Review

Hardware, Algorithms, and Applications of the Neuromorphic Vision Sensor: A Review

Claudio Cimarelli [1,†] , Jose Andres Millan-Romera [1] , Holger Voos [1,2] and Jose Luis Sanchez-Lopez [1,*]

1   Automation and Robotics Research Group, Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, 1855 Luxembourg, Luxembourg; claudio.cimarelli@alumni.uni.lu (C.C.); jose.millan@uni.lu (J.A.M.-R.); holger.voos@uni.lu (H.V.)
2   Faculty of Science, Technology, and Medicine, University of Luxembourg, 1359 Luxembourg, Luxembourg
*   Correspondence: joseluis.sanchezlopez@uni.lu
†   Current address: Independent Researcher, 1111 Luxembourg, Luxembourg.

Abstract

Event-based (neuromorphic) cameras depart from frame-based sensing by reporting asynchronous per-pixel brightness changes. This produces sparse, low-latency data streams with extreme temporal resolution but demands new processing paradigms. In this survey, we systematically examine neuromorphic vision along three main dimensions. First, we highlight the technological evolution and distinctive hardware features of neuromorphic cameras from their inception to recent models. Second, we review image-processing algorithms developed explicitly for event-based data, covering works on feature detection, tracking, optical flow, depth and pose estimation, and object recognition. These techniques, drawn from classical computer vision and modern data-driven approaches, illustrate the breadth of applications enabled by event-based cameras. Third, we present practical application case studies demonstrating how event cameras have been successfully used across various scenarios. Distinct from prior reviews, our survey provides a broader overview by uniquely integrating hardware developments, algorithmic progressions, and real-world applications into a structured, cohesive framework. This explicitly addresses the needs of researchers entering the field or those requiring a balanced synthesis of foundational and recent advancements, without overly specializing in niche areas. Finally, we analyze the challenges limiting widespread adoption, identify research gaps compared to standard imaging techniques, and outline promising directions for future developments.

Keywords: neuromorphic sensor; event cameras; event-based image processing; neuromorphic vision applications

1. Introduction

Standard RGB cameras face considerable limitations, particularly in dynamic environments. The principle of these sensors involves capturing visual information as a sequence of frames at specific time intervals. Time-quantizing visual data at predetermined frame rates often results in temporal resolution limitations, as the frame rate is not aligned with the dynamic evolution of the scene. Consequently, significant details can be missed, especially in rapidly changing environments. Moreover, recording every pixel in each frame, regardless of changes since the last capture, leads to data redundancy, which affects the data rate and volume [1].

Instead, the limited dynamic range of standard RGB cameras often causes under- or overexposure in scenes with rapidly varying lighting conditions [2]. In addition, motion

blur is another common problem in high-speed movement scenarios. Then, the latency inherent in a fixed frame rate and power consumption for processing a large amount of data, e.g., resulting from redundant information, poses an obstacle when real-time responsiveness and energy efficiency are required.

In response to these limitations, neuromorphic cameras (NCs), or event cameras, as they are named more frequently in the robotic vision research domain, represent a paradigm shift [3] in acquiring visual information compared to conventional frame-based cameras. In particular, each event camera's pixel operates independently, i.e., with its analog circuit, continuously comparing the current brightness to a reference level [4]. When the difference exceeds a certain threshold, the pixel generates a sparse stream of event packets, e.g., the pixel's address, timestamp, and the polarity of the brightness change, labeled with a high temporal resolution. This method allows for capturing visual information that mimics the human retina [5], approaching image acquisition with a biologically inspired process that responds more to real-world dynamics.

Unlike conventional frame-based images, event streams produced by neuromorphic cameras are fundamentally sparse and asynchronous and encode only changes in logarithmic brightness, rather than absolute intensity values. These properties introduce a range of unique challenges that require rethinking core computer vision paradigms. First, the temporal irregularity of event data, in contrast to the uniform frame rate of conventional cameras, necessitates novel representations (e.g., voxel grids and time surfaces) that preserve high temporal fidelity while supporting context aggregation. Second, the sparse and noise-prone nature of event output, driven by local contrast thresholds and sensor-level artifacts, complicates feature extraction, data association, and motion estimation, particularly in low-contrast or static scenes, limiting the reliability of tasks like SLAM and motion tracking. Third, the absence of absolute intensity makes it difficult to apply standard photometric algorithms or leverage traditional learning models. Recent methods attempt to reconstruct intensity indirectly via noise correlations or require dedicated architectures, such as spiking neural networks or asynchronous convolutional modules. These constraints and opportunities extend into the hardware domain, driving interest in event-driven processors and hybrid sensor-compute systems optimized for low-power, low-latency inference. As a result, event-based vision has driven innovations in both algorithm design, particularly for high-speed, low-power applications such as robotics, HDR imaging, and neuromorphic computing hardware itself.

Nevertheless, due to their unique capabilities, NCs are highly suitable for various applications where real-time processing, adaptability to diverse lighting conditions, and energy efficiency are critical. These include robotics, surveillance, autonomous vehicles, and other areas that require robust and efficient visual sensing. For example, NCs can provide low-latency obstacle detection, even in challenging lighting or weather conditions, which is crucial in autonomous vehicle navigation. In robotics, event cameras enable more responsive situational awareness to changes in the dynamic environment. Due to their radically different sensor modality, NCs can offer non-invasive monitoring systems, which is helpful in data-privacy-preserving scenarios like healthcare. Moreover, their low power consumption and small data volumes, i.e., sparse event packets vs. dense image frames, make them ideal for remote surveillance monitoring systems or search and rescue missions, where energy efficiency is paramount.

To date, neuromorphic vision technology and event-camera image processing have been the focus of multiple reviews, each exploring the topic from a unique perspective. These surveys encompass a wide range of topics, from sensor technology to image-processing methodologies. In Tables 1 and 2, we provide an overview of their focus topic and highlight their key contributions. The multitude of surveys reflects the rapid develop-

ment of neuromorphic vision research in recent years and the traction the field has gained, especially in the robotics and computer vision communities.

Table 1. Overview of previous surveys on neuromorphic vision (Part 1).

Reference	Year	Focus Topic	Key Highlights
This survey	2025	Comprehensive Overview (Hardware, Algorithms, Applications)	Provides a structured and up-to-date survey of neuromorphic vision, covering the evolution of hardware, key developments in event-based image processing, practical application case studies, and current limitations. Highlights challenges and opportunities for broader adoption across domains.
Kudithipudi et al. [6]	2025	Neuromorphic Computing at Scale	Proposes a forward-looking framework for scaling neuromorphic systems from lab prototypes to real-world deployments. Offers high-level guidance on architecture, tools, ecosystem needs, and application potential. It is especially relevant for stakeholders aiming at system-level integration.
Ghosh and Gallego [7]	2025	Event-based Stereo Depth Estimation	Extensive survey of stereo depth estimation with event cameras, encompassing both model-based and deep learning techniques. Compares instantaneous vs. long-term methods, highlights the role of neuromorphic hardware and SNNs for stereo, benchmarks methods and datasets, and proposes future directions for real-world deployment of event-based depth estimation.
Adra et al. [8]	2025	Human-Centered Event-Based Applications	Provides the first comprehensive survey unifying event-based vision applications for body and face analysis. Discusses challenges, opportunities, and less explored topics such as event compression and simulation frameworks.
AliAkbarpour et al. [9]	2024	Emerging Trends and Applications	Reviews a wide spectrum of emerging event-based processing techniques and niche applications (e.g., SSA, RS correction, SR, robotics). Focuses on technical breadth and dataset/tool support, with limited emphasis on foundational hardware or unified algorithmic frameworks.
Shariff et al. [10]	2024	Automotive Sensing (In-Cabin and Out-of-Cabin)	Presents a comprehensive review of event cameras for automotive sensing, covering both in-cabin (driver/passenger monitoring) and out-of-cabin (object detection, SLAM, obstacle avoidance). Details hardware architecture, data processing, datasets, noise filtering, sensor fusion, and transformer-based approaches.
Cazzato and Bono [11]	2024	Application-Driven Event-Based Vision	Reviews event-based neuromorphic vision sensors from an application perspective. Categorizes computer vision problems by field and discusses each application area's key challenges, major achievements, and unique characteristics.
Chakravarthi et al. [12]	2024	Event Camera Innovations	Traces the evolution of event cameras, comparing them with traditional sensors. Reviews technological milestones, major camera models, datasets, and simulators while consolidating research resources for further innovation.
Tenzin et al. [13]	2024	Event-Based VSLAM and Neuromorphic Computing	Surveys the integration of event cameras and neuromorphic processors into VSLAM systems. Discusses feature extraction, motion estimation, and map reconstruction while highlighting energy efficiency, robustness, and real-time performance improvements.
Becattini et al. [14]	2024	Face Analysis	Examines novel applications such as expression and emotion recognition, face detection, identity verification, and gaze tracking for AR/VR, areas not previously covered by event cameras surveys. The paper emphasizes the significant gap in standardized datasets and benchmarks, stressing the importance of using real data over simulations.
Zheng et al. [15]	2023	Deep Learning Approaches	Extensively surveys deep learning approaches for event-based vision, focusing on advancements in data representation and processing techniques. It systematically categorizes and evaluates methods across multiple computer vision topics. The paper discusses the unique advantages of event cameras, particularly under challenging conditions, and suggests future directions for integrating deep learning to exploit these benefits further.

Table 2. Overview of previous surveys on neuromorphic vision (Part 2).

Reference	Year	Focus Topic	Key Highlights
Huang et al. [16]	2023	Self-Localization and Mapping	Discusses various event-based vSLAM methods, including feature-based, direct, motion-compensation, and deep learning approaches. Evaluates these methods on different benchmarks, underscoring their unique properties and advantages with respect to one another. Then, it gives deep reasons for the challenges inherent to sensors and the task of SLAM, drawing future directions for research.
Schuman et al. [17]	2022	Neuromorphic Algorithmic Directions	Highlights algorithmic opportunities for neuromorphic computing, covering SNN training methods, non-ML models, and application potential. Emphasis is placed on open challenges and co-design needs rather than exhaustive architectural or vision-specific coverage.
Shi et al. [18]	2022	Motion and Depth Estimation for Indoor Positioning	Reviews notable techniques for ego-motion estimation, tracking, and depth estimation utilizing event-based sensing. Then, it suggests further research directions for real-world applications to indoor positioning.
Furmonas et al. [19]	2022	Depth Estimation Techniques	Discusses various depth estimation approaches, including monocular and stereo methods, detailing the strengths and challenges of each. It advocates integrating these sensors with neuromorphic computing platforms to enhance depth perception accuracy and processing efficiency.
Cho et al. [20]	2022	Material Innovations and Computing Paradigms	Highlights the evolution from traditional designs to innovative in-sensor and near-sensor computing that optimizes processing speed and energy efficiency. It addresses the challenge of complex manufacturing processes, suggesting directions for future research and application in flexible electronics.
Liao et al. [21]	2021	Technologies and Biological Principles	Reviews advancements in neuromorphic vision sensors, contrasting silicon-based CMOS technologies such as DVS, DAVIS, and ATIS with emerging technologies in analogical devices.
Gallego et al. [3]	2020	Sensor Working Principle and Vision Algorithms	Thoroughly reviews the advancements in event-based vision, emphasizing its unique properties. The survey spans various vision tasks, including feature detection, optical flow, and object recognition, and discusses innovative processing techniques. It also outlines significant challenges and future opportunities in this rapidly evolving field.
Steffen et al. [22]	2019	Stereo Vision and Sensor Principles	Performs a comparative analysis of event-based sensors, focusing on technologies such as DVS, DAVIS, and ATIS. It reviews the biological principles underlying depth perception and explores the approaches to stereoscopy using event-based sensors.
Lakshmi et al. [23]	2019	Object Motion and SLAM	Reviews state-of-the-art event-based vision algorithms for object detection/recognition, object tracking, localization, and mapping. Highlights the necessity of adapting conventional vision algorithms. Also provides an overview of publicly available event datasets and their applications.
Vanarse et al. [24]	2016	Neuromorphic Vision, Auditory, and Olfactory Sensors	Highlights low power consumption in the prototypical developments of DVS and DAVIS using asynchronous spiking output. Suggests future research directions in neuro-biological emulating sensors for vision, audition, and olfaction with multi-sensor integration.

While most recent surveys tend to specialize in emerging niches or focus on specific technical components, there remains a clear need for a broader yet integrative perspective. Our work fills this gap by methodically bridging the evolution of neuromorphic vision hardware, algorithmic strategies, and real-world deployment scenarios. We aim to connect technological advancements with their practical adoption across industries by outlining the

key challenges, limitations, and opportunities in this rapidly evolving field. This structured synthesis enables both newcomers and experts to grasp the interplay between sensor design, data processing, and application-driven innovation.

To provide quantitative insight into the technological landscape of neuromorphic vision, we analyzed the keywords, titles, and abstracts from our bibliography. It consists of entries collected via Zotero and manually curated through open APIs such as arXiv, CrossRef, and SemanticScholar. Approximately 40% of the entries include abstracts. We developed a Python-based (version 3.13.2) script that parses these BibTeX-style entries, extracts the relevant fields, and performs tokenization and filtering. Common English stopwords and non-scientific terms are excluded using a predefined filter. The resulting set of keywords is further refined by applying heuristics informed by domain-specific LLM models to focus only on meaningful technical terms, finally extracting the 30 most frequent scientific keywords. Lastly, we performed a deep analysis of the research database at https://www.dimensions.ai/ (accessed on 31 July 2025) using the exact keywords coupled with "neuromorphic camera" and "event camera" strings to restrict the search to only those related to our current review scope, and we present our findings in Figure 1. In particular, the picture shows our balanced coverage on each sub-topic of the neuromorphic camera in the literature in the present review, with only two outliers in the coverage percentage for which the keyword was the least frequent in the database.
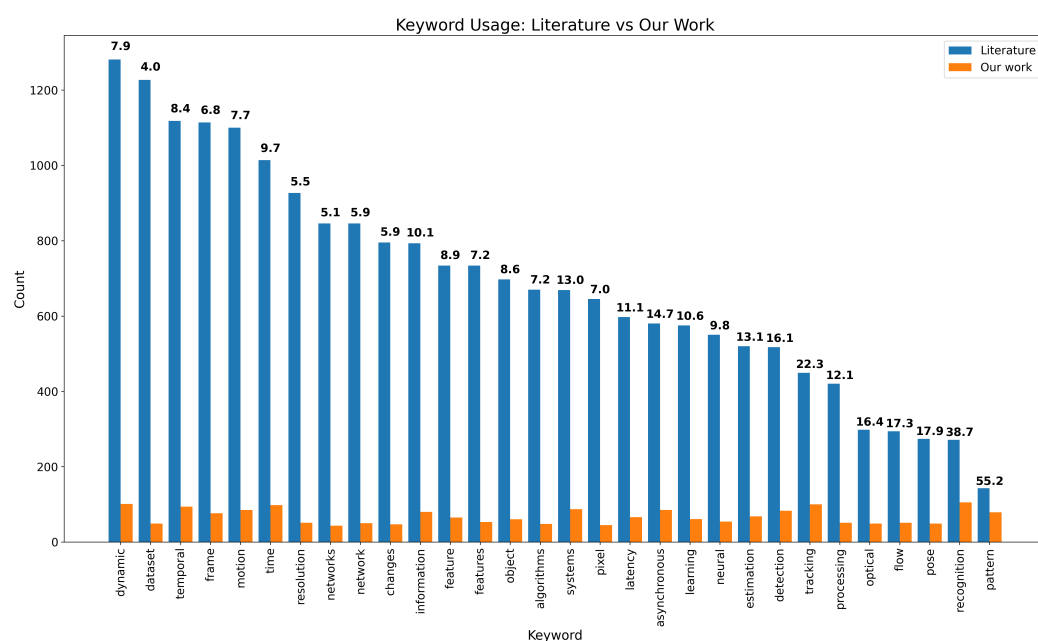

Figure 1. Keyword frequency analysis chart. We compare the accuracy of the keywords in the literature retrieved from https://www.dimensions.ai/ (accessed on 31 July 2025) with those used in our review. The percentage over the bars indicates the percentage covered.

To guide the reader through the organization of this survey, we illustrate the paper's structure in Figure 2. This visual abstract outlines the main thematic areas—hardware, algorithms, and applications—each contributing to the overall discussion and analysis in the final section.
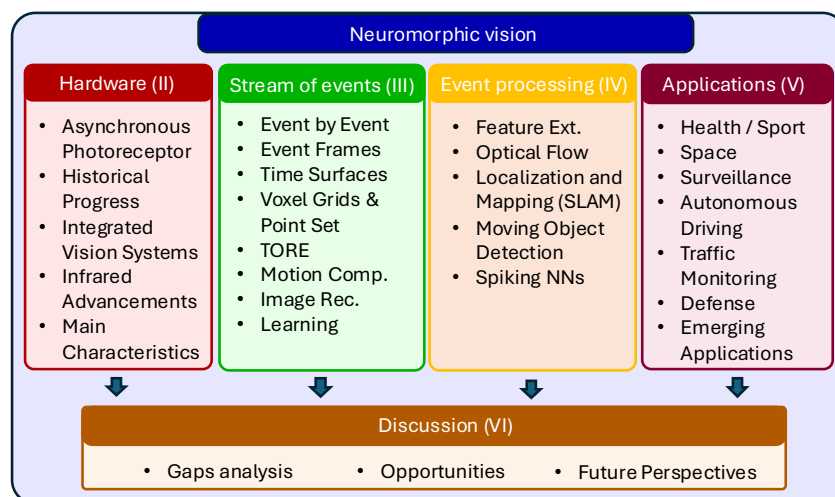
Figure 2. Overview of the paper's structure, illustrating the main sections, in parentheses, and their relationships.

In summary, with this survey, we make several distinct contributions:

- Neuromorphic Cameras' Hardware Evolution: We present a timeline of the evolution of neuromorphic vision sensor technology, revealing the chronological progress of the hardware, how it differs from standard vision systems, and why these differences matter (Section 2).
- Event-based Image Processing and Algorithms: We examine the progression of image-processing techniques from classical methods to advanced deep learning approaches (Sections 3 and 4).
- Application Focus: We discuss key application case studies demonstrating how the unique properties of neuromorphic cameras impact real-world solutions (Section 5).
- Gaps, Limitations, and Future Opportunities: We analyze the key challenges hindering the adoption of neuromorphic vision sensors, from hardware constraints to algorithmic gaps and real-world application barriers, while highlighting the opportunities unlocked by this radical shift in visual sensing modality (Section 6).

2. The Neuromorphic Vision Sensor

The initial concept of the NC invention arrived from the research group of Professor Carver Mead at Caltech with the publication of the book "*Analog VLSI and Neural Systems*" in 1989 [25]. Notably, Misha Mahowald, Mead's student, developed during her Ph.D. from 1986 to 1992 at Caltech the first neuromorphic chip to spike events resulting from detected light intensity variations [26–28]. The first specialized commercial application inspired by these ideas was a motion detection system for pointing devices designed by Xavier Arreguit of CSEM for Logitech in 1996 [29].

The following discussion clarifies the pixel design and its asynchronous data output; illustrates camera models that came to the market, from visible light to the infrared spectrum; and concludes by examining the unique characteristics of event cameras that make them a distinctive technology.

2.1. The Neuromorphic Camera's Asynchronous Photoreceptor

Neuromorphic cameras leverage asynchronous photoreceptors to efficiently mimic the responsiveness and energy efficiency of the human visual system, a concept that has been explored since the early developments of the silicon retina [26]. These receptors detect changes in light intensity and encode this information into discrete events, resulting in an array of pixels operating independently. This mechanism contrasts with the conventional

camera's approach of capturing entire frames regularly, thus processing and transmitting large volumes of redundant data.

Central to the operation of neuromorphic cameras is the Asynchronous Address-Event Representation (AER), an innovative communication protocol developed from the pioneering research by the Caltech group led by Carver Mead [27] and refined through subsequent research [30]. AER uses time-coded addresses to encode and dynamically transmit events between the silicon photoreceptor and the computing processor. Each event, whether an ON-event, indicating an increase in light, or an OFF-event, indicating a decrease, is defined by its pixel reference, timestamp, and polarity.

The neuromorphic photoreceptors respond to light intensity variations on a logarithmic scale. This capability allows the sensor to handle various lighting conditions effectively, from dim to bright. Each pixel analog circuit, as the primary designed in [31], detects changes that surpass a voltage threshold encoded in the photoreceptor, triggering the transmission of the AE packet, as illustrated in Figure 3.
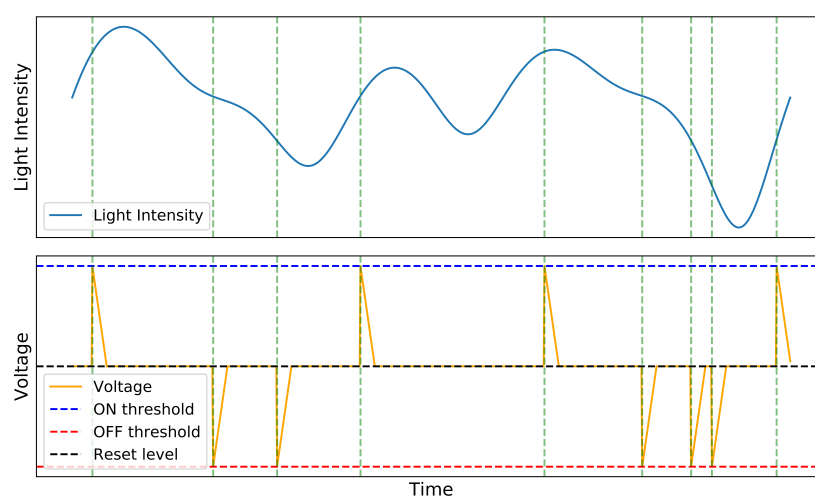

Figure 3. The asynchronous output of an operating event-camera photodiode. Inspired from an image in [4]. Dashed green lines indicate when a change exceeding the voltage threshold encoded in the photoreceptor is detected.

