

Review

An Application-Driven Survey on Event-Based Neuromorphic Computer Vision

Dario Cazzato *  and Flavio Bono 

Joint Research Centre (JRC), European Commission, 21027 Ispra, Italy; flavio.bono@ec.europa.eu

* Correspondence: dario.cazzato@ec.europa.eu

Abstract: Traditional frame-based cameras, despite their effectiveness and usage in computer vision, exhibit limitations such as high latency, low dynamic range, high power consumption, and motion blur. For two decades, researchers have explored neuromorphic cameras, which operate differently from traditional frame-based types, mimicking biological vision systems for enhanced data acquisition and spatio-temporal resolution. Each pixel asynchronously captures intensity changes in the scene above certain user-defined thresholds, and streams of events are captured. However, the distinct characteristics of these sensors mean that traditional computer vision methods are not directly applicable, necessitating the investigation of new approaches before being applied in real applications. This work aims to fill existing gaps in the literature by providing a survey and a discussion centered on the different application domains, differentiating between computer vision problems and whether solutions are better suited for or have been applied to a specific field. Moreover, an extensive discussion highlights the major achievements and challenges, in addition to the unique characteristics, of each application field.

Keywords: event cameras; dynamic vision sensors; event-based cameras; neuromorphic cameras; computer vision; survey; review



Citation: Cazzato, D.; Bono, F. An Application-Driven Survey on Event-Based Neuromorphic Computer Vision. *Information* **2024**, *15*, 472. <https://doi.org/10.3390/info15080472>

Academic Editor: Alessandra Lumini

Received: 6 June 2024

Revised: 7 July 2024

Accepted: 5 August 2024

Published: 9 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Computer vision has reached a mature stage and finds applications across a wide range of fields. Computer vision applications heavily rely on image sensors, which detect electromagnetic radiation, typically in the form of visible or infrared light. Indeed, image acquisition plays a pivotal role in machine vision as long as high-quality images allow for consequent higher-quality processing and analysis. This stage involves both hardware and software components, and selecting the right components is crucial for the following processing and analysis.

In general, images can be generated with active or passive methods. In the active method, a light source illuminates the object, while in the passive method, natural sunlight serves as the illumination. Usually, in production lines and industrial applications, the active method is more suitable, since it leads to more reproducible conditions; however, this method means that the choice of light source and illumination conditions becomes critical. Depending on the application, typical light sources can be lamps, LEDs, and/or laser sources. The wavelength of the electromagnetic wave also matters for image recording, with visible, infrared, or X-ray regions of the spectrum used for scene illumination [1].

In both cases, computer vision has relied on frame-based cameras for a long time; they produce data corresponding to the captured light intensity at each selected pixel in a synchronous manner (i.e., the whole sensor records the scene at defined frame rates). While this technology has been effective and superior to other camera types for many years, frame-based cameras still exhibit limitations that impact performance and accuracy. Some of the challenges associated with frame-based cameras include

- **High latency:** the latency between the presentation of physical stimuli, the transduction into analog values, and the encoding time into a digital representation are high if the data are sampled at a fixed rate, independent from the stimuli.
- **Low-dynamic range:** frame-based cameras have difficulties in handling scenes with very high variations in brightness.
- **Power consumption:** high-quality and feature-rich frame-based cameras present high consumption, making them unpractical in many resource-constrained environments.
- **Motion blur:** when capturing high-speed motion, frame-based cameras introduce motion blur, affecting subsequent image processing accuracy.
- **Limited frame rates:** traditional sensors allow for slow frame rates (typically in the order of 20–240 fps), whereas for high-speed recordings, specialized highly complex and expensive equipment is needed.

Typical frame-based cameras acquire incident light at a frequency indicative of the temporal resolution and latency of the sensor. Two major reading schemes are global and rolling shutter. In global shutter schemes, the entire array is read simultaneously at fixed timestamps; in this way, the camera exposure is synchronized to be the same for all the pixels of the array. Consequently, the sensor latency depends on the number of pixels on the sensor to transfer. To decrease latency, rolling shutter cameras read the array row by row; the time to read out a single row is known as *line time* and can be of the order of 10–20 microseconds. This reading scheme reduces latency to the product of the line time and the number of rows but, as a consequence, the occurrence of motion blur artifacts increases. In general, lower exposition times mitigate blur at the cost of causing overly dark photos and loss of details. Higher frame rate cameras can mitigate motion blur, but this leads to increased power consumption and overheating issues, and the associated image processor or digital signal processor must handle exponentially larger data volumes [2].

The aforementioned limitations impact the possibility of applying traditional computer vision in certain domains, depending on the very specific requirements of each application field. To address these challenges, researchers have started to explore other solutions; among them, in the last 60 years, neuromorphic cameras (also known as event cameras) have emerged, taking inspiration from biological vision systems. If attempts to electronically model the mammal visual system date to 1970 [3] the seminal work on such sensor was introduced in 1988 [4] and constructed an analog model of the early stages of retinal processing on a single silicon chip. Differently from traditional frame-based cameras, neuromorphic image sensors operate asynchronously, mimicking the spatio-temporal information extraction of biological vision systems to enhance data acquisition, sensitivity, dynamic range, and spatio-temporal resolution. Unlike traditional frame-based cameras, these sensors adhere to a different paradigm, resulting in sparse event data output. In this way, their application can benefit from reduced temporal latency, high dynamic range, robustness to lighting conditions, data transmission speed, and reduced power consumption. Moreover, neuromorphic cameras have the advantage of working under privacy-reservation conditions since they capture pixel brightness variations instead of static imagery [5]. As evidence of the potential of such sensors, both academia and industry have shown interest in neuromorphic cameras and, since their original introduction, several breakthroughs have improved performance and provided different working alternatives, leading to commercial solutions and sensors ready for the market. The state of the art company case studies show how event cameras have been successfully applied in several fields, from robotics to intelligent transportation systems, from surveillance and security to healthcare, from facial analysis to space imaging. Figure 1 reports two commercial examples of neuromorphic cameras, the Prophesee EVK4 [6] (a) and the Inivation DAVIS346 [7] (b).

Nevertheless, due to the distinct characteristics of such sensors, the methods and algorithms developed for frame-based computer vision are no longer directly applicable to event-based vision systems. Furthermore, the extensive datasets used for training and characterizing frame-based systems cannot be seamlessly transferred to their event-based counterparts due to the fundamental differences in data output. As a consequence, event

cameras have garnered significant interest from researchers, and academic investigation in this domain has flourished as experts seek to address fundamental questions related to event camera technology. Given the transformative potential of event cameras, it is not surprising that a wealth of work has been proposed to unlock their full capabilities and overcome associated challenges.

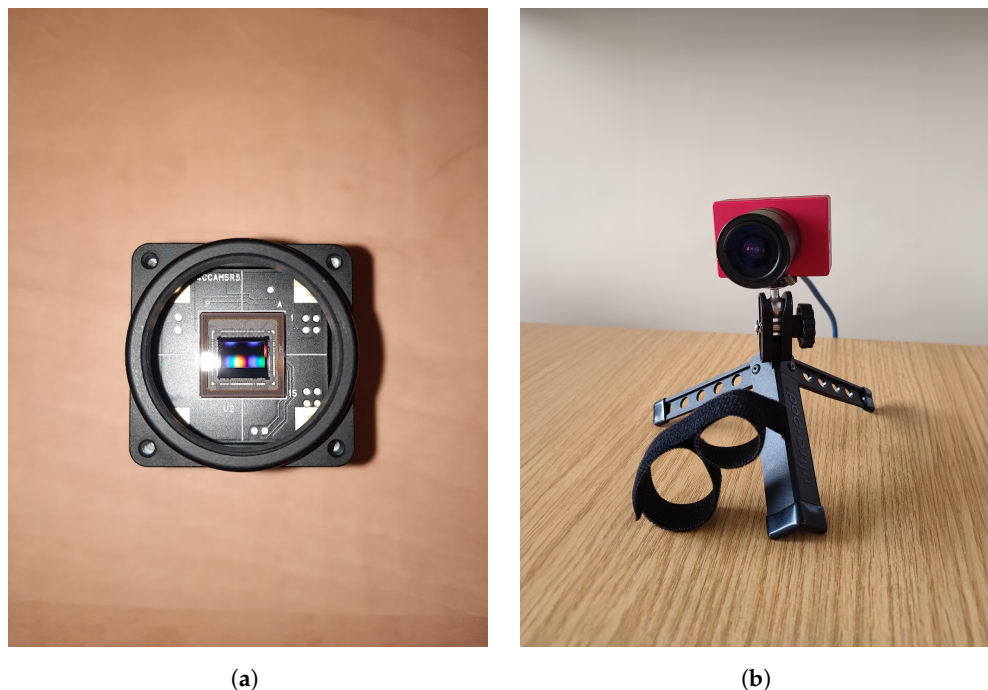


Figure 1. Two commercial examples of neuromorphic sensors. (a) The Prophesee EVK4 that uses the Sony IMX636 CMOS. (b) The Inivation DAVIS346. Courtesy of Prof. Maria Martini, Kingston University London, UK.

Additionally, surveys and comprehensive analyses that consolidate existing knowledge, highlight trends, and provide valuable insights for both practitioners and academics have been proposed. Nevertheless, it is possible to observe that this, albeit extensive, literature is missing a critical analysis based on different application domains (see Section 4), frequently preferring to list certain computer vision tasks instead. Often, a solution for a low-level computer vision task can serve as a building element for tasks hierarchically classified at a higher-level. However, while it is acknowledged that the traditional hierarchical classification of computer vision tasks may evolve with the latest progress in deep learning, and in particular with the advent of end-to-end systems, it is still crucial to separate computer vision tasks from systems that have been implemented in a specific application domain, which has unique features and challenges. This is particularly true for the case of the relatively recent neuromorphic sensor that has less well-established research compared to classic cameras and, thus, less clear links between results in specific tasks and applications.

Finally, the reduction in the cost of producing event cameras, the availability of event camera simulators and datasets, as well as the very rapid advances in deep learning, are leading to a rapid diffusion of such technology; and this widens the number of works in this field. Similarly, the number of workshops, academic conferences, and special issues in journals has increased tremendously over the years, with the necessity to re-organize works presented in previous surveys to include the most recent research outcomes.

This work aims to fill the aforementioned gaps with the following contributions:

1. a collection of past surveys about neuromorphic cameras and computer vision with the introduction of a specific taxonomy to easily classify and refer to them;

2. a critical analysis driven by the different application domains, instead of low-, medium-, or high-level computer vision tasks;
3. an updated review that includes recent works and research outcomes.

The first two contributions, to the best of our knowledge, are introduced for the first time in the context of reviews about neuromorphic cameras.

This work is organized as summed up by the scheme shown in Figure 2. Section 2 briefly sums up the working principle of a neuromorphic camera and introduces the event-based paradigm. The methodology used to select the papers forming the subject of this survey and the reason behind the selected Scopus search queries are given in Section 3. In Section 4, existing surveys in the literature are discussed and cataloged using a proposed taxonomy that considers the main focus of each survey. Section 5 briefly puts in context the proposed analysis by application domain with respect to low-, middle-, and high-level computer vision tasks. In Section 6, a review analyzes works that are presented by their application domain, divided into groups, and each subsection analyzes one of these groups. The amount of information and works is very high, so a discussion that takes into account all the aspects presented in this manuscript highlighting common outcomes and challenges, as well as peculiarities of each application field and future directions, can be found in Section 7. Section 8 has the conclusions.

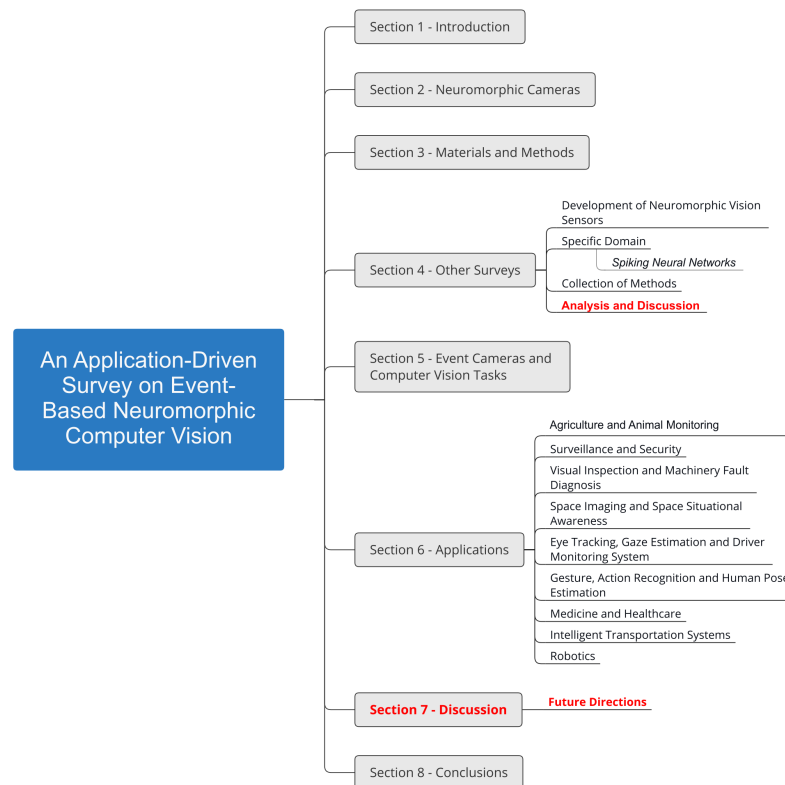


Figure 2. A summary of the paper organization. Critical analysis and discussions are highlighted with red text.

2. Neuromorphic Cameras

The biological vision system has been optimized over hundreds of millions of years of evolution and has excellent image information perception and highly parallel information processing capabilities. The retina transmits information in the form of images, shadows, and colors to the brainstem through a crossover, where the final visual information is extracted and useless visual information is discarded, recognizing visual information in the brain when the final processing is complete. With the continuous development of vision

sensor arrays, the performance of traditional imaging systems that capture brightness at a fixed rate is constantly being improved. However, the subsequent amount of raw data collected is also increasing, making data transmission and processing more and more complex and demanding. In contrast, living organisms efficiently process sensory information in complex environments due to a well-established hierarchical structure, co-location of computation and storage, and very complex neural networks. Several types of image-sensing systems that simulate the biological visual structures of humans and animals have been proposed [8]. We use *neuromorphic sensors* to refer to those bio-inspired devices that try to mimic the sensing and early visual-processing characteristics of living organisms [9].

Neuromorphic vision is generally divided into three levels: the *structural level*, which imitates the retina; the *device functional level*, which approaches the retina; and the *intelligence level*, which surpasses the retina. In neuromorphic vision sensors, an optoelectronic nanodevice simulates the biological vision sampling model and information processing function, and a perception system with or beyond the biological vision ability is constructed under limited physical space and power consumption constraints using simulation engineering techniques, such as device function-level approximation [10]. A schematic diagram that shows the analogies between the human visual system (top) and neuromorphic vision (bottom) is reported in Figure 3.

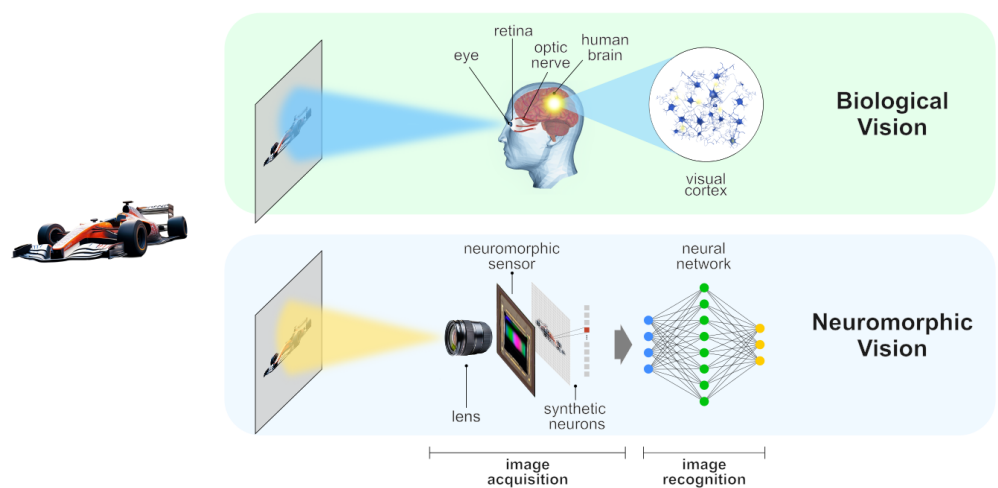


Figure 3. Analogies between the human visual system (top) and the neuromorphic vision sensor (bottom). “Neuron” (<https://skfb.ly/oyUVY>, accessed on 4 August 2024) by mmarynguyen is licensed under Creative Commons Attribution-NonCommercial. “Human Head” (<https://skfb.ly/ouFsp>, accessed on 4 August 2024) by VistaPrime is licensed under Creative Commons Attribution. Car and lens generated with Adobe Firefly©.

Similarly to conventional cameras, a neuromorphic vision sensor, also known as an *event camera*, *address-event representation (AER)*, or *silicon retina*, is composed of pixels, but, instead of capturing intensity, each pixel asynchronously captures intensity changes in the scene that exceed user-defined thresholds. The camera outputs streams of events, where the i – th event e_i is defined by

$$e_i = (x_i, t_i, p_i), \quad (1)$$

with x_i denoting the pixel position with spatial coordinates (x_i, y_i) where the event is triggered, with t_i the timestamp, and with p_i the polarity of the event that can be defined as ON (positive)/ OFF (negative), or also, $p_i \in \{1, -1\}$, to distinguish an increase or decrease in intensity from darker to brighter values or vice-versa, exceeding given thresholds. The event data in a time sequence are recorded as $\{e_i\}_{i=1}^{N_e}$, with N_e being the number of total events in the interval of time t_e .

In other words, a single event occurs if there is a change in the brightness magnitude (from the last event or the initial brightness condition) that reaches a threshold C^+ for positive or ON changes and C^- for negative (OFF) events. Events are triggered in an asynchronous way and timestamped with microsecond resolution and can be transmitted with very low latency (in general, in the order of sub-milliseconds). Figure 4 shows an example of different sensor outputs when recording a white rotating PC fan in the xyt -plan: the frame-based camera (right) grabs a frame depending on its internal clock, and thus on the established exposition time and as a function of the frame rate. This means that blur will be present when the rotation speed is high with respect to the camera frame rate, and that the same scene will be captured when the image is still. In the case of event cameras, a continuous stream will be captured, where the only pixels that activate will capture the movement, either observing a change from dark to bright or vice-versa. In the case of high rotation speeds, the motion will still be completely captured. On the contrary, in case of no motion and no noise, no events will be streamed.

