic computing and event-driven DNNs, hardware acceleration is essential for unlocking the full potential of event-based vision, particularly in power-constrained and efficiency-critical scenarios. One optimization strategy aims to reduce power consumption directly at the sensor interface. For example, [182] proposes a dedicated on-chip DVS interface that aggregates asynchronous event streams into ternary event frames, substantially lowering the energy demands of subsequent processing stages.

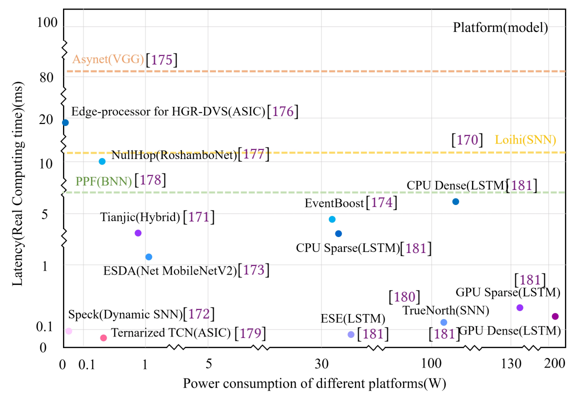


Fig. 7. Power consumption of different platforms(W) vs. Latency(Real Computing time)(ms)

In addition to such front-end innovations, a variety of FPGA-based accelerators have been explored to enhance efficiency across the entire event-vision pipeline by exploiting the inherent parallelism of event data. For instance, [176] presents a reconfigurable processing element architecture that integrates both a median filter core and an AI accelerator core for CNN inference within a system-on-chip (SoC). Leveraging Reconfigurable Multiple Constant Multiplication (RMCM) for efficient resource sharing, this design achieves an energy consumption of only 593.4 nJ per inference under 65 nm technology, significantly reducing computational cost. Similarly, [183] develops a complete event-driven optical flow camera system with FPGA-based acceleration for key modules such as event-driven corner detection and adaptive block matching optical flow. Expanding toward heterogeneous architectures, [184] demonstrates an FPGA/ARM platform for Event-based Monocular Multi-View Stereo (EMVS), where algorithm restructuring and mixed-precision quantization boost throughput while minimizing memory footprint. Finally, [174] proposes a Zynq SoC–based event–vision fusion accelerator that employs hardware–software co-design to distribute tasks between FPGA and CPU, maximizing parallelism and efficiency in computationally demanding fusion algorithms.

Application-specific hardware acceleration design. The development of specialized hardware architectures tailored to accelerate sThe development of specialized hardware architectures tailored to specific tasks has become a central focus in event-based vision research. By optimizing hardware resources for domain applications, these solutions achieve substantial performance gains. For instance, [176] introduces an energy-efficient processor-on-chip for hand gesture recognition. A key driver is aerial robotics, where stringent constraints on size, weight, and power demand highly optimized systems. [39] presents FPGA-based hardware–software co-designs that exploit FPGA reconfigurability to implement efficient pipelines for drone tasks such as navigation and obstacle avoidance. Complementarily, [185] demonstrates a neuromorphic approach using Intel Loihi, where event data feed directly into SNNs for end-to-end drone control, achieving high responsiveness with low power. These directions highlight the synergy between specialized hardware and event-based vision. As edge applications—such as autonomous driving, robotics, and AR—demand low-power, real-time performance, domain-optimized accelerators will be pivotal for practical deployment.

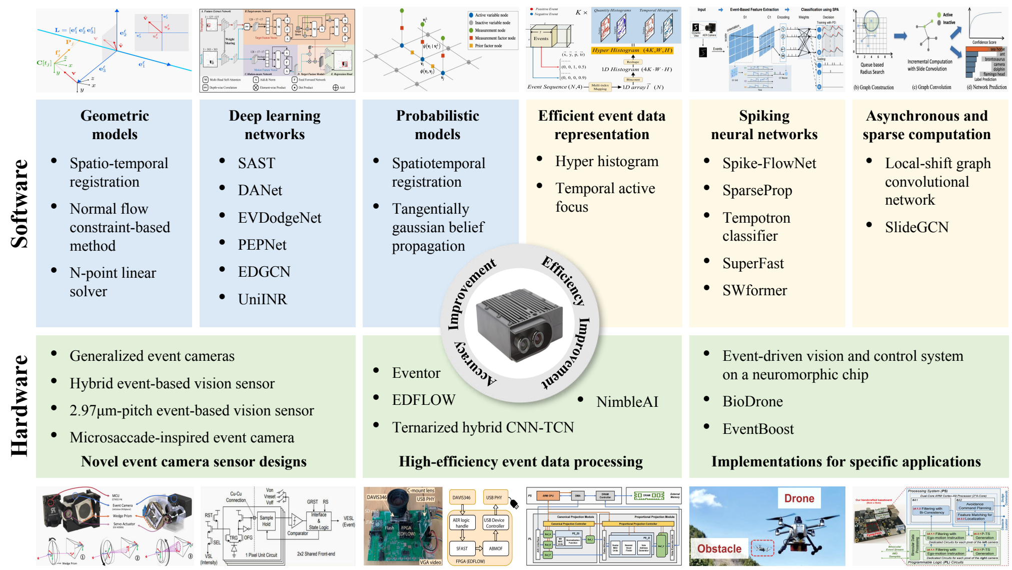


Fig. 8. Specific hardware and software design for acceleration.

A key consideration in hardware acceleration is the trade-off between accuracy and computational latency. Hardware-specific accelerators often reduce latency substantially by exploiting task-specific parallelism and sparse dataflow, but this comes at the cost of reduced architectural flexibility and higher development overhead. Recent case studies illustrate this balance clearly. For instance, EventBoost [174] employs a software-hardware co-design on a Zynq SoC to accelerate event-visual fusion for UAV localization, achieving a 24.33% improvement in accuracy compared to state-of-the-art systems while maintaining only 30 ms end-to-end latency, thereby meeting real-time constraints on resource-constrained platforms. Similarly, BioDrone [186] integrates an FPGA-based processing pipeline for autonomous drone navigation. Experimental results show that the FPGA accelerator reduces per-frame processing latency from nearly 20 ms on CPU to 2.2 ms, a nearly 10× speed-up, while sustaining almost identical perception accuracy (within 1–2% deviation) compared to CPU baselines. These studies demonstrate the fundamental trade-off in hardware-specific acceleration: latency can be drastically reduced without significant accuracy loss, but achieving this requires tight co-design between algorithms and hardware. As a result, while FPGA and ASIC solutions provide a practical pathway toward real-time, power-efficient event-based vision deployment, their specialized nature may limit general applicability across tasks.

5.2 Software acceleration

Motivation. Deploying event cameras on mobile platforms requires software acceleration to handle sparse, asynchronous data in real time under strict resource limits. By transforming raw events into structured representations through clustering, compression, or sparse matrix operations, software acceleration streamlines denoising, filtering, feature extraction, and inference. Exploiting sparsity, reducing redundancy, and applying techniques like adaptive sampling and dynamic memory management are key to achieving energy-efficient, real-world mobile applications.

Challenge. Effective software acceleration for event cameras on mobile platforms faces several challenges. First, sparse and asynchronous nature of event data makes dense-input algorithms inefficient, requiring specialized designs. Second, integration with deep learning frameworks is nontrivial, as most models are tailored for frame-based inputs. Third, limited on-device resources demand optimized memory management and throughput to balance speed and power. Finally, current benchmarks overlook unique characteristics of event data, calling for dedicated evaluation metrics.

Literature review. In event-based vision systems, software acceleration is essential for improving the efficiency and performance of the event processing pipeline. By optimizing key stages of event handling, researchers can significantly enhance computational speed, reduce latency, and better utilize limited resources. Software acceleration techniques target various pipeline stages, such as event sampling, preprocessing, feature extraction, and event analysis. Each stage presents distinct challenges, and numerous specialized acceleration methods have been developed to address them:

Event sampling. Event cameras capture raw data with precise timestamps and spatial locations. Unlike uniform sampling, adaptive sampling prioritizes significant scene changes to improve data efficiency. The key challenge is balancing temporal resolution with sparsity to prevent information loss or redundancy. [72] addresses this by introducing a recurrent convolutional SNN-based adaptive sampling module that dynamically adjusts rates based on spatio-temporal event patterns, enhancing overall efficiency.

Event pre-processing. Events need to be denoised, filtered, and formatted to improve subsequent processing. Common challenges include false positives and inefficient filtering. Enhanced filtering algorithms, such as adaptive median filters, can retain important features while reducing computational overhead. A lightweight, hardware-friendly neural network architecture, 2-D CNN, is introduced in [176] for DVS gesture recognition, using a customized median filter to enhance signal-to-noise ratio and reduce hardware complexity.

Feature extraction. Useful features, including spatial, temporal, and frequency features, are extracted from preprocessed data. Traditional techniques may not effectively capture event data's unique properties. Software acceleration, particularly through deep learning models tailored for sparse data, can optimize feature extraction while reducing processing time. The FARSE-CNN model, proposed in [187], integrates sparse convolutional and asynchronous LSTM modules for efficient event data processing. [188] introduces SWformer, an attention-free architecture leveraging sparse wavelet transforms to capture high-frequency patterns, resulting in improved energy efficiency and performance.

Event analysis. This step involves pattern recognition, classification, or regression, commonly for tasks like object detection or tracking. The sparsity of event data can lead to high computational costs. Optimized CNNs and SNNs can enhance the efficiency and effectiveness of analysis. A fast linear solver for camera motion restoration, developed in [189], addresses geometric problems and adapts to sudden motion changes, providing a robust solution for event data.

Moreover, existing methods improve event processing by optimizing event representations, integrating geometric and probabilistic models, and employing deep learning techniques. They also leverage the asynchronous and sparse characteristics of event data to accelerate computation without sacrificing accuracy.

Efficient event data representation. Efficient event data representation methods convert sparse, asynchronous event streams into structured formats for effective processing with reduced overhead. Recent approaches include event stacking [190], Temporal Activity Focus (TAF) [191], and HyperHistogram (HH) [192]. TAF adaptively adjusts time window length and resolution based on spatial and polarity cues, enhancing flexibility. HH builds multiple histograms from event polarity and temporal statistics, integrating them into 3D tensors to preserve fine-grained details.

Geometric models-based methods. Geometric models leverage spatial relationships within scenes to improve the performance of event-based visual algorithms. By integrating geometric constraints from event data, these models reduce error accumulation, enhancing both accuracy and computational efficiency. Recent progress has been notable in event-based visual odometry [164], motion estimation [189, 193–196], and time-to-collision estimation [197].

Probabilistic models-based methods. Probabilistic models offer a principled way to represent data and quantify uncertainty, making them well-suited for the noisy, asynchronous nature of event streams. By leveraging probabilistic inference and optimization, they improve both robustness and accuracy in event-based vision. Recent studies demonstrate their effectiveness in object classification [198], where class likelihoods are directly maximized from event data, and in optical flow estimation [199], where noise modeling and belief propagation enhance precision.