As explored by Steffen et al. [22], this sophisticated protocol incorporates a digital bus system and multiplexing strategies that allow all pixels to transmit their information over the same line efficiently and asynchronously, significantly reducing power consumption and data volume.

Further refining the process, the AER's implementation via address encoders generates unique binary addresses for each pixel event. This overall strategy highlights the role of AER in transmitting only essential visual information while discarding irrelevant static scenes and ensuring effective responses to rapid changes in the environment [32].

2.2. Progress in Visible-Light Event-Camera Models

With the theoretical progress made in the 1990s, several research organizations started producing the first neuromorphic sensors. Toby Delbruck proposed the first generic event camera in 2008 in collaboration with Patrick Lichtsteiner and Christoph Posch under the dynamic vision sensor (DVS), the earliest event-camera technology. Lichtsteiner, Posch, and Delbruck [4,31] proposed a novel silicon retina design that outputs AER in a 128×128 pixel grid. Since the DVS's inception, many companies have commercialized VGA to megapixel resolution event cameras, from a small reality like CelePixel or Insightness to a technological giant like Samsung, producing further innovations of the original

sensor. Event-camera expression gradually took place in the last few years to highlight the AER output of neuromorphic vision devices and differentiate them from their standard camera counterparts. A summary of key hardware and software advancements in neuromorphic vision is illustrated in the timeline shown in Figure 4.
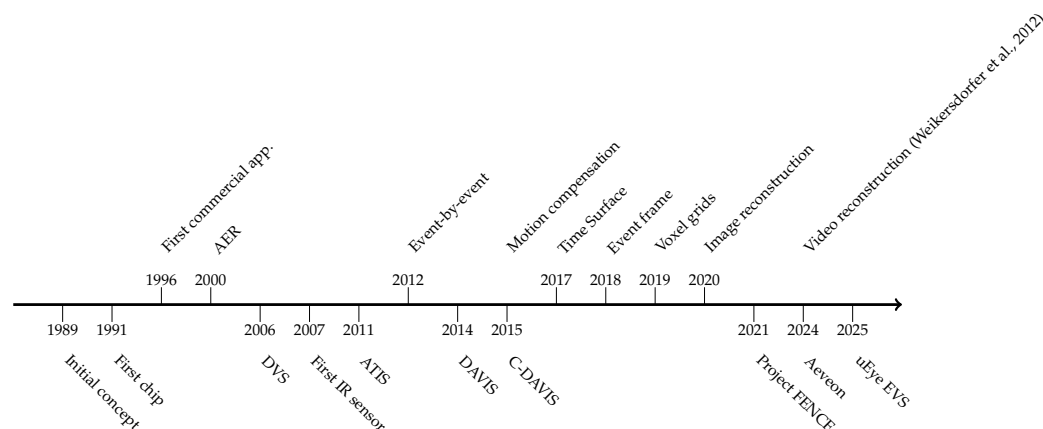

Figure 4. Timeline of pivotal milestones in developing the neuromorphic vision sensor. Events related to software developments are depicted above the axis, whereas hardware-related milestones are positioned below it [33].

In 2014, Christoph Posch co-founded Chronocam (now Prophesee) in France, focusing on the development and commercialization of Asynchronous Time-based Image Sensor (ATIS) technology, outlined in Posch's 2011 research at the Austrian Institute of Technology [34]. The ATIS marks a significant advancement in event-camera technology, merging the temporal contrast-detection capabilities of the DVS with innovative time-based intensity measurement pixels. This integration allows the ATIS to capture event-based data and provide absolute brightness measurements with high accuracy. However, incorporating a pulse-width-modulated (PWM) intensity readout mechanism for each DVS pixel, aimed at enhancing reconstruction and recognition capabilities, necessitated an extra photodiode per pixel, effectively doubling the pixel size. Moreover, because the PWM readout process required the transfer of triple the data amount, the ATIS latency is significantly increased, particularly impacting the sensor's ability to capture fast-moving or dimly lit objects. Despite this, the sensor's QVGA resolution (e.g., 304 × 240) substantially improves detail and image quality over the original DVS. Furthermore, the ATIS addresses critical limitations of traditional imaging systems by significantly reducing temporal redundancy and delivering a high dynamic range (143 dB static and 125 dB at 30 FPS). Further developments of ATIS technology have led to multiple sensor generations, such as Prophesee Metavision GEN3, including collaboration with Sony on the IMX636 and IMX637 sensors. These sensors, featuring stacked CMOS technology, underscore the ATIS's ongoing evolution and potential in various high-performance imaging applications.

The Dynamic and Active-pixel Vision Sensor (DAVIS) [35], developed by IniVation, represents a significant advancement in vision sensor technology. Unlike its predecessors, the DAVIS integrates neuromorphic event-driven and active-pixel sensor (APS) functionality within the same photodiode. This innovative design enables the DAVIS to interleave event data with conventional intensity frames, using a shared pixel to generate grayscale and event data. The pixel architecture of the DAVIS offers several benefits: it achieves a dynamic range of 130 dB for event detection and 51 dB for grayscale intensity frames. Additionally, it features a minimized latency of just 3 μs. Despite its dual functionality, the pixel area in the DAVIS is only marginally larger (about 5%) than that of a standard

DVS pixel, resulting in a slightly reduced high dynamic range compared to the ATIS but in a more compact form factor.

The Color Dynamic and Active-Pixel Vision Sensor (C-DAVIS) represents a significant advancement, building upon the foundations of the earlier DAVIS model [36]. This sensor combines monochrome event-based pixels with a five-transistor APS architecture integrated under a Red, Green, Blue, and White (RGBW) color filter array. Capable of outputting both rolling or global shutter RGBW-coded VGA resolution frames and asynchronous monochrome QVGA temporal contrast events, the C-DAVIS excels in capturing vibrant color details as well as tracking swift movements with remarkable temporal precision. This blend of capabilities is efficiently packed into a compact design, featuring a 2 × 2-pixel RGBW unit with dimensions of merely 20 μm × 20 μm, showcasing the C-DAVIS's ability to combine high-resolution color imaging with fast, event-based motion detection.

In 2023, IniVation introduced the Aeveon sensor, an advancement in neuromorphic vision technologies, to address the limitations of previous models like the DAVIS. The Aeveon is designed to allow each pixel to generate several event types, including full pixel value (RGB), multi-bit and single-bit change events, and area events. Moreover, it employs a stacked sensor design with Adaptive Event Cores, merging characteristics of neuromorphic sensors with frame-based sensors. This design is compatible with various pixel types, from standard RGB to infrared. Furthermore, the sensor offers the flexibility to select an adaptable region of interest (ROI), where the user can focus the event stream reception, similar to an attention mechanism. With its unified solution, the Aeveon should facilitate integration with existing systems while providing an immediate replacement for the current vision modules and a pathway to introduce new event-based features gradually.

In Table 3, we provide a list of currently commercialized event cameras. Other sensors have been described in the literature, such as in [3,19]. However, not all are available for purchase, e.g., the early DVS128 and DVS240 from IniVation, or are not easy to procure online, such as the models produced by Insightness, Samsung [37], and CelePixel [38,39]. Recently, IDS and Prophesee partnered to create the new uEye EVS camera series [40], enabling ultra-fast imaging (sub-100 μs resolution) and significantly reducing data processing and power consumption.

Table 3. Comparison of currently commercially available event cameras. Meps is millions of events per second.

Manufacturer	Model	Resolution	Latency	Temporal Resolution	Max Throughput	Dynamic Range	Power	Image Frames
IniVation	DAVIS346 (also, Color)	346 × 260	<1 ms	1 μs	12 Meps	120 dB	<180 mA	Graysc./Color
IniVation	DVXplorer	640 × 480	<1 ms	65–200 μs	165 Meps	90–110 dB	<140 mA	No
IniVation	DVXplorer Lite	320 × 320	<1 ms	65–200 μs	100 Meps	90–110 dB	<140 mA	No
IniVation	DVXplorer Micro	640 × 480	<1 ms	65–200 μs	450 Meps	90–110 dB	<140 mA	No
Prophesee	Gen 3 VGA CD	640 × 480	40–200 μs	NA	66 Meps	>120 dB	NA	No
Prophesee	GENX320	320 × 320	<150 μs	1 μs	NA	>120 dB	>36 μW	No
Sony/Prophesee	IMX636	1280 × 720	100–220 μs	NA	1060 Meps	86 dB	NA	Grayscale
Sony/Prophesee	IMX637	640 × 512	100–220 μs	NA	1060 Meps	86 dB	NA	Grayscale
Sony/Prophesee	IMX646	1280 × 720	800–9000 μs	NA	1060 Meps	110 dB	NA	Grayscale
Sony/Prophesee	IMX647	640 × 512	800–9000 μs	NA	1060 Meps	110 dB	NA	Grayscale
Imago Tech./Prophesee	Vision Cam EB	640 × 480	200 μs	NA	30 Meps	>120 dB	NA	No
IDS/Sony/Prophesee	uEye EVS	1280 × 720	<100 μs	<100 μs		>120 dB	10 μW	No

Notably, the price of such devices is still a few thousand dollars, making them currently functional only for industrial purposes. The cost of production is the main obstacle to the diffusion of event cameras in larger commercial markets until mass production of the sensor's silicon. However, recent partnerships, such as the announced collaboration between Prophesee and Qualcomm, allow us to foresee that event cameras may soon be adopted for mobile platform imaging. As an example of this trend, Google is adding event-based vision to its Visual Intelligence and Android XR platforms—paving the way for advanced AR glasses [41,42]. As adoption expands and commercial partnerships grow, the focus is increasingly on the technical challenges that remain, motivating further innovation in sensor integration and system design.

2.3. Toward Fully Integrated Neuromorphic Vision Systems

Recent progress has not only advanced pixel architectures, system integration, and commercial availability but also highlighted several remaining technical bottlenecks. Many visible-light event cameras, while achieving improved resolution, dynamic range, and frame fusion capabilities, continue to rely heavily on external or off-chip processing, which can limit their potential in latency-critical, power-constrained, or edge-driven applications. As the field matures, an emerging line of research shifts from simply refining sensing hardware to fundamentally rethinking the coupling of sensing and computation, aiming to achieve real-time, ultra-efficient, and adaptive vision in complex environments. The latest examples of this paradigm are seen in neuromorphic architectures that tightly merge event-driven sensing and in-sensor computation, enabling a leap beyond the traditional boundaries of event-camera design.

In addition to recent advances in pixel design and readout circuits, a new wave of neuromorphic sensors is fundamentally redefining the event-camera landscape by tightly merging sensing and computation at the hardware level. For example, the Speck chip [43] exemplifies a fully integrated neuromorphic sensing and computing system on a chip, which unites a dynamic vision sensor with a fully asynchronous event-driven spike neural network processor. This architecture enables truly "always-on" operation where the computation is activated only by incoming events, pushing idle power consumption to below one milliwatt. As a result, Speck delivers robust, real-time, ultra-low-power processing for high-speed event streams, opening up practical opportunities for edge intelligence in severely power-constrained scenarios.

Complementing this approach, the Tianmouc chip [44] brings a dual-path design inspired by biological science that mimics the parallel ventral (cognitive) and dorsal (action) streams of the primate visual system. Using a hybrid pixel array and parallel, heterogeneous readout chains, Tianmouc achieves high-resolution, high-precision perception and simultaneously delivers rapid, sparse, event-based responses in highly dynamic and unpredictable scenes. Notable system capabilities include operation at up to 10,000 frames per second, a dynamic range reaching 130 dB, and considerable bandwidth reduction through sparse event encoding. This complementary vision paradigm, validated on demanding open-world tasks such as autonomous driving, underlines the importance of accurate scene understanding and rapid recognition of rare and safety-critical events.

Together, these integrated sensing–computing systems exemplify a clear trend of moving beyond conventional event-camera designs. This means switching to a highly adaptive, energy-efficient, and genuinely edge-capable neuromorphic vision platform, where sensing and visual perception reasoning are seamlessly fused at the hardware level.

2.4. Development of the Infrared Neuromorphic Vision Sensor

In the field of neuromorphic chip research, the primary emphasis has been on the visible waveband. However, there are certain situations when objects of interest become challenging to perceive due to fluctuations in scene illumination. This problem becomes more pronounced when the photons of interest are not emitted in the visible waveband, such as during nighttime or when the atmospheric conditions are unsuitable for the visible waveband.

To this aim, one approach is to shift or extend the measurable light spectrum toward the infrared region [45]. These sensors are typically categorized based on the wavelength range they are sensitive to, which includes short-wave infrared (SWIR), mid-wave infrared (MWIR), and long-wave infrared (LWIR). Each sensor type has its advantages and is suitable for different applications. SWIR cameras typically operate in the wavelength range of 1 μm to 2.5 μm. SWIR sensors can distinguish between organic and inorganic materials, making them ideal for various applications, including recycling, the food industry, drones in agriculture to detect water shortages, military applications involving lasers, and improving atmospheric transmission. MWIR cameras usually use a wavelength range of 3 to 5 μm. MWIR sensors are known for their ability to detect thermal radiation emitted by objects at high temperatures, making them suitable for airborne and ground-based surveillance, thermography, and gas detection applications. LWIR cameras typically operate in the wavelength range of 8 to 14 μm, commonly referred to as the thermal imaging region, as it allows the detection of the thermal radiation emitted by objects or materials at ambient temperature. LWIR sensors suit thermal imaging, night vision, and medical diagnostics applications. See Table 4 for a comparison of infrared neuromorphic cameras by wavelength.

Posch et al. [46] developed the first IR event-based sensor by coupling a microbolometer array with typical DVS readout circuitry [47]. A microbolometer is a thermal sensor that detects thermal infrared radiation based on the variation in its temperature-dependent electrical resistance, and it is sensitive in the LWIR range. It can be integrated with complementary metal–oxide–semiconductor (CMOS) readout circuitry. However, the time constant of current microbolometers (around 10 ms) is relatively slow and does not allow us to take full advantage of the NC technology. Alternative IR technologies, such as cryogenic IR quantum sensors for MWIR and LWIR or InGaAs for the SWIR region, seem more promising.

Furthermore, SCD is announcing the preparation of a product in SWIR [48]. This product, with a resolution of VGA (15 μm pitch) for the imaging mode and quarter-VGA for event-based output, should be available shortly under the name SWIFT-EI. In addition, DARPA has launched the FENCE project [49] to develop event-based infrared cameras sensitive to the infrared band above 3 μm, supposedly including MWIR and LWIR.

Table 4. Comparison of infrared neuromorphic cameras by wavelength.

Category	Wavelength	Capability	Applications
Short-wave (SWIR)	[1 μm–2.5 μ]	Organic vs. inorganic	Recycling, food, agriculture, and military
Mid-wave (MWIR)	[3 μm–5 μ]	Thermal radiation	Surveillance, thermography, and gas detection
Long-wave (LWIR)	[8 μm–14 μ]	Thermal radiation and ambient temperature	Imaging, night vision, and medical diagnosis

2.5. Main Characteristics of Event Cameras

The unique design and operational characteristics of neuromorphic vision sensors offer several advantages over traditional vision technologies. Here, we list the most remarkable:

- **High temporal resolution**: NCs can capture fast-moving objects and obtain greater detail of the evolution of the motion without having to interpolate between frames. Light intensity change is detected by analog circuits with high-speed response. Then, a digital readout with a 1MHz clock timestamps the event with microsecond resolution [3].

- **Low latency**: Event cameras have low latency, meaning they can respond quickly to environmental changes. Contrary to the traditional camera, NCs do not have a shutter, so there is no exposure time to wait for before transmitting brightness change events. Therefore, latency is often tens to hundreds of microseconds under laboratory conditions to a few milliseconds under real conditions [3].

- **High dynamic range**: NCs can capture bright and dark scenes without losing detail. This property is particularly beneficial for sudden changes in illumination that can cause overexposure or in low-light environments where the scene may appear too dark. This property is due to the logarithmic response at the photoreceptors. Hence, whereas static vision sensors have a dynamic range limited to 60 dB because all the pixels share the same measurement integration time (dictated by the shutter), event cameras can go over 120 dB for their independent pixel operations.

- **Low power**: NCs consume much less energy than traditional cameras, making them suitable for low-power devices and applications. Notably, all pixels are activated independently based on the illumination changes each one detects, and the analog circuit is very efficient. As a result, the NC's power demand can go as low as a few milliwatts.

- **Sparsity**: NCs provide data only when there is a change in the scene. Hence, the amount of data that needs to be processed is reduced. As a result, the event cameras may output up to $100\times$ less data than traditional cameras with similar resolution. To fully exploit this technology, NCs must be coupled with chips capable of processing the events with algorithms such as spiking neural networks designed to maintain the low-power premises and keep events' intrinsic asynchronous nature intact. SynSense is one such company that develops neuromorphic chips, such as the DYNAP-CNN chip, for ultra-low-power applications, e.g., IoT devices, and can be integrated with the same chip with the DVS pixel array, as demonstrated by IniVation Speck.

3. Working with Stream of Events

In this section, we explore various representations of the event stream and how they are processed for specific applications. Figure 5 illustrates the trend in event representation methods over time, based on a search with target queries in the Dimensions.ai database, highlighting the increasing popularity and development of these methods. This chart provides a quantitative view of the evolution of event representations, as discussed in the subsections of this section.

The intrinsically different nature of the neuromorphic sensor compared to traditional cameras necessitates new approaches to represent the information captured and to process it in a format suitable for input into specific image-processing algorithms. Neuromorphic cameras (NCs), also known as event cameras, trigger asynchronous events for those pixels that detect a brightness change exceeding a certain threshold—caused either by camera motion or by moving objects in the scene. As a result, the sensor outputs so-called events

that encode not only the spatial location of the pixel (*x* and *y* coordinates) but also the polarity of the brightness change (positive or negative) and a precise timestamp.

An example of such an event stream is shown in Figure 6, captured alongside fixed-rate grayscale frames. This visual juxtaposition highlights the drastic difference from the classic image input format.
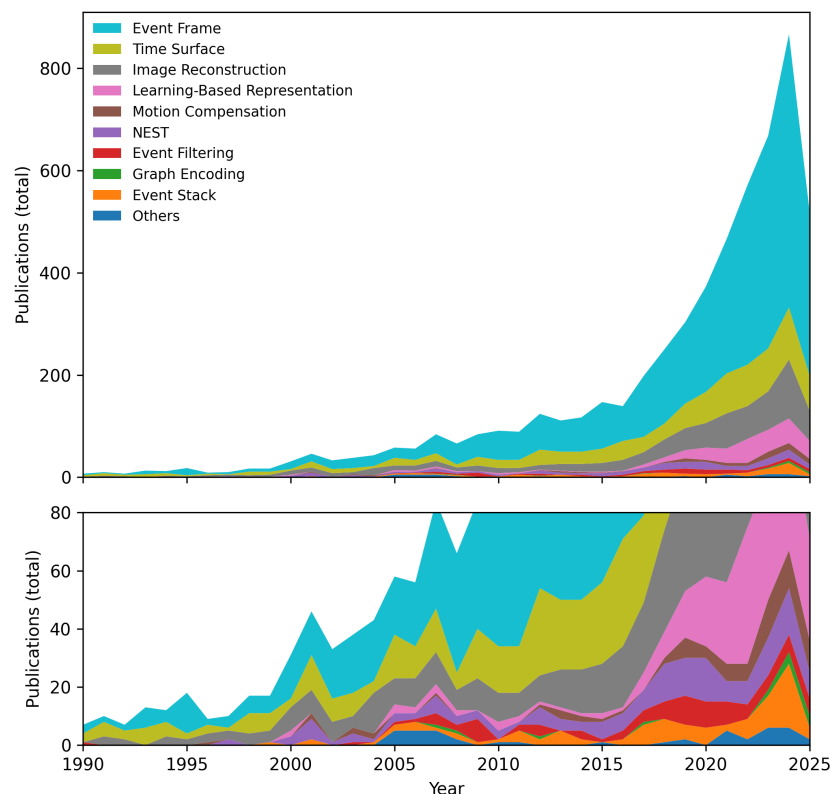

Figure 5. Event representation trends over time. The figure's legend entries are ordered by descending number of publications in 2025. "Others" includes voxel grid, Spike Tensor, TORE volume, and Cloud Encoding.
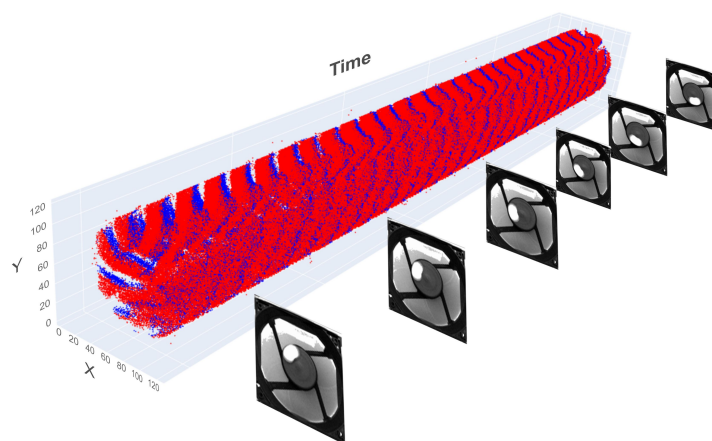

Figure 6. The stream of events as blue points interleaved with grayscale image frames at a fixed rate. While ON events are represented in red, OFF events are represented in blue. Image credits to [50].