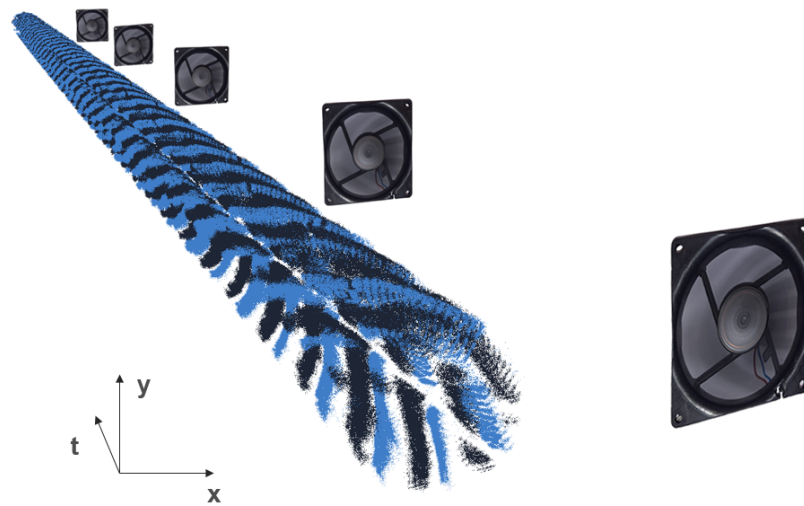
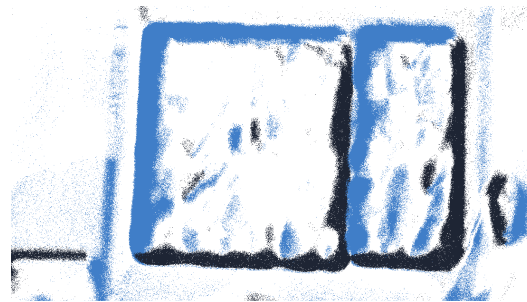


Figure 4. Output from a neuromorphic sensor (left) and a frame-based camera (right) while recording a rotating PC fan in the xyt -plan. ON and OFF events are rendered respectively as blue and black pixels on a white background.

The fast response of a sensor capable of capturing events at a high rate makes it suitable for considering accumulating events over time, to better grasp the scene. When creating an output representation comparable to images of frame-based cameras, it is possible to perform the following: all the events that occurred in a time interval are moved to image coordinates depending on the pixel position, the most recent event can be kept in case of multiple events at the same pixel, and thus the image can be formed at a frame rate that is a function of the accumulation time. The case of “no event” for a given pixel is usually modeled by giving the value 0 to that pixel. In this way, a video stream can be created. An example of an obtained image by accumulating events over time from an event camera obtained in an indoor navigation scenario is reported in Figure 5a. ON and OFF events are rendered respectively as blue and black pixels on a white background. At this point, it is fundamental to highlight how the functioning of event cameras poses a new paradigm shift, where the output is sparse and asynchronous instead of dense and synchronized. Moreover, the output is no longer a set of grayscale intensities, but a stream of events e_i , as defined in Equation (1). This poses new challenges in terms of camera setup and computer vision algorithm design: first of all, classic computer vision algorithms are not easily or directly applicable, since they are designed to work with a fundamentally different information source. This has opened a new research area that investigates alternative

representations that are more suitable for algorithms dedicated to event processing and/or that can facilitate the feature extraction phase of computer vision pipelines. Refer to [11] for more details on event representation. Moreover, the captured stream of events strongly depends on the values of the different configurable thresholds (*biases*) associated with the event camera acquisition scheme, making an experimental setup stage necessary, where the values depend on the specific application context and the environmental conditions.



(a) Accumulation image obtained with an accumulation time of 33 ms.



(b) The same scene recorded with a frame-based camera.

Figure 5. (a): An example accumulation image from an event camera while moving in an indoor environment. ON and OFF events are rendered respectively as blue and black pixels on a white background. (b): the same scene taken from a RGB camera.

Early event-based sensors presented high noise levels, unsuitable for real and commercial applications [12]. The development of optoelectronic materials, together with advancements in robotics and biomedical fields, has led to a new generation of event-based cameras that have been adopted by research and industry. The major differentiating parameters of event cameras with respect to frame-based cameras are their latency, dynamic range, power consumption, and bandwidth.

The dynamic vision sensor (DVS) [13] emulates a simplified three-layer retina to realize an abstraction of the information flow through the photoreceptor, bipolar, and ganglion cells. Internally, a differentiation circuit amplifies the changes with high precision and outputs a voltage logarithmically encoding the photo-current. This output voltage is compared against global thresholds that are offset from the reset voltage to detect increasing (ON) and decreasing (OFF) changes.

In practice, DVS has become a synonym for event camera, although it only represents a possible technology and thus a type of neuromorphic camera. In fact, other families of event cameras are also available. For example, an asynchronous time-based image sensor (ATIS) [14] encodes both relative changes and absolute exposure information, thanks to an exposure measurement unit. Absolute instantaneous pixel illuminance is acquired by converting the integrated photo charge into the timing of asynchronous pulse edges. The disadvantage is that the pixels are at least double the area of DVS pixels. Another way to combine relative changes with static information is in a dynamic and active pixel vision sensor (DAVIS) [15]. The difference is that ATIS are still an asynchronous event-driven technology, while in DAVIS an active pixel sensor (APS) retrieves static scene information like in a frame-based camera, and events can be placed on top of this representation. Two frames with both color and event information from the dataset released in [16] are reported in Figure 6. Being based on frame-based principles, DAVIS sensors have limited dynamic range compared to DVS, and they display redundancies in the case of static scenes. Refer to [17] for more information on these families of sensors, and to [12] for a discussion that also includes other emerging technologies.

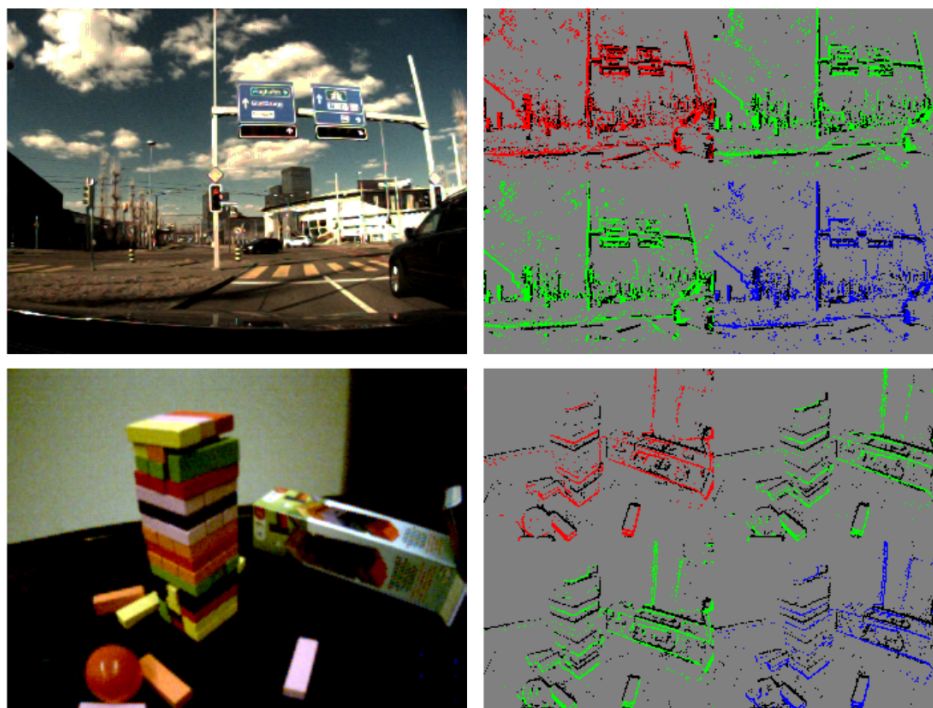


Figure 6. Two frames with respective color and event information from the CED: Color Event Dataset in [16].

3. Materials and Methods

First, a pilot investigation established the required foundation and understanding of the domain, as well as the selection of appropriate keywords to search and select the papers. Data were gathered from Scopus, because of its large coverage [18]. The output of this phase was the following Scopus query, used with the Scopus Search API to extract the paper list. The aforementioned search was carried out in April 2024.

```
TITLE(
  ("neuromorphic" AND ("vision" OR "camera" OR "sensor"))
  OR "event camera" OR "event based camera" OR "event triggered camera"
  OR "dynamic vision sensor" OR "event based sensor" OR "event sensor"
  OR "address event representation" OR "event vision sensor"
  OR "event based vision sensor" OR "silicon retina")
  OR KEY(
  ("neuromorphic" AND ("vision" OR "camera" OR "sensor"))
  OR "event camera" OR "event based camera" OR "event triggered camera"
  OR "dynamic vision sensor" OR "event based sensor" OR "event sensor"
  OR "address event representation" OR "event vision sensor"
  OR "event based vision sensor" OR "silicon retina")
  AND LIMIT-TO(LANGUAGE, "English")
```

The selected field codes were TITLE and KEY (that includes AUTHKEY, INDEXTERMS, TRADENAME, and CHEMNAME); the field code ABS that also searched in the abstract text was excluded because the keywords search in abstracts was misleading and included many works not related with the scope of this manuscript. The selected attributes and Boolean relations used for the TITLE field code were replicated for the KEY and were selected to include all names by which event cameras are known; we only added the constraint that the word *neuromorphic* must be present together with *vision*, *camera* or *sensor*. The word *event* alone can lead to non-relevant results; thus, it was only used in combination with the correct and complete names associated with event cameras. Spaces or dashes were automatically handled by the search engine. Finally, the language was limited to English.

Criteria for selecting primary studies and data collection were defined to reduce selection bias and provide a mean for ascertaining the validity of the review process. The initial search result produced 2667 records. After removing possible duplicates and similar issues (e.g., corrigendum, live demonstrations, etc.), this number fell to 2603. After this, studies whose abstract did not suggest any relation with our scope were excluded. At the end of this phase, about 1500 works were eligible. We separated surveys (the main ones are summed up and discussed in the next section) from other works, and divided the remaining works by primary application field (if any), reviewing and citing most relevant works in the different subsections of Section 6. We considered relevance depending on the publication year, number of citations, if the article addressed specific issues for the first time, and how much the work focused on the application field; about the latter, this meant that if many works focused on the same problem, we could not report all of them, even if they were very relevant, while a very unique attempt to use event cameras in a new field was cited, to show the potential of such sensors. The works that were eligible but were not classifiable in any application field were excluded from the analysis of Section 6 but kept for our overall analysis and cited if they contributed to a certain definition/citation during the manuscript and/or during the discussions. At this point, we repeated the research with Google Scholar to extend the research to thesis works and arXiv pre- and post-prints, to avoid missing certain very specific documents. The Google Scholar query system has fewer degrees of freedom than Scopus, so different queries with the specific keywords of each single application field were repeated. In this way, eight papers were included in our review. Finally, in seven cases, works that were cited in the selected papers were also included because of their significance. A summary of the complete search process is shown in Figure 7. The documents added after the Scopus query filtering phase were represented by the “injection” block. The same occurred with surveys, but this was done intentionally to extend the overview of other aspects not related to the main scope of this survey (see the next section).

As a final note, while the criteria employed in the selection of works inherently lent a qualitative aspect to our analysis of Section 6, they offered valuable insights pertinent to the scope of this review. Furthermore, Section 7 will facilitate a discussion aimed at deriving objective insights and drawing definitive conclusions.

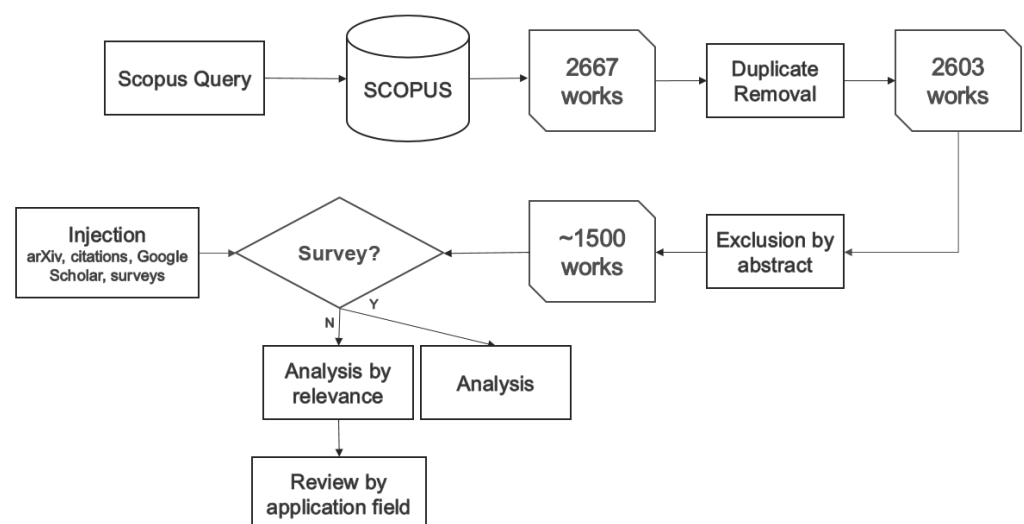


Figure 7. A scheme of the manuscript selection process. We report the documents added after filtering the output from the Scopus search as “injection” in the diagram.

4. Other Surveys

There is a wide range of emerging scenarios to which event cameras can contribute. Due to the influence that event cameras have in academia and in industry, different surveys

and reviews have been proposed in recent years. From their analysis, it emerged that it is convenient to divide these works into three main sets:

- **Surveys based on the development of neuromorphic vision sensors:** these works focus on physical sensor design and hardware aspects, ranging from conventional devices based on integrated circuits to new emerging technologies;
- **Surveys based on a specific application domain:** these works mainly focus on a very specific topic, reconstructing the evolution of proposed methods to address the challenges related to the specific domain;
- **Surveys based on a collection of methods:** these works consider how classic computer vision problems have been redesigned when the input comes from a neuromorphic sensor.

Herein, the most recent surveys are presented and discussed.

4.1. Development of Neuromorphic Vision Sensors

To the best of our knowledge, the first reviews that considered event cameras date from 1996 and are represented by the work of [9,19]. Here, circuit design principles are discussed, also reviewing early visual processing capabilities. From that moment on, different surveys have been proposed over the years [20,21], discussing hardware aspects and/or the design of very-large-scale integration (VLSI) neuromorphic circuits for event-camera-based signal processing. In [22], the analysis of sensors was extended to other sensors mimicking other senses like silicon cochleas. Other similar works analyzed the hardware aspects of optical sensing devices [17], suggesting how such new sensors can outperform classic frame cameras in many application fields, due to their numerous advantages. A comparison of hardware aspects of event and frame-based cameras was given in [23]. The work in [12] specialized in the design aspects of neuromorphic sensors, also focusing on emerging devices including optoelectronic random-access memory, neural networks, and hemispherically shaped vision sensors. In [24], a survey of event cameras in a more generic context of bio-inspired vision systems was reported, while the work in [8] presented an in-depth analysis of optoelectronic materials and perovskites to design modern bio-inspired vision sensors, envisioning event-based cameras in a broader context of bionic sensors.

4.2. Specific Domain

The literature analysis showed that there is an impressive amount of work that has exploited event camera design or employed such sensors to address vision tasks; thus, it is not surprising that many surveys directly tackled a specific task or restricted the analysis to a single application domain. In [25], the problem of binocular stereo vision was analyzed, both from a sensor and algorithm perspective, while a similar approach was taken in [2] for the task of depth estimation. In [26], the focus was on autonomous driving and assistance systems, while in [27] it was robotics and autonomous systems. The application field of medicine was tackled in the work in [28], not limited to imaging but also involving the processing of bio-signals for diagnosis, and biomedical interfaces. The work in [29] introduced a literature review of event-based data-driven technology considering available datasets and simulators, showing the potential of this technology over other event-based algorithms. A focus on robotics, with an analysis that involved both perception and control, was presented in the work in [30], while [31] tackled the problem of semantic segmentation.

Spiking Neural Networks

A growing computational paradigm for efficient neural architectures comes from spiking neural networks (SNNs). They consist of neurons interconnected by synapses that determine information propagation from the source to target neurons. Unlike conventional neural networks, the information is encoded and transmitted in the form of spiking neurons [32–34]. In SNNs, a spike takes the form of a single binary bit. If a neuron receives input spikes (from an event or other neurons since they are hierarchically organized), it modifies its internal state

and produces an output spike if the computed state is higher than a threshold. SNNs are suitable for implementation on neuromorphic hardware, to provide even more efficient and low-latency solutions [35].

Examples of comprehensive overviews on SNNs are found in [36–41]. Guided by a hierarchical classification of SNN learning rules, they reviewed recent trends in learning rules and network architectures. Although their application ranges and the different sets of architectures can easily spread towards the third group, i.e., *collection of methods*, we still classified them as *specific domain* surveys because of the unique characteristics of SNNs, starting from their basic building blocks, spiking neurons.

Nevertheless, specialized surveys that depended on the specific application are available [42,43], or with a focus on hardware aspects [35,44]. Considering the availability of surveys on SNN and how SNN often goes beyond computer vision, touching perception theory, neuromorphic computing, and neuroscience, we did not directly consider SNNs in our survey, unless they directly involved computer vision systems for a specific application domain.

4.3. Collection of Methods

The work in [11] represents a seminal work about event cameras from a computer vision perspective, with a very well-structured analysis about how to meaningfully represent events, as well as a taxonomy for the different event processing solutions after the paradigm shift of dealing with events instead of frames. Afterwards, a review of different tasks was performed, ranging from optical flow estimation to 3D reconstruction and simultaneous localization and mapping (SLAM). An in-depth survey of current trends in deep-learning for event-based vision was presented in [45]. The authors divided the state-of-the-art works by the event representation chosen (surface, voxel, spike, etc.) and by task (image restoration and enhancement vs. scene understanding and 3D vision). Challenges introduced by the different computational paradigms for both camera design and computational models were also considered in [46]. Recently, a survey that dealt with several classic computer vision tasks was given in [10]. The authors also explored hardware and circuit design aspects in detail.

Event cameras, their applications, and associated tasks were also considered in [47,48], the first with an analysis of possible tasks, the latter with a focus on how conventional vision algorithms have to be reformulated to adapt to this paradigm shift, also considering datasets and simulators. The work in [49] expanded the focus from vision to neuromorphic auditory and olfactory sensors. Very recently, the work in [50] briefly introduced event cameras, exploring methods for tasks like autofocus, 3D reconstruction, and super-resolution imaging.