Deep learning-based methods. Deep learning has greatly advanced event-based vision by extracting complex spatio-temporal patterns for accurate prediction and decision-making. Recent efforts focus on efficiency, such as lightweight networks and sparse Transformers [200] for object detection, and deep models for dynamic obstacle avoidance [201] and tracking [202]. Advanced methods like GCNs and cross-representation distillation [203] further improve scene understanding, while point cloud-based networks enhance pose relocalization [132], and unified implicit neural representations support rolling shutter image restoration [204].

A promising direction lies in lightweight neural architectures tailored to the sparse and asynchronous nature of event streams. Rather than compressing conventional models, these approaches co-design network structures and learning principles with the sensing modality itself. IDNet [144] replaces expensive 4D correlation volumes with an iterative motion-compensation loop, where a lightweight ConvGRU progressively refines residual flow, achieving competitive accuracy with far fewer parameters and memory. FARSE-CNN [187] proposes a fully asynchronous recurrent sparse-CNN, incorporating spatio-temporal compression modules to learn hierarchical features directly from events, yielding state-of-the-art efficiency at low computational cost. Ultralight Polarity-Split SNN [205] leverages polarity-split encoding and a learnable spatio-temporal loss for event-stream super-resolution, providing low-latency inference with minimal model size and enabling on-sensor deployment.

Asynchronous and sparse computation-based acceleration. Asynchronous and sparse computation aligns with event cameras' data characteristics by updating only when changes occur, selectively processing active regions. This reduces computational cost and boosts efficiency. Typical approaches focus on pixel brightness changes or region-specific features. Recent advances include sparse convolutional networks for asynchronous streams [206], graph-based frameworks [207], and local shift operations for optimized event handling [208].

6 APPLICATION: MOBILE PLATFORM-BASED TASK

This section explores the various tasks associated with mobile platforms, emphasizing the use of event cameras and other sensing technologies to improve performance in dynamic environments. These tasks are categorized into intrinsic sensing, external sensing, and event-based SLAM, each offering distinct applications and advantages, as shown in Fig.9.

6.1 Intrinsic sensing

Vision odometry. Event-based visual odometry exploits asynchronous event streams to estimate motion with higher precision and lower computational cost than frame-based methods [209]. Early work focused on direct motion extraction, such as angular velocity estimation via contrast maximization on event edges [210] and improved attitude estimation under high rotation using enhanced aggregation functions [211]. More advanced systems integrate complementary modalities: [212] incorporates depth for robust odometry in challenging conditions, while first event-based stereo visual-inertial system [213] fuses stereo events, standard frames, and IMUs through spatiotemporal correlation and motion compensation, significantly boosting accuracy in dynamic scenes. Although these methods suggest proof-of-concept readiness, their large-scale deployment in unstructured real-world environments remains limited. Advanced methods have taken a step forward by attempting GPS-denied navigation for unmanned systems, comparing real-time terrain fingerprints generated from event camera outputs with pre-stored fingerprints derived from satellite imagery [214].

Optical flow. Event cameras are highly suitable for optical flow estimation, supporting real-time processing in HDR environments critical for motion tracking. Existing methods fall into two main categories: *(i) Deep learning-based approaches*, which leverage neural networks to infer flow from event streams—for example, a hierarchical SNN for global motion perception [215], though they often require event-to-frame preprocessing. *(ii) Traditional computer vision approaches*, which exploit the spatio-temporal structure of event data. These include normal flow estimation fused with IMU data for high-dynamic velocity estimation [216] and contrast maximization methods with tailored reward functions [143]. Event-based optical flow has shown promise in drone status estimation, and deep learning methods perform well in lab settings; however, their robustness to sensor noise and illumination variability in real-world deployments remains an open research problem.

6.2 External sensing

Mapping. Event cameras improve 3D map construction by capturing rapid environmental changes with high temporal resolution. Feature-based methods, such as extracting points or lines, remain common—for instance, [217] detects straight lines in railway settings and combines them with odometry for infrastructure mapping. However, these methods depend on stable, repeatable features, which may be unreliable in sparse or occluded scenes. To address this, recent works adopt alternative strategies: BeNeRF [218] reconstructs NeRFs from a single blurred image and an event stream, reducing reliance on extensive pose data, while AsynHDR [219] leverages LCD modulation for HDR imaging, enriching scene information for more accurate 3D reconstruction. In summary, feature-based and neural approaches show promise for 3D reconstruction in dynamic environments, but many rely on assumptions about feature stability or controlled settings, indicating they are largely at the proof-of-concept stage rather than fully deployment-ready.

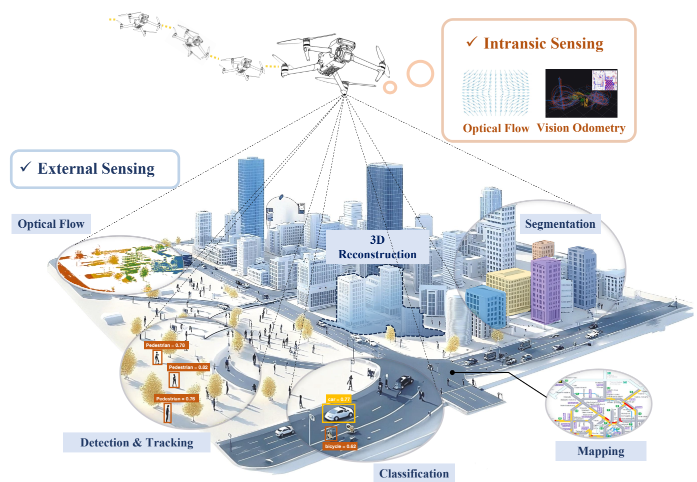


Fig. 9. Event camera-based sensing tasks on mobile platforms

Object detection & tracking. The high temporal resolution and low latency of event cameras make them particularly well-suited for precise object detection and tracking in dynamic environments [32, 220]. Existing methods can be broadly categorized into two groups: *(i) Event stream-based detection and tracking methods.* These approaches primarily leverage the spatio-temporal characteristics of event data. For example, [221] proposes a stereo event-based tracking algorithm that addresses occlusion by combining 3D reconstruction with cluster tracking. [222] proposes an end-to-end event cloud-based object tracking framework using density-insensitive key-event sampling, graph-based embedding, and motion-aware likelihood prediction. EVPropNet [223] tracks drone propellers using event data, while EDOPT [224] performs six-degree-of-freedom object pose tracking solely with event cameras. *(ii) Fusion-based detection and tracking methods.* These methods combine event data with additional sensors to enhance accuracy and robustness. [225] proposes a two-stage gaze estimation framework using event and frame data with anchor-based state shifts and denoising distillation. [226] explores object tracking from RGB and event data by leveraging a pre-trained ViT with mask modeling and orthogonal high-rank loss to enhance inter-modal token interaction. [227] proposes a motion-adaptive event sampling and bidirectional-enhanced fusion framework to align event and image data for more accurate object tracking. High-frequency drone localization using mmWave radar and event streams is demonstrated in [5], and [228] integrates depth camera data with events for obstacle tracking. In summary, some event-only methods and fusion-based approaches demonstrate impressive performance in dynamic and high-speed conditions. Fusion-based systems that combine RGB, depth, or radar data appear closest to real-world deployment, especially in drone navigation and obstacle tracking. In the industry, Meituan has explored combining event cameras with mmWave radar for UAV localization [5]. Tobii, a global leader in eye-tracking, together with Meta, has investigated event-camera-based gaze tracking [229, 230]. Meanwhile, SAAZ Micro Inc. and Neurobus have applied event cameras to drone detection [231, 232].

Optical flow. Event cameras excel in external optical flow estimation, enabling robust motion tracking of dynamic objects in high-speed scenarios [233, 234]. Unlike internal perception, which focuses on self-motion estimation, external optical flow targets the movement of surrounding objects, supporting applications such as object capture and robotic control. Existing approaches can be broadly categorized into multimodal fusion, contrast maximization, and event-specific dense flow estimation. For instance, RPEFlow [235] enhances accuracy by fusing events with RGB images and point clouds via cross-modal attention. Contrast maximization (CM)-based methods [236, 237] provide a principled formulation and achieve state-of-the-art results with innovations like multi-reference focus loss and time-aware flow modeling. Complementarily, TEGBP [199] introduces an efficient framework for deriving dense optical flow directly from sparse event data. These developments highlight the growing versatility of event cameras for external motion analysis across diverse environments. Despite these advances, most methods remain at the proof-of-concept stage, validated primarily in controlled laboratory settings. It is worth noting that this technology shows significant potential for applications such as fall detection, crowd detection and tracking, and traffic data acquisition [238].

Classification. Event cameras' ability to capture rapid changes in lighting and motion greatly improves image classification [239, 240]. Event-based classification methods leverage subtle motion and illumination variations with specialized models. For instance, [241] uses graph convolutional networks on event-derived graph structures for classification, while [242] applies deep learning to recognize individuals from gait patterns. Event cameras also aid microscopic object classification using SNNs [243], and [244] captures facial micro-expressions for emotion recognition, showcasing event data's sensitivity to subtle cues. These methods have been validated on specific datasets. While promising, their generalization to large-scale, unconstrained environments remains a challenge. It is worth noting that such approaches hold significant potential for applications in medical diagnostics [245].

3D Reconstruction. Event cameras enable precise 3D reconstruction in dynamic scenes by leveraging high temporal resolution and sparse data. Real-time monocular reconstruction is shown in [246], while [247] used stereo event cameras for semi-dense 3D reconstruction, balancing accuracy and efficiency. Combining structured light with event cameras allowed high-speed 3D scanning with less data redundancy [248]. A polarization-based method in [249] maintains precision even at low event rates. Specific applications include real-time 3D hand gesture estimation [250], 3D human pose and shape estimation [251], and non-rigid object reconstruction handling complex motions [252]. In summary, real-time monocular and semi-dense stereo reconstructions have been achieved in controlled scenarios. Non-rigid object reconstruction and human pose estimation highlight potential, but deployment in complex scenes is still limited.

Segmentation. Event cameras improve image segmentation performance by enabling accurate and efficient segmentation in rapidly changing environments [253, 254]. Segmentation approaches based on event data can be divided into: *(i) Motion-based segmentation methods* leverage the motion information captured by event cameras to segment scenes. For example, [255] proposes a motion compensation-based iterative optimization algorithm that segments scenes into independently moving objects, effectively exploiting the event camera's high sensitivity to motion changes for dynamic target segmentation. *(ii) Deep learning-based semantic segmentation methods* utilize deep neural networks for feature extraction and semantic understanding. Existing works in this category explore how to effectively combine event data with RGB frames or adapt pre-trained segmentation models to the event domain. For instance, CMDA [256] exploits HDR of event cameras to complement the limited dynamic range of frame cameras, achieving robust semantic segmentation in challenging nighttime scenes. More recently, SAM-Event-Adapter [257] introduces a lightweight adapter to bridge event data with the Segment Anything Model (SAM), enabling zero-shot semantic segmentation and demonstrating strong generalization across diverse event datasets. Similarly, [258] proposes a multi-scale feature distillation method to align embeddings from event data with RGB images, further facilitating the adaptation of SAM for robust and universal object segmentation in the event domain. Although mostly evaluated on datasets, these methods have strong potential in IoT applications, such as crowd detection and traffic data acquisition, as well as in medical contexts like high-speed particle tracking in microfluidic devices, where accurate object segmentation is essential.