This fundamentally different data format calls for new analysis methods. One direction is to develop algorithms that operate directly on the sparse and asynchronous event stream. Alternatively, the event stream can be converted into more conventional representations, which require fewer or no modifications to existing algorithms. Another design choice is

whether to process individual events to minimize latency or to group events into packets, which can then be transformed into other formats and offer more contextual information. In either case, prior context must be considered, as a single event alone lacks sufficient information [51].

To translate the distinctiveness of each preprocessing strategy to explicitly grouped families, Table 5 contrasts them by *what they preserve vs. discard*, typical tasks, and core limitations. Methodologically, we normalize each representation family along five axes: (i) the family itself, (ii) which signal properties are *preserved* (temporal precision, polarity, sparsity, local spatial context), (iii) which are *lost* (e.g., microsecond timing, long-term history), (iv) typical downstream tasks, and (v) core limitations. This axis set is distilled from recent surveys (e.g., [9,15,16]) and the present review, so that readers can understand which preprocessing is more appropriate for the use case and navigate the approaches, even though they might have been treated with slightly different naming by different authors.

Table 5. Comparative analysis of event representations/preprocessing methods.

Method Family	Preserved Properties	What is Lost	Typical Tasks	Limitations	Representative References
Event-by-Event Processing	Max. temporal precision, polarity, sparsity	Local spatial context (unless modeled)	Low-latency tracking, control, VO, gesture	High per-event cost, poor GPU batching	[33,52]
Temporal Binary Representation (TBR)	On/off activity, memory efficiency, sparsity	Counts, time precision	Gesture/activity recognition	Severe info loss for regression tasks	[53]
Event Frames	Spatial context, polarity	Microsecond timing, sparsity	Optical flow, stereo, CNN steering	Motion blur for long windows	[54,55]
Time Surfaces (SAE, TS, SITS)	Newest timestamp per pixel, sparsity	Older timestamps	Corner detection, vSLAM	Normalization tuning; slow/low-light degradation	[56–58]
Voxel Grids	Polarity, temporal bins	Intra-bin timing	DL pipelines, flow, depth	Quantization, memory grows with bins	[59–61]
TORE Volumes/ Temporal Queues	K-recent timestamps/pixel	Older than K	Denoising, classification	Extra memory/bandwidth for K	[62]
Graph/ Event-Cloud Encodings	Full sparsity, exact timestamps, polarity; local topology	Dense-grid regularity (less GPU-friendly)	GNN-based classification	Graph construction cost/fixed size	[63,64]
Learned Grids (ESTs, Matrix-LSTM, NEST, GWD)	Task-optimized mix of timing, polarity, spatial context	Some interpretability, ultra-low latency	SOTA detection, reconstruction, SR, deblurring	Needs data, more compute	[60,61,65–67]

3.1. Event-by-Event Processing

Among the methods that process event-by-event, we recognize probabilistic Bayesian filters, e.g., Kalman filters (KFs) or particle filters (PFs), and spiking neural networks (SNN). Bayesian filtering is a statistical method that relies on Bayesian theory to maintain a probability distribution over the possible states and update this distribution as new data becomes available. Examples of this approach are found especially concerning pose estimation [33,68]. Gallego et al. [69] demonstrate the tracking of the 6-DoF pose of a DVS camera from an existing photometric depth map. Kim et al. [70] provide accurate rotational motion tracking while reconstructing high dynamic range spherical mosaic views from

gradient images using Poisson solvers [71]. Scheerlinck et al. [72] introduced a continuous-time formulation for intensity estimation and fusion of events with image frames using a complementary filter. Additionally, the paper provides a new dataset for evaluating image reconstruction. Later, they proposed a method to compute the spatial convolution of a linear kernel with the output of an event camera, using an internal state that encodes the convolved image information and demonstrates the application of the proposed method to Harris corner detection [73].

Instead, the spiking neural network (SNN) [52] is a type of neural network that models the behavior of biological neurons and the communication between them using discrete spike signals. Hence, they can process asynchronous inputs by encoding the spike's timing. Typical use of these networks involves character recognition, as demonstrated in [74], where an SNN architecture named HFirst exploits the event's temporal information, integrating Integrate-and-Fire neurons with a Winner-Take-All selection strategy. While in HFirst, the network comprises handcrafted kernels such as the Gabor filter [75], the SLAYER algorithm [76] demonstrates how to handle the non-differentiable nature of the spike signal. It performs a modified backpropagation to learn the weights and axonal delay parameters of SNNs. The SNN's asynchronous and sparse spiking pattern can be exploited by specific neuromorphic hardware, such as the Intel Loihi [77], to achieve highly power-efficient models compared to the traditional Deep Neural Networks (DNNs) running on GPUs. Unfortunately, this type of hardware is not yet commercially available, so we have to rely on conventional chips, on which SNNs do not have the same energy efficiency properties. Therefore, converting packets of events into 2D or 3D representations is often convenient to process by computer vision algorithms that can better use the currently available hardware resources. In the following, we will explore the most common ways to preprocess event packets and transform them into a format that allows standard image-processing algorithms to analyze them.

3.2. Event Frame

One of the first 2D representations of event packets is the event frame, which helps process event streams using traditional computer vision techniques, algorithms, and tools not explicitly designed for event-based data. Furthermore, the event frame representation can be used to visualize the events in a way familiar to human observers. In this representation, the events accumulate over time and are used to update a brightness increment image. The advantage of event frames is that their frame rate can be adapted to the use case. However, they have severe limitations compared to other representations, such as time surfaces or voxel grids, in capturing the temporal dynamics of the events. Examples of event frame applications include optical flow [54], stereo vision [78], and deep learning applied to steering angle prediction [55].

3.3. Temporal Binary Representation

A particularly compact and efficient event aggregation strategy is the Temporal Binary Representation (TBR) [53]. TBR first stacks short-interval binary event frames (presence or absence of an event per pixel per time slice) and then losslessly encodes these as decimal values via binary-to-decimal conversion. This enables preserving fine-grained temporal information over a longer window in a single, compact frame while allowing standard CNNs to process the result. While TBR is highly effective for gesture/activity/action recognition with low memory and computation, the loss of precise event count and timing can limit its suitability for tasks such as optical flow or fine time-resolved regression.

3.4. Time Surface and Surface of Active Events (TS/SAE/SITS)

Another popular representation is the time surface (TS) [56,57]. A TS is a spatiotemporal representation of an event and its surrounding activities that uses the arrival time of events from nearby pixels. It is a 2D array where each pixel stores the time of the most recent event at that location, with the pixel's intensity indicating the event's time. Time surfaces are a time-resolved version of an image and can be used to analyze the dynamics of an event stream over time. Recent events are emphasized over past events using an exponential kernel, and normalization is used to achieve invariance to motion speed. Each pixel value can be computed by filtering events within a space–time window to reduce sensitivity to noise. However, time surfaces compress information by keeping only one timestamp per pixel, which can reduce their effectiveness in scenes with frequent events or textures.

Surface of Active Events (SAE) [79] is similar to the time surface but typically focuses on the active regions within the event stream. While the time surface is a direct encoding of the most recent event per pixel, SAE can highlight areas of activity and capture faster dynamic events more effectively. This makes it particularly suited for tasks like corner detection and visual SLAM, where distinguishing rapid changes is crucial.

In addition to these, the Speed-Invariant Time Surface (SITS) [58] is an extension of the time surface that addresses the limitation of sensitivity to object speed in typical TS representations. SITS introduces a novel formulation that makes the time surface invariant to object speed, which is crucial for corner detection and motion tracking, especially under fast and abrupt motion. This approach allows the surface to remain robust even when the direction of motion changes rapidly, a typical challenge for event-based cameras in real-world conditions. Thus, while TS and SAE capture the basic temporal and spatial relationships in event data, SITS adds an extra layer of speed invariance, offering enhanced stability and performance for dynamic and fast-moving scenarios, such as corner detection and real-time motion tracking.

3.5. Voxel Grids and Point Sets

Voxel grids involve dividing a 3D space into a regular grid of voxels, essentially 3D pixels, associated with a value representing the features or characteristics of the object or scene at that location. For example, each voxel contains the number of events within a spatiotemporal volume in the event-camera context. The temporal dimension is discretized in multiple bins, and the voxel value is found by bilinear interpolation [59]. Voxel grids help represent volumetric data in a structured way that neural networks can process efficiently.

A similar approach represents events directly as 3D point sets, where each point is associated with an event fired at a particular time. For example, Benosman et al. [79] employ this representation to estimate motion velocity as a vector proportional to the slope of a plane fitted on the set of points.

3.6. TORE Volumes

Event cameras generate asynchronous, sparse event data that must be transformed into a format usable by conventional computer vision algorithms. Traditional representations, such as voxel grids, often suffer from information loss due to temporal binning. The *Time-Ordered Recent Event (TORE) volumes* [62] aim to address this limitation by retaining the *K* most recent events (timestamps and polarity) for each pixel, thereby maintaining the high temporal fidelity of the original event stream. This bio-inspired approach utilizes a compact 4D buffer that efficiently stores recent events without resorting to fixed temporal windows. TORE volumes preserve the sparse nature of event data, making them computationally efficient and suitable for GPU processing. They have demonstrated significant improvements

across a range of tasks, including denoising, event-based image reconstruction, and object classification, and have become a promising alternative to discretized voxel stacks in deep learning pipelines.

3.7. Graph/Event-Cloud Encodings

Traditional event-processing approaches often convert sparse event data into dense-grid-like representations, which can lose valuable temporal and spatial precision. Instead, graph/event-cloud representations aim to preserve this sparsity and fine-grained temporal resolution by treating each event (or group of events) as a node in a dynamic graph. In this approach, events are represented as nodes with attributes such as spatial coordinates, temporal information, and polarity, and edges are established between neighboring nodes based on their spatiotemporal relationships. This enables the use of Graph Neural Networks (GNNs) to operate directly on asynchronous, sparse data, significantly reducing computational costs by only updating relevant parts of the graph as new events arrive [63,64]. This method aligns event data with established point-cloud networks, preserving both sparsity and precise timing, two key advantages of event-based vision systems. However, the trade-off lies in the added graph construction overhead and the need for techniques like fixed-size sampling or clustering to enable efficient batch processing during training.

3.8. Motion Compensation

Motion compensation [69] is a technique that represents events as image frames to reduce the motion blur and visualize sharp edges. It involves accumulating events over a certain period and using them to update an image representation of the scene that considers the motion. The intuition is that event cameras capture how edges move in the scene and are used to align the events that trigger them. Hence, we optimize an objective function called the focus function [80] to find the trajectories that warp the event back to a reference time to maximize the visual sharpness. As a result, the resulting image is sharper, making it more informative and interpretable than a raw event stream. Hence, the feature extraction [81] and visual odometry [82] tasks are easier to approach using the produced sharp-edge map. In addition, motion compensation can be used with other event representations, such as time surfaces [83] or 3D point sets [84].

3.9. Image Reconstruction

Image reconstruction involves obtaining grayscale frames of the scene from accumulated events. If standard camera images are available, they can be fused to add more information and overcome visual defects such as motion blur and limited dynamic range. Some of the most relevant image reconstruction techniques for these types of cameras include the following:

- **Spike-based image reconstruction** [85] involves accumulating spike data over time and reconstructing an image of the scene. One common approach is to use a spike-based reconstruction algorithm that considers the spatial and temporal patterns of the spikes to recreate a picture that closely approximates the original scene.
- **Adaptive filtering** [86] filters out noise and artifacts in data captured by event cameras. Because these cameras capture data asynchronously and at high temporal resolution, there is often a lot of noise in the data that can interfere with image reconstruction. Adaptive filtering techniques use a combination of statistical analysis and machine learning algorithms to filter out the noise and improve the quality of the reconstructed image.
- **Compressed sensing** [87] can reconstruct high-quality images from a relatively small quantity of data using a combination of algorithms and mathematical models.

- **Deep learning** [88] methods entail using neural networks to learn patterns in visual data and generate high-quality reconstructed images. This technique involves training a neural network on a large dataset of visual data and using it to reconstruct images from the sparse data captured by NCs and event cameras. Deep learning has shown promising results in improving the quality of reconstructed images from these types of cameras.

3.10. Learning-Based Representations

Learning-based approaches have emerged to address the limitations of handcrafted event stream representations. Rather than relying on fixed grids or time surfaces, these methods explicitly learn representations that are tailored to the target task, maximizing the retention of relevant visual or temporal information. For example, Event Spike Tensors (ESTs) [61], Matrix-LSTM [60], and Neural Event Stacks (NEST) [65] are all end-to-end architectures that learn optimal spatiotemporal features for classification or reconstruction. In particular, Teng et al. [65] propose Neural Event Stacks (NEST), a novel spatiotemporal encoding that respects physical constraints while effectively capturing motion dynamics. Their learned representation achieves state-of-the-art performance in image enhancement tasks such as deblurring and super-resolution.

Annamalai et al. [89] introduce a deep learning memory surface, which encodes temporal motion history directly from sparse events. Designed for anomaly detection, this representation preserves the asynchronous nature of the data while enabling efficient spatiotemporal analysis. Building on this, Vemprala et al. [90] use event variational autoencoders (VAEs) to handle environmental changes effectively. Schaefer et al. [64] process event data as evolving spatiotemporal graphs, named AEGNN. Unlike previous methods that convert event streams into dense representations, AEGNN treats events as sparse data, updating only relevant parts of the graph.

Guo et al. [91] address the efficient representation of volumetric videos with feature grids and introduce dynamic codebooks for storage optimization. Wang et al. [92] develop an adaptive sampling approach that dynamically selects the most relevant events in the input stream. They also introduce EAS-SNN, a spiking neural network (SNN), to enhance temporal learning by using recurrent connections that preserve context over time. Gu et al. [67] improve event-based video reconstruction by learning contrast-threshold-adaptive parameter representations, addressing issues like blurry outputs and artifacts.

Furthermore, new metrics such as the Gromov–Wasserstein discrepancy (GWD) [66] have been proposed to efficiently select among handcrafted representations based on how well they preserve task-relevant information, without requiring expensive retraining. These learning-based and metric-driven approaches enable fine-grained management of what information is preserved or discarded, at the cost of increased computational complexity and a reliance on sufficient annotated training data.

4. Event Stream Processing Algorithms

In this section, we delve into the algorithms used for processing event streams, focusing on their capabilities, challenges, and applications. Figure 7 illustrates the trends in event stream processing algorithms over time, based on a search with targeted queries in the *Dimensions.ai* database. This chart provides insight into how the research landscape has shifted towards more efficient and specialized processing algorithms.
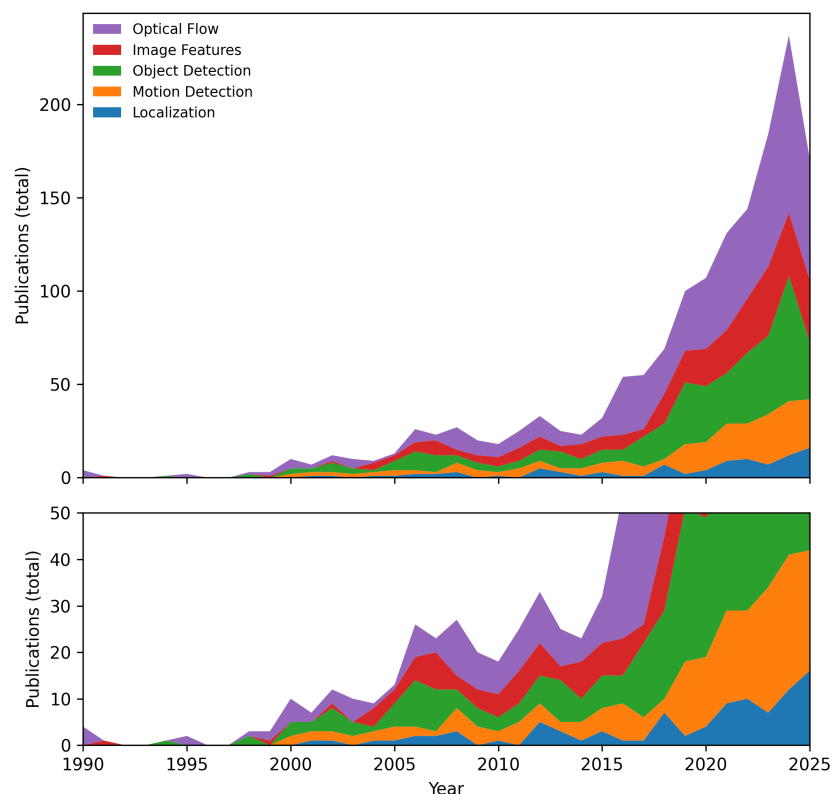
Figure 7. Processing algorithm trends over time. The figure's legend entries are ordered by descending number of publications in 2025.

4.1. Extraction and Tracking of Image Features

Identifying distinctive and informative features in visual data is the first step to further analyzing and understanding the surrounding world through the eyes of the camera. Therefore, extracting significant features is critical to distill sensorial input into condensed information that more complex algorithms can use. In practice, features enable visual tasks for a higher level of comprehension or situational awareness [93], like object recognition, image retrieval, camera localization, or 3D reconstruction. In traditional image analysis, features are extracted from pixel intensity patterns corresponding to geometric structures like corners and edges. Classical methods such as Harris [94], HOG [95], FAST [96], SIFT [97], SURF [98], and ORB [99] enable feature detection and description, ensuring robustness to transformations in scale, rotation, and illumination. These methods rely on dense image frames, where feature vectors encode the appearance of local pixel neighborhoods for matching and tracking.

In contrast, event-based vision requires adapting feature extraction methodologies due to its sparse and asynchronous nature. This necessitates feature detection techniques that leverage the temporal structure of event streams rather than relying on fixed-frame representations. Vasco et al. [100] propose an adaptation of Harris, while [101] advanced a version of the FAST corner detection developed to work on time-surface representations of event streams. Instead, Clady et al. [102] find corners as the intersection of planes fitted on the time surface. Alzugaray and Chli [103] presented an efficient version of eFAST for asynchronous corner detection called Arc. Subsequently, they built the ACE tracker [104] that uses a normalized local region descriptor applied to corners. FA-Harris [105] is a faster corner detection method inspired by the Harris detector. To achieve speed, they introduce a Global Surface of Active Events (G-SAE) unit and corner candidate selection and determine detection scores, showing improved accuracy performance. Li et al. [106] move towards more complex descriptors constructed using the gradient information from

Speed-Invariant Time Surfaces (SITS) [58]. DART [107] uses a log-polar grid to obtain a robust descriptor valuable for object detection and tracking. Recently, deep learning-based descriptors have started to appear. Huang et al. [108] propose a variation of the TS representation, Tencode, that considers polarities and creates a multitemporal resolution input for training a deep network inspired by the Superpoint architecture [109]. Their approach, EventPoint, shows promising results concerning previous methods, e.g., [110], where Harris corners are extracted from predicted image gradients instead.

Extraction and tracking are intertwined tasks, as features are good if we can track them for long frame sequences [111]. Feature tracking refers to establishing correspondences or matches between visual features over time in a sequence of images or video frames through a process usually referred to as data association. Feature tracking is typically performed by detecting the key points in the first frame of the sequence and then matching them with key points in subsequent frames. Matching can be achieved using various techniques, such as nearest-neighbor matching. The objective of tracking is to obtain a model of the motion between frames of the visual features, usually obtained by minimizing an objective function, such as reprojection or photometric error functions.

Matching features detected in consecutive frames is usually achieved using Iterative Closest Point (ICP) [112], as in [113], where large polygonal shapes are tracked. Notably, Tedaldi et al. [114] and Kueng et al. [115] similarly perform tracking with ICP binary templates obtained with the Canny edge detection algorithm [116] centered around Harris corners. In contrast, in previous approaches, the model template is generated from predefined patterns, for example, complex-shaped objects tracked by gradient descent [117] or by multiple kernels, e.g., Gabor filters, feeding a Gaussian tracker [118]. Glover and Bartolozzi [119] use a particle filter to improve over the previous approach applying the Hough transform to track a fast-moving ball [120].

Zhu et al. [121] approach the data association problem with a probabilistic framework that jointly optimizes matching with feature displacements in an Expectation–Maximization scheme. Gehrig et al. [81] propose EKLT that resolves the data association challenge with a generative model to predict the future appearance of generic features. Hence, they use Maximum Likelihood Estimation (MLE) to optimize the warp parameters and brightness increment velocity. While most previous techniques operate on intermediate representations that accumulate events in a traditional frame format, HASTE [122] aims to track on an event-by-event basis. Hence, they revisit a previous tracker formulation [123] with an efficient evaluation of the alignment score function that determines the transition among a discretized space of hypothetical states. Event Clustering-based Detection and Tracking (eCDT) [124] solves detection and tracking simultaneously with a novel clustering method that separates event groups based on the neighboring polarity and spatiotemporal adjacency. Finally, Messikommer et al. [125] trained a neural network to predict displacements by employing a correlation layer.