4.4. Analysis and Discussion

To guide readers towards the works that best align with their research interest, and to foster interdisciplinary research, a summary of the works is provided in Table 1. For each work, we present the year of publication, the category to which the work belongs following our previous distinction, and a brief description of the work. Note that, in some cases, an analyzed work discusses different aspects of neuromorphic vision together: in this case, the classification was carried out following the primary focus, or based on which parts are more exhaustive. Summing up, an exhaustive analysis of these works makes it possible to state some observations. If surveys on sensor development reflect the evolution of such a sensor over the years, for the case of the domain-specific surveys, it is also possible to observe how the topics addressed by the analyzed surveys have tended to evolve and adapt in response to new insights and developments. This highlights the relevance and timeliness of these surveys, reflective of the current trends and advancements in the respective domains. Moreover, apart from works that reviewed hardware and circuit design or those on spiking neural networks, the existing surveys, on the one hand, tend to present low-, middle-, and high-level computer vision tasks without distinction; on the other hand, they often also

combine together tasks with the final field of application, such as autonomous navigation, robotics, or medicine. In many cases, a solution for a low-level computer vision task can be a single composing block of a high-level task. So that this categorization can change with new advances in deep learning and with the advent of end-to-end systems, it becomes crucial to differentiate between the specific challenges of tasks and their suitability for addressing specific industrial problems and/or satisfying the needs of application fields. As stated in the introduction, this work, among its contributions, is an attempt to fill this gap.

Table 1. Summary of relevant surveys on neuromorphic cameras. For spiking neural networks, since these are outside of the scope of this survey, we only report the year and indicate the topic with *SNN* for further reading.

Typology	Authors	Publication Year	Brief Description
Sensors development	Etienne-Cummings and Van der Spiegel [9]	1996	First surveys on neuromorphic cameras
	Indiveri et al. [19]	1996	First surveys on neuromorphic cameras
	Kramer and Indiveri [20]	1998	Sensor analysis and two robotics applications
	Indiveri [21]	2008	Neuromorphic circuits and selective attention chip pixel analysis
	Liu and Delbruck [22]	2010	Sensor analysis
	Posch et al. [17]	2014	Sensor design
	Wu [23]	2018	Hardware design aspects and neural network-oriented vision chips
	Liao et al. [12]	2021	Sensor analysis
Domain Specific	Kim et al. [24]	2022	Sensor analysis
	Li et al. [8]	2023	Sensor design with focus on materials
	Steffen et al. [25]	2019	Stereo Vision
	Chen et al. [26]	2020	Autonomous driving
	Sun et al. [29]	2021	Data-driven approaches
	Furmonas et al. [2]	2022	Depth estimation
	Sandamirskaya et al. [27]	2022	Robotics
	Bartolozzi et al. [30]	2022	Robotics
	Jia [31]	2022	Semantic segmentation
	Aboumerhi et al. [28]	2023	Medicine
	Bing et al. [43]	2018	SNN
	Pfeiffer and Pfeil [41]	2019	SNN
	Paredes et al. [42]	2019	SNN
	Bouvier et al. [35]	2019	SNN
	Tavanaei et al. [38]	2019	SNN
	Nunes et al. [36]	2022	SNN
Yamazaki et al. [39]	2022	SNN	
Wang et al. [40]	2022	SNN	
Basu et al. [44]	2022	SNN	
Yi et al. [37]	2023	SNN	
Collection of Methods	Dong-il and Tae-jae et al. [47]	2015	Task-driven analysis
	Lakshmi et al. [48]	2019	Paradigm shift, dataset and simulators
	Gallego et al. [11]	2020	Event representation, computational approaches, applications
	Tayarani-Najaran et al. [49]	2021	Visual, auditory, and olfactory domains
	Zou et al. [46]	2022	Analysis driven by paradigm shift
	Zheng et al. [45]	2023	Computer vision tasks with deep learning focus
	Li and Sun [10]	2023	Data
Zhu et al. [50]	2024	Task-driven analysis	

5. Event Cameras and Computer Vision Tasks

Although there are no clear-cut boundaries, it is very useful to classify computer vision processes into a hierarchy of high-, middle-, (or intermediate-) and low-level tasks [51]. Low-level processes, also called retinotopic processes, involve primitive operations to reduce noise, contrast enhancement, and image sharpening, and their output is usually another image. Middle-level or regional processing involves tasks such as image partitioning into regions or objects (segmentation), or their description to reduce them to a form suitable for further processing, like the classification of individual objects. High-level processing deals with the interpretation and use of what is seen in the image and, at the far end of the continuum, the cognitive functions normally associated with vision. In other words, high-level vision deals with the interpretation and use of what is seen in the image, whereas middle-level vision deals with how this information is organized into what we experience as objects and surfaces [52]. In classic computer vision pipelines, this hierarchy is necessary to distinguish the different computational steps of a complete solution, which starts with image pre-processing (low-level tasks), tunes data for feature extraction (middle-level tasks), and uses the latter to extract meaningful object or scene properties (high-level tasks). Note that the amount of data tends to decrease when passing from the full image to a set of descriptors or regions of interest, up to the final output that can be, for example, pixel coordinates of the object bounding box and/or the belonging class [53]. The data flow concerning the hierarchical level representation and meaning of the task is summed up in the diagram in Figure 8. Each column represents, respectively, the hierarchical level, the goal, the total data decrease associated with levels, and a visual example in a classic computer vision pipeline for the case of human hand gesture recognition. About the latter, it is possible to observe how the raw data acquisition can lead to noisy data, which are cleaned with low-level processing algorithms (e.g., median filters to remove salt and pepper noise) and compressed to speed up computation. A segmentation process separates useful information from the background, and these data are used as the input for a classifier, which outputs the final label. In this way, the amount of associated data also reduces with each step.

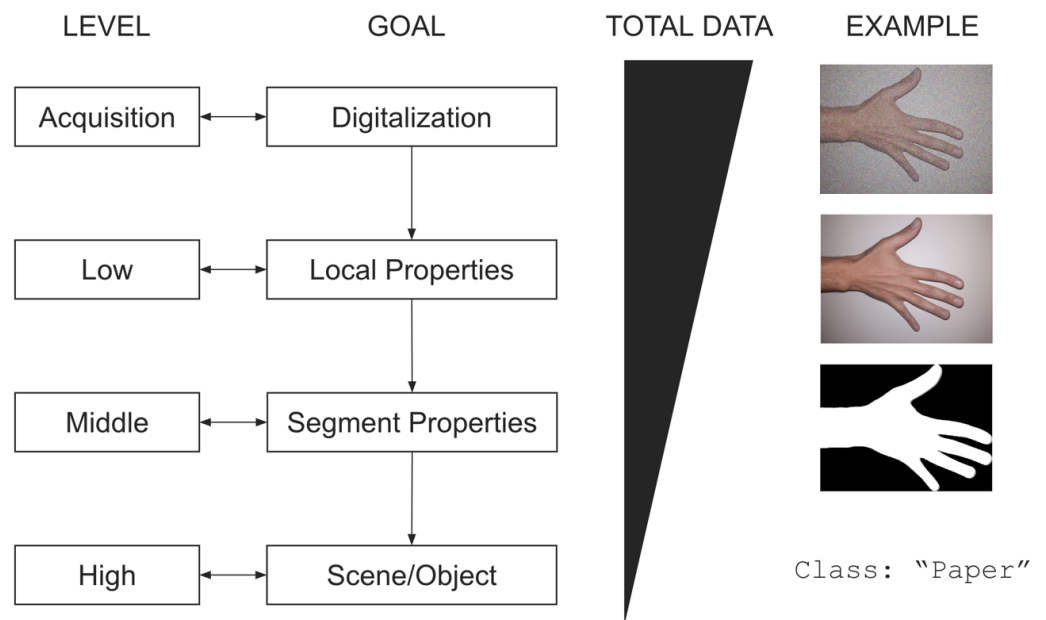


Figure 8. The three-level hierarchical organization to classify computer vision tasks. Usually, the amount of data tends to be lower with higher-level presentations.

Even the recent deep learning approaches that enabled end-to-end learning to address various computer vision tasks are capable of higher-level abstract features that represent powerful semantic representations and are usually located in deeper layers, while the

first layers tend to represent low- (color regions) and middle-level features (like edges, corner) [54]. Independently from this analogy, this hierarchical organization is also useful to highlight how low-level tasks are the building blocks of a higher-level task. For example, a high-level task could be human activity recognition. Low-level tasks could be feature extraction and image segmentation, and middle-level computer vision tasks could be object detection, tracking, and recognition [55].

Computer vision pipelines may achieve great levels of accuracy and understanding of a visual input by integrating low- and high-level information, and event cameras do not represent an exception. We discussed how other surveys usually focused on such tasks as the building blocks of complete applications (see Section 3); a further logical step is to classify the state of the art with a direct focus on the application domain, as will be done in the next section.

6. Applications

In the following, a review of state-of-the-art works categorized by their different application fields and following the methodology presented in Section 3 is presented. The reviewed works employed neuromorphic cameras, either as the unique data source, as well as in multi-sensor systems, and/or with other information sources (e.g., LEDs, measuring instruments, etc.). Application domains that are similar or strongly related were grouped; each of the following subsections will focus on a different group.

At the end of each subsection, a table summarizing the reviewed works is reported (see Tables 2–10); for each work, we report the publication year, the specific final application the work tried to address, details on the sensors employed in the work, notes on data to experimentally evaluate the system, the main computer vision task(s) involved, and publication type; e.g., journal, conference, thesis work. Details on the conference/journal name are also reported. In the tables, with DVS, we denote a pure event stream without grayscale/color information. The heterogeneity of the analyzed works makes it necessary to state several considerations on the content of these tables. First of all, regarding data availability, if authors employed data that have not been released, we report this as not available (*n.a.*). Several works employed both synthetic and real data, coupled in different ways: in some cases, training and tests were performed on datasets converted with the usage of event camera simulators, but a real scenario was added at the end of the experimental phase; in other cases, both approaches were used for different sub-tasks. We tried to give priority to the usage of event cameras, without reporting all the single cases considered in the manuscript, unless we considered the scientific relevance of the section using synthetic data more relevant compared to the experimental test with the event camera. This means that the label *n.a.* does not necessarily imply a non-replicable experimental evaluation. Note that, in certain cases, the authors created data with various sensors or with a DAVIS device, but then the solution purely employed the event stream (other information was used for the evaluation or as part of the released data). In this case, the following convention was adopted: the type of event camera used in the work is always reported, independently from the proposed system; other sensors are reported if data are available as part of the released dataset. This was also applied in the cases where we did not find data reported as available, or available soon; we apologize in advance to those authors who eventually made data available online, but not reported in this manuscript. If data are available, a link to the dataset website/download page is reported. Regarding the column named *main computer vision tasks*, a task-based treatment was beyond our scope and has already been widely covered in the literature; however, it was useful to report computer vision tasks involved, to highlight their association with the final application, so we reported the most important ones for each work, although many tasks are not discussed in this work.

6.1. Agriculture and Animal Monitoring

The key technology in this field is certainly represented by autonomous robots and drones, which are already extensively applied in precision agriculture and farm monitor-

ing, independently from the particular sensor employed [56]. Often, unmanned aerial vehicles (UAVs) for agriculture and farming still make extensive usage of frame-based cameras to perform 2D object detection and tracking in the visible spectrum [57], although hyperspectral imaging is also an emerging technology [58]. Herein, we review systems directly developed to be applied in agriculture and animal monitoring, as well as scene interpretation for autonomous navigation in agricultural and dense vegetation scenarios.

The thesis work in [59] proposed a pipeline that performs fruit detection and classification from event camera inputs. Real-world scenarios like object overlap or variation in size and appearance are taken into account. The specific context was oriented towards investigating the usage of a color DAVIS sensor to extract spatio-temporal patterns in each color filter, having identified color as a feature more relevant than shape and appearance. The complete color information can only be obtained using either an APS readout or computationally demanding events-to-frames reconstruction techniques, thus the focus moved towards event signals for the classification stage. In [60], a dataset for autonomous navigation in different agricultural environments was proposed. The authors highlighted how the currently available datasets tend to privilege the data acquisition process in cities, offices, roads, etc., all cases that are very different from the visual appearance of data coming from the agricultural environment, motivating the introduction of a specific dataset. The aforementioned work proposed the Agri-EBV-autumn dataset, composed of 26 sequences (in 5 different scenarios) of event-based cameras, LiDAR, RGB, and depth information, with additional data for sensor calibration and temporal synchronization. A similar desideratum on the availability of specific data was expressed by the authors in [61]: in fact, the task of autonomous navigation in dense vegetation scenarios presents different conditions like lighting changes, terrains, and/or the effects of wind on leaves. The authors employed a ground robot to collect visual sequences in natural outdoor scenarios with an event camera and applied a bio-inspired neural algorithm for spatio-temporal memory. Finally, they encoded memory in a spiking neural network running on a neuromorphic computer to obtain real-time performance.

Tracking movements of fish for animal behavior analysis was proposed in [62]. A frame-based camera was coupled with an event camera to acquire a dataset used to test classic tracking methods on the bounding boxes detected by a convolutional neural network (CNN). A beamsplitter was used to approximately align the field of view of both cameras, and then the homography was estimated and used to refine the alignment. The work in [63] used event cameras to identify a behavior called *ecstatic display* in nesting Chinstrap penguins; this is characterized by an animal that stands upright, points its head upwards, beats its wings back and forth, and emits a loud call. The problem was formulated as a temporal action detection task and solved in two steps: firstly, temporal region proposals were generated; afterwards, they were classified as ecstatic display or not.

Table 2. Summary of reviewed works in agriculture and animal monitoring.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[60] (2021)	Autonomous navigation in agricultural environments	DVS, Depth, Color, LiDAR	Released [64]	SLAM	Conference (IEEE ICRA)
[59] (2022)	Fruit detection	DAVIS	n.a.	Segmentation	MPhil thesis work
[62] (2022)	Fish trajectory tracking	DVS, Frame-based	n.a.	Detection, tracking	arXiv preprint
[61] (2023)	Autonomous navigation in dense vegetation environments	DAVIS	Released [65]	SLAM	Journal (Science Robotics)
[63] (2023)	Penguin behavior analysis	DAVIS	Released [66]	Classification	Conference (IEEE/CVF CVPR)

6.2. Surveillance and Security

One of the first works using event cameras for detecting vehicles and pedestrians in traffic surveillance scenarios was in [67]. Here, tracking by mean shift and clustering were opportunely tied down to detect and track objects of interest. This approach was extended in [68], leading to an embedded platform for visual surveillance that performs tracking, vehicle speed and length estimation, and vehicle type classification (car and trucks). In [69], event cameras were employed to detect and track multiple persons applying and evaluating Gaussian mixture models. In [70], a single event camera was used to detect the location information of UAVs using their own onboard circle-shaped blinking LEDs. The authors employed a temporal band-pass filter to detect the marker and then used this output to compute relative 3D coordinates of the UAV with respect to the camera. Finally, the work in [71] dealt with the problem of crowd monitoring, and it released the first specific dataset together with a GAN architecture to detect anomalous behavior in crowds. However, no public data were made available.

In [72,73], two solutions to detect human intrusion using an UAV with an event camera as the unique image device were presented. Both systems detect clusters of events caused by moving objects in a static background. Then, in the first work, a CNN was used to estimate the probability that a cluster corresponds to a person. In the second, an attention priority map modeled the fact that the event camera was not static to capture the regions that triggered more events within a time frame and, together with a corner tracker, events with higher priority were selected and the corners were correctly tracked. This let the system filter out clusters that were not updated in a consistent manner or that contained a number of corner tracks lower than a threshold. A computationally efficient approach for intrusion detection was proposed in [74]. A module targeted moving objects using a multivariate normal distribution that was updated event by event using their spatio-temporal information; this update was quick, since it only involved computing the mean and covariance. When the probabilistic distribution converged, a custom CNN was applied to a local region of interest in the image detects if the movement came from a person.

In [75], event cameras were selected to achieve person re-identification, because they can better guarantee the anonymity of subjects, without having the privacy concerns of traditional camera usage in public spaces. The proposed system was evaluated in synthetic datasets and used groups of events to perform person matching in non-overlapping event data using a CNN. The authors stressed that, with traditional images, it is common to apply face blurring or masking, as well as encryption techniques, but if this guarantees privacy for the vision component, it does not necessarily ensure end-to-end privacy. A problem with event cameras is that images reconstructed from a stream of events might constrain event-based privacy-preserving person re-id. Thus, they tested the accuracy with reconstructed images, which was in general lower, concluding how event sensors can represent a step towards privacy-preserving person re-id.

Table 3. Summary of reviewed works in surveillance and security.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[67] (2006)	Traffic surveillance	DVS	n.a.	Detection, tracking	Conference (IEEE DSP)
[68] (2007)	Traffic surveillance	DVS	n.a.	Classification, tracking	Conference (ACM ICDSC)
[69] (2012)	People detection	DVS	n.a.	Tracking	Conference (IEEE/CVF CVPRW)
[71] (2019)	Anomalous behavior analysis	DAVIS	n.a.	Classification	arXiv preprint

Table 3. Cont.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[73] (2020)	Intrusion detection	DAVIS, IMU	Released [76]	Detection, tracking	Conference (IEEE ICRA)
[70] (2021)	UAV tracking	DVS	n.a.	Tracking	Conference (IEEE IROS)
[72] (2021)	Intrusion detection	DAVIS	n.a.	Detection, tracking	Conference (IEEE ICUAS)
[74] (2022)	Intrusion detection	DAVIS	n.a.	Detection, tracking	Conference (IEEE SSRR)
[75] (2022)	Person re-identification	Synthetic data	SAIVT [77], DukeMTMC-reid [78]	Detection, Classification	Conference (IEEE/CVF WACV)

6.3. Visual Inspection and Machinery Fault Diagnosis

The intrinsic nature of event cameras certainly makes them attractive for typical industrial machine vision applications like object counting and particle trajectory tracking.

The first work to characterize and find the upper limits of different event cameras in front of a lathe and a computerized numerical control was proposed in [79]. In [80], the Hough Circle Transform was developed to track microspheres and estimate Brownian motion, i.e., the thermal agitation of micro/nanosized particles in a fluid. The results showed real-time position detection at a frequency of several kHz with a low computational cost. Brownian motion was precisely detected with high speed. In [81], an event camera was used for particle tracking measurements. The experimental setup was built in a way such that only a few pixels registered changes, greatly reducing the bandwidth, as well as storage and processing requirements. The system was tested with a solid-liquid two-phase pipe flow, investigating Reynolds numbers based on pipe diameter and bulk velocity.