6.3 Simultaneous Localization and Mapping (SLAM)

Event-based SLAM leverages the high temporal resolution and low latency of event cameras for robust localization and mapping in dynamic environments [259, 260]. A key enhancement is multimodal fusion, combining complementary sensors to overcome individual limitations. For instance, Ultimate SLAM [261] fuses events, frames, and IMU data for HDR and high-speed scenarios, while Implicit Event-RGBD Neural SLAM [260] integrates event and RGB-D data to handle motion blur and lighting changes. Advances in feature representation further boost robustness. Line-based SLAM methods [158] mitigate feature loss in fast motion or low-light using PTAM frameworks, while optimization-based approaches like CMax-SLAM [262] apply contrast maximization for precise rotational motion estimation through event-based global bundle adjustment. Despite these promising results, most event-based SLAM methods remain at the proof-of-concept stage, validated mainly in controlled or semi-controlled scenarios. Only a few methods have been demonstrated in real-world dynamic environments [259, 263], indicating that full deployment readiness is still limited.

7 FUTURE DIRECTION AND DISCUSSION

Despite extensive research on event cameras, their application in mobile sensing remains in its early stages, with significant opportunities for advancement while balancing trade-offs in latency, power consumption, and accuracy.

(1) Improving event cameras using optics devices. Event cameras respond only to illumination changes and remain inactive in fully static scenes, limiting continuous perception. While some hardware-based solutions exist, they often reduce energy efficiency due to added mechanical components [35, 264]. Future work could explore dynamic optical elements, such as electro-optic materials, to induce controlled illumination changes via optical phase arrays, enabling detection in static environments. Hybrid optical-electronic systems, combining event cameras with active illumination or computational imaging, may further enhance performance [249, 265]. Key challenges remain, including designing diverse scanning patterns without blind spots, developing easily implementable non-mechanical illumination devices, and improving event signal quality under high-speed illumination.

(2) Designing neuromorphic hardware to facilitate event processing. Current hardware faces challenges in processing event-based data. CPUs suffer from frequent context switching, while GPUs are ill-suited for asynchronous, high-frequency events. FPGAs provide parallelism and low latency but lack end-to-end pipeline optimization. Although some efforts have developed dedicated accelerators [39, 171, 174], the massive data volume in event stream (e.g., thousands of events in milliseconds) can easily overwhelm I/O bandwidth, and end-to-end latency often violates worst-case budgets under event bursts, causing jitter that disrupts control loops. Moreover, mapping high-level neural models to neuromorphic ISAs remains manual and brittle, limiting programmability and debugging. Future work should pursue specialized neuromorphic hardware tailored for event-driven computing. Architectures based on SNNs or asynchronous pipelines can better exploit event sparsity, while higher-throughput interconnects with QoS, priority channels for control streams, and traffic shaping can alleviate I/O bottlenecks. Hard real-time scheduling, bounded-latency NoCs, and deadline-aware routers are needed to guarantee temporal determinism. Finally, stable intermediate representations (IRs) and compiler toolchains are essential to improve programmability across diverse neuromorphic platforms.

(3) Leveraging the complementary strengths of event camera and other sensors. Event cameras complement frame cameras, radar, and LiDAR by providing high temporal resolution and low latency, which mitigate motion blur and enhance perception in dynamic environments. While frame cameras offer texture and color [32, 266], radar ensures robustness under poor lighting and weather [5, 267], and LiDAR supplies precise 3D geometry [268, 269], their integration with event streams enables improved visual odometry, detection, reconstruction, and SLAM. Yet, most existing fusion methods remain loosely coupled and overlook the intrinsic properties of raw multimodal data—for instance, frame images are spatially dense but temporally sparse, event and radar data are spatially sparse yet temporally dense, and LiDAR point clouds are sparse in both space and time. Future research should therefore develop tightly coupled deep learning– or optimization-based frameworks that explicitly exploit these raw data complementary characteristics, alongside optimized hardware designs for real-time, energy-efficient robotic applications.

(4) Bio-inspired algorithm design. Event cameras inherently exhibit neuromorphic traits, making them well-suited for bio-inspired algorithms. Integrating SNNs, which process discrete spikes like the brain, enables sparse, low-power, and precise event-driven perception for tasks such as recognition and tracking [270]. Bio-inspired models based on primate vision improve segmentation [39], while neuromorphic control systems emulate sensorimotor loops for fast decisions [271]. Nevertheless, most current innovations primarily focus on mimicking the biological mechanisms, developing sophisticated algorithms or hardware for high-level recognition and scene understanding. While inspiring, such approaches may not be optimal for tasks requiring strict real-time performance due to their significant computational overhead, or in some cases, may be impractical to realize. Future research should therefore explore alternative nature-inspired strategies—rather than strictly replicating biological mechanisms—by designing bio-inspired models that enable efficient event processing, and by optimizing neuromorphic hardware to effectively support these systems.

(5) Hardware-software co-optimization techniques. Efficient event data processing requires tight hardware-software co-design. Traditional hardware struggles with asynchronous, sparse event streams. Dedicated neuromorphic hardware—such as SNN chips, FPGA-based processors, and ASICs—can better support event-driven computation, improving efficiency [39, 174]. On the software side, event-driven programming models and lightweight architectures are needed to reduce latency and redundancy, as conventional deep learning frameworks are not well-suited for event data [32]. However, most existing approaches focus on either hardware or software in isolation, overlooking cross-layer optimizations. Techniques such as dynamic resource allocation, memory access optimization, and hardware-aware model compression could further improve speed and energy efficiency. Future research should therefore emphasize holistic co-design of algorithms and neuromorphic hardware to realize real-time, low-power event-based vision systems for robotics, autonomous driving, and edge intelligence.

8 CONCLUSION

Event-based vision is a transformative approach for mobile sensing, offering high temporal resolution, low latency, and energy efficiency. This survey reviews event camera principles, event representations, algorithms, hardware/software acceleration, and diverse mobile applications. Despite advantages, challenges remain in event processing, sensor fusion, and real-time use on resource-limited platforms. Future work should enhance hardware with advanced optics, develop neuromorphic processors for asynchronous data, and apply bio-inspired algorithms to boost perception. Integrating event cameras with LiDAR and radar will further broaden applications in dynamic settings. We hope this survey inspires research and practical deployment in the mobile sensing field.

REFERENCES

[1] Yunfan Ren, Fangcheng Zhu, Guozheng Lu, Yixi Cai, Longji Yin, Fanze Kong, Jiarong Lin, Nan Chen, and Fu Zhang. Safety-assured high-speed navigation for mavs. *Science Robotics*, 10(98):eado6187, 2025.

[2] Xuecheng Chen*, Haoyang Wang*, Yuhan Cheng, Haohao Fu, Yuxuan Liu, Fan Dang, Yunhao Liu, Jinqiang Cui, and Xinlei Chen. Ddl: Empowering delivery drones with large-scale urban sensing capability. *IEEE Journal of Selected Topics in Signal Processing*, 2024.

[3] Xin Zhou, Xiangyong Wen, Zhepei Wang, Yuman Gao, Haojia Li, Qianhao Wang, Tiankai Yang, Haojian Lu, Yanjun Cao, Chao Xu, et al. Swarm of micro flying robots in the wild. *Science Robotics*, 7(66):eabm5954, 2022.

[4] Zhuozhu Jian, Zejia Liu, Haoyu Shao, Xueqian Wang, Xinlei Chen, and Bin Liang. Path generation for wheeled robots autonomous navigation on vegetated terrain. *IEEE RA-L*, 9(2):1764–1771, 2023.

[5] Haoyang Wang, Jingao Xu, Xinyu Luo, Xuecheng Chen, Ting Zhang, Ruiyang Duan, Yunhao Liu, and Xinlei Chen. Ultra-high-frequency harmony: mmwave radar and event camera orchestrate accurate drone landing. In *Proceedings of the 23th ACM SenSys*, 2025.

[6] Haoyang Wang, Jingao Xu, Chenyu Zhao, Zihong Lu, Yuhan Cheng, Xuecheng Chen, Xiao-Ping Zhang, Yunhao Liu, and Xinlei Chen. Transformloc: Transforming mavs into mobile localization infrastructures in heterogeneous swarms. *Proceedings of the IEEE INFOCOM*, 2024.

[7] Jingao Xu, Hao Cao, Zheng Yang, Longfei Shangguan, Jialin Zhang, Xiaowu He, and Yunhao Liu. {SwarmMap}: Scaling up real-time collaborative visual {SLAM} at the edge. In *19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22)*, pages 977–993, 2022.

[8] Yuxuan Liu, Haoyang Wang, Fanhang Man, Jingao Xu, Fan Dang, Yunhao Liu, Xiao-Ping Zhang, and Xinlei Chen. Mobiair: Unleashing sensor mobility for city-scale and fine-grained air-quality monitoring with airbert. In *Proceedings of the ACM MobiSys*, 2024.

[9] Xuecheng Chen, Zijian Xiao, Yuhan Cheng, ChenChun Hsia, Haoyang Wang, Jingao Xu, Susu Xu, Fan Dang, Xiao-Ping Zhang, Yunhao Liu, et al. Soscheduler: Toward proactive and adaptive wildfire suppression via multi-uav collaborative scheduling. *IEEE Internet of Things Journal*, 2024.

[10] Jiaxu Leng, Yongming Ye, Mengjingcheng Mo, Chenqiang Gao, Ji Gan, Bin Xiao, and Xinbo Gao. Recent advances for aerial object detection: A survey. *ACM Computing Surveys*, 56(12):1–36, 2024.

[11] Yanggang Xu, Jirong Zha, Jiyuan Ren, Xintao Jiang, Hongfei Zhang, and Xinlei Chen. Scalable multi-agent reinforcement learning for effective uav scheduling in multi-hop emergency networks. In *Proceedings of the 30th ACM MobiCom*, pages 2028–2033, 2024.

[12] Haoyang Wang, Jingao Xu, Chenyu Zhao, Yuhan Cheng, Xuecheng Chen, Chaopeng Hong, Xiao-Ping Zhang, Yunhao Liu, and Xinlei Chen. Aerial shepherds: Enabling hierarchical localization in heterogeneous mav swarms. *arXiv preprint arXiv:2506.08408*, 2025.

[13] Nan Chen, Fanze Kong, Wei Xu, Yixi Cai, Haotian Li, Dongjiao He, Youming Qin, and Fu Zhang. A self-rotating, single-actuated uav with extended sensor field of view for autonomous navigation. *Science Robotics*, 8(76):eade4538, 2023.

[14] Haoyang Wang*, Xuecheng Chen*, Yuhan Cheng, Chenye Wu, Fan Dang, and Xinlei Chen. H-swarmloc: efficient scheduling for localization of heterogeneous mav swarm with deep reinforcement learning. In *Proceedings of the 20th ACM SenSys CML-IoT*, pages 1148–1154. 2022.