4.2. Optical Flow

Optical flow [126] is a technique for estimating the motion of objects or scenes in a sequence of images or video frames based on the apparent movement of pixels between frames. Optical flow computes a dense vector field that represents the displacement of each pixel in the image over time. Also, it can be used to track objects or scenes by identifying regions of similar optical flow using clustering techniques, such as mean shift. The task of optical flow is closely related to feature tracking, differing in that it computes a displacement vector for every pixel in the input frame rather than sparse keypoints, regardless of the detection algorithm. However, due to incomplete information, e.g., the lack of knowledge of the scene geometry, ambiguity, noise, and occlusion, the problem is ill-defined and

requires additional constraints. For example, brightness constancy assumptions [127] or a local smoothness prior [128] are usually applied.

Unlike image frames, events do not contain the same amount of information that can be extracted from observing the absolute brightness directly on an image plane. Hence, early methods, such as [79,129–131], start with testing optical flow reconstruction on the simple motion vector field created by a rotating black bar pattern, which triggers events on a continuous spiral in the $x - y - t$ space. Benosman et al. [129] propose an algorithm based on the Lucas-Kanade [128] coarse-to-fine iterative approach by computing the partial derivative over a small neighborhood of events. Later, Benosman et al. [79] refined this approach with an alternative formulation that finds the flow as the slope of a plane fitted on a spatiotemporal region of the event stream. Brosch et al. [130] considers multiple issues with previous approaches, such as the numerical instability of the gradient approximation approach or plane fitting that requires either too small or too many events for robust estimation. Hence, they suggest a methodology that measures velocity as the response of a family of Gabor filters tuned to different velocities and directions by fitting their frequency sensitivity on the experimental data [132]. These early approaches have been compared on a common benchmark where the optical flow was generated from a camera rotating on its three axes, and an Inertial Measurement Unit (IMU) was used to generate ground truth from the gyro angular rates [133].

Similarly to the tuned Gabor filter, the SNN in [131] forms layers of neurons responding to eight speeds, eight directions, and on/off events on a 5×5-pixel region. The approach mimics the classic Lucas–Kanade in a bio-inspired framework. In contrast, Paredes-Valles et al. [134] demonstrate learning the neurons' connection parameters from unsupervised data with a hierarchical SNN architecture. To this aim, they introduce a novel adaptive mechanism for the Leaky Integrate-and-Fire neurons and a stable implementation of the Spike-Timing-Dependent Plasticity (STDP) learning protocol. Additionally, they released the code for simulating a large SNN on GPU-accelerated hardware in an open-source library, cuSNN. More recent approaches, e.g., [135–137], have drastically improved the accuracy performance either by combining artificial neural networks for extending to deeper layers or by adopting more complex architectures [138].

Instead of computing flow on the raw events, Bardow et al. [139] propose jointly estimating the image log intensity with the velocity field in a sliding window variational optimization scheme. Besides demonstrating high dynamic range frame reconstruction, this approach can obtain a dense optical flow field. However, in areas where events have not been received, the optical flow is less reliable, as it results only from the constraints in the optimization equations, such as smoothing regularization terms.

Contrary to Lucas–Kanade-inspired works, Liu and Delbrück [54] propose a method based on block matching, a technique widely used for video compression. They extend their previous FPGA implementation [140] with more efficient computations for real-time operation. Remarkably, they accumulate 2D histograms of events in three adaptive time slices that are continuously rotated. Then, they found the best matching block of a region centered around an incoming event using the Sum of Absolute Difference (SAD) function. Subsequently, Liu and Delbruck [141] proposed a further improvement based on a novel corner detection algorithm implemented in hardware, SFAST, which allows for skipping computations for non-keypoint events.

Furthermore, deep learning has been applied to leverage the large availability of data. Due to the lack of ground-truth optical flow in the event domain, initial work approached the problem following a self-supervised learning paradigm [59,142,143]. They adopt models, e.g., U-Net, and loss functions from the standard camera deep learning research that can learn the 3D structure and the motion of the camera together with the flow. More-

over, the diverse nature of events requires finding the input representation that preserves the most information [144]. Hence, while exploring slight variations in the input format, recent methods introduce correlation cost volumes [145], recurrent units [146], and transformer blocks [147], usually in an encoder–decoder architecture fashion. More recently, BlinkSim [148], a simulator of actual event data and optical flow ground truth based on the Blender 3D engine, has been released, allowing further tuning of deep learning models.

4.3. Camera Localization and Mapping

Estimating a camera's 6-DoF pose is fundamental to enabling autonomy in robotics and vision systems, underpinning tasks such as navigation, mapping, and interaction with the environment [93]. When both mapping and localization occur simultaneously in an unknown environment, the task is referred to as Simultaneous Localization and Mapping (SLAM) [149,150].

Although SLAM can rely on various sensors, including LiDAR, IMU, RADAR, or even radio-based methods [151], event cameras naturally align with visual SLAM (VSLAM). Their low latency, high temporal resolution, and robustness to motion blur and lighting changes make them attractive alternatives. However, the asynchronous data stream they produce challenges conventional SLAM pipelines, which typically assume a fixed frame input.

Event cameras output sparse, high-frequency brightness changes rather than global frames. As a result, SLAM algorithms must be restructured to handle this format, often using feature-based or direct methods. Many systems represent the scene with semi-dense edge maps co-estimated with camera pose, leveraging the fact that events are primarily triggered by edge motion [152–154].

Early event-based SLAM systems were predominantly feature-based, extracting and tracking corners or lines to estimate motion and build sparse 3D reconstructions [70,155,156]. Corner detectors such as eHarris [100], eFAST [101], and FA-Harris [105] were adapted to event data but often struggled with noise and motion variation. More recent methods improved stability by incorporating learning-based feature extractors, including recurrent networks and time-surface representations. Line-based tracking also added geometric constraints, and feature positions were typically optimized using probabilistic filters or bundle adjustment [82,157].

Direct approaches avoid explicit features and instead align event data with geometric or photometric models. Common strategies involve transforming events into image-like representations such as time surfaces and then aligning them against known scene structure or intensity maps [70,158]. Bayesian filtering is often used for incremental motion estimation, while methods like EVO [154] align event images with semi-dense maps. EMVS [159] introduced an efficient back-projection method to accumulate events in 3D space and recover depth from multiple viewpoints.

To improve performance in low-texture regions or during fast motion, many systems integrate IMU measurements. Visual–Inertial Odometry (VIO) pipelines such as Ultimate SLAM [160] or ESVIO [161] fuse event and inertial data, often using continuous-time trajectory models. Stereo event cameras have also been employed to recover depth through temporal and spatial consistency [162,163], while RGB-D setups like DEVO [164] combine event streams with depth sensors to enhance mapping fidelity.

Motion compensation remains key to improving spatial coherence. Techniques such as contrast maximization [152] or event-cloud alignment aim to sharpen accumulated events, supporting robust tracking even under fast motion or extreme lighting.

Loop closure and long-term consistency, while less explored, are gaining traction. Recent work applies spatiotemporal descriptors and graph-based optimization to reduce drift and improve global accuracy.

Deep learning has also become central to event-based SLAM. Early self-supervised approaches by Zhu et al. [59] and Ye et al. [143] showed that depth, optical flow, and ego-motion can be learned jointly from voxel-grid or time-surface representations. These models typically use CNN encoder–decoders trained with photometric or warping losses.

Subsequent work improved monocular depth estimation. Hidalgo-Carrio et al. [165] used recurrent CNNs to accumulate spatiotemporal information and predict dense depth from events alone. EMoDepth [166] refined this with a cross-modal training strategy: using aligned frames only during training while operating with events alone at inference, achieving state-of-the-art accuracy on MVSEC and DSEC.

Pose relocalization also benefited from deep models. CNN–LSTM networks [167] and transformer-based approaches like AECRN [168] exploit entropy-based event representations to regress 6-DoF pose. PEPNet [169] introduced a point-based model that processes raw event streams as 4D point clouds, outperforming prior work while remaining lightweight. Spiking neural networks (SNNs) have been explored for their potential efficiency on neuromorphic hardware. Spike-FlowNet [135] combined ANN and SNN layers for optical flow, while a fully spiking approach by Hagenaars et al. [170] achieved comparable results with much lower energy cost. Although many learned methods still rely on auxiliary frames or depth maps during training, the trend is moving toward fully event-driven models. Progress in spatiotemporal event representations—such as entropy frames, voxel grids, or point clouds—alongside attention modules, recurrent encoders, and spiking networks, is making real-time, frame-free SLAM increasingly feasible.

For evaluation, most approaches rely on public benchmarks such as MVSEC [171], DSEC [172], the IJRR Event Camera Dataset [173], and M3ED [174].

Event-based SLAM remains an evolving frontier. While feature-based and direct methods offer complementary strengths, major challenges persist in scalability, robustness, and fusion. Continued development of hybrid pipelines, neuromorphic hardware, and self-supervised learning is likely to drive future advances in autonomous event-based systems.

4.4. Moving Object Detection

Motion detection clearly highlights the advantages of neuromorphic photoreceptors compared to standard cameras. Thanks to their event-driven nature, neuromorphic sensors offer higher temporal resolution and faster responses, providing a more efficient way to detect moving objects. Unlike standard cameras, which rely on sequences of intensity frames and indirect measurements (such as optical flow), event cameras directly sense motion as changes occur in the scene. Under constant lighting and stationary camera conditions, segmenting moving objects becomes relatively straightforward, as only moving elements trigger events [175]. However, when the camera itself is moving, separating object motion from the camera's ego-motion becomes more complex.

Initial efforts to tackle this challenge relied on classical computer vision techniques adapted to neuromorphic sensing. For instance, Glover and Bartolozzi [120] successfully tracked a fast-moving ball with an event camera mounted on the iCub robot by integrating Hough-transform circle detection with optical flow techniques, achieving robust detection at 500 Hz despite significant background clutter caused by robot movement. Similarly, Vasco et al. [176] leveraged the joint velocities of the robot to distinguish the motion of independent objects from the motion induced by the camera, effectively tracking the general shapes of objects.

To improve robustness under ego-motion, researchers explored motion-compensated representations of event data [69], which align events into sharp images by estimating and removing the camera's motion. This approach enabled Mitrokhin et al. [84] to detect moving objects through motion inconsistencies and was further extended by Stoffregen et al. [177], who introduced a clustering method that jointly estimates object motions to refine segmentation results.

As deep learning entered the field, early models were adapted specifically for object detection using event data. Cannici et al. [178] proposed YOLE and fcYOLE, two neural architectures designed to process events either through integrated surfaces or in a fully asynchronous manner. These models demonstrated the feasibility of adapting frame-based convolutional techniques to sparse event streams. Building on these ideas, Liang et al. [179] introduced GFA-Net and CGFA-Net—transformer-based detectors evaluated on the EventKITTI dataset that combine local feature extraction with global context through edge-aware position encoding.

Expanding on these foundational approaches, Mitrokhin et al. [180] presented a more integrated neural network-based pipeline for motion segmentation. Their model simultaneously estimated depth, ego-motion, segmentation masks, and object velocities. They also introduced the EV-IMO dataset, providing detailed pixel-wise annotations in challenging indoor scenes. Later, the EVIMO2 dataset [181] expanded these benchmarks with greater complexity and more extensive annotations, facilitating robust training for both supervised and semi-supervised methods.

In parallel, neural architectures were refined to exploit the asynchronous nature of events better. For instance, spiking neural networks, previously explored for optical flow tasks, have also shown promise in segmentation. SpikeMS [138] applied a deep spiking encoder–decoder architecture to motion segmentation using DVS input, achieving performance comparable to artificial neural networks while significantly reducing energy consumption. Recent approaches, like the Recurrent Vision Transformer (RVT) by Gehrig and Scaramuzza [182], began employing transformer architectures to fully leverage event data's temporal and spatial properties. The RVT reached state-of-the-art results on automotive detection benchmarks (Prophesee GEN1), achieving extremely low latency detection and demonstrating that transformer models could significantly enhance event-based object detection.

To address complex outdoor scenes where ego-motion plays a dominant role, methods like EmoFormer by Zhou et al. [183] have emerged. EmoFormer cleverly uses events only during training to inject strong motion awareness into a segmentation network, which then performs segmentation using only standard images at inference. They introduced the DSEC-MOS dataset, providing pixel-wise motion annotations for driving scenarios and addressing a critical gap in available training data. A complementary approach by Georgoulis et al. [184], called "Out of the Room", explicitly compensates for ego-motion using monocular depth estimation before segmenting independently moving objects, further setting new benchmarks on EV-IMO and DSEC-MOTS datasets.

Given the difficulty and cost of labeling event data, recent methods also explored unsupervised or semi-supervised strategies. Un-EvMoSeg by Wang et al. [185] introduced an entirely unsupervised method using geometric constraints to detect independently moving objects without needing labeled data, achieving competitive results compared to supervised approaches. Similarly, LEOD, proposed by Wu et al. [186], uses pseudo-labels and temporal consistency to train detectors effectively with minimal supervision, demonstrating strong results with very few annotations.

Beyond neural networks, researchers have also drawn inspiration from biology. The retina-inspired Object Motion Sensitivity (OMS) framework, initially introduced

by Snyder et al. [187] and further advanced into a low-overhead segmentation algorithm by Clerico et al. [188], emulates retinal circuits to discriminate object motion from ego-motion. This bio-inspired approach enables lightweight, efficient segmentation of moving objects directly from event-camera data, avoiding the need for explicit ego-motion compensation. Another non-learning approach, JSTR by Zhou et al. [189], combined IMU measurements and geometric reasoning to segment moving objects effectively, showcasing robust results without relying on heavy learning frameworks.

Hybrid methods combining event data with other modalities, particularly RGB frames, have also proved valuable. For instance, RENet [190] fuses event data and standard RGB images using attention mechanisms, greatly improving object detection accuracy under diverse conditions, including challenging lighting and rapid motion scenarios. Another notable hybrid approach, FlexEvent, introduced by Lu et al. [191], focuses on adapting object detection to arbitrary event frequencies. It combines event data with RGB frames using an adaptive fusion module (FlexFuser) and a frequency-adaptive learning strategy (FAL), achieving robust object detection performance across frequencies ranging from 20 Hz up to 180 Hz. This flexibility makes it particularly suitable for dynamic, real-world scenarios where event rates vary significantly.

As new datasets expand the range of evaluation scenarios, the field steadily bridges the gap between low-level motion cues and high-level scene understanding. With approaches ranging from fully event-based models to hybrid and unsupervised methods, current systems are increasingly capable of accurate, real-time segmentation, even in challenging, dynamic environments.

4.5. Spiking Neural Networks for Event-Based Processing

Spiking neural networks (SNNs) have emerged as an effective computational paradigm for processing the sparse, asynchronous data streams generated by neuromorphic sensors. Their spike-based coding schemes naturally exploit temporal dynamics and enable low-power computation, making them well suited for event-driven tasks. Recent advances have focused on scaling SNN architectures by incorporating design principles from deep learning, notably convolutional residual networks and transformers, to improve representational capacity, training stability, and hardware compatibility.

A core challenge in deep SNNs is preserving temporal features across many layers without experiencing vanishing or exploding gradients. Conventional deep CNNs solve this via residual connections, which protect important temporal information through the layers, enabling deeper network architectures that can efficiently process both spatial and temporal features in event-driven data. However, spiking analogs struggle due to the discrete, non-linear dynamics of spiking neurons. Fang et al. [192] addressed this with the Spike-Element-Wise (SEW) residual block, which adapts the identity mapping to spike-based signals through element-wise operations (ADD, AND, IAND) between the residual and shortcut paths. This design guarantees gradient flow across hundreds of layers, enabling direct training of ultra-deep SNNs (e.g., >100 layers) without degradation. Empirical results show that SEW-ResNet outperforms previous directly trained SNNs in both accuracy and temporal efficiency. Other CNN–SNN hybrids build on these ideas, combining spiking convolutions with temporal pooling or biologically inspired kernels to further enhance spatial–temporal feature extraction. The inclusion of residual connections has been shown to significantly improve the performance of spiking CNNs in tasks like dynamic scene recognition and object tracking [192].

While CNN-based SNNs excel at local spatial–temporal feature extraction, they lack global context modeling. Transformer [193] models have gained attention for their ability to capture long-range dependencies in event data. A significant challenge in transpos-

ing these models to event data is the design of spike-driven self-attention mechanisms that can effectively process sparse and asynchronous spike-based inputs, as conventional self-attention is ill-suited to sparse, asynchronous spikes. Recent works [194–197] have introduced novel spike-based approximations to attention, improving both efficiency and scalability. Yao et al. [194] proposed an efficient spike-driven self-attention mechanism that replaces multiplications with addition-only spike operations, making it hardware-friendly and energy-efficient. Their Spike Firing Approximation (SFA) training strategy bridges the representation gap between ANN and SNN attention layers, enabling large-scale training and competitive performance with ANN Vision Transformers on ImageNet-scale classification, detection, and segmentation. Related works refine spike-based attention for long-range temporal modeling in spikes. The Spiking Transformer [195] introduces an addition-only self-attention (A2OS2A) that removes softmax/scaling and mixes binary/ReLU/ternary units to reduce information loss while maintaining non-multiplicative computation, reaching 78.66% ImageNet-1k accuracy. Spatial–temporal attention [197] integrates time and space within spiking self-attention via block-wise processing at the same complexity, improving static and neuromorphic benchmarks. Hybrid Spiking Vision Transformer [196] targets event-camera detection by coupling ANN spatial modules with SNN temporal modules, showing gains on multiple benchmarks. Spatial–temporal spiking transformers can be inserted after spiking CNN stages to supply global context with bounded complexity, supporting the front-end/back-end split for event streams. Methods for conversion from artificial neural networks (ANNs) to SNNs have also seen significant progress. TTFSFormer [198] leverages Time-To-First-Spike (TTFS) coding to transfer pretrained ANN transformers into the spiking domain with minimal accuracy loss, reducing the need for costly retraining. These advancements enable spiking transformers to achieve state-of-the-art performance in tasks such as event-driven object detection and sequence modeling by accurately capturing temporal dependencies in complex event streams.

SNN architectures have demonstrated strong results in a variety of event-driven vision tasks: from image segmentation, where Spike2Former [199] performs segmentation on event data, combining high spike efficiency with performance comparable to frame-based methods, to dynamic scene super-resolution, with examples such as Spk2SRImgNet [200], which addresses the problem by leveraging motion-aligned collaborative filtering techniques. Beyond these task-specific advances, the development of dedicated datasets further supports the evaluation and comparison of SNN-based methods. Datasets such as UCF-Crime-DVS [201] provide valuable benchmarks for anomaly detection, complementing established neuromorphic datasets (e.g., Shifted MNIST [178], DVS Gesture [202], N-Caltech101-DVS [203], CIFAR10-DVS [204]) for systematic model evaluation.

Despite these advances, key challenges remain: efficient large-scale training, improved interpretability, and the integration of CNN and transformer strengths into unified spiking architectures. However, the strengths of SNNs, particularly their temporal dynamics, sparsity-driven efficiency, and compatibility with neuromorphic hardware, position them as key enablers for future event-driven sensor technologies. Neuromorphic vision chips such as the Speck and Tianmouc [43,44] platforms underscore that algorithm–hardware co-design for spiking vision is feasible today and motivates SNN backbones tailored to such chips. Continued progress in spiking architectures, including hybrid models that combine the strengths of CNNs and transformers, and the development of more efficient training techniques will be crucial for unlocking the full potential of SNNs in real-time, event-based tasks.

5. Applications

This section discusses the various applications of event cameras across different domains, emphasizing their advantages and challenges in real-world settings. Figure 8 below shows the trend in applications over time, based on a search with specific queries in the *Dimensions.ai* database. It reflects the growth and diversification of event-camera applications in fields such as healthcare, autonomous driving, and surveillance.

The principal qualities of event cameras, namely, microsecond-level temporal resolution, low latency derived from independent pixel circuits, minimal power consumption (around 1 mW), reduced memory usage due to sparse output, and high dynamic range for visual sensing under sunlight or moonlight [205], align well with the functional needs of several domains. These include search and rescue, surveillance, autonomous driving [206], traffic monitoring [207], power line inspections [208], Industry 4.0, and situational awareness in space [209–211]. In these scenarios, event cameras enable robust vision despite motion, extreme lighting, or power constraints. For example, high temporal resolution is essential for fast-moving robotics; high dynamic range benefits outdoor and orbital observation; and low power consumption is key in mobile or embedded platforms. Augmented and Virtual Reality (AR/VR) applications, such as eye tracking and gesture recognition [212], benefit similarly from low latency and data sparsity. In Table 6, we associate specific sensor strengths with the application domains where they provide measurable technical benefits and advantages compared to classical vision systems based on insights from the literature review summarized herein.
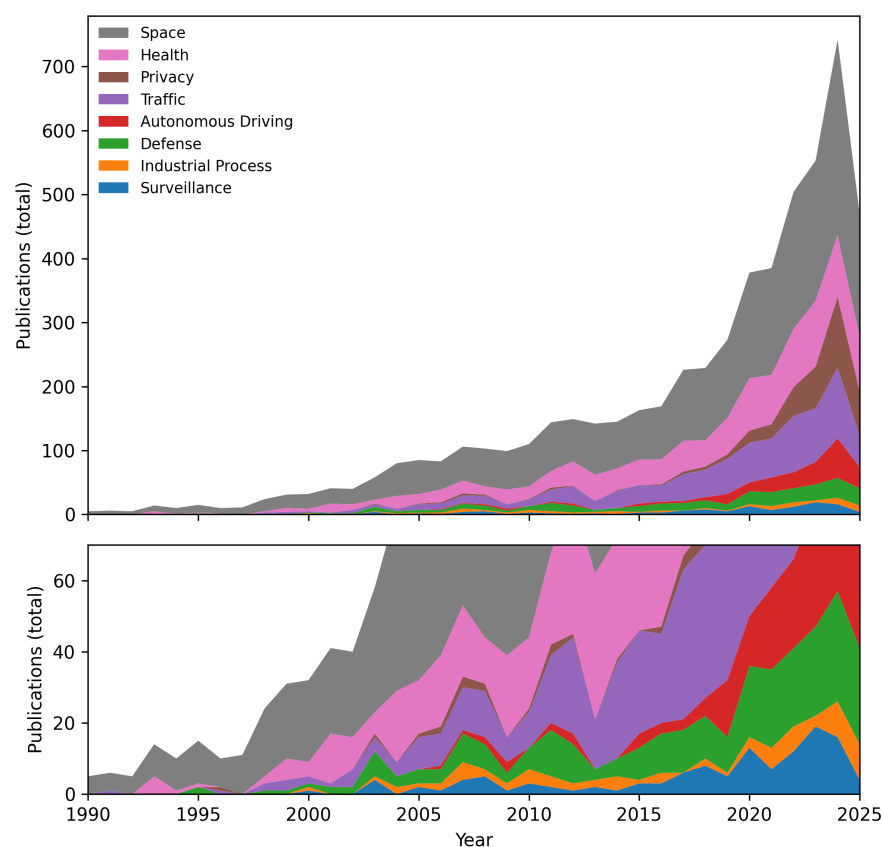

Figure 8. Application trends over time. The figure's legend entries are ordered by descending number of publications in 2025.