In [82], an event camera was used as an add-on to Magneto-optic Kerr effect (MOKE) microscopy, used to observe magnetic domains and various other micro-structures in magnetic materials. The authors reconstructed videos from events by frame stacking and used time surface methods to evaluate the results. A system to count falling corn grains using an event camera was proposed in [83]. The system was composed of a vibrating feeder to allow the number of falling corn grains to be controlled. Images made with an event accumulation time set to 2 ms were used (i.e., an image generation rate of 500 fps). Corn grains were segmented using morphological classic image processing and, by using horizontal count lines and tracking of the same grain over time, they were able to correctly count each fast-moving single corn grain only once. In [84], event cameras were used for machine fault diagnosis. In particular, a two-channel 2D representation was created by considering positive and negative events as two single channels, and this image was used as the input of a CNN architecture to detect four types of rotating machine faults of industrial bearings. Three rotating speeds of 1200, 1800, and 2400 r/min were evaluated in the experiments.

The work in [85] investigated the possibility of detecting the rotational speed from event data, with applications in engine monitoring during car repairs, fault detection in electrical appliances, and more. The solution proposed several image processing steps to estimate the rotational speed; first of all, clustering was used to detect the different rotating targets, while an outlier removal component removed the background noise. The K-means was employed to separate clusters, using the Davis–Bouldin index [86] to work without prior knowledge of the number of clusters. Streams of the same downsampled cluster were registered with the iterative closest point (ICP), to obtain the angle of rotation in a known period of time, and this knowledge was used to achieve accurate estimation of rotational speed.

In [87], event cameras were used as optical disdrometers, i.e., devices to measure the diameter and speed of hydrometeors at ground level. The system was tested with a droplet generator in a setup where droplets were occupying between 10 and 20 pixels of the sensor area. The datasets have been made publicly available. In the context of Schlieren imaging, a technique to estimate the flow of transparent media (with several applications in industry, from aerodynamics to gas leakage detection), event cameras were combined with frame cameras to visualize gas streams in [88]. Here, the authors also created and published a dataset due to the lack of available data. In [89], event data were used to classify delivery packages and detect falling events in a ring sorting belt scenario. The authors employed YOLO [90] as the backbone network for testing different event representations. A dataset for the two tasks was also proposed. The system classifies three types of packages and detects the event of a delivery package falling.

Table 4. Summary of reviewed works in visual inspection and machinery fault diagnosis.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[79] (2011)	Sensor characterization in machine vision	DVS	n.a.	- (raw data analysis)	Conference (IEEE SIGMAP)
[81] (2011)	Particle tracking	DVS, ultra high-speed camera	n.a.	Tracking	Journal (Springer Experiments in Fluids)
[80] (2012)	Particle tracking	DVS, frame-based camera	n.a.	Detection, tracking	Journal (Wiley, Journal of microscopy)
[82] (2022)	Magnetic materials analysis	DVS, microscopy	n.a.	- (raw data analysis)	Journal (AIP Advances)
[83] (2022)	Corn grain counting	DVS	n.a.	Tracking	arXiv preprint
[89] (2022)	Fault detection in industrial pipeline	DAVIS	n.a.	Detection, classification	Journal (Frontiers in Neurorobotics)
[84] (2023)	Machine fault detection	DVS	n.a.	Classification	Journal (IEEE Transactions on Industrial Informatics)
[85] (2023)	Rotational speed estimation	DAVIS	n.a.	Segmentation	Journal (IEEE Transactions on Mobile Computing)
[88] (2023)	Schlieren imaging	DVS, frame-based camera	Released [91]	- (trajectory estimation by optical flow)	Journal (IEEE-TPAMI)
[87] (2024)	Optical disdrometers	DAVIS	n.a.	- (raw data analysis)	Journal (Copernicus Publications, Atmospheric Measurement Techniques)

6.4. Space Imaging and Space Situational Awareness

Works in this field tend to exploit event cameras to obtain space situational awareness (SSA); i.e. the discipline that keeps track of objects in orbit and predicts their position over time. Indeed, the advantages of lower bandwidth and power requirements make them suitable for use in remote locations and space-based platforms.

The work of [92] was the first attempt to employ event cameras in SSA applications. The authors modified a robotic electro-optic telescope to support the event-based sensors and pre-existing equipment simultaneously, to develop and test object detection algorithms. Several trials were conducted, tracking and detecting objects in low earth orbit and in geosynchronous orbit, proving the viability of event-based sensing for space situational awareness. The work in [93] instead proposed a feature-based detection and tracking

method for SSA applications implemented via a cascade of increasingly selective event filters, to isolate events associated with space objects without losing the high temporal resolution of the sensors. Moreover, the authors presented a dataset composed of multiple event sensor data in both daytime and nighttime recordings and containing 572 labeled targets with a wide range of sizes, trajectories, and signal-to-noise ratios.

A star tracker is a vision system with an image processing algorithm used to estimate the attitude of a spacecraft by recognizing star patterns. In [94], the authors used an event camera instead of frame-based imaging devices to build a star tracker. A simple heuristic selected the best event images for star identification, generating a set of rotation estimations. Furthermore, relative rotations were estimated from the event images, and the different measurements were fused in an optimization step with rotation averaging and a final bundle adjustment, showing the feasibility of implementing a star tracker with event cameras. In [95], the possibility of using event cameras for spaceflight and the effects of neutron radiation on camera performance were investigated. An event-based sensor was irradiated under wide-spectrum neutrons and its effects were classified. The authors analyzed radiation-induced damage to the sensor under wide-spectrum neutrons and the radiative effect on the signal-to-noise ratio of the output at different angles of incidence from the beam source. The results showed fast recovery during radiation and a high correlation of noise event bursts with respect to source macro-pulses. Significant differences were found in the spatial structure of noise events at different angles. Finally, the Event-based Radiation-Induced Noise Simulation Environment (Event-RINSE), a simulation environment based on the noise modeling presented in the manuscript and capable of simulating the effects of radiation-induced noise from the collected data on any stream of events, was introduced. The work in [96] dealt with the problem of calibration of event cameras for reliable and accurate measurement acquisition. The authors produced a star mapping and source-finding algorithm to generate resolved images of event sources at varying speeds to calibrate the event camera, finding relationships to project the events' pixel coordinates to coordinates in a target resident space object reference frame. In [97], SSA was obtained using event cameras to detect, localize, and track resident space objects. A computationally efficient estimation with sub-pixel detection using an unsupervised tracking-by-detection algorithm was achieved.

Table 5. Summary of reviewed works in space imaging and space situational awareness.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[92] (2019)	Space imaging	ATIS, DAVIS, telescope	n.a.	Detection	Journal (Springer, The Journal of the Astronautical Sciences)
[94] (2019)	Star tracker	DAVIS	Released [98]	- (modeling, parameter estimation)	Conference (IEEE/CVF CVPRW)
[93](2020)	Space imaging	ATIS, DAVIS, telescope	Released [99]	Detection, tracking	Journal (IEEE Sensors Journal)
[95] (2021)	Sensor characterization for space applications	DVS	n.a.	- (signal processing techniques)	Journal (IEEE Access)
[97] (2022)	Resident space objects analysis	Simulated data	Released [100]	Detection, tracking	Journal (Frontiers in neuroscience)
[96] (2023)	Calibration for space imaging	DVS, telescope	Released [101]	Tracking	Journal (Springer Astrophysics)

6.5. Eye Tracking, Gaze Estimation, and Driver Monitoring Systems

In this section, we group together eye tracking and gaze estimation with driver monitoring systems (DMS), i.e., the in-cabin safety system that detects a driver's dangerous driving behaviors. Such systems, in general, analyze the driver's physical and cognitive state to reveal fatigue, distraction, and other dangerous driving conditions, warning the driver to correct his/her behavior. This grouping is motivated by the fact that the event of driver distraction is detected by direct analysis of visual facial cues like slow eyelid movement, blinking, narrowed eyes, yawning, gaze, head pose, etc. [102].

The automatic detection of eye positions, their temporal consistency, and their mapping into a line of sight in the real world have become a hot topic in the field of computer vision during the last decades, and different sensors, from expensive ad hoc devices to low-cost consumer cameras, from visible to near-infrared (NIR), from invasive to unconstrained devices, have been experimented on and tested over the years [103]. Different works considering the unique features of event cameras have also been presented. In [104], event camera simulated data were used to improve the eye segmentation step of a gaze estimation algorithm designed for infrared cameras; in particular, to extract temporal correlation across frames that, in turn, guided a lightweight deep neural network that predicted the region of interest in the actual frame and a predictive algorithm that decided whether the current eye frame required going through a full-fledged segmentation algorithm or could be extrapolated from the previous segmentation map. In [105], simulated event camera data were used for dataset creation, in order to train a neural network based on the YOLOv3-tiny architecture [106], to detect and track faces and eyes. To exploit the natural sparsity of events, the architecture was modified to integrate a fully convolutional gated recurrent unit (GRU) layer, defining a *gated recurrent YOLO* (GR-YOLO). The performance was then evaluated on real event camera data. Moreover, blinking was detected, having a direct application to driver attention monitoring systems.

In [107], event camera data were simulated to improve eye tracking detection from color images using an image domain translation. A cross-modal neural network was then employed using both RGB and simulated event data and then tested with a real DAVIS camera. A hybrid system of frame and event camera was proposed in [108]. An online 2D pupil fitting method updated a parametric model for each event or a few grouped events, while the final point of gaze was estimated in real time with a polynomial regressor. A classic camera anchored the pupil tracking system with traditional pupil detection algorithms, while the events allowed it to update the location of the pupil at a very high speed, equivalent to 10,000 Hz of a frame-based camera. A dataset composed of event and image data from 24 subjects watching a video stimulus, with their saccadic motions and smooth pursuits, was also created. The dataset was recorded by mounting two DAVIS sensors with NIR lenses and a NIR illumination source close to the user's head. Moreover, in [109], event camera images were paired with grayscale frames, although coming from the same DAVIS device. An event-to-image encoding technique captured event data in temporal bins of 33 milliseconds, fusing events with the nearest corresponding grayscale frames. A neural network performed temporal fusion of the information of grayscale and event images, taking as input two consecutive temporally encoded frames to predict gaze centroids. The authors introduced the Gaze-FELL dataset, recorded with a low-light setup and made with five subjects wearing a head-mounted display with a DVS and a frame-based camera, to store eye patches while watching a video stimulus. In [110], instead, IR LEDs were combined with event camera data to implement eye tracking. In particular, *coded differential lighting*, a novel dual-LED design to enhance event camera data, while suppressing non-specular background events, was introduced. The events triggered by the flashing lights were filtered to calculate glint locations at high frequency. Since glints have a unique binary sequence of pulses, the correspondence of each calculated glint with the illumination sequence was used to infer the position of the corneal sphere with respect to the camera.

A system that fully employs event cameras was proposed in [111]. A real-time pipeline to extract pupil features in the form of ellipses and a recurrent neural network accurately estimated gaze from a pupil feature sequence. The network was trained with the same dataset as [108], but only using the event cameras data and opportunely adding event labels. In [112], instead, an event-based eye-tracking system that extracts pupil features was proposed. It takes data from an event camera on top of a head-mounted display and only uses event data. The events triggered by eye motion are converted into three-channel frames with an event-to-frame conversion method, where each channel represents positive events, negative events, and their combination. Events representing the pupil are classified by a CNN; a tracker is finally employed to reduce the amount of CNN inference.

A system that extracts several visual cues with applications in driving monitoring systems was proposed in [113]. Here, a network trained on synthetic event camera data estimated head pose, eye gaze, and the presence of facial occlusions. An event integration method was capable of handling both short- and long-term temporal dependencies to compensate for global head motion. Regarding driver monitoring systems not directly based on gaze analysis, the work in [114] extracted the events of yawns and seat belts (fastening or unfastening). A neural network designed for both tasks combined a CNN backbone with a self-attention module and a recurrent head, and it was trained on synthetic data derived from RGB and NIR video from both private and public datasets.

In [115], the problem of driver face detection was addressed by first constructing an event representation by making the time domain discrete and presenting a lightweight translation-invariant backbone to extract multi-scale features. The authors introduced a shift feature pyramid network and shift context modules to take into account the feature extraction step with limited computational cost. In [116], a dataset recorded with DAVIS and a depth sensor for driving monitoring systems was proposed and tested with some state-of-the-art architectures. The dataset contained data for driver drowsiness detection, driver gaze-zone recognition, and driver hand-gesture recognition. Before the testing phase, three popular event-encoding methods used to convert asynchronous event slices to event frames were presented. In [117], event cameras were augmented by submanifold sparse neural network models (SSNN) and integrated into a DMS, with the particular use case of driver distraction monitoring. An SSNN model was trained with synthetic event data generated from driver monitoring public datasets.

Table 6. Summary of reviewed works in eye tracking, gaze estimation, and driver monitoring systems.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[116] (2020)	DMS	DAVIS, RGBD, IR	Released [118]	Detection, classification	Journal (IEEE Transactions on Intelligent Transportation Systems)
[105] (2021)	DMS	Simulated data	n.a.	Detection, tracking	Journal (Elsevier Neural Networks)
[108] (2021)	Gaze estimation	DAVIS with NIR lenses, video stimulus	Released [119]	Detection, tracking	Journal (IEEE Transactions on Visualization and Computer Graphics)
[104] (2022)	Gaze estimation	Simulated data	OpenEDS TEyeD [121]	[120], Segmentation	Conference (IEEE VR)
[110] (2022)	Eye tracking	DVS	n.a.	Tracking	Conference (IEEE/CVF WACV)

Table 6. Cont.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[115] (2022)	Face detection for DMS	DAVIS	NeuroIV [116]	Detection	Conference (IEEE ICARM)
[107] (2023)	Pupil localization	Simulated data	WIDER FACE [122]	Detection	Journal (IEEE Access)
[112] (2023)	Eye tracking	DAVIS	Angelopoulos et al. [108]	Detection	Conference (IEEE AICAS)
[113] (2023)	DMS	Synthetic data	BIWI [123]	Detection, tracking	Journal (IEEE Access)
[114] (2023)	DMS	DVS	YawDD [124]	Classification	Journal (IEEE Access)
[117] (2023)	DMS	DVS	n.a.	Classification	Journal (IEEE Open Journal of Vehicular Technology)
[109] (2024)	Gaze estimation	DAVIS, video stimulus	Available on request	Detection	arXiv preprint
[111] (2024)	Gaze estimation	DAVIS	Angelopoulos et al. [108]	Segmentation, Detection	Journal (IEEE TPAMI)

6.6. Gesture, Action Recognition, and Human Pose Estimation

The ability to capture moving object dynamics as events makes neuromorphic cameras an emerging sensor type for human and/or human body part pose and shape estimation. A first complete solution to estimate 2D human pose was proposed in [125]. The system employed a combination of event cameras and a bio-inspired software architecture. First, size and position invariant line features were extracted and organized into vectorial segments. The extracted line segments were used to measure the similarity with a known set of posture libraries; the classification was based on a modified line segment Hausdorff-distance scheme. To achieve size and position invariance, event-cluster-based methods computed from individual pixel events were used. Subsequently, several works have been proposed over the years, often exploiting the emerging outcomes of deep learning.

Two-dimensional shape from event cameras was estimated using a single CNN in [126]. Here, a benchmark dataset of human body movements, made using data from four synchronized DVS cameras with a resolution of 346×260 was introduced. A total of 33 movements of 17 subjects were recorded in total. Conversely, 3D human motion was estimated with the help of grayscale information in [127]. The method combined model-based optimization with CNN-based human pose detection. The method relied on a pre-processing step to reconstruct a template mesh; the skeleton parameters of the template were optimized to match data from the event camera. A three-step tracking algorithm takes in the input intensity, events, and a textured mesh to obtain the final pose.

The work in [128] presented a two-stage system for estimating human poses: firstly, a modified version of the U-Net [129] mask prediction network was employed to eliminate moving backgrounds; then, to facilitate information flow between frames, a deep learning architecture was used. A time-ordered recent event volume representation was used to construct denser and more informative input tensors. The paper also addressed the problem of upper body motion with a stationary lower body. In addition, in [130], human pose was estimated with event cameras: here, the key idea was the usage of a lightweight image-like event representation; in this way, the system can deal with the disappearance of static body parts; moreover, it allows using the higher quantity of data available with frame-based datasets in the pre-training step, followed by fine-tuning with event camera datasets.

Similarly as for body pose, gesture recognition has been investigated for more than a decade. Since works like [131,132] where, respectively, event camera-based detection of rock, paper, and scissors, and finger tracking and hand swipe direction motion detection, were proposed, more works on gesture recognition have been proposed. In [133], a complete software/hardware system that used a neurosynaptic processor and an event camera to recognize 11 different hand gestures was proposed. The core algorithm consisted of a CNN trained offline. The resulting stream of instantaneous classification outputs was filtered using a sliding window.

In [134], event data were modeled as a set of 3D points in space-time. This representation was used to adapt PointNet [135] and PointNet++ [136], neural network architectures designed for 3D point cloud matching and recognition. In this way, the gesture recognition problem was modeled and solved with 3D object recognition. Similarly, in [137], a stream was represented as spatio-temporal 3D event clouds, but a dynamic graph CNN (DGCNN) [138], a neural network architecture designed for classification and segmentation tasks with point clouds, was adapted. The difference with PointNet or PointNet++ is that DGCNN incorporates the information of neighbors of each point during the feature embedding computation phase, making full usage of the local structure, capturing geometric features with high resolution. In [139], DVS was integrated with a wearable glove with five high-frequency active LED markers, to reduce global latency and enhance the robustness of the recognition performance. A restricted spatio-temporal particle filter was used to estimate the trajectories of the markers. The work in [140] introduced a dataset and three classification methods, implementing SNN architectures to test performance on the dataset for the problem of sign language gesture recognition. Similarly, in [141], the authors proposed a system to detect sign language gestures based on SNN with spatio-temporal back propagation training methods and a specific dataset composed of frame-based and event data for the experimental evaluation.

Recently, human action recognition exploiting neuromorphic vision has also shown promising results. In [142], a *compressed event tensor* was used to represent event data. This representation projected the quantization of events into a fine voxel grid structure with a high temporal resolution, to address low temporal resolution and blurring. Moreover, a framework called *branched event net* was introduced to deal with both static and dynamic scenes. The system also works for object detection tasks. Another example of specific event data representation can be found in [143]. Here, the introduced representation is named *Compact Event Image* and it was generated using a module based on self-attention in a learnable way. This module summarizes the long-term dynamics and temporal patterns of the events into a frame set, and this is combined with backbone architectures to achieve end-to-end action recognition.

The video transformer network [144], a transformer-based architecture, was used to acquire spatial embedding from events using a temporal self-attention mechanism in the work presented in [145]. Spatial and temporal operations are separated to obtain a more computationally efficient architecture. A specific loss function that contrasts temporally misaligned frames is used to learn fine-grained spatial cues in the spatial backbone network. This approach was tested on actions from both egocentric vision and from hand/arm gestures (with DVS data).