[15] Jiahe Cui, Yuze He, Jianwei Niu, Zhenchao Ouyang, and Guoliang Xing. αlidar: An adaptive high-resolution panoramic lidar system. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 1515–1529, 2024.

[16] Yuze He, Li Ma, Jiahe Cui, Zhenyu Yan, Guoliang Xing, Sen Wang, Qintao Hu, and Chen Pan. Automatch: Leveraging traffic camera to improve perception and localization of autonomous vehicles. In *Proceedings of the 20th ACM SenSys*, pages 16–30, 2022.

[17] Adam Francis, Shuai Li, Christian Griffiths, and Johann Sienz. Gas source localization and mapping with mobile robots: A review. *Journal of Field Robotics*, 39(8):1341–1373, 2022.

[18] Dario Floreano and Robert J Wood. Science, technology and the future of small autonomous drones. *nature*, 521(7553):460–466, 2015.

[19] Dji mavic 4. [Online]. https://enterprise.dji.com/matrice-4-series?site=enterprise&from=nav.

[20] Wing's drone. [Online]. https://wing.com/technology.

[21] Fangcheng Zhu, Yunfan Ren, Longji Yin, Fanze Kong, Qingbo Liu, Ruize Xue, Wenyi Liu, Yixi Cai, Guozheng Lu, Haotian Li, et al. Swarm-lio2: Decentralized, efficient lidar-inertial odometry for uav swarms. *IEEE Transactions on Robotics*, 2024.

[22] Emerson Sie, Xinyu Wu, Heyu Guo, and Deepak Vasisht. Radarize: Enhancing radar slam with generalizable doppler-based odometry. In *The 22nd ACM International Conference on Mobile Systems, Applications, and Services (ACM MobiSys '24)*, 2024.

[23] Jialin Zhang, Xiang Huang, Jingao Xu, Yue Wu, Qiang Ma, Xin Miao, Li Zhang, Pengpeng Chen, and Zheng Yang. Edge assisted real-time instance segmentation on mobile devices. In *Proceedings of the IEEE ICDCS*, pages 537–547, 2022.

[24] Hao Cao, Jingao Xu, Danyang Li, Longfei Shangguan, Yunhao Liu, and Zheng Yang. Edge assisted mobile semantic visual slam. *IEEE TMC*, 2023.

[25] Tong Qin, Jie Pan, Shaozu Cao, and Shaojie Shen. A general optimization-based framework for local odometry estimation with multiple sensors, 2019.

[26] Danyang Li, Yishujie Zhao, Jingao Xu, Shengkai Zhang, Longfei Shangguan, and Zheng Yang. Reshaping edge-assisted visual slam by embracing on-chip intelligence. *IEEE Transactions on Mobile Computing*, 2024.

[27] 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.

[28] Xu Zheng, Yexin Liu, Yunfan Lu, Tongyan Hua, Tianbo Pan, Weiming Zhang, Dacheng Tao, and Lin Wang. Deep learning for event-based vision: A comprehensive survey and benchmarks. *arXiv preprint arXiv:2302.08890*, 2023.

[29] Bharatesh Chakravarthi, Aayush Atul Verma, Kostas Daniilidis, Cornelia Fermuller, and Yezhou Yang. Recent event camera innovations: A survey. *arXiv preprint arXiv:2408.13627*, 2024.

[30] Alvaro Novo, Francisco Lobon, Hector Garcia de Marina, Samuel Romero, and Francisco Barranco. Neuromorphic perception and navigation for mobile robots: a review. *ACM Computing Surveys*, 56(10):1–37, 2024.

[31] Waseem Shariff, Mehdi Sefidgar Dilmaghani, Paul Kielty, Mohamed Moustafa, Joe Lemley, and Peter Corcoran. Event cameras in automotive sensing: A review. *IEEE Access*, 2024.

[32] Daniel Gehrig and Davide Scaramuzza. Low-latency automotive vision with event cameras. *Nature*, 629(8014):1034–1040, 2024.

[33] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):1964–1980, 2019.

[34] Davide Falanga, Kevin Kleber, and Davide Scaramuzza. Dynamic obstacle avoidance for quadrotors with event cameras. *Science Robotics*, 5(40):9712, 2020.

[35] Botao He, Ze Wang, Yuan Zhou, Jingxi Chen, Chahat Deep Singh, Haojia Li, Yuman Gao, Shaojie Shen, Kaiwei Wang, Yanjun Cao, et al. Microsaccade-inspired event camera for robotics. *Science Robotics*, 9(90):eadj8124, 2024.

[36] Yunhao Zou, Ying Fu, Tsuyoshi Takatani, and Yinqiang Zheng. Eventhdr: From event to high-speed hdr videos and beyond. *IEEE T-PAMI*, 2024.

[37] Cedric Scheerlinck, Henri Rebecq, Daniel Gehrig, Nick Barnes, Robert Mahony, and Davide Scaramuzza. Fast image reconstruction with an event camera. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 156–163, 2020.

[38] G.gallego, recent papers on event-based vision. [Online]. https://docs.google.com/spreadsheets/d/1_OBbSz10CkxXNDHQd-Mn_ui3OmymMFvm-lW316uvxy8/edit?pli=1&gid=0#gid=0.

[39] Jingao Xu, Danyang Li, Zheng Yang, Yishujie Zhao, Hao Cao, Yunhao Liu, and Longfei Shangguan. Taming event cameras with bio-inspired architecture and algorithm: A case for drone obstacle avoidance. In *Proceedings of the 29th ACM MobiCom*, pages 1–16, 2023.

[40] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128 x 128 120 db 15us latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits*, 43(2):566–576, 2008.

[41] Christoph Posch, Daniel Matolin, and Rainer Wohlgenannt. A qvga 143 db dynamic range frame-free pwm image sensor with lossless pixel-level video compression and time-domain cds. *IEEE Journal of Solid-State Circuits*, 46(1):259–275, 2010.

[42] Misha Mahowald. The silicon retina. In *An Analog VLSI System for Stereoscopic Vision*, pages 4–65. Springer, 1994.

[43] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240× 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341, 2014.

[44] Shoushun Chen and Menghan Guo. Live demonstration: Celex-v: A 1m pixel multi-mode event-based sensor. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1682–1683. IEEE, 2019.

[45] Celepixel celex5-mipi-stereo sdk github repository. [Online]. https://github.com/CelePixel/CeleX5-MIPI-Stereo.

[46] Prophesee genx320 metavision sensor product page (china). [Online]. https://www.prophesee-cn.com/event-based-sensor-genx320/.

[47] inivation. [Online]. https://inivation.com/.

[48] prophesee. [Online]. https://www.prophesee-cn.com/.

[49] Lucid vision labs - official website (china). [Online]. http://thinklucid.cn/.

[50] Alex Zihao Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. *IEEE Robotics and Automation Letters*, 3(3):2032–2039, 2018.

[51] Mustafa Sakhai, Szymon Mazurek, Jakub Caputa, Jan K. Argasiński, and Maciej Wielgosz. Pedestrian intention prediction in adverse weather conditions with spiking neural networks and dynamic vision sensors, 2024.

[52] Yuhuang Hu, Jonathan Binas, Daniel Neil, Shih-Chii Liu, and Tobi Delbruck. Ddd20 end-to-end event camera driving dataset: Fusing frames and events with deep learning for improved steering prediction. In *2020 IEEE 23rd ITSC*, pages 1–6, 2020.

[53] Mathias Gehrig, Willem Aarents, Daniel Gehrig, and Davide Scaramuzza. Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, 6(3):4947–4954, 2021.

[54] Simon Klenk, Jason Chui, Nikolaus Demmel, and Daniel Cremers. Tum-vie: The tum stereo visual-inertial event dataset. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8601–8608, 2021.

[55] Ling Gao, Yuxuan Liang, Jiaqi Yang, Shaoxun Wu, Chenyu Wang, Jiaben Chen, and Laurent Kneip. Vector: A versatile event-centric benchmark for multi-sensor slam. *IEEE Robotics and Automation Letters*, 7(3):8217–8224, July 2022.

[56] Etienne Perot, Pierre De Tournemire, Davide Nitti, Jonathan Masci, and Amos Sironi. Learning to detect objects with a 1 megapixel event camera. *Advances in Neural Information Processing Systems*, 33:16639–16652, 2020.

[57] Jiqing Zhang, Xin Yang, Yingkai Fu, Xiaopeng Wei, Baocai Yin, and Bo Dong. Object tracking by jointly exploiting frame and event domain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13043–13052, 2021.

[58] Aayush Atul Verma, Bharatesh Chakravarthi, Arpitsinh Vaghela, Hua Wei, and Yezhou Yang. etram: Event-based traffic monitoring dataset, 2024.

[59] Hebei Li, Jin Wang, Jiahui Yuan, Yue Li, Wenming Weng, Yansong Peng, Yueyi Zhang, Zhiwei Xiong, and Xiaoyan Sun. Event-assisted low-light video object segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3250–3259, 2024.

[60] Taewoo Kim, Hoonhee Cho, and Kuk-Jin Yoon. Frequency-aware event-based video deblurring for real-world motion blur. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24966–24976, 2024.

[61] Guoqiang Liang, Kanghao Chen, Hangyu Li, Yunfan Lu, and Lin Wang. Towards robust event-guided low-light image enhancement: a large-scale real-world event-image dataset and novel approach. In *Proceedings of the IEEE/CVF CVPR*, pages 23–33, 2024.

[62] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. *The International journal of robotics research*, 36(2):142–149, 2017.

[63] Wenbin Li, Sajad Saeedi, John McCormac, Ronald Clark, Dimos Tzoumanikas, Qing Ye, Yuzhong Huang, Rui Tang, and Stefan Leutenegger. Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset. *arXiv preprint arXiv:1809.00716*, 2018.

[64] Jacques Kaiser, J Camilo Vasquez Tieck, Christian Hubschneider, Peter Wolf, Michael Weber, Michael Hoff, Alexander Friedrich, Konrad Wojtasik, Arne Roennau, Ralf Kohlhaas, et al. Towards a framework for end-to-end control of a simulated vehicle with spiking neural networks. In *2016 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR)*, pages 127–134. IEEE, 2016.

[65] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic dvs events. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1312–1321, 2021.

[66] Zhongyang Zhang, Shuyang Cui, Kaidong Chai, Haowen Yu, Subhasis Dasgupta, Upal Mahbub, and Tauhidur Rahman. V2ce: Video to continuous events simulator. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 12455–12461. IEEE, 2024.

[67] Video to Event Simulator. [Online]. https://docs.prophesee.ai/stable/samples/modules/core_ml/viz_video_to_event_simulator.html.

[68] Anish Bhattacharya, Ratnesh Madaan, Fernando Cladera, Sai Vemprala, Rogerio Bonatti, Kostas Daniilidis, Ashish Kapoor, Vijay Kumar, Nikolai Matni, and Jayesh K Gupta. Evdnerf: Reconstructing event data with dynamic neural radiance fields. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5846–5855, 2024.

[69] Joachim Ott, Zuowen Wang, and Shih-Chii Liu. Text-to-events: Synthetic event camera streams from conditional text input. In *2024 Neuro Inspired Computational Elements Conference (NICE)*, pages 1–10. IEEE, 2024.