Table 6. Event-camera applications by domain vs. strengths leveraged. The checkmarks represent sensor capabilities that are technically beneficial to each application.

Application Domain	High Temp. Res.	Low Latency	High Dyn. Range	Low Power	Data Sparsity	Motion Robustness
Health and Sport Monitoring	✓			✓		✓
Industrial Monitoring	✓	✓	✓	✓	✓	✓
Space Sector	✓		✓	✓		✓
Surveillance and Rescue	✓	✓	✓	✓	✓	✓
Autonomous Driving	✓	✓	✓		✓	✓
Traffic Monitoring		✓		✓		
Defense	✓	✓	✓	✓	✓	✓
Others	✓			✓	✓	

5.1. Health and Sport-Activity Monitoring

Event cameras can capture detailed information about an individual's activity, including detecting falls, tracking movements, and analyzing gait patterns, and could be applied to provide early warning signs of health issues or injuries. Several works have recently been proposed to estimate the human body pose from event-camera measurements, e.g., [161,213–217], a task for which a dedicated dataset has been released [218]. In healthcare, event cameras are being explored for applications such as surgical monitoring and neural imaging, where the ability to capture subtle, fast physiological motions can improve diagnostic accuracy [219,220]. Furthermore, combining high-temporal-resolution events with color images allows for interpolating new frames faster than the original video stream, reducing ghosting and other artifacts caused by non-linear motions [221,222].

5.2. Industrial Process Monitoring and Agriculture

Event cameras are emerging as powerful tools for industrial environments that demand high-speed, high-precision monitoring. Their low latency, high temporal resolution, and robustness to lighting variations make them particularly well suited for real-time quality control, equipment diagnostics, and predictive maintenance. In agriculture, their asynchronous operation and high temporal resolution open new avenues for real-time crop monitoring and precise field management, enhancing precision farming techniques [223].

One illustrative use case is high-speed object counting. For example, Bialik et al. [224] demonstrated a Prophesee EVK1 event camera successfully counting corn grains on a fast-moving feeder line, showcasing the potential of neuromorphic vision in manufacturing and logistics applications.

Beyond object counting, NCs have shown promise in broader industrial process monitoring tasks. For instance, Dold et al. [225] investigated the use of event cameras for laser welding, a domain where conventional photodiodes and high-speed cameras are typically used. Their study demonstrated that event cameras could visualize welding dynamics with superior temporal fidelity and detect production anomalies using learned representations.

In vibration monitoring, a critical task for predictive maintenance and structural diagnostics, Baldini et al. [226] used event cameras to track mechanical vibrations with an accuracy comparable to expensive laser Doppler vibrometers. Their system combined stereo event tracking and video reconstruction (via E2VID [227,228]) to measure subtle displacement patterns at sub-pixel resolution.

These examples reflect the increasing adoption of event cameras for industrial process monitoring, which requires high-frequency observation and fast decision-making from precision manufacturing to large-scale industrial systems.

5.3. Space Sector

Neuromorphic sensors can be applied to telescopes to track stars [209], satellites [229], or debris in orbit [230] from the ground to avoid potential damage to other infrastructures. Recent research suggests that NCs with high spatial and temporal resolution may be exploited to identify the material of satellites [231]. Furthermore, Jawaid et al. [232] leverage the high dynamic range of the event sensor to estimate satellite pose to ensure robustness to drastic illumination changes. Also, Mahlknecht et al. [233] demonstrate that event cameras are suitable for planetary explorations where challenging scenarios, such as the Mars landscape, pose many challenges in estimating the autonomous robot self-position. The International Space Station (ISS) has an event-based sensor to detect lightning and sprite events in the mesosphere. These events can occur in as little as 100 microseconds [234].

5.4. Surveillance and Search and Rescue

NCs are well suited for monitoring public spaces or securing buildings due to their low power consumption and real-time processing capabilities. These features enable them to monitor large areas without frequent maintenance or battery replacements. Furthermore, traditional surveillance cameras can be limited in accurately detecting and tracking objects in complex and dynamic environments where the HDR capabilities of these sensors compensate for lighting variability factors and nighttime operations. Research has shown that NCs can detect and track multiple moving objects in real time, even if many challenges of complex environments must still be addressed [235]. Ganan et al. [236] propose an event-based processing scheme for efficient intrusion detection and tracking of people using a probabilistic distribution and CNN, which has been validated in various scenarios on a DJI F450 drone.

Aerial robots are the primary platform for surveillance and search and rescue applications. Mainly, event-based cameras' high temporal resolution and dynamic range help handle the motion blur caused by UAVs while detecting and tracking possible intruders [237]. Recent work by Rodriguez-Gomez et al. [238] introduces an asynchronous event-based clustering and tracking method for intrusion monitoring in a UAS. Their approach leverages efficient event clustering and feature tracking while incorporating a sampling mechanism to adapt to hardware constraints, demonstrating improved accuracy and robustness in real-world scenarios. Deep learning methods for event-based human intrusion detection in UAV surveillance have also been explored to gain more confidence in determining the type of moving object. Pérez-Cutiño et al. [239] present a fully event-based processing scheme that detects intrusions as clusters of events and classifies them using a CNN to determine whether they correspond to a person. In particular, this method eliminates the need for additional onboard sensors and fully exploits the asynchronous nature of event cameras.

Similarly to surveillance, event cameras are useful in search and rescue operations, especially in environments like forests or mountains, because of their high temporal resolution and low power consumption, allowing them to capture detailed information about the environment and providing real-time feedback to rescuers while moving fast and for extended missions. More importantly, the low latency of NCs can allow remote UAV pilots to perform more aggressive flights [240], which is critical to reducing operation time while safely avoiding obstacles in unknown cluttered environments [241].

5.5. Autonomous Driving

Autonomous driving applications like collision avoidance could benefit from event cameras. With their low latency and high temporal resolution, they can capture detailed information about the environment and provide real-time feedback to the autonomous

driving system. Hence, they will be an essential resource for implementing advanced driver assistance systems (ADASs) and self-driving cars in the future. To this aim, Wzorek and Kryjak [242] recently demonstrated how a neural network could detect traffic signs. Not only that, but event cameras have also been tested on the driver distraction detection task by Yang et al. [243], where the authors evaluated the proposed approach by converting standard video clips with an event simulation tool [244].

5.6. Traffic Monitoring

Event cameras may be helpful for traffic monitoring applications, such as estimating car speed [207]. Their low power consumption and real-time processing capabilities make them well suited for monitoring large areas without frequent maintenance or battery replacements. Therefore, event cameras can detect and track multiple cars on the road simultaneously [245] or pedestrians and cyclists [246].

5.7. Defense

Event cameras offer significant advantages for defense applications due to their low power consumption, high temporal resolution, and ultra-low latency. These features make neuromorphic cameras ideal for embedded systems in UAVs and other autonomous platforms, enhancing obstacle detection, target tracking, and surveillance while maintaining power efficiency. In reconnaissance and battlefield monitoring, NCs provide continuous high-speed data streams that improve situational awareness in real time [247]. Their ability to track fast-moving targets is particularly valuable for Unmanned Ground and Underwater Vehicles (UGVs and UUVs), where reaction time is critical [248].

Recent studies have also explored using NCs for laser warning and detect-before-launch (DBL) capabilities. For instance, Boehrer et al. [249] demonstrate how the high temporal resolution of event cameras can be leveraged to detect laser emissions and retro-reflections from pointed optics, key indicators of hostile intent. Their system was evaluated in operational scenarios during the DEBELA trial, showing that event-based sensing enables early and reliable threat detection. Complementing this work, the DEBELA project [250] investigates electro-optical technologies for future self-protection systems, focusing on within-visual-range missile threats that are difficult to detect using conventional sensors.

NCs also show promise in Counter-Unmanned Aerial Systems (C-UASs), where their ability to capture fast-moving drones or hypersonic missiles can aid early warning systems. Their sensitivity in the infrared and short-wave infrared bands allows for enhanced night vision and detection of low-signature propellants [251,252]. Together, these capabilities position event-based sensors as powerful tools for modern defense, offering real-time threat perception, reduced false alarms, and greater autonomy in decision-making.

5.8. Other Emerging Applications

The range of industries exploiting neuromorphic cameras is unlimited. The last years have produced a handful of proof-of-concept studies that push neuromorphic vision into domains largely absent from the mainstream literature. We outline three particularly promising, yet still underexplored, directions that complement the existing catalog of applications of the survey.

Low latency and high dynamic range permit detection of the subtle chromatic and geometric changes caused by blood flow or respiration. A recent proof-of-concept study showed accurate heart rate estimation from facial events while using dramatically less data and power than frame-based approaches [253]. Public multi-subject datasets and standard evaluation protocols are still missing, making this a ripe area for first-mover contributions.

Because they record only brightness changes, event cameras discard most static texture information that fuels face or scene re-identification. Indoor-localization prototype

algorithms have demonstrated that event-only maps can enable accurate positioning while transmitting almost no personally identifiable visual content [254]. The next step is to formalize threat models and privacy metrics so that lightweight encryption and anonymization schemes can be co-designed with resource-constrained inference.

Ultra-low-power spiking neural network chips are now being paired directly with event sensors in sense–process–act stacks that run entirely on the edge. Field trials in battery-powered drones and smart home nodes report faster response times and substantial energy savings compared to CNN frame-based baselines [255]. Open challenges remain in tailoring event representations to tiny on-chip memories and scheduling heterogeneous cores under single-digit milliwatt power caps.

Finally, neuromorphic vision systems can be integrated into sensor fusion frameworks, combining modalities such as inertial sensors, microphones, or bio-signals to enhance situational awareness. For instance, Kiselev et al. [256] demonstrate a real-time FPGA-based system combining a DVS with a Dynamic Audio Sensor (DAS), achieving significantly higher classification accuracy through multi-modal input. Similarly, O'Connor et al. [257] present a spiking Deep Belief Network that fuses input from a silicon retina and cochlea, achieving robust performance even under sensory noise. These examples highlight how event-based fusion can enrich perceptual systems in fields like mobile robotics, smart wearables, and embedded AI.

6. Discussion

In this section, we delve deeper into the identified gaps in the current stage of development in the field of neuromorphic vision, drawing insights into the directions that research and industry could take to capitalize on the multiple opportunities this sensor offers. Hence, in Table 7, we provide an overview of the gaps and future directions that we analyze in more detail herewith.

Table 7. Gap analysis and future directions overview.

Level	Gap Analysis	Opportunities and Future Directions
	Sensor availability	Lower manufacturing costs, mass production
	Manufacturing complexity	Industrial collaborations (e.g., Prophesee)
Hardware	High sensor cost	Infrared, non-visible spectrum
	Low spatial resolution	Neuromorphic chips for consumer electronics
	Limited spectral range	Edge devices, low power, real-time processing
	Immature event-based algorithms	Improve SNNs, LSTMs
	Lack of universal data representation	Neural Event Stacks (NEST)
Algorithmic	Real-time processing challenges	Transformer models (object detection, etc.)
	Lack of benchmarks	Standardized event-based benchmarks
	Sparse data management	Synthetic event-data generation (e.g., v2e)
	Lab prototypes	Real-world industrial solutions
	Poor performance in dynamic environments	Integration with traditional sensors
Applications	Poor event-based SLAM performance	Event-based SLAM in complex environments
	Limited commercial solutions	Autonomous driving, robotics, surveillance
	Drift, loop closure issues	Low latency, high temporal resolution

6.1. Gap Analysis

Despite considerable advancements in neuromorphic sensors and algorithms, several gaps remain that prevent their widespread adoption and limit their ability to fully replace classical vision sensors.

At the hardware level, the primary limitations are sensor availability, manufacturing complexity, and cost. Neuromorphic sensors remain expensive due to their specialized

manufacturing processes, restricting broad commercial availability. Additionally, current event cameras typically provide lower spatial resolution compared to traditional frame-based sensors, limiting their effectiveness in applications demanding high detail. Another significant hardware constraint is the limited spectral range, with most sensors operating only in the visible spectrum. Although early initiatives like the DARPA FENCE program and recent developments of infrared-sensitive neuromorphic sensors (e.g., SWIR-sensitive cameras) exist, these efforts are still at an early stage, limiting widespread implementation.

Algorithmically, a major challenge arises from fundamental differences between event-based and conventional visual data. Event-based vision algorithms are comparatively less mature and require new data representation methods and processing approaches. Although methods like voxel grids, time surfaces, and event histograms have emerged, a universally accepted approach adaptable across multiple vision tasks is still lacking. The continuous and asynchronous nature of sparse event streams poses significant challenges to developing robust algorithms and represents a substantial paradigm shift from traditional computer vision techniques. Notwithstanding their sparse nature, real-time processing and intelligent clustering of event streams remain challenging, as managing the high volume of events and extracting relevant information is nontrivial. Additionally, benchmarks and standardized evaluation frameworks designed explicitly for event-based data remain limited, impeding progress in algorithm validation.

At the application level, neuromorphic vision systems are primarily limited to laboratory prototypes, with few robust, commercially viable solutions available. Achieving consistent performance in uncontrolled, dynamic environments remains difficult; particularly significant challenges arise from environmental noise such as intermittent lighting variations and sensor-induced noise, requiring more advanced noise filtering methods. For example, critical tasks such as event-based visual SLAM, which is paramount for the future of autonomous driving or other robotic contexts, still struggle with drift reduction, effective loop closure detection, and reliable operation in complex real-world scenarios involving rapid movements or significant scene aspect changes.

Finally, fully exploiting the inherent energy efficiency advantages of neuromorphic sensors in practical deployments demands integration with specialized neuromorphic computing hardware, optimized explicitly for processing sparse and asynchronous event data. Current general-purpose hardware, such as CPUs and GPUs, lacks the efficiency for event-based processing, while dedicated neuromorphic computing platforms that support SNNs remain limited in commercial availability. Achieving widespread industrial use will require concerted efforts toward hardware innovation and software maturity, a challenge most companies are currently unable to tackle without greater standardization and market maturity.

6.2. Opportunities and Future Directions

However, despite the gaps discussed, several promising opportunities exist for further advancing neuromorphic vision technology.

In hardware, key opportunities include reducing sensor manufacturing costs through mass production and strategic industrial collaborations. Recent partnerships, such as that between Prophesee and Qualcomm, which aim at integrating event-based cameras into smartphones, and Google's integration of neuromorphic sensors into Android XR for augmented reality, are paving the way for broader market adoption. Furthermore, neuromorphic chips like SynSense Speck, designed for ultra-low-power and high-speed imaging, can potentially extend event-based sensing to consumer electronics and affordable machine vision solutions. Expanding into infrared and non-visible spectral domains also presents significant potential, particularly for security, defense, and environmental

monitoring applications. Additionally, integrating neuromorphic sensors into edge devices paired with neuromorphic processors, which significantly reduce power consumption and enhance real-time processing capabilities, presents another critical opportunity for practical implementations, especially in energy-constrained environments.

On the event-processing side, improving temporal neural networks remains important, including architectures like spiking neural networks (SNNs) and Long Short-Term Memory (LSTM), which naturally handle time-based event data. A growing body of work also explores learned event-based representations, which encode spatiotemporal patterns in formats better suited to downstream processing, signaling space for improvement in this area. Moreover, transformer-based models, initially developed for language processing [193] and later adapted to traditional computer vision [258], are starting to show potential for event-based vision tasks such as object detection, video reconstruction, and pose estimation. These models effectively capture long-term temporal dependencies in event data, offering advantages over conventional convolutional networks. In this context, sparse-aware transformer designs like the Event Transformer (EvT) [259] further improve computational efficiency by leveraging the unique sparsity of event streams, making them more suitable for real-time, resource-constrained applications.

Developing standardized deep learning benchmarks and datasets specifically for event-based vision tasks is critical to accelerating algorithmic maturity and adoption [15]. Advances in synthetic event-data generation tools (e.g., v2e [244]) that accurately emulate sensor behavior under varying conditions also offer significant potential to facilitate algorithm development and training, reducing dependency on extensive real-world data collection. These tools can further enhance algorithm robustness to environmental factors, such as noise, varying illumination, and complex scenes, by providing an extensive and controllable source of training data for neural network-based methods. Lastly, developing computationally efficient algorithms optimized for specialized neuromorphic hardware accelerators remains essential for enabling practical and widespread adoption.

Regarding applications, moving from laboratory prototypes to real-world industrial solutions remains a significant opportunity. Integrating neuromorphic sensors with traditional cameras and other sensor types (such as IMUs, LiDAR, and microphones) can combine strengths and significantly enhance system performance. In particular, event-based SLAM systems that leverage both neuromorphic sensing and neuromorphic computing represent an immediate opportunity, especially in complex environments where conventional sensors struggle, such as autonomous vehicles navigating dynamic urban settings, drones operating under variable lighting conditions, or robotic systems employed in search and rescue and defense applications.

Furthermore, several less explored application domains could notably benefit from neuromorphic sensors, opening new opportunities for adoption. For instance, agriculture and precision farming can leverage event-based vision, e.g., for real-time crop monitoring. Healthcare applications, particularly surgical assistance, patient monitoring, or even microexpression analysis for telemedicine, could exploit the sensitivity of neuromorphic sensors to rapid and subtle physiological changes. Additionally, applications in sports analytics, such as real-time ball tracking or athlete movement analysis, present another promising use case, given the sensor's ability to precisely track high-speed objects without motion blur. Even in heavy industry, where traditional high-speed cameras are already used for equipment inspection and wear monitoring [260], a transition to event-based vision could improve temporal resolution and data efficiency under harsh, dynamic conditions.

Increased awareness and dissemination efforts are crucial to facilitating industrial adoption. Initiatives like the 4th International Workshop on Event-based Vision at CVPR 2025 and the NeVi 2024 Workshop at ECCV 2024 are already helping to connect academia

and industry by highlighting practical benefits and driving interest in neuromorphic sensors. Similarly, industry-focused events such as the VISION Fair provide valuable opportunities to reach broader industrial stakeholders. Expanding participation in these events, supported by targeted promotional activities and strategic partnerships, will further encourage market adoption and raise industry awareness of event-based vision technologies.

These advances across event-driven hardware, algorithmic innovation, and industrial collaboration underscore the rapid progression and diversification of neuromorphic vision technologies. To contextualize these developments further, the next section synthesizes broader insights and key perspectives highlighted in recent comprehensive reviews [6,17].

6.3. Perspectives on Neuromorphic Computing

Recent comprehensive reviews emphasize that neuromorphic computing is entering a pivotal new phase, driven by cross-disciplinary advancements in hardware, algorithms, and practical deployments [6,17]. Unlike traditional von Neumann architectures, neuromorphic systems physically integrate memory and computation, exploiting event-driven operations, local plasticity, and extreme parallelism to achieve ultra-low-power and real-time performance. These characteristics uniquely position neuromorphic platforms as ideal solutions for edge computing, robotics, distributed sensor networks, and resource-constrained, time-critical applications.

Two complementary trajectories characterize the current evolution of the field. First, neuromorphic hardware is rapidly scaling in both neuron count and functional specialization. Crucially, a prominent trend identified in the latest literature is the tight integration of sensors and computation on single devices, enabling ultra-efficient, event-driven inference directly at the data capture point. Recent implementations, such as the Speck chip [43]—integrating an event sensor with a fully asynchronous spike-based processor—and the biologically inspired dual-pathway Tianmouc chip [44], exemplify this convergence, providing practical demonstrations of real-time, energy-efficient perception in challenging dynamic environments. Second, advances in specialized spiking neural network (SNN) algorithms complement these hardware innovations. Modern SNN architectures, particularly transformer-inspired models and sparse, event-driven learning approaches, are now significantly better aligned with emerging neuromorphic hardware capabilities [6,17]. These algorithmic innovations take advantage of the strengths of sensor-integrated hardware to address real-world challenges such as sparse data management, temporal misalignment, and efficient real-time processing.