Table 7. Summary of reviewed works in gesture, action recognition, and human pose estimation.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[125] (2011)	Gesture recognition	DVS	n.a.	Classification	Journal (IEEE-TPAMI)
[131] (2011)	Gesture recognition	DVS	n.a.	Classification	Conference (IEEE CIMSIVP)
[132] (2012)	Hand gesture UI	DVS	n.a.	Classification	Conference (IEEE ICIP)
[133] (2017)	Gesture recognition	DVS	Released [146]	Classification	Conference (IEEE/CVF CVPR)
[126] (2019)	Pose estimation	DAVIS, motion capture	Released [147]	Pose estimation	Conference (IEEE/CVF CVPRW)
[134] (2019)	Gesture recognition	DVS	DVS128 Gesture [133]	Classification	Conference (IEEE/CVF WACV)
[137] (2020)	Gesture recognition	DVS	DVS128 Gesture [133], DHP19 [126]	Classification	Conference (IEEE IS-CAS)
[127] (2020)	Pose estimation	DAVIS	n.a.	Pose estimation, tracking	Conference (IEEE/CVF CVPR)
[139] (2021)	Gesture recognition	DAVIS	n.a.	Classification, tracking	Journal (IEEE Transactions on Automation Science and Engineering)
[142] (2022)	Action recognition	DAVIS	DVS128 Gesture [133], N-Caltech101 [148], DVSAAction [149], NeuroVI [116]	Classification	Journal (IEEE access)
[140] (2022)	Sign language	DVS	Released [150]	Classification	Journal (Springer, Pattern Analysis and Applications)
[143] (2022)	Action recognition	DVS	DVS128 Gesture [133], UCF101-DVS [151], HMDB51-DVS [151]	Classification	Conference (IEEE CRC)
[128] (2023)	Pose estimation for dancing	DAVIS, RGB (HD), motion capture	Released	Pose estimation	Journal (Elsevier Neurocomputing)
[130] (2023)	Pose estimation	DVS	DHP19 [126], Human3.6m [152]	Pose estimation	Conference (IEEE/CVF CVPR)
[141] (2023)	Sign language	DAVIS	Released [153]	Classification	Journal (MDPI Electronics)
[145] (2023)	Action recognition	Synthetic data	N-EPIC-Kitchens [154]	Classification	Conference (IEEE IROS)

6.7. Medicine and Healthcare

The work of [155] proposed an event camera-based system to detect accidental falls in elderly home care applications. The implemented algorithm estimates the instantaneous motion vector and reports fall events. Fall events are distinguished from walking, crouching down, and sitting down. The work in [156] employed a DAVIS sensor to detect falls. First, a set of temporal range proposals relevant to the event of a fall were extracted. Each temporal proposal was processed by a feature extraction backbone network, and the candidate proposal was classified together with a phase of temporal boundary refinement.

In [157], event cameras were used to capture subtle changes in the surface of the skin caused by the pulsatile flow of blood in the wrist and to estimate the patient's heart rate. The authors used a colored dot on the patient's wrist in order to compute, on 5×5 pixel regions of the area of interest, the dominant frequency from a periodogram, which is an estimate of power spectral density, to estimate the final pulse rate. In [158], a lightweight wearable to support visually impaired people with navigation and obstacle avoidance tasks was proposed. Two event cameras were used to extract depth information in real time, which was translated into acoustic signals. Spatial auditory signals were simulated at the computed origin of visual events. A device that can be used as a visual-to-auditory sensory substitution device, as well as a component of a real-time retinal prosthesis or vision augmentation system, was proposed in [159]. The sensory block is composed of an event camera, and data are then processed in another system component that treats events like post-synaptic potentials. A final block emulates the temporal contrast sensitive retinal ganglion cells. Another medical application can be found in the doctoral thesis in [160], which coupled event cameras with medical imaging devices to compute red blood cell velocities and densities within capillaries, proposing a system that can estimate deregulation of the micro-circulation within minutes.

Furthermore, in [161], an event camera was used to detect and localize molecules with the single-molecule localization microscopy technique. An optical setup was prepared to detect and localize single molecules, focusing on blinking labels. The response of the sensor was characterized with respect to a fluorescent signal, and organic dyes for single-molecule fluorescence imaging were then detected. The authors compared the results with frame-based vision, demonstrating how event cameras are suited to the extraction of biological dynamics, particularly for monitoring processes with a wide range of dynamic scales and for blinking-based images.

Table 8. Summary of reviewed works in medicine and healthcare.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[155] (2008)	Fall detection	DVS	n.a.	Classification	Journal (IEEE Transactions on Biomedical Circuits and Systems)
[158] (2016)	Assisting device	DVS	n.a.	Detection	Conference (IEEE Healthcom)
[159] (2016)	Assisting device	DVS	n.a.	- (raw data analysis)	Conference (IEEE BioCAS)
[160] (2018)	Cellular analysis	ATIS, Frame-based camera	n.a.	Detection, tracking	PhD Thesis
[156] (2022)	Fall detection	DAVIS	n.a.	Classification	Journal (IEEE Transactions on Cybernetics)
[157] (2023)	Heart rate detection	DVS	n.a.	- (raw data analysis)	Conference (IEEE ISM)
[161] (2023)	Cellular analysis (SMLM)	DVS, microscope	Released [162]	Detection	Journal (Nature Photonics)

6.8. Intelligent Transportation Systems

In this section, we analyze the field of scene analysis for intelligent vehicles; i.e., vehicles that can detect the scene and send the proper outcomes to the driver (in terms of dashboard/audio messages), as well as an autonomous driving system. This industrial sector necessarily overlaps with the autonomous navigation pursued in many robotics applications (see Section 6.9), as well as some scene analysis work previously introduced

in Section 6.2; regarding the first point, to keep the focus on computer vision, only works related to scene understanding are reviewed in this part; regarding the second, we consider the final application scenario proposed by the authors, although the underlying challenges are often shared by both domains. In [163], a vehicle detection and tracking system was proposed. The authors compared three different classical clustering methods and four tracking approaches. A dataset recorded by a DAVIS sensor mounted on a highway bridge was used to test the system. The results showed the possibility of detecting and tracking multiple vehicles at a high frame rate. In [164], the same problem was addressed by using an SNN trained on synthetic data from roadside-event-based cameras with multiple weather conditions.

Neuromorphic cameras were used to collect the point cloud data of moving targets and detect motor/non-motor vehicles and pedestrians using the geometrical, quantitative, and Gaussian projection characteristics of the captured point clouds in [165]. The system also implemented a shadow removal step based on feature similarity, again based on point cloud distribution characteristics.

A dataset for lane extraction composed of event frames (with a total of labeled 17103 lane instances) was released and tested with state-of-the-art semantic and instance segmentation methods in [166]. In [167], a CNN with a feature attention gate component (FAGC) for vehicle detection was introduced. The system integrated grayscale and event features fed into the FAGC to generate pixel-level attention feature coefficients that improved the performance.

In [168], YOLO was tested with different representations of event data and with a deep learning-based video frame reconstruction technique to detect traffic signs. Similarly, a system of visible light positioning that used event cameras to localize their position based on multiple flickering LEDs was proposed in [169]. The authors employed an LED detector when flickering was visible and a Gaussian mixture probability hypothesis density filter for tracking, without requiring data association. LED flickering frequency and position were detected.

Table 9. Summary of reviewed works in intelligent transportation systems.

Work	Application	Sensors	Datasets	Main Computer Vision Tasks	Publication Type
[163] (2018)	Vehicle detection	DAVIS	n.a.	Detection, tracking	Journal (Hindawi, Journal of advanced transportation)
[164] (2023)	Vehicle detection	Simulated data	n.a.	Detection	Conference (IEEE IV)
[165] (2021)	Vehicle detection	DVS	n.a.	Detection, classification	Journal (IEEE Sensors Journal)
[166] (2019)	Lane detection	DVS	Released [170]	Classification, segmentation	Conference (IEEE/CVF CVPRW)
[167] (2021)	Vehicle detection for autonomous navigation	DVS, Grayscale	DDT-17 [171]	Detection	Journal (IEEE Sensors Journal)
[168] (2022)	Traffic sign detection	DVS	The 1 Megapixel Automotive Detection Dataset [172]	Detection	Conference (IEEE SPA)
[169] (2020)	Light positioning system	DAVIS	n.a.	- (image processing techniques)	Journal (IEEE Sensors Journal)

6.9. Robotics

The possibility of providing robots with event cameras can allow solving robotic tasks that range from situational awareness to vision-based control, and from object grasping to terrain reconstruction. Considering the quantity of computer vision tasks that are connected with the field of robotics, as well as the availability of surveys that directly focus on robotic applications [11,27,30], this section illustrates recent advancements directly focusing, as done up to now, on the final application. Before doing that, it is important to highlight how one of the most investigated components for event camera-based processing in robotics is situational awareness, which is a key element in the success of autonomous tasks. For example, in [173], an event camera was used to obtain obstacle avoidance in UAV systems. The work used this data source to distinguish between static and dynamic objects and presented a fast strategy to generate the motor commands necessary to avoid obstacles. The proposed system was able to avoid multiple obstacles of different sizes and shapes during navigation at relative speeds of up to 10 m/s. The UAV could operate in both indoor and outdoor scenarios. A SLAM system for ground robots that exploited a single event camera was proposed in [174]. The authors introduced a method that applied optimization over motion and structure: in particular, the representation image of warped events was extended with contrast maximization for the 3D case. In this way, the robot could perceive non-planar environments under arbitrary motion.

Multiple robot detection and trajectory tracking methods (with experimental results for up to 4 robots operating at the same time) were proposed in [175]. The task of detecting robots in an indoor arena was achieved. The DBSCAN algorithm [176] was used to detect the robots, with a single k -dimensional tree to track them. RGB data were used to provide the ground truth. The authors explored the performance of the camera in the case of different event accumulation times and light conditions in the indoor arena. The work in [177] introduced a grasping framework for multiple known and unknown objects based on an event camera (with a model-based and model-free approach, respectively). For the known objects, the camera was used to localize them in the scene, while point cloud processing detected and registered them. In the case of unknown objects, event-based object segmentation was proposed to localize the objects, as well as visual servoing with grasp planning to localize, align, and grasp the targeting object. Tests were performed with a system composed of a UR10 robot, a neuromorphic camera, and a Barrett hand gripper.

In [178], a UAV with a cable-suspended load employed an event camera to achieve a robust and fast estimation of the state of the cable during transportation. The authors achieved a faster estimation when compared with frame-based cameras, a fundamental requirement for a controller for feedback on the state of a cable. A point cloud representation was used to model event data, and the respective measurements were fitted to a Bézier curve to estimate the cable angle and angular velocity. The work in [179] proposed visual tracking of beacons made of LEDs as an optical camera communication system based on neuromorphic cameras for robotic applications. High-frequency visual intensity signals from visual beacons were captured by an event camera, and a robust demodulation algorithm to decode the transmitted data was presented. Detection was achieved by conventional blob detection and tracking methods, taking advantage of the nature of event data.

Very recently, in [180], event cameras were used to implement an LED-based communication system for multi-agent robots, proposing an alternative to fiducial markers with frame-based cameras. The implemented multi-agent system verified the system functionality with experiments using physical robots. In [181], the first study to fuse frame-based and event-based data with a deep learning approach to estimate the instantaneous human steering wheel angle in driving scenarios was presented. To achieve this, the authors introduced a dataset of both data sources captured with a DAVIS camera and that integrated human control data, for a total of 51 hours of driving on 4000 km of highway and urban driving, with different illumination conditions. In [182], a prototype using a pulsed line laser and a DVS was designed for fast terrain reconstruction. Temporal histograms were used at each

laser pulse instance to adapt a scoring function. The score of each event was calculated and mapped on the score maps. Afterwards, the maps were averaged and the laser stripe was extracted by selecting the maximum scoring pixel for each column when above a threshold.

Table 10. Summary of reviewed works in robotics.

Work	Application	Sensors	Datasets	Computer Tasks	Vision	Publication Type
[182] (2014)	Terrain reconstruction	DVS, laser scanner	n.a.	- (raw data analysis)		Journal (Frontiers in neuroscience)
[173] (2020)	Obstacle avoidance	DVS	n.a.	Segmentation		Journal (Science Robotics)
[181] (2020)	Steering prediction	DAVIS	Released [183]	- (raw data and steering angles prediction)		Conference (IEEE ITSC)
[175] (2021)	Robot detection	DVS	n.a.	Tracking		Journal (IEEE Access)
[174] (2022)	Visual Odometry	DVS	KITTI [184], ORB-SLAM2 [185]	SLAM		Journal (MDPI Sensors)
[177] (2022)	Robot grasping	DAVIS	n.a.	Segmentation		Journal (Springer, Journal of Intelligent Manufacturing)
[178] (2023)	Load transportation with UAVs	DAVIS	n.a.	Segmentation		Journal (IEEE Robotics and Automation Letters)
[180] (2024)	LED-based communication system	DVS	n.a.	Detection		Conference (ACM AAMAS)

7. Discussion

From the analysis of the state of the art in the different application fields, it is possible to derive findings that are application-domain-dependent, as well as observations that are directly related to the technology under examination. Before stating the conclusions related to the different application fields, it is important to analyze common challenges and opportunities.

First of all, initial consideration must be given to low- and high-level tasks. We already highlighted how a low-level component addresses a simple task that can be leveraged for downstream applications. In this survey, we did not focus on tasks, because of the wide availability in the literature, moving the center of attention to the application domain instead. It is useful anyway to link computer vision tasks with the final application. A diagram showing this association, limited to the works introduced in Section 6, is reported in Figure 9. It is possible to observe how, apart from very specific tasks like pose estimation in the context of human pose analysis, single tasks serve as building components that transcend multiple application domains, illustrating the principle that a core computational function can be both pivotal and ubiquitous across analyzed application domains. As for traditional imaging, this interconnectivity highlights the integral role of computer vision tasks in advancing the capabilities of technological systems.

Generally speaking, a critical analysis of the progress made by the research community in neuromorphic computer vision in single lower-level tasks makes it possible to state how the application of event cameras in both research and industry can only grow with time, considering that the field of application of such sensors will, in practice, depend only on the creativity in combining approaches or adapting a research outcome from a specific context to a different application field. Thus, integrating the application-driven analysis and the specific application outcomes from this work with other surveys that showed tasks and/or

the internal tasks associated with a single application domain is crucial to obtaining a deep understanding of challenges and opportunities.

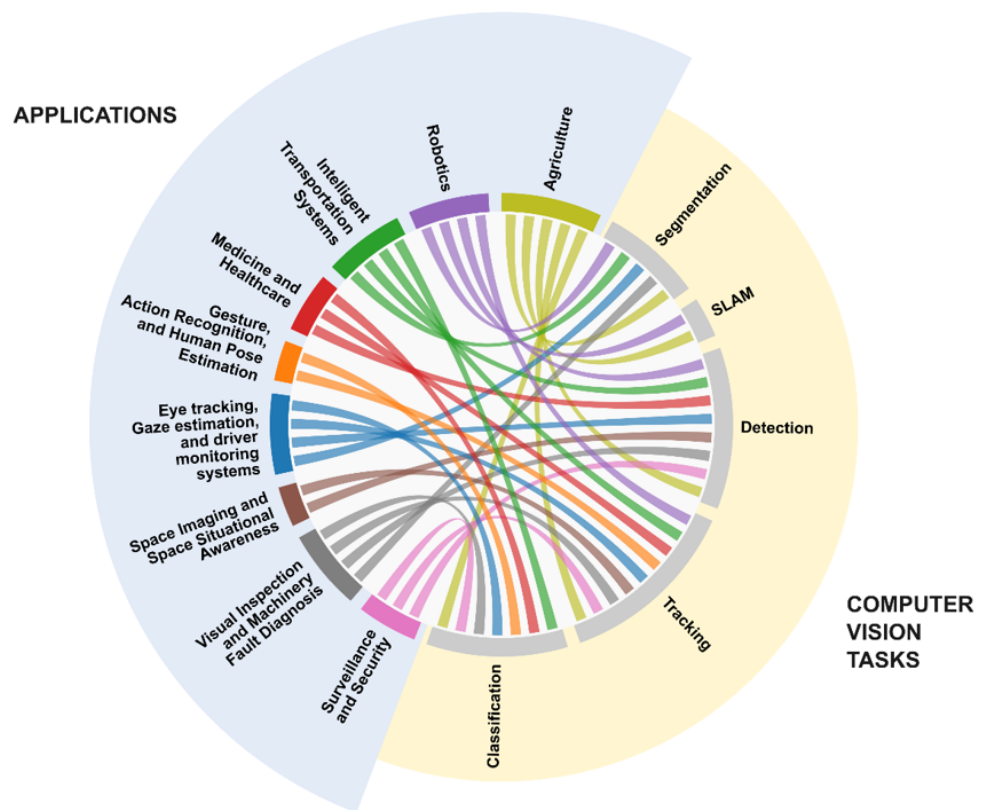


Figure 9. The association between computer vision tasks and application domains for the works analyzed in Section 6.

The computer vision tasks required for the applications summed up in Figure 9 also pose the interesting question of how the analyzed state of the art compares with traditional imaging systems. A quantitative comparison with traditional sensors has often been regarded as *unfair* in the literature [11,26], due to the different time scales over which these tasks have been investigated by the neuromorphic engineering community and the traditional computer vision community, and the consequent different maturity level of the respective solutions. In addition, the advantages of event cameras make neuromorphic sensors more or less suitable depending on the working conditions, so it is essential to consider the final application context and environment. Apart from these considerations, some qualitative observations can still be made. For instance, for detection and classification, traditional imaging algorithms applied in conventional full scene acquisition tend to perform better, due to a richer information content. In contrast, in the case of tracking, event cameras can offer advantages due to their high temporal resolution and low latency, and this becomes particularly true in high-speed or high-dynamic-range scenarios. As for segmentation, the performance can be highly dependent on the specific context. Finally, for SLAM (but very often this is also valid for the other aforementioned tasks), the best outcomes can be achieved when event cameras and traditional cameras are integrated together or as part of multi-sensor solutions (e.g., LiDAR, GPS, etc.), leveraging the strengths of both technologies. In all the cases, however, the application requirements, e.g., the acquisition conditions, such as low-dynamic-range scenarios or the necessity to acquire data at very high frame rates, are fundamental to deciding on the most suitable sensor. To conclude, generally speaking, frame-based solutions currently lead in terms of algorithm maturity and performance in many computer vision tasks. Hence, there is the need to develop new algorithms and analysis methods to fully exploit the potential of event-based acquisitions.