[70] Joseph L Greene, Adrish Kar, Ignacio Galindo, Elijah Quiles, Elliott Chen, and Matthew Anderson. A pytorch-enabled tool for synthetic event camera data generation and algorithm development. In *Synthetic Data for Artificial Intelligence and Machine Learning: Tools, Techniques, and Applications III*, volume 13459, pages 117–137. SPIE, 2025.

[71] Kaustav Chanda, Aayush Verma, Arpitsinh Vaghela, Yezhou Yang, and Bharatesh Chakravarthi. Event quality score (eqs): Assessing the realism of simulated event camera streams via distance in latent space. In *Proceedings of the CVPR*, pages 5105–5113, 2025.

[72] Ziming Wang, Ziling Wang, Huaning Li, Lang Qin, Runhao Jiang, De Ma, and Huajin Tang. Eas-snn: End-to-end adaptive sampling and representation for event-based detection with recurrent spiking neural networks. In *ECCV*, pages 310–328. Springer, 2025.

[73] Shuang Guo and Guillermo Gallego. Event-based mosaicing bundle adjustment. In *ECCV*, pages 479–496. Springer, 2025.

[74] Songnan Lin, Ye Ma, Jing Chen, and Bihan Wen. Compressed event sensing (ces) volumes for event cameras. *International Journal of Computer Vision*, pages 1–21, 2024.

[75] Nealson Li, Muya Chang, and Arijit Raychowdhury. E-gaze: Gaze estimation with event camera. *IEEE T-PAMI*, 2024.

[76] Jianxiong Tang, Jian-Huang Lai, Lingxiao Yang, and Xiaohua Xie. Spike-temporal latent representation for energy-efficient event-to-video reconstruction. In *European Conference on Computer Vision*, pages 163–179. Springer, 2025.

[77] Haram Kim, Sangil Lee, Junha Kim, and H Jin Kim. Real-time hetero-stereo matching for event and frame camera with aligned events using maximum shift distance. *IEEE Robotics and Automation Letters*, 8(1):416–423, 2022.

[78] Bochen Xie, Yongjian Deng, Zhanpeng Shao, and Youfu Li. Eisnet: A multi-modal fusion network for semantic segmentation with events and images. *IEEE Transactions on Multimedia*, 26:8639–8650, 2024.

[79] Yi-Fan Zuo, Wanting Xu, Xia Wang, Yifu Wang, and Laurent Kneip. Cross-modal semidense 6-dof tracking of an event camera in challenging conditions. *IEEE Transactions on Robotics*, 40:1600–1616, 2024.

[80] Antoine Grimaldi, Victor Boutin, Sio-Hoi Ieng, Ryad Benosman, and Laurent U Perrinet. A robust event-driven approach to always-on object recognition. *Neural Networks*, page 106415, 2024.

[81] Xueyan Huang, Yueyi Zhang, and Zhiwei Xiong. Progressive spatio-temporal alignment for efficient event-based motion estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1537–1546, 2023.

[82] Mathias Gehrig, Manasi Muglikar, and Davide Scaramuzza. Dense continuous-time optical flow from event cameras. *IEEE T-PAMI*, 2024.

[83] Hu Cao, Zehua Zhang, Yan Xia, Xinyi Li, Jiahao Xia, Guang Chen, and Alois Knoll. Embracing events and frames with hierarchical feature refinement network for object detection. In *European Conference on Computer Vision*, pages 161–177. Springer, 2025.

[84] Haoran Xu, Peixi Peng, Guang Tan, Yuan Li, Xinhai Xu, and Yonghong Tian. Dmr: Decomposed multi-modality representations for frames and events fusion in visual reinforcement learning. In *Proceedings of the IEEE/CVF CVPR*, pages 26508–26518, 2024.

[85] Ziwei Wang, Timothy Molloy, Pieter Van Goor, and Robert Mahony. Asynchronous blob tracker for event cameras. *IEEE T-RO*, 2024.

[86] Zhiwen Chen, Jinjian Wu, Junhui Hou, Leida Li, Weisheng Dong, and Guangming Shi. Ecsnet: Spatio-temporal feature learning for event camera. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(2):701–712, 2022.

[87] Daniel Czech and Garrick Orchard. Evaluating noise filtering for event-based asynchronous change detection image sensors. In *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 19–24. IEEE, 2016.

[88] Xavier Lagorce, Garrick Orchard, Francesco Galluppi, Bertram E Shi, and Ryad B Benosman. Hots: a hierarchy of event-based time-surfaces for pattern recognition. *IEEE transactions on pattern analysis and machine intelligence*, 39(7):1346–1359, 2016.

[89] Alireza Khodamoradi and Ryan Kastner. $o(n)$ o (n)-space spatiotemporal filter for reducing noise in neuromorphic vision sensors. *IEEE Transactions on Emerging Topics in Computing*, 9(1):15–23, 2018.

[90] Shasha Guo and Tobi Delbruck. Low cost and latency event camera background activity denoising. *IEEE T-PAMI*, 45(1):785–795, 2022.

[91] Yanxiang Wang, Bowen Du, Yiran Shen, Kai Wu, Guangrong Zhao, Jianguo Sun, and Hongkai Wen. Ev-gait: Event-based robust gait recognition using dynamic vision sensors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6358–6367, 2019.

[92] R Baldwin, Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. Event probability mask (epm) and event denoising convolutional neural network (edncnn) for neuromorphic cameras. In *Proceedings of the IEEE/CVF CVPR*, pages 1701–1710, 2020.

[93] Huachen Fang, Jinjian Wu, Leida Li, Junhui Hou, Weisheng Dong, and Guangming Shi. Aednet: Asynchronous event denoising with spatial-temporal correlation among irregular data. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1427–1435, 2022.

[94] Yuxing Duan. Led: A large-scale real-world paired dataset for event camera denoising. In *Proceedings of the IEEE CVPR*, pages 25637–25647, 2024.

[95] Yang Feng, Hengyi Lv, Hailong Liu, Yisa Zhang, Yuyao Xiao, and Chengshan Han. Event density based denoising method for dynamic vision sensor. *Applied Sciences*, 10(6):2024, 2020.

[96] Tobi Delbruck et al. Frame-free dynamic digital vision. In *Proceedings of Intl. Symp. on Secure-Life Electronics, Advanced Electronics for Quality Life and Society*, volume 1, pages 21–26. Citeseer, 2008.

[97] R Wes Baldwin, Mohammed Almatrafi, Jason R Kaufman, Vijayan Asari, and Keigo Hirakawa. Inceptive event time-surfaces for object classification using neuromorphic cameras. In *Image Analysis and Recognition: 16th International Conference, ICIAR 2019*, pages 395–403.

[98] Sio-Hoi Ieng, Christoph Posch, and Ryad Benosman. Asynchronous neuromorphic event-driven image filtering. *Proceedings of the IEEE*, 102(10):1485–1499, 2014.

[99] Shasha Guo, Ziyang Kang, Lei Wang, Limeng Zhang, Xiaofan Chen, Shiming Li, and Weixia Xu. Hashheat: A hashing-based spatiotemporal filter for dynamic vision sensor. *Integration*, 81:99–107, 2021.

[100] Hongjie Liu, Christian Brandli, Chenghan Li, Shih-Chii Liu, and Tobi Delbruck. Design of a spatiotemporal correlation filter for event-based sensors. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 722–725. IEEE, 2015.

[101] Peiqi Duan, Zihao W Wang, Boxin Shi, Oliver Cossairt, Tiejun Huang, and Aggelos K Katsaggelos. Guided event filtering: Synergy between intensity images and neuromorphic events for high performance imaging. *IEEE T-PAMI*, 44(11):8261–8275, 2021.

[102] Ciyu Ruan, Ruishan Guo, Zihang Gong, Jingao Xu, Wenhan Yang, and Xinlei Chen. Pre-mamba: A 4d state space model for ultra-high-frequent event camera deraining. *arXiv preprint arXiv:2505.05307*, 2025.

[103] Ciyu Ruan, Chenyu Zhao, Chenxin Liang, Xinyu Luo, Jingao Xu, and Xinlei Chen. Distill drops into data: Event-based rain-background decomposition network. In *Proceedings of the 30th ACM MobiCom*, pages 2072–2077, 2024.

[104] Xuemei Xie, Jiang Du, Guangming Shi, Jianxiu Yang, Wan Liu, and Wang Li. Dvs image noise removal using k-svd method. In *Ninth International Conference on Graphic and Image Processing (ICGIP 2017)*, volume 10615, pages 1099–1107. SPIE, 2018.

[105] Peiqi Duan, Zihao W Wang, Xinyu Zhou, Yi Ma, and Boxin Shi. Eventzoom: Learning to denoise and super resolve neuromorphic events. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12824–12833, 2021.

[106] Bin Jiang, Bo Xiong, Bohan Qu, M Salman Asif, You Zhou, and Zhan Ma. Edformer: Transformer-based event denoising across varied noise levels. In *European Conference on Computer Vision*, pages 200–216. Springer, 2024.

[107] Ciyu Ruan, Zihang Gong, Ruishan Guo, Jingao Xu, and Xinlei Chen. Edmamba: A simple yet effective event denoising method with state space model. *arXiv preprint arXiv:2505.05391*, 2025.

[108] Chris Harris, Mike Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.

[109] Valentina Vasco, Arren Glover, and Chiara Bartolozzi. Fast event-based harris corner detection exploiting the advantages of event-driven cameras. In *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 4144–4149. IEEE, 2016.

[110] Ryad Benosman, Charles Clercq, Xavier Lagorce, Sio-Hoi Ieng, and Chiara Bartolozzi. Event-based visual flow. *IEEE T-NNLS*, 25(2):407–417, 2013.

[111] Ignacio Alzugaray and Margarita Chli. Ace: An efficient asynchronous corner tracker for event cameras. In *2018 International Conference on 3D Vision (3DV)*, pages 653–661, 2018.

[112] Sumin Hu, Yeeun Kim, Hyungtae Lim, Alex Junho Lee, and Hyun Myung. ecdt: Event clustering for simultaneous feature detection and tracking. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3808–3815. IEEE, 2022.

[113] Yuan Gao, Yuqing Zhu, Xinjun Li, Yimin Du, and Tianzhu Zhang. Sd2event: Self-supervised learning of dynamic detectors and contextual descriptors for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3055–3064, 2024.

[114] Daniel Gehrig, Henri Rebecq, Guillermo Gallego, and Davide Scaramuzza. Eklt: Asynchronous photometric feature tracking using events and frames. *International Journal of Computer Vision*, 128(3):601–618, 2020.

[115] Hochang Seok and Jongwoo Lim. Robust feature tracking in dvs event stream using bézier mapping. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1658–1667, 2020.

[116] Jason Chui, Simon Klenk, and Daniel Cremers. Event-based feature tracking in continuous time with sliding window optimization. *arXiv preprint arXiv:2107.04536*, 2021.