However, the community underscores the critical need for tighter hardware–algorithm co-design, standardized benchmarking datasets, and accessible software toolchains. Open questions remain regarding the seamless integration of neuromorphic systems with conventional computing environments, scalable software ecosystems, and tangible demonstrations of application-level advantages compared to traditional AI solutions. The convergence of integrated neuromorphic hardware and sophisticated SNN algorithms represents a defining theme for the coming decade. We expect substantial advancements in adaptive, robust, and energy-efficient applications driven by efficient event-based biologically inspired architectures.

7. Summary

Neuromorphic vision sensors, or event-based cameras, represent a fundamental shift from traditional frame-based imaging. Unlike conventional cameras that record full images at fixed intervals, neuromorphic cameras capture visual information asynchronously by registering changes in brightness at each pixel. This results in sparse, low-latency, and highly efficient data regarding power consumption and storage. In addition, neuromorphic sen-

sors have exceptionally high temporal resolution and dynamic range, allowing them to operate effectively in challenging lighting conditions and rapidly changing environments.

This review systematically covered the key dimensions of neuromorphic vision technology: the evolution of sensor hardware, the specialized algorithms developed to process event-based data, and their diverse applications. Hardware advancements highlighted include sensor architectures that evolved from early silicon-retina concepts to increasingly sophisticated designs that capture richer visual information, including colors and absolute light intensity, at higher resolution. Algorithmically, event-based processing has adapted and extended classical image-processing tasks, e.g., feature detection, optical flow estimation, visual odometry, and object tracking, to handle asynchronous event streams efficiently. Finally, neuromorphic cameras have demonstrated substantial potential in various practical fields, including robotics, autonomous vehicles, industrial automation, and surveillance, taking advantage of their unique capabilities to enhance real-time responsiveness and robustness to environmental dynamics.

References

1. Posch, C.; Serrano-Gotarredona, T.; Linares-Barranco, B.; Delbruck, T. Retinomorphic Event-Based Vision Sensors: Bioinspired Cameras with Spiking Output. *Proc. IEEE* **2014**, *102*, 1470–1484. [CrossRef]
2. Ceccarelli, A.; Secci, F. RGB Cameras Failures and Their Effects in Autonomous Driving Applications. *IEEE Trans. Dependable Secur. Comput.* **2023**, *20*, 2731–2745. [CrossRef]
3. Gallego, G.; Delbrück, T.; Orchard, G.; Bartolozzi, C.; Taba, B.; Censi, A.; Leutenegger, S.; Davison, A.J.; Conradt, J.; Daniilidis, K.; et al. Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 154–180. [CrossRef] [PubMed]
4. Lichtsteiner, P.; Posch, C.; Delbruck, T. A 128 × 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE J. Solid-State Circuits* **2008**, *43*, 566–576. [CrossRef]
5. Boahen, K. Retinomorphic vision systems. In Proceedings of the Fifth International Conference on Microelectronics for Neural Networks, Lausanne, Switzerland, 12–14 February 1996; MNNFS-96. [CrossRef]
6. Kudithipudi, D.; Schuman, C.; Vineyard, C.M.; Pandit, T.; Merkel, C.; Kubendran, R.; Aimone, J.B.; Orchard, G.; Mayr, C.; Benosman, R.; et al. Neuromorphic computing at scale. *Nature* **2025**, *637*, 801–812. [CrossRef]
7. Ghosh, S.; Gallego, G. Event-Based Stereo Depth Estimation: A Survey. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–20. [CrossRef]
8. Adra, M.; Melcarne, S.; Mirabet-Herranz, N.; Dugelay, J.L. Event-Based Solutions for Human-centered Applications: A Comprehensive Review. *arXiv* **2025**, arXiv:2502.18490. [CrossRef]
9. AliAkbarpour, H.; Moori, A.; Khorramdel, J.; Blasch, E.; Tahri, O. Emerging Trends and Applications of Neuromorphic Dynamic Vision Sensors: A Survey. *IEEE Sens. Rev.* **2024**, *1*, 14–63. [CrossRef]

10. Shariff, W.; Dilmaghani, M.S.; Kielty, P.; Moustafa, M.; Lemley, J.; Corcoran, P. Event Cameras in Automotive Sensing: A Review. *IEEE Access* **2024**, *12*, 51275–51306. [CrossRef]

11. Cazzato, D.; Bono, F. An Application-Driven Survey on Event-Based Neuromorphic Computer Vision. *Information* **2024**, *15*, 472. [CrossRef]

12. Chakravarthi, B.; Verma, A.A.; Daniilidis, K.; Fermuller, C.; Yang, Y. Recent event camera innovations: A survey. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 342–376.

13. Tenzin, S.; Rassau, A.; Chai, D. Application of event cameras and neuromorphic computing to VSLAM: A survey. *Biomimetics* **2024**, *9*, 444. [CrossRef]

14. Becattini, F.; Berlincioni, L.; Cultrera, L.; Del Bimbo, A. Neuromorphic Face Analysis: A Survey. *arXiv* **2024**, arXiv:2402.11631. [CrossRef]

15. Zheng, X.; Liu, Y.; Lu, Y.; Hua, T.; Pan, T.; Zhang, W.; Tao, D.; Wang, L. Deep learning for event-based vision: A comprehensive survey and benchmarks. *arXiv* **2023**, arXiv:2302.08890.

16. Huang, K.; Zhang, S.; Zhang, J.; Tao, D. Event-based simultaneous localization and mapping: A comprehensive survey. *arXiv* **2023**, arXiv:2304.09793.

17. Schuman, C.D.; Kulkarni, S.R.; Parsa, M.; Mitchell, J.P.; Date, P.; Kay, B. Opportunities for neuromorphic computing algorithms and applications. *Nat. Comput. Sci.* **2022**, *2*, 10–19. [CrossRef]

18. Shi, C.; Song, N.; Li, W.; Li, Y.; Wei, B.; Liu, H.; Jin, J. A Review of Event-Based Indoor Positioning and Navigation. In Proceedings of the WiP Twelfth International Conference on Indoor Positioning and Indoor Navigation, CEUR Workshop Proceedings, Beijing, China, 5–7 September 2022.

19. Furmonas, J.; Liobe, J.; Barzdenas, V. Analytical Review of Event-Based Camera Depth Estimation Methods and Systems. *Sensors* **2022**, *22*, 1201. [CrossRef]

20. Cho, S.W.; Jo, C.; Kim, Y.H.; Park, S.K. Progress of Materials and Devices for Neuromorphic Vision Sensors. *Nano-Micro Lett.* **2022**, *14*, 203. [CrossRef] [PubMed]

21. Liao, F.; Zhou, F.; Chai, Y. Neuromorphic vision sensors: Principle, progress and perspectives. *J. Semicond.* **2021**, *42*, 013105. [CrossRef]

22. Steffen, L.; Reichard, D.; Weinland, J.; Kaiser, J.; Roennau, A.; Dillmann, R. Neuromorphic Stereo Vision: A Survey of Bio-Inspired Sensors and Algorithms. *Front. Neurorobot.* **2019**, *13*, 28. [CrossRef] [PubMed]

23. Lakshmi, A.; Chakraborty, A.; Thakur, C.S. Neuromorphic vision: From sensors to event-based algorithms. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1310. [CrossRef]

24. Vanarse, A.; Osseiran, A.; Rassau, A. A Review of Current Neuromorphic Approaches for Vision, Auditory, and Olfactory Sensors. *Front. Neurosci.* **2016**, *10*, 115. [CrossRef]

25. Mead, C. *Analog VLSI and Neural Systems*; Addison-Wesley: Boston, MA, USA, 1989; p. 371.

26. Mahowald, M.; Douglas, R. A silicon neuron. *Nature* **1991**, *354*, 515–518. [CrossRef]

27. Mahowald, M.A.; Mead, C. The Silicon Retina. *Sci. Am.* **1991**, *264*, 76–82. [CrossRef]

28. Mahowald, M. VLSI Analogs of Neuronal Visual Processing: A Synthesis of Form and Function. Ph.D. Thesis, California Institute of Technology, Pasadena, CA, USA, 1992. [CrossRef]

29. Arreguit, X.; Schaik, F.V.; Bauduin, F.; Bidiville, M.; Raeber, E. A CMOS motion detector system for pointing devices. In Proceedings of the 1996 IEEE International Solid-State Circuits Conference, Digest of Technical Papers, ISSCC, San Francisco, CA, USA, 10 February 1996; IEEE: Piscataway, NJ, USA, 1996. [CrossRef]

30. Boahen, K. Point-to-point connectivity between neuromorphic chips using address events. *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.* **2000**, *47*, 416–434. [CrossRef]

31. Lichtsteiner, P.; Posch, C.; Delbruck, T. A 128 × 128 120db 30 mw asynchronous vision sensor that responds to relative intensity change. In Proceedings of the 2006 IEEE International Solid State Circuits Conference-Digest of Technical Papers, San Francisco, CA, USA, 6–9 February 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 2060–2069.

32. Lazzaro, J.; Wawrzynek, J. A Multi-Sender Asynchronous Extension to the AER Protocol. In Proceedings of the Proceedings Sixteenth Conference on Advanced Research in VLSI, Chapel Hill, NC, USA, 27–29 March 1995; pp. 158–169. [CrossRef]

33. Weikersdorfer, D.; Conradt, J. Event-Based Particle Filtering For Robot Self-localization. In Proceedings of the 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO), Guangzhou, China, 11–14 December 2012; pp. 866–870. [CrossRef]

34. Posch, C.; Matolin, D.; Wohlgenannt, R. A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor with Lossless Pixel-Level Video Compression and Time-Domain CDS. *IEEE J. Solid-State Circuits* **2011**, *46*, 259–275. [CrossRef]

35. Brandli, C.; Berner, R.; Yang, M.; Liu, S.C.; Delbruck, T. A 240 × 180 130 dB 3 μs Latency Global Shutter Spatiotemporal Vision Sensor. *IEEE J. Solid-State Circuits* **2014**, *49*, 2333–2341. [CrossRef]

36. Li, C.; Brandli, C.; Berner, R.; Liu, H.; Yang, M.; Liu, S.C.; Delbruck, T. Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor. In Proceedings of the 2015 IEEE International Symposium on Circuits and Systems (ISCAS), Lisbon, Portugal, 24–27 May 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 718–721.

37. Son, B.; Suh, Y.; Kim, S.; Jung, H.; Kim, J.S.; Shin, C.; Park, K.; Lee, K.; Park, J.; Woo, J.; et al. A 640 × 480 dynamic vision sensor with a 9 μm pixel and 300 Meps address-event representation. In Proceedings of the 2017 IEEE International Solid-State Circuits Conference (ISSCC), Piscataway, NJ, USA, 5–9 February 2017. [CrossRef]

38. Chen, S.; Tang, W.; Zhang, X.; Culurciello, E. A 64 × 64 Pixels UWB Wireless Temporal-Difference Digital Image Sensor. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2012**, *20*, 2232–2240. [CrossRef]

39. Chen, S.; Guo, M. Live Demonstration: CeleX-V: A 1M Pixel Multi-Mode Event-Based Sensor. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; IEEE: Piscataway, NJ, USA, 2019. [CrossRef]

40. IDS Imaging Development Systems GmbH; Prophesee. IDS Launches New Industrial Camera Series Featuring Prophesee Event-Based Metavision® Sensing and Processing Technologies. 2025. Available online: https://www.prophesee.ai/2025/03/05/ids-launches-new-industrial-camera-series-featuring-prophesee-event-based-metavision-sensing-and-processing-technologies/ (accessed on 28 March 2025).

41. Hands On with Android XR and Google's AI-Powered Smart Glasses. 2024. Available online: https://www.wired.com/story/google-android-xr-demo-smart-glasses-mixed-reality-headset-project-moohan (accessed on 28 March 2025).

42. SynSense. Speck™: Event-Driven Neuromorphic Vision SoC. 2022. Available online: https://www.synsense.ai/products/speck-2/ (accessed on 28 March 2025).

43. Yao, M.; Richter, O.; Zhao, G.; Qiao, N.; Xing, Y.; Wang, D.; Hu, T.; Fang, W.; Demirci, T.; De Marchi, M.; et al. Spike-based dynamic computing with asynchronous sensing-computing neuromorphic chip. *Nat. Commun.* **2024**, *15*, 4464. [CrossRef] [PubMed]

44. Yang, Z.; Wang, T.; Lin, Y.; Chen, Y.; Zeng, H.; Pei, J.; Wang, J.; Liu, X.; Zhou, Y.; Zhang, J.; et al. A vision chip with complementary pathways for open-world sensing. *Nature* **2024**, *629*, 1027–1033. [CrossRef]

45. Boettiger, J.P. A Comparative Evaluation of the Detection and Tracking Capability Between Novel Event-Based and Conventional Frame-Based Sensors. Master's Thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, USA, 2020.

46. Posch, C.; Matolin, D.; Wohlgenannt, R. A Two-Stage Capacitive-Feedback Differencing Amplifier for Temporal Contrast IR Sensors. In Proceedings of the 2007 14th IEEE International Conference on Electronics, Circuits and Systems, Marrakech, Morocco, 11–14 December 2007; IEEE: Piscataway, NJ, USA, 2007. [CrossRef]

47. Posch, C.; Matolin, D.; Wohlgenannt, R.; Maier, T.; Litzenberger, M. A Microbolometer Asynchronous Dynamic Vision Sensor for LWIR. *IEEE Sens. J.* **2009**, *9*, 654–664. [CrossRef]

48. Jakobson, C.; Fraenkel, R.; Ben-Ari, N.; Dobromislin, R.; Shiloah, N.; Argov, T.; Freiman, W.; Zohar, G.; Langof, L.; Ofer, O.; et al. Event-Based SWIR Sensor. In *Proceedings of the Infrared Technology and Applications XLVIII, May 2022*; Fulop, G.F., Kimata, M., Zheng, L., Andresen, B.F., Miller, J.L., Kim, Y.H., Eds.; SPIE: Bellingham, WA, USA, 2022. [CrossRef]

49. DARPA. DARPA Announces Research Teams to Develop Intelligent Event-Based Imagers. 2021. Available online: https://www.darpa.mil/news/2021/intelligent-event-based-imagers (accessed on 7 June 2023).

50. Cazzato, D.; Renaldi, G.; Bono, F. A Systematic Parametric Campaign to Benchmark Event Cameras in Computer Vision Tasks. *Electronics* **2025**, *14*, 2603. [CrossRef]

51. Alevi, D.; Stimberg, M.; Sprekeler, H.; Obermayer, K.; Augustin, M. Brian2CUDA: Flexible and Efficient Simulation of Spiking Neural Network Models on GPUs. *Front. Neuroinform.* **2022**, *16*, 883700. [CrossRef]

52. Lee, J.H.; Delbruck, T.; Pfeiffer, M. Training Deep Spiking Neural Networks Using Backpropagation. *Front. Neurosci.* **2016**, *10*, 508. [CrossRef] [PubMed]

53. Innocenti, S.U.; Becattini, F.; Pernici, F.; Bimbo, A.D. Temporal Binary Representation for Event-Based Action Recognition. *arXiv* **2020**, arXiv:2010.08946. [CrossRef]

54. Liu, M.; Delbrück, T. Adaptive Time-Slice Block-Matching Optical Flow Algorithm for Dynamic Vision Sensors. In Proceedings of the British Machine Vision Conference, BMVC, Newcastle, UK, 3–6 September 2018. [CrossRef]

55. Maqueda, A.I.; Loquercio, A.; Gallego, G.; Garcia, N.; Scaramuzza, D. Event-Based Vision Meets Deep Learning on Steering Prediction for Self-Driving Cars. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018. [CrossRef]

56. Lagorce, X.; Orchard, G.; Galluppi, F.; Shi, B.E.; Benosman, R.B. HOTS: A Hierarchy of Event-Based Time-Surfaces for Pattern Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1346–1359. [CrossRef] [PubMed]

57. Sironi, A.; Brambilla, M.; Bourdis, N.; Lagorce, X.; Benosman, R. HATS: Histograms of Averaged Time Surfaces for Robust Event-Based Object Classification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018. [CrossRef]

58. Manderscheid, J.; Sironi, A.; Bourdis, N.; Migliore, D.; Lepetit, V. Speed Invariant Time Surface for Learning to Detect Corner Points with Event-Based Cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.

59. Zhu, A.Z.; Yuan, L.; Chaney, K.; Daniilidis, K. Unsupervised Event-Based Learning of Optical Flow, Depth, and Egomotion. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–17 June 2019; IEEE: Piscataway, NJ, USA, 2019. [CrossRef]

60. Cannici, M.; Ciccone, M.; Romanoni, A.; Matteucci, M. A differentiable recurrent surface for asynchronous event-based data. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 136–152.

61. Gehrig, D.; Loquercio, A.; Derpanis, K.; Scaramuzza, D. End-to-End Learning of Representations for Asynchronous Event-Based Data. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 5632–5642. [CrossRef]

62. Baldwin, R.W.; Liu, R.; Almatrafi, M.; Asari, V.; Hirakawa, K. Time-Ordered Recent Event (TORE) Volumes for Event Cameras. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 2519–2532. [CrossRef] [PubMed]

63. Deng, Y.; Chen, H.; Liu, H.; Li, Y. A Voxel Graph CNN for Object Classification with Event Cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1172–1181.

64. Schaefer, S.; Gehrig, D.; Scaramuzza, D. AEGNN: Asynchronous Event-Based Graph Neural Networks. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 12361–12371. [CrossRef]

65. Teng, M.; Zhou, C.; Lou, H.; Shi, B. NEST: Neural Event Stack for Event-Based Image Enhancement. In *Lecture Notes in Computer Science*; Springer Nature: Cham, Switzerland, 2022; pp. 660–676. [CrossRef]

66. Zubić, N.; Gehrig, D.; Gehrig, M.; Scaramuzza, D. From chaos comes order: Ordering event representations for object recognition and detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 12846–12856.

67. Gu, D.; Li, J.; Zhu, L. Learning Adaptive Parameter Representation for Event-Based Video Reconstruction. *IEEE Signal Process. Lett.* **2024**, *31*, 1950–1954. [CrossRef]

68. Censi, A.; Scaramuzza, D. Low-latency event-based visual odometry. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 703–710. [CrossRef]

69. Gallego, G.; Rebecq, H.; Scaramuzza, D. A Unifying Contrast Maximization Framework for Event Cameras, with Applications to Motion, Depth, and Optical Flow Estimation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3867–3876. [CrossRef]

70. Kim, H.; Handa, A.; Benosman, R.; Ieng, S.H.; Davison, A. Simultaneous Mosaicing and Tracking with an Event Camera. In Proceedings of the British Machine Vision Conference 2014, Nottingham, Nottingham, UK, 1–5 September 2014; pp. 26.1–26.12. [CrossRef]

71. Tumblin, J.; Agrawal, A.; Raskar, R. Why I Want a Gradient Camera. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 103–110. [CrossRef]

72. Scheerlinck, C.; Barnes, N.; Mahony, R. Continuous-time intensity estimation using event cameras. In *Proceedings of the Asian Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 308–324.

73. Scheerlinck, C.; Barnes, N.; Mahony, R. Asynchronous spatial image convolutions for event cameras. *IEEE Robot. Autom. Lett.* **2019**, *4*, 816–822. [CrossRef]

74. Orchard, G.; Meyer, C.; Etienne-Cummings, R.; Posch, C.; Thakor, N.; Benosman, R. HFirst: A Temporal Approach to Object Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 2028–2040. [CrossRef]

75. Granlund, G.H. In search of a general picture processing operator. *Comput. Graph. Image Process.* **1978**, *8*, 155–173. [CrossRef]

76. Shrestha, S.B.; Orchard, G. SLAYER: Spike Layer Error Reassignment in Time. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2018; Volume 31.

77. Davies, M.; Srinivasa, N.; Lin, T.H.; Chinya, G.; Cao, Y.; Choday, S.H.; Dimou, G.; Joshi, P.; Imam, N.; Jain, S.; et al. Loihi: A Neuromorphic Manycore Processor with On-Chip Learning. *IEEE Micro* **2018**, *38*, 82–99. [CrossRef]

78. Kogler, J.; Sulzbachner, C.; Kubinger, W. Bio-inspired Stereo Vision System with Silicon Retina Imagers. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 174–183. [CrossRef]

79. Benosman, R.; Clercq, C.; Lagorce, X.; Ieng, S.-H.; Bartolozzi, C. Event-Based Visual Flow. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 407–417. [CrossRef] [PubMed]

80. Gallego, G.; Gehrig, M.; Scaramuzza, D. Focus Is All You Need: Loss Functions for Event-Based Vision. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 12272–12281. [CrossRef]

81. Gehrig, D.; Rebecq, H.; Gallego, G.; Scaramuzza, D. EKLT: Asynchronous Photometric Feature Tracking Using Events and Frames. *Int. J. Comput. Vis.* **2020**, *128*, 601–618. [CrossRef]

82. Rebecq, H.; Horstschaefer, T.; Scaramuzza, D. Real-Time Visual-Inertial Odometry for Event Cameras Using Keyframe-Based Nonlinear Optimization. In Proceedings of the Procedings of the British Machine Vision Conference, London, UK, 4–7 September 2017; p. 16. [CrossRef]

83. Zhu, A.Z.; Atanasov, N.; Daniilidis, K. Event-Based Visual Inertial Odometry. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5816–5824. [CrossRef]

84. Mitrokhin, A.; Fermüller, C.; Parameshwara, C.; Aloimonos, Y. Event-Based Moving Object Detection and Tracking. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1–9. [CrossRef]

85. Zhu, L.; Li, J.; Wang, X.; Huang, T.; Tian, Y. NeuSpike-Net: High Speed Video Reconstruction via Bio-inspired Neuromorphic Cameras. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; IEEE: Piscataway, NJ, USA, 2021. [CrossRef]

86. Wang, Z.W.; Duan, P.; Cossairt, O.; Katsaggelos, A.; Huang, T.; Shi, B. Joint Filtering of Intensity Images and Neuromorphic Events for High-Resolution Noise-Robust Imaging. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: Piscataway, NJ, USA, 2020. [CrossRef]

87. Duan, P.; Wang, Z.; Shi, B.; Cossairt, O.; Huang, T.; Katsaggelos, A. Guided Event Filtering: Synergy between Intensity Images and Neuromorphic Events for High Performance Imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 8261–8275. [CrossRef]

88. Han, J.; Zhou, C.; Duan, P.; Tang, Y.; Xu, C.; Xu, C.; Huang, T.; Shi, B. Neuromorphic Camera Guided High Dynamic Range Imaging. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: Piscataway, NJ, USA, 2020. [CrossRef]

89. Annamalai, L.; Chakraborty, A.; Thakur, C.S. EvAn: Neuromorphic Event-Based Sparse Anomaly Detection. *Front. Neurosci.* **2021**, *15*, 699003. [CrossRef]

90. Vemprala, S.; Mian, S.; Kapoor, A. Representation learning for event-based visuomotor policies. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 4712–4724.