On the other hand, application-domain-based analysis alone lets us observe how each application tries to exploit the different advantages of switching to the event-based paradigm. It is not easy to make a systematic comparison with other sensors, since they have been investigated for decades, while the off-the-shelf availability of event cameras is very recent in comparison. Moreover, it is possible to observe the absence of an outstanding neuromorphic sensor technology: the best option to employ from the different families of neuromorphic sensors strongly depends on the application specifications and constraints, without any privileged choice. The introduced paradigm shift has led to several ways to represent events and to model them in a machine-efficient way. As such, it is quite evident that there is no best representation, and very often the best choice depends on the specific application, as extensively analyzed in [11,186,187]. More research is necessary, to investigate new event stacking models and encoding schemes that can use existing solutions and architecture to process event data, achieving a better performance. The biggest bottleneck in this field is possibly the lack of a comprehensive theoretical framework to formally describe and analyze event-based sensing and algorithms [49]. As for event representation, this gap makes the translation of traditional machine learning algorithms difficult and, as a consequence, deep learning architectures must be adapted and modified to efficiently process events. Nevertheless, these sensors introduce a set of trade-offs, such as the optimal balance between latency and power consumption, and present several parameters that can suddenly change the acquired data. We think that these issues cannot be mitigated without an eye to the application domain, working conditions, and specific priorities, e.g., processing time vs. accuracy.

All the analyzed application domains certainly share the necessity of obtaining ad hoc data to train and test neural network architectures. It can be observed how researchers building event-based approaches often have to start from scratch and have to acquire their own data due to a lack of available datasets. Very few task-oriented datasets are available with full-frame sequences. Having common datasets is fundamental to comparing and evaluating methods, as well as for implementing machine learning-based solutions, as happens with traditional imaging. In many cases, we observed how the dataset was particularly associated with the specific goal of the paper, making it impossible to summarize the datasets introduced in the analyzed literature. It is possible that the availability of such sensors will favor the introduction of shared data that researchers will use to evaluate and test their solutions. At the same time, many event camera simulators have been proposed in the literature [188]. Simulators of event cameras based on data coming from frame-based cameras work by imitating changes in intensity with respect to time from standard imaging system data. However, upsampling strategies are usually employed to simulate the high temporal resolution of event cameras when using traditional sensors working at lower frame rates. Since continuous pixel visual signal data are not available, an interpolation step is necessary to reconstruct a linear approximation of the continuous underlying visual signal. Examples of simulators with available implementations can be found in [188–190]. To this end, creating synthetic datasets with event camera simulators represents a unique opportunity that has a two-fold advantage. First of all, it leads to converting frame-camera datasets, which are much more numerous than event datasets; in this way, it would be possible to obtain the relevant quantity of information necessary to use deep neural networks or to use backbone networks pre-trained with (many) simulated data and fine-tuned with (fewer) real data. On the other hand, simulators allow performing fruitful research, even in the absence of such sensor availability. This does not detract from the fact that it is crucial for the scientific community to propose and make available specific datasets based on event data, to establish a common framework for algorithm evaluation, as well as in the light of the limitations of simulators in terms of realism, event camera biases/noise modeling, and how well neural networks trained on synthetic events generalize to real data.

Specific observations can instead be stated by separately analyzing the application domains.

For animal and environmental monitoring, it is possible to conclude that, although cameras and inertial measurement units (IMUs) are fundamental for autonomous navigation in GPS-denied conditions, it is still necessary to explore new sensors to deal with problems like direct sunlight or darkness [191]. It is true that weight and cost are two obstacles to the adoption of event cameras in scenarios like UAVs; however, considering how recent advancements are moving such chips to sensor sizes comparable to smartphones, it appears that exponential expansion in this application field is only a matter of time.

The principal challenges related to the adoption of event cameras, as well as with other vision sensors for surveillance and security applications, are certainly related to privacy issues. In fact, interconnected and ubiquitous data acquisition systems not only become more powerful but also more vulnerable. If it is true that the increased adoption of AI presents an opportunity to address various challenges, it is also true that AI models are also vulnerable to advanced and sophisticated attacks, like evasion [192] or poisoning [193]. These vulnerabilities represent a challenge in the adoption of event cameras in security systems. The usage of sensors that by design preserve the person's identity could certainly help to address these privacy-related constraints. One notable example shows how event cameras can be adopted to achieve anonymized re-identification [194], because they can better guarantee the anonymity of subjects, without having the privacy concerns of frame-based cameras in public spaces. With traditional images, it is common to apply face blurring or masking, as well as encryption techniques, but this far from ensures end-to-end privacy. Moreover, research has shown that it is possible to reconstruct high-quality video frames from the event streams produced by these cameras [195,196], and this might constrain event-based privacy-preserving person re-identification. Apart from the aforementioned weaknesses, if it is crucial to address these issues (e.g., protection from reconstruction attacks [197]) before achieving robust privacy-preserving systems, neuromorphic computer vision certainly represents a step towards privacy-oriented surveillance.

For industrial applications and machine vision, event cameras are a unique opportunity to obtain unprecedented quality control, even in complex lighting environments. It is important to highlight how our analysis showed that the nature of the sensor alone, although promising for such tasks, is not sufficient, and more complex algorithms are required before having performance comparable to multi-sensor systems or industrial ad hoc setups. Moreover, further research should address the monitoring of areas where workers and machines cooperate, to achieve next-generation safety levels; in this case, similar considerations as those previously stated for privacy issues are valid: the advantage of event cameras in this application field is that they capture only a fraction of the visual information compared to frame-based vision, naturally hiding sensitive visual details.

For space exploration and SSA, preliminary works have shown very promising results. Undoubtedly, the key features for the adoption of neuromorphic computer vision for such applications include lower bandwidth and power requirements, thus being suitable for remote processing, e.g., space-based platforms. This often implies new challenges, such as the necessity to modify electro-optic instruments to support the event-based sensors and the pre-existing equipment simultaneously, and/or a specific calibration setup for the different reference systems. The results are often very exploratory, showing that a lot of research must still be performed; considering the unique characteristics of such cameras, we can expect a growth in such applications in the next few years.

Eye tracking, gaze estimation, and driver monitoring systems are receiving growing interest in the neuromorphic computer vision research community. In fact, the nature of these problems related to the detection and tracking of specific features can certainly benefit from event cameras. From the analysis of the state of the art, it emerged how, as this new potential application is only in its initial stages, there is plenty of room for improvement; moreover, it was observed how works are progressively switching from multi-sensor systems to being purely event-driven. The future of such systems will also depend on the final application scenario; for example, for eye segmentation or an unconstrained gaze tracker, the brightness information that DAVIS sensors deliver can lead to single-sensor

systems, while, when the objects of tracking are saccadic movements or the requirements demand keeping the frame rate high, multi-sensor systems seem to be privileged. However, both event and traditional camera solutions still have several gaps to fill. State-of-the-art analysis shows that frame-based and event cameras can bring together the best of both worlds and provide multiple modalities to deal with problems that appear to be better addressed in this way, rather than in the single data-source domain alone.

Similar considerations are also valid for gesture, action recognition, and human pose estimation. Moreover, it is interesting to observe how the core problem here is how to learn spatio-temporal contextual information from events. In fact, using a predefined accumulation interval may result in adverse effects like motion blur and a weak texture [143]. While research outcomes are showing promising results, the problem of converting event data into a proper representation remains an open issue, as widely observed.

An analysis of state-of-the-art works for medicine and healthcare showed the feasibility of using neuromorphic sensors for elderly activity monitoring, and many works on fall detection have been proposed. Furthermore, an active and growing research field is investigating the usage of event cameras coupled with other data sources to realize assistive devices, in particular for visual impairment. As already observed, when sensitive data are transmitted, it is clear that realizing privacy-preserving systems is crucial for the adoption of such sensors in real scenarios, as investigated in [198,199]. As a final note, the authors in [157] stated how there are no established studies that have delved into the utilization of event cameras for specific biomedical purposes. However, their work suggested how event cameras possess attributes conducive to innovative applications in healthcare and medicine. Together with our proposed analysis, we can conclude that, as for other application fields with few works, an impressive growth in works in these domains can be expected.

The most investigated application field for neuromorphic cameras is certainly robotics and, as a consequence, intelligent transportation systems, where the primary research focus is on tasks associated with scene interpretation. Endowing robots with neuromorphic technologies represents a very promising solution for the creation of robots that can be seamlessly integrated into many automated tasks. Neuromorphic cameras could revolutionize the robot sensing landscape. In particular, these sensors are privileged because of their fast reaction time, which leads to low-latency algorithms for perception and decision making. Nevertheless, the development of end-to-end systems that fully integrate event-based processing from the perception step to the actuation step still needs much research.

Future Directions

Event cameras hold significant potential in many scenarios, and their algorithms are rapidly evolving. Nevertheless, the conventional frame-based vision has made unprecedented achievements, been investigated for a longer time, and is continuously improving. The application-oriented analysis in previous paragraphs provided evidence of how the field of event-based vision represents a vibrant area of ongoing research. Where this was evident from the state-of-the-art analysis, future developments related to a specific application field were reported. It is desirable to state some generic observations about the future directions of neuromorphic cameras. The capacity to revolutionize both analyzed and new application fields will necessarily depend on several factors. First of all, the adoption of neuromorphic sensors will grow as the manufacturing costs of these cameras continue to decrease. Secondly, as researchers delve deeper into their potential, more innovative and successful uses will be identified. The future of event cameras also lies in addressing the challenges of deep-learning-based computer vision techniques, a consequence of the paradigm shift of dealing with streams of events. From these premises, the future of event cameras looks encouraging, with steady development and potential mass production on the horizon.

Moreover, to date, neuromorphic cameras have only implemented a small subset of the principles of the biological vision system. For example, a complete understanding of all

ganglion cells is still lacking [200]. New computational models and SNNs that can efficiently process the spiking nature of output data have been proposed and represent a very active research field. Implementing SNNs in conventional Von Neumann machines limits their computing efficiency, due to an asynchronous network activity that leads to quasi-random access of synaptic weights in time and space, as each spiking neuron should ideally be its own processor, without a central clock [41], and because of their highly parallel nature with amalgamated memory and computational units. Ongoing research in neuromorphic hardware is targeted towards the optimization of the execution of SNNs, to close the gap between the simulations of SNNs on Von Neumann machines and biological SNNs [201]. Eventually, neuromorphic cameras realized on silicon retina will face limitations due to circuit complexity, large pixel area, low fill factors, and poor pixel response uniformity [12]. This shows how breakthroughs in neuroscience and neuromorphic research, from both algorithm and hardware perspectives, remain crucial.

A report from Gartner© from 2020 estimated that *“by 2025, traditional computing technologies will hit a digital wall, forcing the shift to new computing paradigms such as neuromorphic computing”* [202]. In light of this, it can be cautiously stated that neuromorphic cameras hold revolutionary potential for widespread application in fields requiring computer vision. However, this prediction underscores the need for continued research from both industry and academia.

8. Conclusions

In this work, a detailed analysis of the literature was given, discussing how neuromorphic vision sensors have impacted computer vision in the last few years. We tried to fill gaps in other surveys with a critical analysis based on the different application domains, instead of the set low- or high-level computer vision tasks themselves. In addition, our comprehensive analysis of the existing literature revealed numerous surveys, which we subsequently discussed and categorized. The presented discussion illustrated different outcomes by application field. It also analyzed the problem of dataset availability: the lack of available datasets can be partially addressed with event camera simulators, but specific datasets are still often necessary to use such sensors in real-life applications.

An application-oriented approach was applied in the review process, in order to illustrate the wider impact of neuromorphic cameras, not limited to the academy. From the analysis of the state of the art, it emerged that event cameras have been applied in practically all application fields of computer vision. Certain fields, like robotics, have a very well-established literature, while new emerging fields like medicine or space exploration have only seen recent application of event cameras, although the advantages of lower bandwidth and power requirements are very appealing, and we can expect a growth in proposed works in the next few years. As a general conclusion, it is possible to state that neuromorphic cameras are a revolutionary technology with the potential to be applied in any scenario where computer vision and image processing algorithms can be utilized. This will be true as long as the production cost of event cameras continues to reduce. Their unique ability to capture dynamic visual information makes them invaluable tools for a wide range of applications, and as researchers continue to explore their capabilities, more innovative applications are likely to be discovered for these versatile sensors.

Author Contributions: Conceptualization, D.C. and F.B.; methodology, D.C. and F.B.; writing—original draft preparation, D.C.; writing—review and editing, D.C. and F.B.; supervision, F.B.; funding acquisition, F.B. All authors have read and agreed to the published version of the manuscript.

Funding: The work of this study was carried out under the European Commission, Joint Research Centre Exploratory Research project INVISIONS (Innovative Neuromorphic Vision Sensors).

Acknowledgments: The authors would like to acknowledge Eugenio Gutiérrez for his valuable contribution in shaping the INVISIONS project.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AER	Address-event representation
APS	Active pixel sensor
CNN	Convolutional neural network
ATIS	Asynchronous time-based image sensor
DAVIS	Dynamic and active pixel vision sensor
DGCNN	Dynamic graph CNN
DMS	Driving monitoring systems
DVS	Dynamic vision sensor
FAGC	Feature attention gate component
MOKE	Magneto-optic Kerr effect
n.a.	Not available
NIR	Near-infrared
SSA	Space situational awareness
SLAM	Simultaneous localization and mapping
SSNN	sparse neural network models
SNN	Spiking neural Network
VLSI	Very large-scale integration
UAV	Unmanned aerial vehicle

References

- Golnabi, H.; Asadpour, A. Design and application of industrial machine vision systems. *Robot. -Comput.-Integr. Manuf.* **2007**, *23*, 630–637. [\[CrossRef\]](#)
- Furmonas, J.; Liobe, J.; Barzdenas, V. Analytical review of event-based camera depth estimation methods and systems. *Sensors* **2022**, *22*, 1201. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fukushima, K.; Yamaguchi, Y.; Yasuda, M.; Nagata, S. An electronic model of the retina. *Proc. IEEE* **1970**, *58*, 1950–1951. [\[CrossRef\]](#)
- Mead, C.A.; Mahowald, M.A. A silicon model of early visual processing. *Neural Netw.* **1988**, *1*, 91–97. [\[CrossRef\]](#)
- Dong, Y.; Li, Y.; Zhao, D.; Shen, G.; Zeng, Y. Bullying10K: A Large-Scale Neuromorphic Dataset towards Privacy-Preserving Bullying Recognition. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 1923–1937.
- Prophesee Evaluation Kit 4. Available online: <https://www.prophesee.ai/event-camera-evk4/> (accessed on 31 May 2024).
- Inivation DAVIS346 Specifications. Available online: <https://inivation.com/wp-content/uploads/2019/08/DAVIS346.pdf> (accessed on 31 May 2024).
- Li, H.; Yu, H.; Wu, D.; Sun, X.; Pan, L. Recent advances in bioinspired vision sensor arrays based on advanced optoelectronic materials. *APL Mater.* **2023**, *11*, 081101. [\[CrossRef\]](#)
- Etienne-Cummings, R.; Van der Spiegel, J. Neuromorphic vision sensors. *Sens. Actuators Phys.* **1996**, *56*, 19–29. [\[CrossRef\]](#)
- Li, Z.; Sun, H. Artificial intelligence-based spatio-temporal vision sensors: Applications and prospects. *Front. Mater.* **2023**, *10*, 1269992. [\[CrossRef\]](#)
- Gallego, G.; Delbrück, T.; Orchard, G.; Bartolozzi, C.; Taba, B.; Censi, A.; Leutenegger, S.; Davison, A.J.; Conrath, J.; Daniilidis, K.; et al. Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 154–180. [\[CrossRef\]](#)
- Liao, F.; Zhou, F.; Chai, Y. Neuromorphic vision sensors: Principle, progress and perspectives. *J. Semicond.* **2021**, *42*, 013105. [\[CrossRef\]](#)
- Lichtsteiner, P.; Posch, C.; Delbruck, T. A 128 x 128 120 db 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: New York, NY, USA, 2006; pp. 2060–2069.
- 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* **2010**, *46*, 259–275. [\[CrossRef\]](#)
- Berner, R.; Brandli, C.; Yang, M.; Liu, S.C.; Delbruck, T. A 240 x 180 10 mw 12 us latency sparse-output vision sensor for mobile applications. In Proceedings of the 2013 Symposium on VLSI Circuits, Kyoto, Japan, 12–14 June 2013; IEEE: New York, NY, USA, 2013; pp. C186–C187.
- Scheerlinck, C.; Rebecq, H.; Stoffregen, T.; Barnes, N.; Mahony, R.; Scaramuzza, D. CED: Color event camera dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; IEEE: New York, NY, USA, 2019.
- 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. [\[CrossRef\]](#)
- Mongeon, P.; Paul-Hus, A. The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics* **2016**, *106*, 213–228. [\[CrossRef\]](#)