[117] Zexiang Yi, Jing Lian, Yunliang Qi, Zhaofei Yu, Huajin Tang, Yide Ma, and Jizhao Liu. Deep pulse-coupled neural networks. *arXiv preprint arXiv:2401.08649*, 2023.

[118] Zengyu Wan, Yang Wang, Zhai Wei, Ganchao Tan, Yang Cao, and Zheng-Jun Zha. Event-based optical flow via transforming into motion-dependent view. *IEEE Transactions on Image Processing*, 2024.

[119] Elias Mueggler, Chiara Bartolozzi, and Davide Scaramuzza. Fast event-based corner detection. 2017.

[120] Ignacio Alzugaray and Margarita Chli. Haste: multi-hypothesis asynchronous speeded-up tracking of events. In *31st British Machine Vision Virtual Conference (BMVC 2020)*, page 744. ETH Zurich, Institute of Robotics and Intelligent Systems, 2020.

[121] Philippe Chiberre, Etienne Perot, Amos Sironi, and Vincent Lepetit. Long-lived accurate keypoints in event streams. *arXiv preprint arXiv:2209.10385*, 2022.

[122] Ruoxiang Li, Dianxi Shi, Yongjun Zhang, Kaiyue Li, and Ruihao Li. Fa-harris: A fast and asynchronous corner detector for event cameras. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6223–6229, 2019.

[123] Sherif AS Mohamed, Jawad N Yasin, Mohammad-hashem Haghbayan, Antonio Miele, Jukka Heikkonen, Hannu Tenhunen, and Juha Plosila. Dynamic resource-aware corner detection for bio-inspired vision sensors. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 10465–10472. IEEE, 2021.

[124] Shane Harrigan, Sonya Coleman, D Ker, Pratheepan Yogarajah, Zheng Fang, and Chengdong Wu. Rot-harris: A dynamic approach to asynchronous interest point detection. In *2021 17th International Conference on Machine Vision and Applications (MVA)*, pages 1–6. IEEE, 2021.

[125] Saeed Afshar, Nicholas Ralph, Ying Xu, Jonathan Tapson, André van Schaik, and Gregory Cohen. Event-based feature extraction using adaptive selection thresholds. *Sensors*, 20(6):1600, 2020.

[126] Sami Barchid, José Mennesson, Jason Eshraghian, Chaabane Djéraba, and Mohammed Bennamoun. Spiking neural networks for frame-based and event-based single object localization. *Neurocomputing*, 559:126805, 2023.

[127] Man Yao, Guangshe Zhao, Hengyu Zhang, Yifan Hu, Lei Deng, Yonghong Tian, Bo Xu, and Guoqi Li. Attention spiking neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 45(8):9393–9410, 2023.

[128] Paul J Best. A method for registration of 3-d shapes. *IEEE Trans Pattern Anal Mach Vision*, 14:239–256, 1992.

[129] Alex Zihao Zhu, Nikolay Atanasov, and Kostas Daniilidis. Event-based feature tracking with probabilistic data association. In *IEEE ICRA*, pages 4465–4470. IEEE, 2017.

[130] Juan Pablo Rodríguez-Gomez, A Gómez Eguíluz, José Ramiro Martínez-de Dios, and Anibal Ollero. Asynchronous event-based clustering and tracking for intrusion monitoring in uas. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8518–8524. IEEE, 2020.

[131] Siqi Li, Zhikuan Zhou, Zhou Xue, Yipeng Li, Shaoyi Du, and Yue Gao. 3d feature tracking via event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18974–18983, June 2024.

[132] Hongwei Ren, Jiadong Zhu, Yue Zhou, Haotian FU, Yulong Huang, and Bojun Cheng. A simple and effective point-based network for event camera 6-dofs pose relocalization, 2024.

[133] Min Seok Lee, Jae Hyung Jung, Ye Jun Kim, and Chan Gook Park. Event-and frame-based visual-inertial odometry with adaptive filtering based on 8-dof warping uncertainty. *IEEE Robotics and Automation Letters*, 9(2):1003–1010, 2024.

[134] Kuanxu Hou, Delei Kong, Junjie Jiang, Hao Zhuang, Xinjie Huang, and Zheng Fang. Fe-fusion-vpr: Attention-based multi-scale network architecture for visual place recognition by fusing frames and events, 2022.

[135] Michael Calonder, Vincent Lepetit, and Pascal Fua. Keypoint signatures for fast learning and recognition. In *ECCV*, pages 58–71. Springer, 2008.

[136] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[137] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *IEEE ICCV*, pages 2564–2571, 2011.

[138] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *IEEE ECCV*, pages 778–792. Springer, 2010.

[139] Xiangyuan Wang, Huai Yu, Lei Yu, Wen Yang, and Gui-Song Xia. Toward robust keypoint detection and tracking: A fusion approach with event-aligned image features. *IEEE Robotics and Automation Letters*, 9(9):8059–8066, 2024.

[140] Xiangyuan Wang, Kuangyi Chen, Wen Yang, Lei Yu, Yannan Xing, and Huai Yu. Fe-detr: Keypoint detection and tracking in low-quality image frames with events, 2024.

[141] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *IEEE/CVF CVPR*, page 3867–3876. IEEE, June 2018.

[142] Daqi Liu, Alvaro Parra, and Tat-Jun Chin. Globally optimal contrast maximisation for event-based motion estimation. In *IEEE/CVF CVPR*, pages 6348–6357, 2020.

[143] Timo Stoffregen and Lindsay Kleeman. Event cameras, contrast maximization and reward functions: An analysis. *IEEE/CVF CVPR*, pages 12292–12300, 2019.

[144] Yilun Wu, Federico Paredes-Vallés, and Guido C. H. E. de Croon. Lightweight event-based optical flow estimation via iterative deblurring, 2024.

[145] Yutian Chen, Shi Guo, Fangzheng Yu, Feng Zhang, Jinwei Gu, and Tianfan Xue. Event-based motion magnification, 2024.

[146] Yongjian Deng, Hao Chen, Hai Liu, and Youfu Li. A voxel graph cnn for object classification with event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1172–1181, 2022.

[147] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. Events-to-video: Bringing modern computer vision to event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3857–3866, 2019.

[148] Simon Schaefer, Daniel Gehrig, and Davide Scaramuzza. Aegnn: Asynchronous event-based graph neural networks, 2022.

[149] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. Keyframe-based visual–inertial odometry using nonlinear optimization. *Int. J. Rob. Res.*, 34(3):314–334, March 2015.

[150] Thomas Whelan, Stefan Leutenegger, Renato F. Salas-Moreno, Ben Glocker, and Andrew J. Davison. Elasticfusion: Dense slam without a pose graph. In *Robotics: Science and Systems*, 2015.

[151] Elias Mueggler, Basil Huber, and Davide Scaramuzza. Event-based, 6-dof pose tracking for high-speed maneuvers. *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2761–2768, 2014.

[152] Wenzhen Yuan and Srikumar Ramalingam. Fast localization and tracking using event sensors. *IEEE ICRA*, pages 4564–4571, 2016.

[153] J. L. Bertrand, Arda Yiğit, and Sylvain Durand. Embedded event-based visual odometry. *2020 6th International Conference on Event-Based Control, Communication, and Signal Processing (EBCCSP)*, pages 1–8, 2020.

[154] Beat Kueng, Elias Mueggler, Guillermo Gallego, and Davide Scaramuzza. Low-latency visual odometry using event-based feature tracks. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 16–23, 2016.

[155] D. Zhu, Jinhu Dong, Zhongcong Xu, Canbo Ye, Yinbai Hu, Hang Su, Zhengfa Liu, and Guang Chen. Neuromorphic visual odometry system for intelligent vehicle application with bio-inspired vision sensor. *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 2225–2232, 2019.

[156] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. Svo: Fast semi-direct monocular visual odometry. *IEEE ICRA*, pages 15–22, 2014.

[157] David Weikersdorfer and Jörg Conradt. Event-based particle filtering for robot self-localization. *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 866–870, 2012.

[158] William Chamorro, Joan Solà, and J. Andrade-Cetto. Event-based line slam in real-time. *IEEE RA-L*, 7:8146–8153, 2022.

[159] Henri Rebecq, Guillermo Gallego, Elias Mueggler, and Davide Scaramuzza. Emvs: Event-based multi-view stereo—3d reconstruction with an event camera in real-time. *International Journal of Computer Vision*, 126:1394–1414, 2018.

[160] David Weikersdorfer, Raoul Hoffmann, and Jörg Conradt. Simultaneous localization and mapping for event-based vision systems. In *International Conference on Virtual Storytelling*, 2013.

[161] Elias Mueggler, Guillermo Gallego, and Davide Scaramuzza. Continuous-time trajectory estimation for event-based vision sensors. *Robotics: Science and Systems XI*, 2015.

[162] Daqi Liu, Álvaro Parra, Yasir Latif, Bo Chen, Tat-Jun Chin, and Ian D. Reid. Asynchronous optimisation for event-based visual odometry. *2022 International Conference on Robotics and Automation (ICRA)*, pages 9432–9438, 2022.

[163] Cédric Le Gentil, Florian Tschopp, Ignacio Alzugaray, Teresa Vidal-Calleja, Roland Y. Siegwart, and Juan I. Nieto. Idol: A framework for imu-dvs odometry using lines. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5863–5870, 2020.

[164] Daqi Liu, Álvaro Parra, and Tat-Jun Chin. Spatiotemporal registration for event-based visual odometry. *IEEE/CVF CVPR*, pages 4935–4944, 2021.

[165] Johannes Schemmel, Daniel Brüderle, Andreas Grübl, Matthias Hock, Karlheinz Meier, and Sebastian Millner. A wafer-scale neuromorphic hardware system for large-scale neural modeling. In *2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1947–1950, 2010.

[166] Eustace Painkras, Luis A. Plana, Jim Garside, Steve Temple, Francesco Galluppi, Cameron Patterson, David R. Lester, Andrew D. Brown, and Steve B. Furber. Spinnaker: A 1-w 18-core system-on-chip for massively-parallel neural network simulation. *IEEE Journal of Solid-State Circuits*, 48(8):1943–1953, 2013.

[167] Ben Varkey Benjamin, Peiran Gao, Emmett McQuinn, Swadesh Choudhary, Anand R. Chandrasekaran, Jean-Marie Bussat, Rodrigo Alvarez-Icaza, John V. Arthur, Paul A. Merolla, and Kwabena Boahen. Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proceedings of the IEEE*, 102(5):699–716, 2014.

[168] Paul A. Merolla, John V. Arthur, and et al. Rodrigo Alvarez-Icaza. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673, 2014.

[169] De Ma, Juncheng Shen, Zonghua Gu, Ming Zhang, Xiaolei Zhu, Xiaoqiang Xu, Qi Xu, Yangjing Shen, and Gang Pan. Darwin: A neuromorphic hardware co-processor based on spiking neural networks. *Journal of Systems Architecture*, 77:43–51, 2017.

[170] Mike Davies, Narayan Srinivasa, Tsung-Han Lin, and et al. Chinya, Gautham. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1):82–99, 2018.