91. Guo, H.; Peng, S.; Yan, Y.; Mou, L.; Shen, Y.; Bao, H.; Zhou, X. Compact neural volumetric video representations with dynamic codebooks. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 75884–75895.

92. Wang, Z.; Wang, Z.; Li, H.; Qin, L.; Jiang, R.; Ma, D.; Tang, H. EAS-SNN: End-to-End Adaptive Sampling and Representation for Event-based Detection with Recurrent Spiking Neural Networks. *arXiv* **2024**, arXiv:2403.12574. [CrossRef]

93. Bavle, H.; Sanchez-Lopez, J.L.; Cimarelli, C.; Tourani, A.; Voos, H. From SLAM to Situational Awareness: Challenges and Survey. *Sensors* **2023**, *23*, 4849. [CrossRef]

94. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988. [CrossRef]

95. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 1, pp. 886–893.

96. Rosten, E.; Drummond, T. Machine Learning for High-Speed Corner Detection. In *Computer Vision—ECCV 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 430–443. [CrossRef]

97. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

98. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. *Lect. Notes Comput. Sci.* **2006**, *3951*, 404–417.

99. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2564–2571.

100. Vasco, V.; Glover, A.; Bartolozzi, C. Fast Event-Based Harris Corner Detection Exploiting the Advantages of Event-Driven Cameras. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; IEEE: Piscataway, NJ, USA, 2016. [CrossRef]

101. Mueggler, E.; Bartolozzi, C.; Scaramuzza, D. Fast Event-based Corner Detection. In Proceedings of the Procedings of the British Machine Vision Conference, London, UK, 4–7 September 2017; p. 33. [CrossRef]

102. Clady, X.; Ieng, S.H.; Benosman, R. Asynchronous event-based corner detection and matching. *Neural Netw.* **2015**, *66*, 91–106. [CrossRef]

103. Alzugaray, I.; Chli, M. Asynchronous Corner Detection and Tracking for Event Cameras in Real Time. *IEEE Robot. Autom. Lett.* **2018**, *3*, 3177–3184. [CrossRef]

104. Alzugaray, I.; Chli, M. ACE: An Efficient Asynchronous Corner Tracker for Event Cameras. In Proceedings of the 2018 International Conference on 3D Vision (3DV), Verona, Italy, 5–8 September 2018; IEEE: Piscataway, NJ, USA, 2018. [CrossRef]

105. Li, R.; Shi, D.; Zhang, Y.; Li, K.; Li, R. FA-Harris: A Fast and Asynchronous Corner Detector for Event Cameras. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; IEEE: Piscataway, NJ, USA, 2019. [CrossRef]

106. Li, R.; Shi, D.; Zhang, Y.; Li, R.; Wang, M. Asynchronous event feature generation and tracking based on gradient descriptor for event cameras. *Int. J. Adv. Robot. Syst.* **2021**, *18*, 1–13. [CrossRef]

107. Ramesh, B.; Yang, H.; Orchard, G.; Le Thi, N.A.; Zhang, S.; Xiang, C. DART: Distribution Aware Retinal Transform for Event-Based Cameras. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2767–2780. [CrossRef]

108. Huang, Z.; Sun, L.; Zhao, C.; Li, S.; Su, S. EventPoint: Self-Supervised Interest Point Detection and Description for Event-based Camera. *arXiv* **2022**, arXiv:2109.00210.

109. DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperPoint: Self-Supervised Interest Point Detection and Description. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018. [CrossRef]

110. Chiberre, P.; Perot, E.; Sironi, A.; Lepetit, V. Detecting Stable Keypoints from Events through Image Gradient Prediction. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; IEEE: Piscataway, NJ, USA, 2021. [CrossRef]

111. Shi, J.; Tomasi. Good features to track. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR-94, Seattle, WA, USA, 21–23 June 1994; IEEE Computer Society Press: Piscataway, NJ, USA, 1994. [CrossRef]

112. Besl, P.; McKay, N.D. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256. [CrossRef]

113. Ni, Z.; Bolopion, A.; Agnus, J.; Benosman, R.; Regnier, S. Asynchronous Event-Based Visual Shape Tracking for Stable Haptic Feedback in Microrobotics. *IEEE Trans. Robot.* **2012**, *28*, 1081–1089. [CrossRef]

114. Tedaldi, D.; Gallego, G.; Mueggler, E.; Scaramuzza, D. Feature detection and tracking with the dynamic and active-pixel vision sensor (DAVIS). In Proceedings of the 2016 Second International Conference on Event-Based Control, Communication, and Signal Processing (EBCCSP), Krakow, Poland, 13–15 June 2016; pp. 1–7. [CrossRef]

115. Kueng, B.; Mueggler, E.; Gallego, G.; Scaramuzza, D. Low-Latency Visual Odometry Using Event-Based Feature Tracks. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; pp. 16–23. [CrossRef]

116. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [CrossRef]

117. Ni, Z.; Ieng, S.H.; Posch, C.; Régnier, S.; Benosman, R. Visual tracking using neuromorphic asynchronous event-based cameras. *Neural Comput.* **2015**, *27*, 925–953. [CrossRef]

118. Lagorce, X.; Ieng, S.H.; Clady, X.; Pfeiffer, M.; Benosman, R.B. Spatiotemporal features for asynchronous event-based data. *Front. Neurosci.* **2015**, *9*, 46. [CrossRef]

119. Glover, A.; Bartolozzi, C. Robust Visual Tracking with a Freely-Moving Event Camera. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 3769–3776. [CrossRef]

120. Glover, A.; Bartolozzi, C. Event-Driven Ball Detection and Gaze Fixation in Clutter. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Repubic of Korea, 9–14 October 2016; IEEE: Piscataway, NJ, USA, 2016. [CrossRef]

121. Zhu, A.Z.; Atanasov, N.; Daniilidis, K. Event-Based Feature Tracking with Probabilistic Data Association. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 4465–4470. [CrossRef]

122. Alzugaray, I.; Chli, M. Haste: Multi-Hypothesis Asynchronous Speeded-Up Tracking of Events. In Proceedings of the 31st British Machine Vision Virtual Conference (BMVC 2020), Virtual Event, UK, 7–10 September 2020; ETH Zurich, Institute of Robotics and Intelligent Systems: Zurich, Switzerland, 2020; p. 744.

123. Alzugaray, I.; Chli, M. Asynchronous Multi-Hypothesis Tracking of Features with Event Cameras. In Proceedings of the 2019 International Conference on 3D Vision (3DV), Québec City, QC, Canada, 16–19 September 2019; IEEE: Piscataway, NJ, USA, 2019. [CrossRef]

124. Hu, S.; Kim, Y.; Lim, H.; Lee, A.J.; Myung, H. eCDT: Event Clustering for Simultaneous Feature Detection and Tracking. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]

125. Messikommer, N.; Fang, C.; Gehrig, M.; Scaramuzza, D. Data-driven Feature Tracking for Event Cameras. *arXiv* **2022**, arXiv:2211.12826. [CrossRef]

126. Horn, B.K.; Schunck, B.G. Determining optical flow. *Artif. Intell.* **1981**, *17*, 185–203. [CrossRef]

127. Barron, J.L.; Fleet, D.J.; Beauchemin, S.S. Performance of optical flow techniques. *Int. J. Comput. Vis.* **1994**, *12*, 43–77. [CrossRef]

128. Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981.

129. Benosman, R.; Ieng, S.H.; Clercq, C.; Bartolozzi, C.; Srinivasan, M. Asynchronous frameless event-based optical flow. *Neural Netw.* **2012**, *27*, 32–37. [CrossRef] [PubMed]

130. Brosch, T.; Tschechne, S.; Neumann, H. On event-based optical flow detection. *Front. Neurosci.* **2015**, *9*, 137. [CrossRef]

131. Orchard, G.; Benosman, R.; Etienne-Cummings, R.; Thakor, N.V. A spiking neural network architecture for visual motion estimation. In Proceedings of the 2013 IEEE Biomedical Circuits and Systems Conference (BioCAS), Rotterdam, The Netherlands, 31 October–2 November 2013; IEEE: Piscataway, NJ, USA, 2013. [CrossRef]

132. Valois, R.L.D.; Cottaris, N.P.; Mahon, L.E.; Elfar, S.D.; Wilson, J. Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vis. Res.* **2000**, *40*, 3685–3702. [CrossRef]

133. Rueckauer, B.; Delbruck, T. Evaluation of Event-Based Algorithms for Optical Flow with Ground-Truth from Inertial Measurement Sensor. *Front. Neurosci.* **2016**, *10*, 176. [CrossRef]

134. Paredes-Valles, F.; Scheper, K.Y.W.; de Croon, G.C.H.E. Unsupervised Learning of a Hierarchical Spiking Neural Network for Optical Flow Estimation: From Events to Global Motion Perception. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2051–2064. [CrossRef]

135. Lee, C.; Kosta, A.K.; Zhu, A.Z.; Chaney, K.; Daniilidis, K.; Roy, K. Spike-FlowNet: Event-Based Optical Flow Estimation with Energy-Efficient Hybrid Neural Networks. In *Computer Vision—ECCV 2020*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 366–382. [CrossRef]

136. Paredes-Valles, F.; de Croon, G.C.H.E. Back to Event Basics: Self-Supervised Learning of Image Reconstruction for Event Cameras via Photometric Constancy. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; IEEE: Piscataway, NJ, USA, 2021. [CrossRef]

137. Zhang, Y.; Lv, H.; Zhao, Y.; Feng, Y.; Liu, H.; Bi, G. Event-Based Optical Flow Estimation with Spatio-Temporal Backpropagation Trained Spiking Neural Network. *Micromachines* **2023**, *14*, 203. [CrossRef]

138. Parameshwara, C.M.; Li, S.; Fermuller, C.; Sanket, N.J.; Evanusa, M.S.; Aloimonos, Y. SpikeMS: Deep Spiking Neural Network for Motion Segmentation. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September 2021; IEEE: Piscataway, NJ, USA, 2021. [CrossRef]

139. Bardow, P.; Davison, A.J.; Leutenegger, S. Simultaneous Optical Flow and Intensity Estimation from an Event Camera. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016. [CrossRef]

140. Liu, M.; Delbruck, T. Block-matching optical flow for dynamic vision sensors: Algorithm and FPGA implementation. In Proceedings of the 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD, USA, 28–31 May 2017; IEEE: Piscataway, NJ, USA, 2017. [CrossRef]

141. Liu, M.; Delbruck, T. EDFLOW: Event Driven Optical Flow Camera with Keypoint Detection and Adaptive Block Matching. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 5776–5789. [CrossRef]

142. Zhu, A.; Yuan, L.; Chaney, K.; Daniilidis, K. EV-FlowNet: Self-Supervised Optical Flow Estimation for Event-based Cameras. In Proceedings of the Robotics: Science and Systems XIV, Robotics: Science and Systems Foundation, Pittsburgh, PA, USA, 26 June 2018. [CrossRef]

143. Ye, C.; Zhu, A.Z.; Daniilidis, K. Unsupervised Learning of Dense Optical Flow, Depth and Egomotion with Event-Based Sensors. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 574–581.

144. Sun, H.; Dao, M.Q.; Fremont, V. 3D-FlowNet: Event-based Optical Flow Estimation with 3D Representation. *arXiv* **2022**, arXiv:2201.12265. [CrossRef]

145. Gehrig, M.; Millhausler, M.; Gehrig, D.; Scaramuzza, D. E-RAFT: Dense Optical Flow from Event Cameras. In Proceedings of the 2021 International Conference on 3D Vision (3DV), London, UK, 1–3 December 2021; IEEE: Piscataway, NJ, USA, 2021. [CrossRef]

146. Ding, Z.; Zhao, R.; Zhang, J.; Gao, T.; Xiong, R.; Yu, Z.; Huang, T. Spatio-Temporal Recurrent Networks for Event-Based Optical Flow Estimation. In Proceedings of the AAAI Conference on Artificial Intelligence, Oline, 22 February–1 March 2022; Volume 36, pp. 525–533.

147. Tian, Y.; Andrade-Cetto, J. Event Transformer FlowNet for optical flow estimation. In Proceedings of the 33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, 21–24 November 2022; BMVA Press: Durham, UK, 2022.

148. Li, Y.; Huang, Z.; Chen, S.; Shi, X.; Li, H.; Bao, H.; Cui, Z.; Zhang, G. Blinkflow: A dataset to push the limits of event-based optical flow estimation. *arXiv* **2023**, arXiv:2303.07716.

149. Taketomi, T.; Uchiyama, H.; Ikeda, S. Visual SLAM algorithms: A survey from 2010 to 2016. *IPSJ Trans. Comput. Vis. Appl.* **2017**, *9*, 16. [CrossRef]

150. Tsintotas, K.A.; Bampis, L.; Gasteratos, A. The Revisiting Problem in Simultaneous Localization and Mapping: A Survey on Visual Loop Closure Detection. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 19929–19953. [CrossRef]

151. Kabiri, M.; Cimarelli, C.; Bavle, H.; Sanchez-Lopez, J.L.; Voos, H. Graph-Based vs. Error State Kalman Filter-Based Fusion of 5G and Inertial Data for MAV Indoor Pose Estimation. *J. Intell. Robot. Syst.* **2024**, *110*, 87. [CrossRef]

152. Gallego, G.; Scaramuzza, D. Accurate Angular Velocity Estimation with an Event Camera. *IEEE Robot. Autom. Lett.* **2017**, *2*, 632–639. [CrossRef]

153. Reinbacher, C.; Munda, G.; Pock, T. Real-time panoramic tracking for event cameras. In Proceedings of the 2017 IEEE International Conference on Computational Photography (ICCP), Stanford, CA, USA, 12–14 May 2017; IEEE: Piscataway, NJ, USA, 2017. [CrossRef]

154. Rebecq, H.; Horstschaefer, T.; Gallego, G.; Scaramuzza, D. EVO: A Geometric Approach to Event-Based 6-DOF Parallel Tracking and Mapping in Real Time. *IEEE Robot. Autom. Lett.* **2017**, *2*, 593–600. [CrossRef]

155. Weikersdorfer, D.; Hoffmann, R.; Conradt, J. Simultaneous Localization and Mapping for Event-Based Vision Systems. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 133–142. [CrossRef]

156. Weikersdorfer, D.; Adrian, D.B.; Cremers, D.; Conradt, J. Event-Based 3D SLAM with a Depth-Augmented Dynamic Vision Sensor. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; IEEE: Piscataway, NJ, USA, 2014. [CrossRef]

157. Kim, H.; Leutenegger, S.; Davison, A.J. Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera. In *Computer Vision – ECCV 2016*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 349–364. [CrossRef]

158. Gallego, G.; Lund, J.E.; Mueggler, E.; Rebecq, H.; Delbruck, T.; Scaramuzza, D. Event-Based, 6-DOF Camera Tracking from Photometric Depth Maps. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2402–2412. [CrossRef]

159. Rebecq, H.; Gallego, G.; Mueggler, E.; Scaramuzza, D. EMVS: Event-Based Multi-View Stereo—3D Reconstruction with an Event Camera in Real-Time. *Int. J. Comput. Vis.* **2018**, *126*, 1394–1414. [CrossRef]

160. Vidal, A.R.; Rebecq, H.; Horstschaefer, T.; Scaramuzza, D. Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High-Speed Scenarios. *IEEE Robot. Autom. Lett.* **2018**, *3*, 994–1001. [CrossRef]

161. Chen, P.; Guan, W.; Lu, P. Esvio: Event-Based Stereo Visual Inertial Odometry. *IEEE Robot. Autom. Lett.* **2023**, *8*, 3661–3668. [CrossRef]

162. Zhou, Y.; Gallego, G.; Shen, S. Event-Based Stereo Visual Odometry. *IEEE Trans. Robot.* **2021**, *37*, 1433–1450. [CrossRef]

163. Liu, Z.; Shi, D.; Li, R.; Yang, S. ESVIO: Event-Based Stereo Visual-Inertial Odometry. *Sensors* **2023**, *23*, 1998. [CrossRef] [PubMed]

164. Zuo, Y.F.; Yang, J.; Chen, J.; Wang, X.; Wang, Y.; Kneip, L. DEVO: Depth-Event Camera Visual Odometry in Challenging Conditions. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]

165. Hidalgo-Carrio, J.; Gehrig, D.; Scaramuzza, D. Learning Monocular Dense Depth from Events. In Proceedings of the 2020 International Conference on 3D Vision (3DV), Fukuoka, Japan, 25–28 November 2020; IEEE: Piscataway, NJ, USA, 2020. [CrossRef]

166. Zhu, J.; Gehrig, D.; Scaramuzza, D. Self-Supervised Event-based Monocular Depth Estimation using Cross-Modal Consistency. *arXiv* **2024**, arXiv:2401.07218.

167. Jin, Y.; Yu, L.; Li, G.; Fei, S. A 6-DOFs event-based camera relocalization system by CNN-LSTM and image denoising. *Expert Syst. Appl.* **2021**, *170*, 114535. [CrossRef]

168. Hu, L.; Song, X.; Wang, Y.; Liang, P. 6-DoF Pose Relocalization for Event Cameras with Entropy Frame and Attention Networks. In Proceedings of the ACM Symposium on Virtual Reality, Visualization and Interaction (VRCAI), Guangzhou, China, 27–29 December 2022.

169. Ren, H.; Zhou, S.; Taylor, C.J.; Daniilidis, K. A Simple and Effective Point-based Network for Event Camera 6-DOFs Pose Relocalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024.

170. Hagenaars, J.; van Gemert, J.C.; Gehrig, D.; Scaramuzza, D. Self-Supervised Learning of Event-Based Optical Flow with Spiking Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Online, 6–14 December 2021.

171. Zhu, A.Z.; Thakur, D.; Ozaslan, T.; Pfrommer, B.; Kumar, V.; Daniilidis, K. The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2032–2039. [CrossRef]

172. Gehrig, M.; Aarents, W.; Gehrig, D.; Scaramuzza, D. DSEC: A Stereo Event Camera Dataset for Driving Scenarios. *IEEE Robot. Autom. Lett.* **2021**, *6*, 4947–4954. [CrossRef]

173. Mueggler, E.; Rebecq, H.; Gallego, G.; Delbruck, T.; Scaramuzza, D. The Event-Camera Dataset and Simulator: Event-based Data for Pose Estimation, Visual Odometry, and SLAM. *Int. J. Robot. Res.* **2017**, *36*, 142–149. [CrossRef]

174. Li, Z.; Chen, Y.; Xie, Y.; Liu, Y.; Yu, L.; Li, J.; Tang, J. M3ED: A Multi-Modal Multi-Motion Event Detection Dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 13624–13634.

175. Litzenberger, M.; Posch, C.; Bauer, D.; Belbachir, A.; Schon, P.; Kohn, B.; Garn, H. Embedded Vision System for Real-Time Object Tracking using an Asynchronous Transient Vision Sensor. In Proceedings of the 2006 IEEE 12th Digital Signal Processing Workshop & 4th IEEE Signal Processing Education Workshop, Napa, CA, USA, 24–27 September 2006; IEEE: Piscataway, NJ, USA, 2006. [CrossRef]

176. Vasco, V.; Glover, A.; Mueggler, E.; Scaramuzza, D.; Natale, L.; Bartolozzi, C. Independent motion detection with event-driven cameras. In Proceedings of the 2017 18th International Conference on Advanced Robotics (ICAR), Hong Kong, China, 10–12 July 2017; IEEE: Piscataway, NJ, USA, 2017. [CrossRef]

177. Stoffregen, T.; Gallego, G.; Drummond, T.; Kleeman, L.; Scaramuzza, D. Event-Based Motion Segmentation by Motion Compensation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republis of Korea, 27 October–2 November 2019; IEEE: Piscataway, NJ, USA, 2019. [CrossRef]

178. Cannici, M.; Ciccone, M.; Romanoni, A.; Matteucci, M. Asynchronous Convolutional Networks for Object Detection in Neuromorphic Cameras. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1656–1665. [CrossRef]

179. Liang, Z.; Cao, H.; Yang, C.; Zhang, Z.; Chen, G. Global-local Feature Aggregation for Event-based Object Detection on EventKITTI. In Proceedings of the 2022 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Bedford, UK, 20–22 September 2022; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]

180. Mitrokhin, A.; Ye, C.; Fermüller, C.; Aloimonos, Y.; Delbruck, T. EV-IMO: Motion Segmentation Dataset and Learning Pipeline for Event Cameras. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 6105–6112. [CrossRef]

181. Burner, L.; Mitrokhin, A.; Fermüller, C.; Aloimonos, Y. EVIMO2: An Event Camera Dataset for Motion Segmentation, Optical Flow, Structure from Motion, and Visual Inertial Odometry in Indoor Scenes with Monocular or Stereo Algorithms. *arXiv* **2022**, arXiv:2205.03467. [CrossRef]

182. Gehrig, M.; Scaramuzza, D. Recurrent Vision Transformers for Object Detection with Event Cameras. In Proceedings of the CVPR, Vancouver, BC, Canada, 17–24 June 2023.