19. Indiveri, G.; Kramer, J.; Koch, C. Neuromorphic Vision Chips: Intelligent sensors for industrial applications. In *Proceedings of Advanced Microsystems for Automotive Applications*; Springer: Berlin/Heidelberg, Germany, 1996.
20. Kramer, J.; Indiveri, G. Neuromorphic vision sensors and preprocessors in system applications. In *Proceedings of the Advanced Focal Plane Arrays and Electronic Cameras II*, SPIE, Zurich, Switzerland, 7 September 1998; Volume 3410, pp. 134–146.
21. Indiveri, G. Neuromorphic VLSI models of selective attention: From single chip vision sensors to multi-chip systems. *Sensors* **2008**, *8*, 5352–5375. [[CrossRef](#)]
22. Liu, S.C.; Delbruck, T. Neuromorphic sensory systems. *Curr. Opin. Neurobiol.* **2010**, *20*, 288–295. [[CrossRef](#)]
23. Wu, N. Neuromorphic vision chips. *Sci. China Inf. Sci.* **2018**, *61*, 1–17. [[CrossRef](#)]
24. Kim, M.S.; Kim, M.S.; Lee, G.J.; Sunwoo, S.H.; Chang, S.; Song, Y.M.; Kim, D.H. Bio-inspired artificial vision and neuromorphic image processing devices. *Adv. Mater. Technol.* **2022**, *7*, 2100144. [[CrossRef](#)]
25. Steffen, L.; Reichard, D.; Weinland, J.; Kaiser, J.; Roennau, A.; Dillmann, R. Neuromorphic stereo vision: A survey of bio-inspired sensors and algorithms. *Front. Neurobot.* **2019**, *13*, 28. [[CrossRef](#)]
26. 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](#)]
27. Sandamirskaya, Y.; Kabloi, M.; Conradt, J.; Celikel, T. Neuromorphic computing hardware and neural architectures for robotics. *Sci. Robot.* **2022**, *7*, eabl8419. [[CrossRef](#)]
28. Aboumerhi, K.; Güemes, A.; Liu, H.; Tenore, F.; Etienne-Cummings, R. Neuromorphic applications in medicine. *J. Neural Eng.* **2023**, *20*, 041004. [[CrossRef](#)] [[PubMed](#)]
29. Sun, R.; Shi, D.; Zhang, Y.; Li, R.; Li, R. Data-driven technology in event-based vision. *Complexity* **2021**, *2021*, 1–19. [[CrossRef](#)]
30. Bartolozzi, C.; Indiveri, G.; Donati, E. Embodied neuromorphic intelligence. *Nat. Commun.* **2022**, *13*, 1024. [[CrossRef](#)] [[PubMed](#)]
31. Jia, S. Event Camera Survey and Extension Application to Semantic Segmentation. In *Proceedings of the 4th International Conference on Image Processing and Machine Vision*, Hong Kong, China, 25–27 March 2022; pp. 115–121.
32. Hodgkin, A.L.; Huxley, A.F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **1952**, *117*, 500. [[CrossRef](#)]
33. Izhikevich, E.M. Simple model of spiking neurons. *IEEE Trans. Neural Netw.* **2003**, *14*, 1569–1572. [[CrossRef](#)] [[PubMed](#)]
34. Gerstner, W. *Spiking Neurons*; Technical Report; MIT-Press: Cambridge, MA, USA, 1998.
35. Bouvier, M.; Valentian, A.; Mesquida, T.; Rummens, F.; Reyboz, M.; Vianello, E.; Beigne, E. Spiking neural networks hardware implementations and challenges: A survey. *Acm J. Emerg. Technol. Comput. Syst. (JETC)* **2019**, *15*, 1–35. [[CrossRef](#)]
36. Nunes, J.D.; Carvalho, M.; Carneiro, D.; Cardoso, J.S. Spiking neural networks: A survey. *IEEE Access* **2022**, *10*, 60738–60764. [[CrossRef](#)]
37. Yi, Z.; Lian, J.; Liu, Q.; Zhu, H.; Liang, D.; Liu, J. Learning rules in spiking neural networks: A survey. *Neurocomputing* **2023**, *531*, 163–179. [[CrossRef](#)]
38. Tavanaei, A.; Ghodrati, M.; Kheradpisheh, S.R.; Masquelier, T.; Maida, A. Deep learning in spiking neural networks. *Neural Netw.* **2019**, *111*, 47–63. [[CrossRef](#)]
39. Yamazaki, K.; Vo-Ho, V.K.; Bulsara, D.; Le, N. Spiking neural networks and their applications: A Review. *Brain Sci.* **2022**, *12*, 863. [[CrossRef](#)]
40. Wang, S.; Cheng, T.H.; Lim, M.H. A hierarchical taxonomic survey of spiking neural networks. *Memetic Comput.* **2022**, *14*, 335–354. [[CrossRef](#)]
41. Pfeiffer, M.; Pfeil, T. Deep learning with spiking neurons: Opportunities and challenges. *Front. Neurosci.* **2018**, *12*, 409662. [[CrossRef](#)] [[PubMed](#)]
42. Paredes-Vallés, F.; Scheper, K.Y.; De Croon, G.C. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2051–2064. [[CrossRef](#)] [[PubMed](#)]
43. Bing, Z.; Meschede, C.; Röhrbein, F.; Huang, K.; Knoll, A.C. A survey of robotics control based on learning-inspired spiking neural networks. *Front. Neurobot.* **2018**, *12*, 35. [[CrossRef](#)] [[PubMed](#)]
44. Basu, A.; Deng, L.; Frenkel, C.; Zhang, X. Spiking neural network integrated circuits: A review of trends and future directions. In *Proceedings of the 2022 IEEE Custom Integrated Circuits Conference (CICC)*, Newport Beach, CA, USA, 24–27 April 2022; IEEE: New York, NY, USA, 2022; pp. 1–8.
45. 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.
46. Zou, X.L.; Huang, T.J.; Wu, S. Towards a new paradigm for brain-inspired computer vision. *Mach. Intell. Res.* **2022**, *19*, 412–424. [[CrossRef](#)]
47. Dong-il, C.; Tae-jae, L. A review of bioinspired vision sensors and their applications. *Sens. Mater* **2015**, *27*, 447–463.
48. Lakshmi, A.; Chakraborty, A.; Thakur, C.S. Neuromorphic vision: From sensors to event-based algorithms. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1310. [[CrossRef](#)]
49. Tayarani-Najaran, M.H.; Schmuker, M. Event-based sensing and signal processing in the visual, auditory, and olfactory domain: A review. *Front. Neural Circuits* **2021**, *15*, 610446. [[CrossRef](#)]

50. Zhu, S.; Wang, C.; Liu, H.; Zhang, P.; Lam, E.Y. Computational neuromorphic imaging: Principles and applications. In Proceedings of the Computational Optical Imaging and Artificial Intelligence in Biomedical Sciences, SPIE, San Francisco, CA, USA, 27 January–1 February 2024; Volume 12857, pp. 4–10.
51. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*; Pearson Education: Upper Saddle River, NJ, USA, 2009.
52. Cavanagh, P. Visual cognition. *Vis. Res.* **2011**, *51*, 1538–1551. [[CrossRef](#)]
53. Cantoni, V.; Ferretti, M. A Taxonomy of Hierarchical Machines for Computer Vision. *Pyramidal Archit. Comput. Vis.* **1994**, *1*, 103–115.
54. Zeiler, M.D.; Taylor, G.W.; Fergus, R. Adaptive deconvolutional networks for mid and high level feature learning. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Washington, DC, USA, 2011; pp. 2018–2025.
55. Ji, Q. *Probabilistic Graphical Models for Computer Vision.*; Academic Press: Cambridge, MA, USA, 2019.
56. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A review on UAV-based applications for precision agriculture. *Information* **2019**, *10*, 349. [[CrossRef](#)]
57. Cazzato, D.; Cimarelli, C.; Sanchez-Lopez, J.L.; Voos, H.; Leo, M. A survey of computer vision methods for 2d object detection from unmanned aerial vehicles. *J. Imaging* **2020**, *6*, 78. [[CrossRef](#)] [[PubMed](#)]
58. Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens.* **2020**, *12*, 2659. [[CrossRef](#)]
59. El Arja, S. Neuromorphic Perception for Greenhouse Technology Using Event-based Sensors. Ph.D. Thesis, Sydney University, Camperdown, NSW, Australia, 2022.
60. Zujevs, A.; Pudzs, M.; Osadcuks, V.; Ardavs, A.; Galauskis, M.; Grundspenkis, J. An event-based vision dataset for visual navigation tasks in agricultural environments. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May 2021–5 June 2021; IEEE: New York, NY, USA, 2021; pp. 13769–13775.
61. Zhu, L.; Mangan, M.; Webb, B. Neuromorphic sequence learning with an event camera on routes through vegetation. *Sci. Robot.* **2023**, *8*, eadg3679. [[CrossRef](#)] [[PubMed](#)]
62. Hamann, F.; Gallego, G. Stereo Co-capture System for Recording and Tracking Fish with Frame-and Event Cameras. *arXiv* **2022**, arXiv:2207.07332.
63. Hamann, F.; Ghosh, S.; Martinez, I.J.; Hart, T.; Kacelnik, A.; Gallego, G. Low-power, Continuous Remote Behavioral Localization with Event Cameras. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle WA, USA, 17 June–21 June 2024; pp. 18612–18621.
64. Dataset. Agri-EVB-Autumn. Available online: <https://iee-dataport.org/open-access/agri-ebv-autumn> (accessed on 4 July 2024).
65. Dataset. Neuromorphic Sequence Learning with an Event Camera on Routes through Vegetation. Available online: <https://zenodo.org/records/8289547> (accessed on 4 July 2024).
66. Dataset. Low-Power, Continuous Remote Behavioral Localization with Event Cameras. Available online: <https://tub-rip.github.io/eventpenguins/> (accessed on 4 July 2024).
67. Litzenberger, M.; Posch, C.; Bauer, D.; Belbachir, A.N.; 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, Teton National Park, WY, USA, 24–27 September 2006; IEEE: New York, NY, USA, 2006; pp. 173–178.
68. Litzenberger, M.; Belbachir, A.N.; Schon, P.; Posch, C. Embedded smart camera for high speed vision. In Proceedings of the 2007 First ACM/IEEE International Conference on Distributed Smart Cameras, Vienna, Austria, 25–28 September 2007; IEEE: New York, NY, USA, 2007; pp. 81–86.
69. Piątkowska, E.; Belbachir, A.N.; Schraml, S.; Gelautz, M. Spatiotemporal multiple persons tracking using dynamic vision sensor. In Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 16–21 June 2012; IEEE: New York, NY, USA, 2012, pp. 35–40.
70. Stuckey, H.; Al-Radaideh, A.; Escamilla, L.; Sun, L.; Carrillo, L.G.; Tang, W. An optical spatial localization system for tracking unmanned aerial vehicles using a single dynamic vision sensor. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; IEEE: New York, NY, USA, 2021; pp. 3093–3100.
71. Annamalai, L.; Chakraborty, A.; Thakur, C.S. Evan: Neuromorphic event-based anomaly detection. *arXiv* **2019**, arXiv:1911.09722.
72. Pérez-Cutiño, M.A.; Eguíluz, A.G.; Martínez-de Dios, J.; 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; IEEE: New York, NY, USA, 2021; pp. 91–100.
73. Rodríguez-Gomez, J.P.; Eguíluz, A.G.; Martínez-de Dios, J.R.; 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; IEEE: New York, NY, USA, 2020; pp. 8518–8524.
74. Gañán, F.J.; Sanchez-Diaz, J.A.; Tapia, R.; Martínez-de Dios, J.; 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: New York, NY, USA, 2022; pp. 211–216.

75. Ahmad, S.; Scarpellini, G.; Morerio, P.; Del Bue, A. Event-driven re-id: A new benchmark and method towards privacy-preserving person re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2022; pp. 459–468.
76. Dataset. Event Camera Dataset For Intruder Monitoring. Available online: <https://grvc.us.es/davis-dataset-for-intrusion-monitoring/> (accessed on 4 July 2024).
77. Bialkowski, A.; Denman, S.; Sridharan, S.; Fookes, C.; Lucey, P. A database for person re-identification in multi-camera surveillance networks. In Proceedings of the 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA), Fremantle, WA, Australia, 3–5 December 2012; IEEE: New York, NY, USA, 2012; pp. 1–8.
78. Ristani, E.; Solera, F.; Zou, R.; Cucchiara, R.; Tomasi, C. Performance measures and a data set for multi-target, multi-camera tracking. In Proceedings of the European conference on computer vision, Amsterdam, The Netherlands, 8–10 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 17–35.
79. Perez-Peña, F.; Morgado-Estevez, A.; Montero-Gonzalez, R.J.; Linares-Barranco, A.; Jimenez-Moreno, G. Video surveillance at an industrial environment using an address event vision sensor: Comparative between two different video sensor based on a bioinspired retina. In Proceedings of the International Conference on Signal Processing and Multimedia Applications, Seville, Spain, 18–21 July 2011; IEEE: New York, NY, USA, 2011, pp. 1–4.
80. Ni, Z.; Pacoret, C.; Benosman, R.; Ieng, S.; Réginer, S. Asynchronous event-based high speed vision for microparticle tracking. *J. Microsc.* **2012**, *245*, 236–244. [[CrossRef](#)]
81. Drazen, D.; Lichtsteiner, P.; Häfliger, P.; Delbrück, T.; Jensen, A. Toward real-time particle tracking using an event-based dynamic vision sensor. *Exp. Fluids* **2011**, *51*, 1465–1469. [[CrossRef](#)]
82. Zhang, K.; Zhao, Y.; Chu, Z.; Zhou, Y. Event-based vision in magneto-optic Kerr effect microscopy. *AIP Adv.* **2022**, *12*. [[CrossRef](#)]
83. Bialik, K.; Kowalczyk, M.; Blachut, K.; Kryjak, T. Fast-moving object counting with an event camera. *arXiv* **2022**, arXiv:2212.08384.
84. Li, X.; Yu, S.; Lei, Y.; Li, N.; Yang, B. Intelligent machinery fault diagnosis with event-based camera. *IEEE Trans. Ind. Inform.* **2023**, *20*, 380–389. [[CrossRef](#)]
85. Zhao, G.; Shen, Y.; Chen, N.; Hu, P.; Liu, L.; Wen, H. EV-Tach: A Handheld Rotational Speed Estimation System With Event Camera. *IEEE Trans. Mob. Comput.* **2023**, *12*, 380–389. [[CrossRef](#)]
86. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, 224–227. [[CrossRef](#)]
87. Micev, K.; Steiner, J.; Aydin, A.; Rieckermann, J.; Delbruck, T. Measuring diameters and velocities of artificial raindrops with a neuromorphic event camera. *Atmos. Meas. Tech.* **2024**, *17*, 335–357. [[CrossRef](#)]
88. Shiba, S.; Hamann, F.; Aoki, Y.; Gallego, G. Event-based background-oriented schlieren. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *46*, 2011–2026. [[CrossRef](#)]
89. Liu, X.; Yang, Z.X.; Xu, Z.; Yan, X. NeuroVI-based new datasets and space attention network for the recognition and falling detection of delivery packages. *Front. Neurobot.* **2022**, *16*, 934260. [[CrossRef](#)]
90. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
91. Dataset. Event-based Background-Oriented Schlieren. Available online: https://github.com/tub-rip/event_based_bos (accessed on 4 July 2024).
92. 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](#)]
93. Afshar, S.; Nicholson, A.P.; Van Schaik, A.; Cohen, G. Event-based object detection and tracking for space situational awareness. *IEEE Sens. J.* **2020**, *20*, 15117–15132. [[CrossRef](#)]
94. Chin, T.J.; Bagchi, S.; Eriksson, A.; Van Schaik, A. Star tracking using an event camera. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
95. Roffe, S.; Akolkar, H.; George, A.D.; Linares-Barranco, B.; Benosman, R.B. Neutron-induced, single-event effects on neuromorphic event-based vision sensor: A first step and tools to space applications. *IEEE Access* **2021**, *9*, 85748–85763. [[CrossRef](#)]
96. Ralph, N.O.; Marcireau, A.; Afshar, S.; Tohill, N.; Van Schaik, A.; Cohen, G. Astrometric calibration and source characterisation of the latest generation neuromorphic event-based cameras for space imaging. *Astrodynamics* **2023**, *7*, 415–443. [[CrossRef](#)]
97. Ralph, N.; Joubert, D.; Jolley, A.; van Schaik, A.; Cohen, G. Real-time event-based unsupervised feature consolidation and tracking for space situational awareness. *Front. Neurosci.* **2022**, *16*, 821157. [[CrossRef](#)]
98. Dataset. Event-Based Star Tracking Dataset. Available online: <https://www.ai4space.group/research/event-based-star-tracking> (accessed on 4 July 2024).
99. Dataset. The Event-Based Space Situational Awareness (EBSSA) Dataset. Available online: https://www.westernsydney.edu.au/icns/resources/reproducible_research3/publication_support_materials2/space_imaging (accessed on 4 July 2024).
100. Dataset. IEBCS. Available online: <https://github.com/neuromorphicsystems/IEBCS> (accessed on 4 July 2024).
101. Dataset. Event Based—Space Imaging—Speed Dataset. Available online: https://github.com/NicRalph213/ICNS_NORALPH_Event_Based-Space_Imaging-Speed_Dataset (accessed on 4 July 2024).
102. Ji, Q.; Yang, X. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging* **2002**, *8*, 357–377. [[CrossRef](#)]

103. Cazzato, D.; Leo, M.; Distanto, C.; Voos, H. When i look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking. *Sensors* **2020**, *20*, 3739. [CrossRef]
104. Feng, Y.; Goulding-Hotta, N.; Khan, A.; Reyserhove, H.; Zhu, Y. Real-time gaze tracking with event-driven eye segmentation. In Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Christchurch, New Zealand, 12–16 March 2022; IEEE: New York, NY, USA, 2022; pp. 399–408.
105. Ryan, C.; O’Sullivan, B.; Elrasad, A.; Cahill, A.; Lemley, J.; Kielty, P.; Posch, C.; Perot, E. Real-time face & eye tracking and blink detection using event cameras. *Neural Netw.* **2021**, *141*, 87–97.
106. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
107. Kang, D.; Kang, D. Event Camera-Based Pupil Localization: Facilitating Training With Event-Style Translation of RGB Faces. *IEEE Access* **2023**, *11*, 142304–142316. [CrossRef]
108. 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]
109. Banerjee, A.; Mehta, N.K.; Prasad, S.S.; Saurav, S.; Singh, S.; Himanshu. Gaze-Vector Estimation in the Dark with Temporally Encoded Event-driven Neural Networks. *arXiv* **2024**, arXiv:2403.02909.
110. Stoffregen, T.; Daraei, H.; Robinson, C.; Fix, A. Event-based kilohertz eye tracking using coded differential lighting. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 2515–2523.
111. Li, N.; Chang, M.; Raychowdhury, A. E-Gaze: Gaze Estimation with Event Camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**. [CrossRef]
112. Li, N.; Bhat, A.; Raychowdhury, A. E-track: Eye tracking with event camera for extended reality (xr) applications. In Proceedings of the 2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems (AICAS), Hangzhou, China, 11–13 June 2023; IEEE: New York, NY, USA; 2023, pp. 1–5.
113. Ryan, C.; Elrasad, A.; Shariff, W.; Lemley, J.; Kielty, P.; Hurney, P.; Corcoran, P. Real-time multi-task facial analytics with event cameras. *IEEE Access* **2023**, *11*, 76964–76976. [CrossRef]
114. Kielty, P.; Dilmaghani, M.S.; Shariff, W.; Ryan, C.; Lemley, J.; Corcoran, P. Neuromorphic driver monitoring systems: A proof-of-concept for yawn detection and seatbelt state detection using an event camera. *IEEE Access* **2023**, *11*, 96363–96373. [CrossRef]
115. Liu, P.; Chen, G.; Li, Z.; Clarke, D.; Liu, Z.; Zhang, R.; Knoll, A. Neurodfd: Towards efficient driver face detection with neuromorphic vision sensor. In Proceedings of the 2022 International Conference on Advanced Robotics and Mechatronics (ICARM), Guilin, China, 9–11 July 2022; IEEE: New York, NY, USA, 2022; pp. 268–273.
116. Chen, G.; Wang, F.; Li, W.; Hong, L.; Conradt, J.; Chen, J.; Zhang, Z.; Lu, Y.; Knoll, A. NeuroIV: Neuromorphic vision meets intelligent vehicle towards safe driving with a new database and baseline evaluations. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 1171–1183. [CrossRef]
117. Shariff, W.; Dilmaghani, M.S.; Kielty, P.; Lemley, J.; Farooq, M.A.; Khan, F.; Corcoran, P. Neuromorphic driver monitoring systems: A computationally efficient proof-of-concept for driver distraction detection. *IEEE Open J. Veh. Technol.* **2023**, *4*, 836–848. [CrossRef]
118. Dataset. NeuroIV. Available online: <https://github.com/ispc-lab/NeuroIV> (accessed on 4 July 2024).
119. Dataset. Event Based, Near Eye Gaze Tracking Beyond 10,000 Hz. Available online: https://github.com/aangelopoulos/event_based_gaze_tracking (accessed on 4 July 2024).
120. Garbin, S.J.; Shen, Y.; Schuetz, I.; Cavin, R.; Hughes, G.; Talathi, S.S. Openeds: Open eye dataset. *arXiv* **2019**, arXiv:1905.03702.
121. Fuhl, W.; Kasneci, G.; Kasneci, E. Teyed: Over 20 million real-world eye images with pupil, eyelid, and iris 2d and 3d segmentations, 2d and 3d landmarks, 3d eyeball, gaze vector, and eye movement types. In Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Bari, Italy, 4–8 October 2021; IEEE: New York, NY, USA, 2021; pp. 367–375.
122. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. Wider face: A face detection benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5525–5533.
123. Fanelli, G.; Dantone, M.; Gall, J.; Fossati, A.; Van Gool, L. Random forests for real time 3d face analysis. *Int. J. Comput. Vis.* **2013**, *101*, 437–458. [CrossRef]
124. Abtahi, S.; Omidyeganeh, M.; Shirmohammadi, S.; Hariri, B. YawDD: A yawning detection dataset. In Proceedings of the 5th ACM multimedia systems conference, Singapore, 19 March 2014; pp. 24–28.
125. Chen, S.; Akselrod, P.; Zhao, B.; Carrasco, J.A.P.; Linares-Barranco, B.; Culurciello, E. Efficient feedforward categorization of objects and human postures with address-event image sensors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 302–314. [CrossRef]
126. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 1695–1704.
127. Xu, L.; Xu, W.; Golyanik, V.; Habermann, M.; Fang, L.; Theobalt, C. Eventcap: Monocular 3d capture of high-speed human motions using an event camera. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4968–4978.