[171] Jing Pei, Lei Deng, Sen Song, Mingguo Zhao, Youhui Zhang, Shuang Wu, Guanrui Wang, Zhe Zou, Zhenzhi Wu, Wei He, Feng Chen, Ning Deng, Si Wu, Yu Wang, Yujie Wu, Zheyu Yang, Cheng Ma, Guoqi Li, Wentao Han, Huanglong Li, Huaqiang Wu, Rong Zhao, Yuan Xie, and Luping Shi. Towards artificial general intelligence with hybrid tianjic chip architecture. *Nature*, 572(7767):106–+, 2019.

[172] Man Yao, Ole Richter, Guangshe Zhao, Ning Qiao, Yannan 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.

[173] Yizhao Gao, Baoheng Zhang, Yuhao Ding, and Hayden Kwok-Hay So. A composable dynamic sparse dataflow architecture for efficient event-based vision processing on fpga. *Computing Research Repository*, pages 246–257, 2024.

[174] Hao Cao, Jingao Xu, Danyang Li, Zheng Yang, and Yunhao Liu. Eventboost: Event-based acceleration platform for real-time drone localization and tracking. *IEEE Conference on Computer Communications*, 2024.

[175] Ahmed Hasssan, Jian Meng, and Jae-Sun Seo. Spiking neural network with learnable threshold for event-based classification and object detection. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2024.

[176] Zongpei Fu and Wenbin Ye. A 593nj/inference dvs hand gesture recognition processor embedded with reconfigurable multiple constant multiplication technique. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 71(6), 2024.

[177] Alessandro Aimar, Hesham Mostafa, Enrico Calabrese, Antonio Rios-Navarro, Ricardo Tapiador-Morales, Iulia-Alexandra Lungu, Moritz B Milde, Federico Corradi, Alejandro Linares-Barranco, Shih-Chii Liu, et al. Nullhop: A flexible convolutional neural network accelerator based on sparse representations of feature maps. *IEEE transactions on neural networks and learning systems*, 30(3):644–656, 2018.

[178] Fernando Cladera, Anthony Bisulco, Daniel Kepple, Volkan Isler, and Daniel D Lee. On-device event filtering with binary neural networks for pedestrian detection using neuromorphic vision sensors. In *IEEE ICIP*, pages 3084–3088. IEEE, 2020.

[179] Georg Rutishauser, Moritz Scherer, Tim Fischer, and Luca Benini. Ternarized tcn for \mu\mathrm {J}/\text {Inference} gesture recognition from dvs event frames. In *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 736–741. IEEE, 2022.

[180] Filipp Akopyan, Jun Sawada, et al. Truenorth: Design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip. *IEEE transactions on computer-aided design of integrated circuits and systems*, 34(10):1537–1557, 2015.

[181] Song Han, Junlong Kang, et al. Ese: Efficient speech recognition engine with sparse lstm on fpga. In *Proceedings of the 2017 ACM/SIGDA international symposium on field-programmable gate arrays*, pages 75–84, 2017.

[182] Georg Rutishauser, Moritz Scherer, Tim Fischer, and Luca Benini. 7 μ j/inference end-to-end gesture recognition from dynamic vision sensor data using ternarized hybrid convolutional neural networks. *Future Generation Computer Systems*, 149:717–731, 2023.

[183] 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.

[184] Mingjun Li, Jianlei Yang, Yingjie Qi, Meng Dong, Yuhao Yang, Runze Liu, Weitao Pan, Bei Yu, and Weisheng Zhao. Eventor: An efficient event-based monocular multi-view stereo accelerator on fpga platform, 2022.

[185] Antonio Vitale, Alpha Renner, Celine Nauer, Davide Scaramuzza, and Yulia Sandamirskaya. Event-driven vision and control for uavs on a neuromorphic chip. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 103–109, 2021.

[186] Xin Zhao, Shiyu Hu, Yipei Wang, Jing Zhang, Yimin Hu, Rongshuai Liu, Haibin Ling, Yin Li, Renshu Li, Kun Liu, et al. Biodrone: A bionic drone-based single object tracking benchmark for robust vision. *International Journal of Computer Vision*, 132(5):1659–1684, 2024.

[187] Riccardo Santambrogio, Marco Cannici, and Matteo Matteucci. Farse-cnn: Fully asynchronous, recurrent and sparse event-based cnn. *ECCV 2024*, 2024.

[188] Yuetong Fang, Ziqing Wang, Lingfeng Zhang, Jiahang Cao, Honglei Chen, and Renjing Xu. Spiking wavelet transformer. *ECCV 2024*, 2024.

[189] Zhongyang Ren, Bangyan Liao, Delei Kong, Jinghang Li, Peidong Liu, Laurent Kneip, Guillermo Gallego, and Yi Zhou. Motion and structure from event-based normal flow. *ECCV 2024*, 2024.

[190] Lin Wang, I.S. Mohammad Mostafavi, Yo-Sung Ho, and Kuk-Jin Yoon. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In *CVPR 2019*, pages 10073–10082, 2019.

[191] Bingde Liu, Chang Xu, Wen Yang, Huai Yu, and Lei Yu. Motion robust high-speed light-weighted object detection with event camera. *IEEE Transactions on Instrumentation and Measurement*, 72:1–13, 2023.

[192] Yan Peng, Yueyi Zhang, Peilin Xiao, Xiaoyan Sun, and Feng Wu. Better and faster: Adaptive event conversion for event-based object detection. In *AAAI Conference on Artificial Intelligence*, 2023.

[193] Ling Gao, Daniel Gehrig, Hang SU, Davide Scaramuzza, and Laurent Kneip. An n-point linear solver for line and motion estimation with event cameras. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14596–14605, 2024.

[194] Song Wu, Zhiyu Zhu, Junhui Hou, Guangming Shi, and Jinjian Wu. E-motion: Future motion simulation via event sequence diffusion. *Advances in Neural Information Processing Systems*, 37:105552–105582, 2024.

[195] Qianang Zhou, Zhiyu Zhu, Junhui Hou, Yongjian Deng, Youfu Li, and Junlin Xiong. Resflow: Fine-tuning residual optical flow for event-based high temporal resolution motion estimation. *arXiv preprint arXiv:2412.09105*, 2024.

[196] Qianang Zhou, Junhui Hou, Meiyi Yang, Yongjian Deng, Youfu Li, and Junlin Xiong. Spatially-guided temporal aggregation for robust event-rgb optical flow estimation. *arXiv preprint arXiv:2501.00838*, 2025.

[197] Jinghang Li, Bangyan Liao, Xiuyuan LU, Peidong Liu, Shaojie Shen, and Yi Zhou. Event-aided time-to-collision estimation for autonomous driving, 2024.

[198] Qianhui Liu, Haibo Ruan, Dong Xing, Huajin Tang, and Gang Pan. Effective aer object classification using segmented probability-maximization learning in spiking neural networks, 2020.

[199] Jun Nagata and Yusuke Sekikawa. Tangentially elongated gaussian belief propagation for event-based incremental optical flow estimation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21940–21949, 2023.

[200] Yansong Peng, Hebei Li, Yueyi Zhang, Xiaoyan Sun, and Feng Wu. Scene adaptive sparse transformer for event-based object detection, 2024.

[201] Nitin J. Sanket, Chethan M. Parameshwara, Chahat Deep Singh, Ashwin V. Kuruttukulam, Cornelia Fermüller, Davide Scaramuzza, and Yiannis Aloimonos. Evdodgenet: Deep dynamic obstacle dodging with event cameras, 2020.

[202] Yingkai Fu, Meng Li, Wenxi Liu, Yuanchen Wang, Jiqing Zhang, Baocai Yin, Xiaopeng Wei, and Xin Yang. Distractor-aware event-based tracking. *IEEE Transactions on Image Processing*, 32:6129–6141, 2023.

[203] Yongjian Deng, Hao Chen, Bochen Xie, Hai Liu, and Youfu Li. A dynamic graph cnn with cross-representation distillation for event-based recognition, 2023.

[204] Yunfan LU, Guoqiang Liang, Yusheng Wang, Lin Wang, and Hui Xiong. Uniinr: Event-guided unified rolling shutter correction, deblurring, and interpolation, 2024.

[205] Chuanzhi Xu, Haoxian Zhou, Langyi Chen, Yuk Ying Chung, and Qiang Qu. Ultralight polarity-split neuromorphic snn for event-stream super-resolution. *arXiv preprint arXiv:2508.03244*, 2025.

[206] Nico Messikommer, Daniel Gehrig, Antonio Loquercio, and Davide Scaramuzza. Event-based asynchronous sparse convolutional networks, 2020.

[207] Yijin Li, Han Zhou, Bangbang Yang, Ye Zhang, Zhaopeng Cui, Hujun Bao, and Guofeng Zhang. Graph-based asynchronous event processing for rapid object recognition. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 914–923, 2021.

[208] Linhui Sun, Yifan Zhang, Jian Cheng, and Hanqing Lu. Asynchronous event processing with local-shift graph convolutional network. In *Proceedings of AAAI*, 2023.

[209] Giuseppe Mollica, Simone Felicioni, Marco Legittimo, Leonardo Meli, Gabriele Costante, and Paolo Valigi. Ma-vied: A multisensor automotive visual inertial event dataset. *IEEE Transactions on Intelligent Transportation Systems*, 25(1):214–224, 2024.

[210] Guillermo Gallego and Davide Scaramuzza. Accurate angular velocity estimation with an event camera. *IEEE RA-L*, 2(2):632–639, 2017.

[211] Matthew Ng, Zi Min Er, Gim Song Soh, and Shaohui Foong. Aggregation functions for simultaneous attitude and image estimation with event cameras at high angular rates. *IEEE Robotics and Automation Letters*, 7(2):4384–4391, 2022.

[212] Yi-Fan Zuo, Jiaqi Yang, Jiaben Chen, Xia Wang, Yifu Wang, and Laurent Kneip. Devo: Depth-event camera visual odometry in challenging conditions, 2022.

[213] Peiyu Chen, Weipeng Guan, and Peng Lu. Esvio: Event-based stereo visual inertial odometry. *IEEE RA-L*, 8(6):3661–3668, 2023.

[214] Edd Gent. Neuromorphic camera helps drones navigate without gps, 2024. https://spectrum.ieee.org/africa-engineering-hardware.

[215] Federico Paredes-Vallés, Kirk Y. W. Scheper, and Guido C. H. E. de Croon. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception. *IEEE T-PAMI*, 42(8):2051–2064, 2020.

[216] Xiuyuan Lu, Yi Zhou, Junkai Niu, Sheng Zhong, and Shaojie Shen. Event-based visual inertial velometer, 2024.

[217] Florian Tschopp, Cornelius von Einem, Andrei Cramariuc, David Hug, Andrew William Palmer, Roland Siegwart, Margarita Chli, and Juan Nieto. Hough^{2}map – iterative event-based hough transform for high-speed railway mapping. *IEEE Robotics and Automation Letters*, 6(2):2745–2752, 2021.

[218] Wenpu Li, Pian Wan, Peng Wang, Jinghang Li, Yi Zhou, and Peidong Liu. Benerf: Neural radiance fields from a single blurry image and event stream, 2024.