183. Zhou, Z.; Wu, Z.; Paudel, D.P.; Boutteau, R.; Yang, F.; Van Gool, L.; Timofte, R.; Ginhac, D. Event-free moving object segmentation from moving ego vehicle. In Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Abu Dhabi, United Arab Emirates, 14–18 October 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 8960–8965.

184. Georgoulis, S.; Ren, W.; Bochicchio, A.; Eckert, D.; Li, Y.; Gawel, A. Out of the room: Generalizing event-based dynamic motion segmentation for complex scenes. In Proceedings of the 2024 International Conference on 3D Vision (3DV), Davos, Switzerland, 18–21 March 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 442–452.

185. Wang, Z.; Guo, J.; Daniilidis, K. Un-evmoseg: Unsupervised event-based independent motion segmentation. *arXiv* **2023**, arXiv:2312.00114.

186. Wu, Z.; Gehrig, M.; Lyu, Q.; Liu, X.; Gilitschenski, I. Leod: Label-efficient object detection for event cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA,16–22 June 2024; pp. 16933–16943.

187. Snyder, S.; Thompson, H.; Kaiser, M.A.A.; Schwartz, G.; Jaiswal, A.; Parsa, M. Object motion sensitivity: A bio-inspired solution to the ego-motion problem for event-based cameras. *arXiv* **2023**, arXiv:2303.14114.

188. Clerico, V.; Snyder, S.; Lohia, A.; Abdullah-Al Kaiser, M.; Schwartz, G.; Jaiswal, A.; Parsa, M. Retina-Inspired Object Motion Segmentation for Event-Cameras. In Proceedings of the 2025 Neuro Inspired Computational Elements (NICE), Heidelberg, Germany, 24–26 March 2025; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6.

189. Zhou, H.; Shi, Z.; Dong, H.; Peng, S.; Chang, Y.; Yan, L. JSTR: Joint spatio-temporal reasoning for event-based moving object detection. In Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 10650–10656.

190. Zhou, Z.; Wu, Z.; Boutteau, R.; Yang, F.; Demonceaux, C.; Ginhac, D. RGB-Event Fusion for Moving Object Detection in Autonomous Driving. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 7808–7815. [CrossRef]

191. Lu, D.; Kong, L.; Lee, G.H.; Chane, C.S.; Ooi, W.T. FlexEvent: Towards Flexible Event-Frame Object Detection at Varying Operational Frequencies. *arXiv* **2024**, arXiv:2412.06708.

192. Fang, W.; Yu, Z.; Chen, Y.; Huang, T.; Masquelier, T.; Tian, Y. Deep residual learning in spiking neural networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 21056–21069.

193. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]

194. Yao, M.; Qiu, X.; Hu, T.; Hu, J.; Chou, Y.; Tian, K.; Liao, J.; Leng, L.; Xu, B.; Li, G. Scaling spike-driven transformer with efficient spike firing approximation training. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 2973–2990. [CrossRef]

195. Guo, Y.; Liu, X.; Chen, Y.; Peng, W.; Zhang, Y.; Ma, Z. Spiking Transformer: Introducing Accurate Addition-Only Spiking Self-Attention for Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, CO, USA, 3–7 June 2025; pp. 24398–24408.

196. Xu, Q.; Deng, J.; Shen, J.; Chen, B.; Tang, H.; Pan, G. Hybrid Spiking Vision Transformer for Object Detection with Event Cameras. In Proceedings of the International Conference on Machine Learning (ICML), Vancouver, BC, Canada, 13–19 July 2025.

197. Lee, D.; Li, Y.; Kim, Y.; Xiao, S.; Panda, P. Spiking Transformer with Spatial-Temporal Attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 13948–13958.

198. Zhao, L.; Huang, Z.; Ding, J.; Yu, Z. TTFSFormer: A TTFS-based Lossless Conversion of Spiking Transformer. In Proceedings of the International Conference on Machine Learning (ICML), Vancouver, BC, Canada, 13–19 July 2025.

199. Lei, Z.; Yao, M.; Hu, J.; Luo, X.; Lu, Y.; Xu, B.; Li, G. Spike2former: Efficient Spiking Transformer for High-performance Image Segmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 1364–1372.

200. Wang, Y.; Zhang, Y.; Xiong, R.; Zhao, J.; Zhang, J.; Fan, X.; Huang, T. Spk2SRImgNet: Super-Resolve Dynamic Scene from Spike Stream via Motion Aligned Collaborative Filtering. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 11416–11426.

201. Qian, Y.; Ye, S.; Wang, C.; Cai, X.; Qian, J.; Wu, J. UCF-Crime-DVS: A Novel Event-Based Dataset for Video Anomaly Detection with Spiking Neural Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 6577–6585.

202. Amir, A.; Taba, B.; Berg, D.; Melano, T.; McKinstry, J.; di Nolfo, C.; Nayak, T.; Andreopoulos, A.; Garreau, G.; Mendoza, M.; et al. A Low Power, Fully Event-Based Gesture Recognition System. In Proceedings of the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

203. Orchard, G.; Jayawant, A.; Cohen, G.K.; Thakor, N. Converting static image datasets to spiking neuromorphic datasets using saccades. *Front. Neurosci.* **2015**, *9*, 437. [CrossRef]

204. Li, H.; Liu, H.; Ji, X.; Li, G.; Shi, L. Cifar10-dvs: An event-stream dataset for object classification. *Front. Neurosci.* **2017**, *11*, 244131. [CrossRef]

205. Sun, S.; Cioffi, G.; de Visser, C.; Scaramuzza, D. Autonomous Quadrotor Flight Despite Rotor Failure with Onboard Vision Sensors: Frames vs. Events. *IEEE Robot. Autom. Lett.* **2021**, *6*, 580–587. [CrossRef]

206. Chen, G.; Cao, H.; Conradt, J.; Tang, H.; Rohrbein, F.; Knoll, A. Event-based neuromorphic vision for autonomous driving: A paradigm shift for bio-inspired visual sensing and perception. *IEEE Signal Process. Mag.* **2020**, *37*, 34–49. [CrossRef]

207. Litzenberger, M.; Kohn, B.; Belbachir, A.; Donath, N.; Gritsch, G.; Garn, H.; Posch, C.; Schraml, S. Estimation of Vehicle Speed Based on Asynchronous Data from a Silicon Retina Optical Sensor. In Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference, Toronto, ON, Canada, 17–20 September 2006; IEEE: Piscataway, NJ, USA, 2006. [CrossRef]

208. Dietsche, A.; Cioffi, G.; Hidalgo-Carrio, J.; Scaramuzza, D. PowerLine Tracking with Event Cameras. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021.

209. Chin, T.J.; Bagchi, S.; Eriksson, A.; van Schaik, A. Star Tracking Using an Event Camera. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; IEEE: Piscataway, NJ, USA, 2019. [CrossRef]

210. Cohen, G.; Afshar, S.; Morreale, B.; Bessell, T.; Wabnitz, A.; Rutten, M.; van Schaik, A. Event-based Sensing for Space Situational Awareness. *J. Astronaut. Sci.* **2019**, *66*, 125–141. [CrossRef]

211. Afshar, S.; Nicholson, A.P.; van Schaik, A.; Cohen, G. Event-Based Object Detection and Tracking for Space Situational Awareness. *IEEE Sensors J.* **2020**, *20*, 15117–15132. [CrossRef]

212. Angelopoulos, A.N.; Martel, J.N.; Kohli, A.P.; Conradt, J.; Wetzstein, G. Event-Based Near-Eye Gaze Tracking Beyond 10,000 Hz. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 2577–2586. [CrossRef]

213. Colonnier, F.; Della Vedova, L.; Orchard, G. ESPEE: Event-Based Sensor Pose Estimation Using an Extended Kalman Filter. *Sensors* **2021**, *21*, 7840. [CrossRef]

214. Zou, S.; Guo, C.; Zuo, X.; Wang, S.; Wang, P.; Hu, X.; Chen, S.; Gong, M.; Cheng, L. EventHPE: Event-based 3D Human Pose and Shape Estimation. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10976–10985. [CrossRef]

215. Scarpellini, G.; Morerio, P.; Bue, A.D. Lifting Monocular Events to 3D Human Poses. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; IEEE: Piscataway, NJ, USA, 2021. [CrossRef]

216. Shao, Z.; Zhou, W.; Wang, W.; Yang, J.; Li, Y. A Temporal Densely Connected Recurrent Network for Event-based Human Pose Estimation. *Pattern Recognit.* **2024**, *147*, 110048. [CrossRef]

217. Zhang, Z.; Chai, K.; Yu, H.; Majaj, R.; Walsh, F.; Wang, E.; Mahbub, U.; Siegelmann, H.; Kim, D.; Rahman, T. Neuromorphic high-frequency 3D dancing pose estimation in dynamic environment. *Neurocomputing* **2023**, *547*, 126388. [CrossRef]

218. Calabrese, E.; Taverni, G.; Awai Easthope, C.; Skriabine, S.; Corradi, F.; Longinotti, L.; Eng, K.; Delbruck, T. DHP19: Dynamic Vision Sensor 3D Human Pose Dataset. In Proceedings of the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019.

219. Moeys, D.P.; Corradi, F.; Li, C.; Bamford, S.A.; Longinotti, L.; Voigt, F.F.; Berry, S.; Taverni, G.; Helmchen, F.; Delbruck, T. A Sensitive Dynamic and Active Pixel Vision Sensor for Color or Neural Imaging Applications. *IEEE Trans. Biomed. Circuits Syst.* **2017**, *12*, 123–136. [CrossRef] [PubMed]

220. Choi, C.; Lee, G.J.; Chang, S.; Kim, D.H. Inspiration from Visual Ecology for Advancing Multifunctional Robotic Vision Systems: Bio-inspired Electronic Eyes and Neuromorphic Image Sensors. *Adv. Mater.* **2024**, *36*, e2412252. [CrossRef] [PubMed]

221. Tulyakov, S.; Gehrig, D.; Georgoulis, S.; Erbach, J.; Gehrig, M.; Li, Y.; Scaramuzza, D. Time Lens: Event-based Video Frame Interpolation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; IEEE: Piscataway, NJ, USA, 2021. [CrossRef]

222. Tulyakov, S.; Bochicchio, A.; Gehrig, D.; Georgoulis, S.; Li, Y.; Scaramuzza, D. Time Lens++: Event-based Frame Interpolation with Parametric Nonlinear Flow and Multi-Scale Fusion. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]

223. Mavridou, E.; Vrochidou, E.; Papakostas, G.A.; Pachidis, T.; Kaburlasos, V.G. Machine Vision Systems in Precision Agriculture for Crop Farming. *J. Imaging* **2019**, *5*, 89. [CrossRef]

224. Bialik, K.; Kowalczyk, M.; Blachut, K.; Kryjak, T. Fast-moving object counting with an event camera. *arXiv* **2022**, arXiv:2212.08384. [CrossRef]

225. Dold, P.M.; Nadkarni, P.; Boley, M.; Schorb, V.; Wu, L.; Steinberg, F.; Burggräf, P.; Mikut, R. Event-based vision in laser welding: An approach for process monitoring. *J. Laser Appl.* **2025**, *37*, 012040. [CrossRef]

226. Baldini, S.; Bernardini, R.; Fusiello, A.; Gardonio, P.; Rinaldo, R. Measuring Vibrations with Event Cameras. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *48*, 9–16. [CrossRef]

227. Rebecq, H.; Ranftl, R.; Koltun, V.; Scaramuzza, D. Events-to-Video: Bringing Modern Computer Vision to Event Cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.

228. Rebecq, H.; Ranftl, R.; Koltun, V.; Scaramuzza, D. High Speed and High Dynamic Range Video with an Event Camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1964–1980. [CrossRef]

229. Bacon, J.G., Jr. Satellite Tracking with Neuromorphic Cameras for Space Domain Awareness. Master's Thesis, Department of Aeronautics and Astronautics, Cambridge, MA, USA, 2021.

230. Żołnowski, M.; Reszelewski, R.; Moeys, D.P.; Delbrück, T.; Kamiński, K. Observational evaluation of event cameras performance in optical space surveillance. In Proceedings of the NEO and Debris Detection Conference, Darmstadt, Germany, 22–24 January 2019.

231. Jolley, A.; Cohen, G.; Lambert, A. Use of neuromorphic sensors for satellite material characterisation. In Proceedings of the Imaging and Applied Optics Congress, Munich, Germany, 24–27 June 2019.

232. Jawaid, M.; Elms, E.; Latif, Y.; Chin, T.J. Towards Bridging the Space Domain Gap for Satellite Pose Estimation using Event Sensing. *arXiv* **2022**, arXiv:2209.11945. [CrossRef]

233. Mahlknecht, F.; Gehrig, D.; Nash, J.; Rockenbauer, F.M.; Morrell, B.; Delaune, J.; Scaramuzza, D. Exploring Event Camera-Based Odometry for Planetary Robots. *IEEE Robot. Autom. Lett.* **2022**, *7*, 8651–8658. [CrossRef]

234. McHarg, M.G.; Balthazor, R.L.; McReynolds, B.J.; Howe, D.H.; Maloney, C.J.; O'Keefe, D.; Bam, R.; Wilson, G.; Karki, P.; Marcireau, A.; et al. Falcon Neuro: An event-based sensor on the International Space Station. *Opt. Eng.* **2022**, *61*, 085105. [CrossRef]

235. Hinz, G.; Chen, G.; Aafaque, M.; Röhrbein, F.; Conradt, J.; Bing, Z.; Qu, Z.; Stechele, W.; Knoll, A. Online Multi-object Tracking-by-Clustering for Intelligent Transportation System with Neuromorphic Vision Sensor. In *KI 2017: Advances in Artificial Intelligence*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 142–154. [CrossRef]

236. Ganan, F.; Sanchez-Diaz, J.; Tapia, R.; de Dios, J.M.; Ollero, A. Efficient Event-based Intrusion Monitoring using Probabilistic Distributions. In Proceedings of the 2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Sevilla, Spain, 8–10 November 2022; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]

237. de Dios, J.M.; Eguiluz, A.G.; Rodriguez-Gomez, J.; Tapia, R.; Ollero, A. Towards UAS Surveillance Using Event Cameras. In Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Abu Dhabi, United Arab Emirates, 4–6 November 2020; IEEE: Piscataway, NJ, USA, 2020. [CrossRef]

238. Rodriguez-Gomez, J.; Eguiluz, A.G.; Martinez-de Dios, J.; Ollero, A. Asynchronous Event-Based Clustering and Tracking for Intrusion Monitoring in UAS. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 8518–8524. [CrossRef]

239. Pérez-Cutiño, M.; Eguíluz, A.G.; Dios, J.M.d.; Ollero, A. Event-Based Human Intrusion Detection in UAS Using Deep Learning. In Proceedings of the 2021 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 15–18 June 2021; pp. 91–100. [CrossRef]

240. Falanga, D.; Kim, S.; Scaramuzza, D. How Fast Is Too Fast? The Role of Perception Latency in High-Speed Sense and Avoid. *IEEE Robot. Autom. Lett.* **2019**, *4*, 1884–1891. [CrossRef]

241. Loquercio, A.; Kaufmann, E.; Ranftl, R.; Müller, M.; Koltun, V.; Scaramuzza, D. Learning high-speed flight in the wild. *Sci. Robot.* **2021**, *6*, abg581. [CrossRef]

242. Wzorek, P.; Kryjak, T. Traffic Sign Detection with Event Cameras and DCNN. In Proceedings of the 2022 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), Poznan, Poland, 21–22 September 2022; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]

243. Yang, C.; Liu, P.; Chen, G.; Liu, Z.; Wu, Y.; Knoll, A. Event-based Driver Distraction Detection and Action Recognition. In Proceedings of the 2022 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Bedford, UK, 20–22 September 2022; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]

244. Hu, Y.; Liu, S.C.; Delbruck, T. v2e: From Video Frames to Realistic DVS Events. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; IEEE: Piscataway, NJ, USA, 2021.

245. Shair, Z.E.; Rawashdeh, S. High-Temporal-Resolution Event-Based Vehicle Detection and Tracking. *Opt. Eng.* **2022**, *62*, 031209. [CrossRef]

246. Belbachir, A.; Schraml, S.; Brandle, N. Real-Time Classification of Pedestrians and Cyclists for Intelligent Counting of Non-Motorized Traffic. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; IEEE: Piscataway, NJ, USA, 2010. [CrossRef]

247. Kirkland, P.; Di Caterina, G.; Soraghan, J.; Matich, G. Neuromorphic technologies for defence and security. In *Proceedings of the Emerging Imaging and Sensing Technologies for Security and Defence V; and Advanced Manufacturing Technologies for Micro-and Nanosystems in Security and Defence III*; SPIE: St Bellingham, WA, USA, 2020; Volume 11540, pp. 113–130.

248. Stewart, T.; Drouin, M.A.; Gagne, G.; Godin, G. Drone Virtual Fence Using a Neuromorphic Camera. In *Proceedings of the International Conference on Neuromorphic Systems*; ACM: New York, NY, USA, 2021; pp. 1–9.

249. Boehrer, N.; Kuijf, H.J.; Dijk, J. Laser Warning and Pointed Optics Detection Using an Event Camera. In Proceedings of the Electro-Optical and Infrared Systems: Technology and Applications XXI, Edinburgh, UK, 16–19 September 2024; SPIE: St Bellingham, WA, USA, 2024; Volume 13200, pp. 314–326. [CrossRef]

250. Eisele, C.; Seiffer, D.; Sucher, E.; Sjöqvist, L.; Henriksson, M.; Lavigne, C.; Domel, R.; Déliot, P.; Dijk, J.; Kuijf, H.; et al. DEBELA: Investigations on Potential Detect-before-Launch Technologies. In Proceedings of the Electro-Optical and Infrared Systems: Technology and Applications XXI, Edinburgh, UK, 16–19 September 2024; SPIE: St Bellingham, WA, USA, 2024; Volume 13200, pp. 300–313. [CrossRef]

251. Kim, E.; Yarnall, J.; Shah, P.; Kenyon, G.T. A neuromorphic sparse coding defense to adversarial images. In *Proceedings of the International Conference on Neuromorphic Systems*; ACM: New York, NY, USA, 2019; pp. 1–8.

252. Cha, J.H.; Abbott, A.L.; Szu, H.H.; Willey, J.; Landa, J.; Krapels, K.A. Neuromorphic implementation of a software-defined camera that can see through fire and smoke in real-time. In *Proceedings of the Independent Component Analyses, Compressive Sampling, Wavelets, Neural Net, Biosystems, and Nanoengineering XII*; SPIE: St Bellingham, WA, USA, 2014; Volume 9118, pp. 25–34.

253. Moustafa, M.; Lemley, J.; Corcoran, P. Contactless Cardiac Pulse Monitoring Using Event Cameras. *arXiv* **2025**, arXiv:2505.09529. [CrossRef]

254. Kim, J.; Kim, Y.M.; Wu, Y.; Zahreddine, R.; Welge, W.A.; Krishnan, G.; Ma, S.; Wang, J. Privacy-preserving visual localization with event cameras. *arXiv* **2022**, arXiv:2212.03177. [CrossRef]

255. Habara, T.; Sato, T.; Awano, H. Zero-Aware Regularization for Energy-Efficient Inference on Akida Neuromorphic Processor. In Proceedings of the 2025 IEEE International Symposium on Circuits and Systems (ISCAS), London, UK, 25–28 May 2025; IEEE: Piscataway, NJ, USA, 2025; pp. 1–5.

256. Kiselev, I.; Neil, D.; Liu, S.C. Event-driven deep neural network hardware system for sensor fusion. In Proceedings of the 2016 IEEE International Symposium on Circuits and Systems (ISCAS), Montréal, QC, Canada, 22–25 May 2016; IEEE: Piscataway, NJ, USA, 2016. [CrossRef]

257. O'Connor, P.; Neil, D.; Liu, S.C.; Delbruck, T.; Pfeiffer, M. Real-time classification and sensor fusion with a spiking deep belief network. *Front. Neurosci.* **2013**, *7*, 178. [CrossRef] [PubMed]

258. Lv, W.; Xu, S.; Zhao, Y.; Wang, G.; Wei, J.; Cui, C.; Du, Y.; Dang, Q.; Liu, Y. DETRs Beat YOLOs on Real-time Object Detection. *arXiv* **2023**, arXiv:2304.08069. [CrossRef]

259. Sabater, A.; Montesano, L.; Murillo, A.C. Event Transformer. A Sparse-Aware Solution for Efficient Event Data Processing. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Piscataway, NJ, USA, 18–24 June 2022. [CrossRef]

260. Kingdon, G.; Coker, R. The eyes have it: Improving mill availability through visual technology. In Proceedings of the International Semi-Autogenous Grinding and High Pressure Grinding Roll Technology, Vancouver, BC, Canada, 20–23 September 2015; Volume 3, pp. 1–15.