128. 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](#)]
129. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; proceedings, part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
130. Goyal, G.; Di Pietro, F.; Carissimi, N.; Glover, A.; Bartolozzi, C. MoveEnet: Online High-Frequency Human Pose Estimation with an Event Camera. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 4023–4032.
131. Ahn, E.Y.; Lee, J.H.; Mullen, T.; Yen, J. Dynamic vision sensor camera based bare hand gesture recognition. In Proceedings of the 2011 IEEE Symposium on Computational Intelligence For Multimedia, Signal And Vision Processing, Paris, France, 11–15 April 2011; IEEE: New York, NY, USA, 2011; pp. 52–59.
132. Lee, J.H.; Park, P.K.; Shin, C.W.; Ryu, H.; Kang, B.C.; Delbruck, T. Touchless hand gesture UI with instantaneous responses. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; IEEE: New York, NY, USA, 2012; pp. 1957–1960.
133. 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 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7243–7252.
134. Wang, Q.; Zhang, Y.; Yuan, J.; Lu, Y. Space-time event clouds for gesture recognition: From RGB cameras to event cameras. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; IEEE: New York, NY, USA, 2019; pp. 1826–1835.
135. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
136. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
137. Chen, J.; Meng, J.; Wang, X.; Yuan, J. Dynamic graph CNN for event-camera based gesture recognition. In Proceedings of the 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Seville, Spain, 12–14 October 2020; IEEE: New York, NY, USA, 2020; pp. 1–5.
138. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.* **2019**, *38*, 1–12. [[CrossRef](#)]
139. Chen, G.; Xu, Z.; Li, Z.; Tang, H.; Qu, S.; Ren, K.; Knoll, A. A novel illumination-robust hand gesture recognition system with event-based neuromorphic vision sensor. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 508–520. [[CrossRef](#)]
140. Vasudevan, A.; Negri, P.; Di Ielsi, C.; Linares-Barranco, B.; Serrano-Gotarredona, T. SL-Animals-DVS: Event-driven sign language animals dataset. *Pattern Anal. Appl.* **2022**, *25*, 505–520. [[CrossRef](#)]
141. Chen, X.; Su, L.; Zhao, J.; Qiu, K.; Jiang, N.; Zhai, G. Sign language gesture recognition and classification based on event camera with spiking neural networks. *Electronics* **2023**, *12*, 786. [[CrossRef](#)]
142. Liu, C.; Qi, X.; Lam, E.Y.; Wong, N. Fast classification and action recognition with event-based imaging. *IEEE Access* **2022**, *10*, 55638–55649. [[CrossRef](#)]
143. Xie, B.; Deng, Y.; Shao, Z.; Liu, H.; Xu, Q.; Li, Y. Event Tubelet Compressor: Generating Compact Representations for Event-Based Action Recognition. In Proceedings of the 2022 7th International Conference on Control, Robotics and Cybernetics (CRC), Virtual, 15–17 December 2022; IEEE: New York, NY, USA, 2022; pp. 12–16.
144. Neimark, D.; Bar, O.; Zohar, M.; Asselmann, D. Video transformer network. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 3163–3172.
145. de Blegiers, T.; Dave, I.R.; Yousaf, A.; Shah, M. EventTransAct: A video transformer-based framework for Event-camera based action recognition. In Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023; IEEE: New York, NY, USA, 2023; pp. 1–7.
146. Dataset. DVS128 Gesture. Available online: <https://ibm.ent.box.com/s/3hiq58ww1pbbjrinh367ykfdf60xsfm8/folder/50167556794> (accessed on 4 July 2024).
147. Dataset. DHP19. Available online: <https://sites.google.com/view/dhp19/home> (accessed on 4 July 2024).
148. Orchard, G.; Jayawant, A.; Cohen, G.K.; Thakor, N. Converting static image datasets to spiking neuromorphic datasets using saccades. *Front. Neurosci.* **2015**, *9*, 159859. [[CrossRef](#)] [[PubMed](#)]
149. Miao, S.; Chen, G.; Ning, X.; Zi, Y.; Ren, K.; Bing, Z.; Knoll, A. Neuromorphic vision datasets for pedestrian detection, action recognition, and fall detection. *Front. Neurobot.* **2019**, *13*, 38. [[CrossRef](#)]
150. Dataset. SL-Animals-DVS. Available online: <http://www2.imse-cnm.csic.es/neuromorphs/index.php/SL-ANIMALS-DVS-Database> (accessed on 4 July 2024).
151. Bi, Y.; Chadha, A.; Abbas, A.; Bourtsoulatzé, E.; Andreopoulos, Y. Graph-based spatio-temporal feature learning for neuromorphic vision sensing. *IEEE Trans. Image Process.* **2020**, *29*, 9084–9098. [[CrossRef](#)]
152. Ionescu, C.; Papava, D.; Olaru, V.; Sminchisescu, C. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 1325–1339. [[CrossRef](#)]

153. Dataset. DVS-Sign. Available online: <https://github.com/najie1314/DVS> (accessed on 4 July 2024).
154. Plizzari, C.; Planamente, M.; Goletto, G.; Cannici, M.; Gusso, E.; Matteucci, M.; Caputo, B. E2 (go) motion: Motion augmented event stream for egocentric action recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 19935–19947.
155. Fu, Z.; Delbruck, T.; Lichtsteiner, P.; Culurciello, E. An address-event fall detector for assisted living applications. *IEEE Trans. Biomed. Circuits Syst.* **2008**, *2*, 88–96. [[CrossRef](#)]
156. Chen, G.; Qu, S.; Li, Z.; Zhu, H.; Dong, J.; Liu, M.; Conradt, J. Neuromorphic vision-based fall localization in event streams with temporal–spatial attention weighted network. *IEEE Trans. Cybern.* **2022**, *52*, 9251–9262. [[CrossRef](#)] [[PubMed](#)]
157. Jagtap, A.; Saripalli, R.V.; Lemley, J.; Shariff, W.; Smeaton, A.F. Heart Rate Detection Using an Event Camera. In Proceedings of the 2023 IEEE International Symposium on Multimedia (ISM), Laguna Hills, CA, USA, 11–13 December 2023; IEEE: New York, NY, USA, 2023; pp. 243–246.
158. Everding, L.; Walger, L.; Ghaderi, V.S.; Conradt, J. A mobility device for the blind with improved vertical resolution using dynamic vision sensors. In Proceedings of the 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Munich, Germany, 14–16 September 2016; IEEE: New York, NY, USA, 2016; pp. 1–5.
159. Gaspar, N.; Sondhi, A.; Evans, B.; Nikolic, K. A low-power neuromorphic system for retinal implants and sensory substitution. In Proceedings of the 2016 IEEE Biomedical Circuits and Systems Conference (BioCAS), Shanghai, China, 17–19 October 2016; IEEE: New York, NY, USA, 2016; pp. 78–81.
160. Berthelon, X. Neuromorphic Analysis of Hemodynamics Using Event-Based Cameras. Ph.D. Thesis, Sorbonne Université, Paris, France, 2018.
161. Cabriel, C.; Monfort, T.; Specht, C.G.; Izeddin, I. Event-based vision sensor for fast and dense single-molecule localization microscopy. *Nat. Photonics* **2023**, *17*, 1105–1113. [[CrossRef](#)]
162. Dataset. Evb-SMLM. Available online: <https://github.com/Clement-Cabriel/Evb-SMLM> (accessed on 4 July 2024).
163. Chen, G.; Cao, H.; Aafaque, M.; Chen, J.; Ye, C.; Röhrbein, F.; Conradt, J.; Chen, K.; Bing, Z.; Liu, X.; et al. Neuromorphic vision based multivehicle detection and tracking for intelligent transportation system. *J. Adv. Transp.* **2018**, *2018*, 1–13. [[CrossRef](#)]
164. Ikura, M.; Walter, F.; Knoll, A. Spiking Neural Networks for Robust and Efficient Object Detection in Intelligent Transportation Systems With Roadside Event-Based Cameras. In Proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 4–7 June 2023; IEEE: New York, NY, USA, 2023; pp. 1–6.
165. Lu, X.; Mao, X.; Liu, H.; Meng, X.; Rai, L. Event camera point cloud feature analysis and shadow removal for road traffic sensing. *IEEE Sens. J.* **2021**, *22*, 3358–3369. [[CrossRef](#)]
166. Cheng, W.; Luo, H.; Yang, W.; Yu, L.; Chen, S.; Li, W. Det: A high-resolution dvs dataset for lane extraction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 1666–1675.
167. Cao, H.; Chen, G.; Xia, J.; Zhuang, G.; Knoll, A. Fusion-based feature attention gate component for vehicle detection based on event camera. *IEEE Sens. J.* **2021**, *21*, 24540–24548. [[CrossRef](#)]
168. 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: New York, NY, USA, 2022; pp. 86–91.
169. Chen, G.; Chen, W.; Yang, Q.; Xu, Z.; Yang, L.; Conradt, J.; Knoll, A. A novel visible light positioning system with event-based neuromorphic vision sensor. *IEEE Sens. J.* **2020**, *20*, 10211–10219. [[CrossRef](#)]
170. Dataset. DET. Available online: <https://spritea.github.io/DET/> (accessed on 4 July 2024).
171. Binas, J.; Neil, D.; Liu, S.C.; Delbruck, T. DDD17: End-to-end DAVIS driving dataset. *arXiv* **2017**, arXiv:1711.01458.
172. Perot, E.; De Tournemire, P.; Nitti, D.; Masci, J.; Sironi, A. Learning to detect objects with a 1 megapixel event camera. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 16639–16652.
173. Falanga, D.; Kleber, K.; Scaramuzza, D. Dynamic obstacle avoidance for quadrotors with event cameras. *Sci. Robot.* **2020**, *5*, eaaz9712. [[CrossRef](#)]
174. Wang, Y.; Yang, J.; Peng, X.; Wu, P.; Gao, L.; Huang, K.; Chen, J.; Kneip, L. Visual odometry with an event camera using continuous ray warping and volumetric contrast maximization. *Sensors* **2022**, *22*, 5687. [[CrossRef](#)]
175. Iaboni, C.; Patel, H.; Lobo, D.; Choi, J.W.; Abichandani, P. Event camera based real-time detection and tracking of indoor ground robots. *IEEE Access* **2021**, *9*, 166588–166602. [[CrossRef](#)]
176. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the KDD, Portland, OR, USA, 2–4 August 1996; Volume 96, pp. 226–231.
177. Huang, X.; Halwani, M.; Muthusamy, R.; Ayyad, A.; Swart, D.; Seneviratne, L.; Gan, D.; Zweiri, Y. Real-time grasping strategies using event camera. *J. Intell. Manuf.* **2022**, *33*, 593–615. [[CrossRef](#)]
178. Panetsos, F.; Karras, G.C.; Kyriakopoulos, K.J. Aerial Transportation of Cable-Suspended Loads With an Event Camera. *IEEE Robot. Autom. Lett.* **2023**, *9*, 231–238. [[CrossRef](#)]
179. Wang, Z.; Ng, Y.; Henderson, J.; Mahony, R. Smart visual beacons with asynchronous optical communications using event cameras. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; IEEE: New York, NY, USA, 2022; pp. 3793–3799.

180. Nakagawa, H.; Miyatani, Y.; Kanezaki, A. Linking Vision and Multi-Agent Communication through Visible Light Communication using Event Cameras. In Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, Auckland, New Zealand, 6–10 May 2024; pp. 1436–1444.
181. Hu, Y.; Binas, J.; Neil, D.; Liu, S.C.; Delbruck, T. Ddd20 end-to-end event camera driving dataset: Fusing frames and events with deep learning for improved steering prediction. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020; IEEE: New York, NY, USA, 2020; pp. 1–6.
182. Brandli, C.; Mantel, T.; Hutter, M.; Höpflinger, M.; Berner, R.; Delbruck, T. Adaptive pulsed laser line extraction for terrain reconstruction using a dynamic vision sensor. *Front. Neurosci.* **2014**, *7*, 65397. [[CrossRef](#)] [[PubMed](#)]
183. Dataset. DDD20. Available online: <https://sites.google.com/view/davis-driving-dataset-2020/home> (accessed on 4 July 2024).
184. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets robotics: The kitti dataset. *Int. J. Robot. Res.* **2013**, *32*, 1231–1237. [[CrossRef](#)]
185. Mur-Artal, R.; Tardós, J.D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [[CrossRef](#)]
186. Amini, A.; Wang, T.H.; Gilitschenski, I.; Schwarting, W.; Liu, Z.; Han, S.; Karaman, S.; Rus, D. Vista 2.0: An open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; IEEE: New York, NY, USA, 2022; pp. 2419–2426.
187. Lin, S.; Ma, Y.; Guo, Z.; Wen, B. DVS-Voltmeter: Stochastic process-based event simulator for dynamic vision sensors. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 578–593.
188. Gehrig, D.; Gehrig, M.; Hidalgo-Carrió, J.; Scaramuzza, D. Video to events: Recycling video datasets for event cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 13–19 June 2020; pp. 3586–3595.
189. Hu, Y.; Liu, S.C.; Delbruck, T. v2e: From video frames to realistic DVS events. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, TN, USA, 19–25 June 2021; pp. 1312–1321.
190. Rebecq, H.; Gehrig, D.; Scaramuzza, D. ESIM: An open event camera simulator. In Proceedings of the Conference on Robot Learning, PMLR, Zurich, Switzerland, 29–31 October 2018; pp. 969–982.
191. Liu, X.; Chen, S.W.; Nardari, G.V.; Qu, C.; Cladera, F.; Taylor, C.J.; Kumar, V. Challenges and opportunities for autonomous micro-uavs in precision agriculture. *IEEE Micro* **2022**, *42*, 61–68. [[CrossRef](#)]
192. Qiu, S.; Liu, Q.; Zhou, S.; Wu, C. Review of artificial intelligence adversarial attack and defense technologies. *Appl. Sci.* **2019**, *9*, 909. [[CrossRef](#)]
193. Zhang, H.; Gao, J.; Su, L. Data poisoning attacks against outcome interpretations of predictive models. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Virtual, 14–18 August 2021, pp. 2165–2173.
194. Ahmad, S.; Morerio, P.; Del Bue, A. Person re-identification without identification via event anonymization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Vancouver, BC, Canada, 18–22 June 2023; pp. 11132–11141.
195. Bardow, P.; Davison, A.J.; Leutenegger, S. Simultaneous optical flow and intensity estimation from an event camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 884–892.
196. Munda, G.; Reinbacher, C.; Pock, T. Real-time intensity-image reconstruction for event cameras using manifold regularisation. *Int. J. Comput. Vis.* **2018**, *126*, 1381–1393. [[CrossRef](#)]
197. Zhang, X.; Wang, Y.; Yang, Q.; Shen, Y.; Wen, H. EV-Perturb: Event-stream perturbation for privacy-preserving classification with dynamic vision sensors. *Multimed. Tools Appl.* **2024**, *83*, 16823–16847. [[CrossRef](#)]
198. Prasad, S.S.; Mehta, N.K.; Banerjee, A.; Kumar, H.; Saurav, S.; Singh, S. Real-Time Privacy-Preserving Fall Detection using Dynamic Vision Sensors. In Proceedings of the 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 24–26 November 2022; IEEE: New York, NY, USA, 2022; pp. 1–6.
199. Prasad, S.S.; Mehta, N.K.; Kumar, H.; Banerjee, A.; Saurav, S.; Singh, S. Hybrid SNN-based Privacy-Preserving Fall Detection using Neuromorphic Sensors. In Proceedings of the Fourteenth Indian Conference on Computer Vision, Graphics and Image Processing, Rupnagar, India, 15–17 December 2023; pp. 1–9.
200. Wang, H.; Sun, B.; Ge, S.S.; Su, J.; Jin, M.L. On non-von Neumann flexible neuromorphic vision sensors. *npj Flex. Electron.* **2024**, *8*, 28. [[CrossRef](#)]
201. Vanarse, A.; Osseiran, A.; Rassau, A. Neuromorphic engineering—A paradigm shift for future im technologies. *IEEE Instrum. Meas. Mag.* **2019**, *22*, 4–9. [[CrossRef](#)]
202. Gartner©. Gartner Top 10 Strategic Predictions for 2021 and beyond. Available online: <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-predictions-for-2021-and-beyond> (accessed on 6 July 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.