[219] Yuliang Wu, Ganchao Tan, Jinze Chen, Wei Zhai, Yang Cao, and Zheng-Jun Zha. Event-based asynchronous hdr imaging by temporal incident light modulation, 2024.

[220] Mathias Gehrig and Davide Scaramuzza. Recurrent vision transformers for object detection with event cameras, 2023.

[221] Luis A. Camuñas-Mesa, Teresa Serrano-Gotarredona, Sio-Hoi Ieng, Ryad Benosman, and Bernabé Linares-Barranco. Event-driven stereo visual tracking algorithm to solve object occlusion. *IEEE Transactions on Neural Networks and Learning Systems*, 29(9):4223–4237, 2018.

[222] Zhiyu Zhu, Junhui Hou, and Xianqiang Lyu. Learning graph-embedded key-event back-tracing for object tracking in event clouds. *Advances in Neural Information Processing Systems*, 35:7462–7476, 2022.

[223] Nitin J. Sanket, Chahat Deep Singh, Chethan M. Parameshwara, Cornelia Fermüller, Guido C. H. E. de Croon, and Yiannis Aloimonos. Evpropnet: Detecting drones by finding propellers for mid-air landing and following, 2021.

[224] Arren Glover, Luna Gava, Zhichao Li, and Chiara Bartolozzi. Edopt: Event-camera 6-dof dynamic object pose tracking. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 18200–18206, 2024.

[225] Zhiyu Zhu, Jinhui Hou, Jiading Li, Jinjian Wu, and Junhui Hou. Modeling state shifting via local-global distillation for event-frame gaze tracking. *IEEE Transactions on Mobile Computing*, 2025.

[226] Zhiyu Zhu, Junhui Hou, and Dapeng Oliver Wu. Cross-modal orthogonal high-rank augmentation for rgb-event transformer-trackers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22045–22055, 2023.

[227] Zhiwen Chen, Jinjian Wu, Weisheng Dong, Leida Li, and Guangming Shi. Crossei: Boosting motion-oriented object tracking with an event camera. *IEEE Transactions on Image Processing*, 2024.

[228] Xinyu Luo, Haoyang Wang, Ciyu Ruan, Chenxin Liang, Jingao Xu, and Xinlei Chen. Eventtracker: 3d localization and tracking of high-speed object with event and depth fusion. In *Proceedings of the 30th ACM MobiCom*, pages 1974–1979, 2024.

[229] Prophesee and tobii partner to develop next-generation event-based eye tracking solution for ar/vr and smart eyewear, 2025. https://www.prophesee.ai/2025/05/20/prophesee-and-tobii-partner-to-develop-next-generation-event-based-eye-tracking-solution-for-ar-vr-and-smart-eyewear/.

[230] Timo Stoffregen, Hossein Daraei, Clare Robinson, and Alexander Fix. Event-based kilohertz eye tracking using coded differential lighting. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2515–2523, 2022.

[231] saaz.com. [Online]. https://saaz.com/.

[232] neurobus.ai. [Online]. https://neurobus.ai/.

[233] Adarsh Kumar Kosta and Kaushik Roy. Adaptive-spikenet: Event-based optical flow estimation using spiking neural networks with learnable neuronal dynamics, 2023.

[234] Wachirawit Ponghiran, Chamika Mihiranga Liyanagedera, and Kaushik Roy. Event-based temporally dense optical flow estimation with sequential learning, 2023.

[235] Zhexiong Wan, Yuxin Mao, Jing Zhang, and Yuchao Dai. Rpeflow: Multimodal fusion of rgb-pointcloud-event for joint optical flow and scene flow estimation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9996–10006, 2023.

[236] Shintaro Shiba, Yoshimitsu Aoki, and Guillermo Gallego. *Secrets of Event-Based Optical Flow*, page 628–645. Springer Nature Switzerland, 2022.

[237] Shintaro Shiba, Yannick Klose, Yoshimitsu Aoki, and Guillermo Gallego. Secrets of event-based optical flow, depth and ego-motion estimation by contrast maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):7742–7759, 2024.

[238] event-based-vision-iot. [Online]. https://www.prophesee.ai/event-based-vision-iot/.

[239] Hoonhee Cho, Hyeonseong Kim, Yujeong Chae, and Kuk-Jin Yoon. Label-free event-based object recognition via joint learning with image reconstruction from events. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 19809–19820, 2023.

[240] Xu Zheng and Lin Wang. Eventdance: Unsupervised source-free cross-modal adaptation for event-based object recognition, 2024.

[241] Yanxiang Wang, Xian Zhang, Yiran Shen, Bowen Du, Guangrong Zhao, Lizhen Cui, and Hongkai Wen. Event-stream representation for human gaits identification using deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3436–3449, 2022.

[242] Yin Bi, Aaron Chadha, Alhabib Abbas, Eirina Bourtsoulatze, and Yiannis Andreopoulos. Graph-based object classification for neuromorphic vision sensing. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 491–501, 2019.

[243] Emre O. Neftci, Hesham Mostafa, and Friedemann Zenke. Surrogate gradient learning in spiking neural networks, 2019.

[244] Federico Becattini, Federico Palai, and Alberto Del Bimbo. Understanding human reactions looking at facial microexpressions with an event camera. *IEEE Transactions on Industrial Informatics*, 18(12):9112–9121, 2022.

[245] event-based-vision-medical. [Online]. https://www.prophesee.ai/event-based-vision-medical/.

[246] Hanme Kim, Stefan Leutenegger, and Andrew J. Davison. Real-time 3d reconstruction and 6-dof tracking with an event camera. In *ECCV*, 2016.

[247] Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza. *Semi-dense 3D Reconstruction with a Stereo Event Camera*, page 242–258. Springer International Publishing, 2018.

[248] Xueyan Huang, Yueyi Zhang, and Zhiwei Xiong. High-speed structured light based 3d scanning using an event camera. *Opt. Express*, 29(22):35864–35876, Oct 2021.

[249] Manasi Muglikar, Leonard Bauersfeld, Diederik Paul Moeys, and Davide Scaramuzza. Event-based shape from polarization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1547–1556, 2023.

[250] Viktor Rudnev, Vladislav Golyanik, Jiayi Wang, Hans-Peter Seidel, Franziska Mueller, Mohamed Elgharib, and Christian Theobalt. Eventhands: Real-time neural 3d hand pose estimation from an event stream. In *IEEE/CVF ICCV*, pages 12365–12375, 2021.

[251] Shihao Zou, Chuan Guo, Xinxin Zuo, Sen Wang, Pengyu Wang, Xiaoqin Hu, Shoushun Chen, Minglun Gong, and Li Cheng. Eventhpe: Event-based 3d human pose and shape estimation, 2021.

[252] Yuxuan Xue, Haolong Li, Stefan Leutenegger, and Jörg Stückler. Event-based non-rigid reconstruction from contours, 2022.

[253] Linglin Jing, Yiming Ding, Yunpeng Gao, Zhigang Wang, Xu Yan, Dong Wang, Gerald Schaefer, Hui Fang, Bin Zhao, and Xuelong Li. Hpl-ess: Hybrid pseudo-labeling for unsupervised event-based semantic segmentation, 2024.

[254] Lingdong Kong, Youquan Liu, Lai Xing Ng, Benoit R. Cottereau, and Wei Tsang Ooi. Openess: Event-based semantic scene understanding with open vocabularies, 2024.

[255] Timo Stoffregen, Guillermo Gallego, Tom Drummond, Lindsay Kleeman, and Davide Scaramuzza. Event-based motion segmentation by motion compensation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7243–7252, 2019.

[256] Ruihao Xia, Chaoqiang Zhao, Meng Zheng, Ziyan Wu, Qiyu Sun, and Yang Tang. Cmda: Cross-modality domain adaptation for nighttime semantic segmentation, 2023.

[257] Bowen Yao, Yongjian Deng, Yuhan Liu, Hao Chen, Youfu Li, and Zhen Yang. Sam-event-adapter: Adapting segment anything model for event-rgb semantic segmentation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9093–9100. IEEE, 2024.

[258] Zhiwen Chen, Zhiyu Zhu, et al. Segment any event streams via weighted adaptation of pivotal tokens. In *Proceedings of CVPR*, 2024.

[259] Junkai Niu, Sheng Zhong, Xiuyuan Lu, Shaojie Shen, Guillermo Gallego, and Yi Zhou. Esvo2: Direct visual-inertial odometry with stereo event cameras. *IEEE Transactions on Robotics*, 2025.

[260] Delin Qu, Chi Yan, Dong Wang, Jie Yin, Qizhi Chen, Dan Xu, Yiting Zhang, Bin Zhao, and Xuelong Li. Implicit event-rgbd neural slam. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19584–19594, 2024.

[261] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios. *IEEE Robotics and Automation Letters*, 3(2):994–1001, 2018.

[262] Shuang Guo and Guillermo Gallego. Cmax-slam: Event-based rotational-motion bundle adjustment and slam system using contrast maximization. *IEEE Transactions on Robotics*, 40:2442–2461, 2024.

[263] Yi Zhou, Guillermo Gallego, and Shaojie Shen. Event-based stereo visual odometry. *IEEE Transactions on Robotics*, 37(5):1433–1450, 2021.

[264] Yuhan Bao, Lei Sun, Yuqin Ma, and Kaiwei Wang. Temporal-mapping photography for event cameras. In *ECCV*, pages 55–72. Springer, 2024.

[265] Bohan Yu, Jieji Ren, Jin Han, et al. Eventps: Real-time photometric stereo using an event camera. In *Proceedings of CVPR*, 2024.

[266] Baining Zhao, Jianjie Fang, Zichao Dai, Ziyou Wang, Jirong Zha, Weichen Zhang, Chen Gao, Yue Wang, Jinqiang Cui, Xinlei Chen, and Yong Li. UrbanVideo-bench: Benchmarking vision-language models on embodied intelligence with video data in urban spaces. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd ACL*, pages 32400–32423, July 2025.

[267] Haoyang Wang, Jingao Xu, Xinyu Luo, Ting Zhang, Xuecheng Chen, Ruiyang Duan, Jialong Chen, Yunhao Liu, Jianfeng Zheng, Weijie Hong, et al. mme-loc: Facilitating accurate drone landing with ultra-high-frequency localization. *arXiv preprint arXiv:2507.09469*, 2025.

[268] Mingyue Cui, Yuzhang Zhu, Yechang Liu, Yunchao Liu, Gang Chen, and Kai Huang. Dense depth-map estimation based on fusion of event camera and sparse lidar. *IEEE Transactions on Instrumentation and Measurement*, 71:1–11, 2022.

[269] Hanyu Zhou, Yi Chang, and Zhiwei Shi. Bring event into rgb and lidar: Hierarchical visual-motion fusion for scene flow. In *CVPR*, 2024.

[270] Samanwoy Ghosh-Dastidar and Hojjat Adeli. Spiking neural networks. *International journal of neural systems*, 19(04):295–308, 2009.

[271] Xinhui Liu, Meiqi Cheng, Dawei Shi, and Ling Shi. Toward event-based state estimation for neuromorphic event cameras. *IEEE Transactions on Automatic Control*, 68(7):4281–4288, 2022.