Artificial intelligence-based spatio-temporal vision sensors: applications and prospects

Zhixue Li and Hongwei Sun*

College of Computer Science and Engineering, Weifang University of Science and Technology, Weifang, China

Introduction: Advances in machine vision and mobile electronics will be accelerated by the creation of sophisticated optoelectronic vision sensors that allow for sophisticated picture recognition of visual information and data pre-processing. Several new types of vision sensors have been devised in the last decade to solve these drawbacks, one of which is neuromorphic vision sensors, which have exciting qualities such as high temporal resolution, broad dynamic range, and low energy consumption. Neuromorphic sensors are inspired by the working principles of biological sensory neurons and would be useful in telemedicine, health surveillance, security monitoring, automatic driving, intelligent robots, and other applications of the Internet of Things.

Methods: This paper provides a comprehensive review of various state-of-the-art AI vision sensors and frameworks.

Results: The fundamental signal processing techniques deployed and the associated challenges were discussed.

Discussion: Finally, the role of vision sensors in computer vision is also discussed.

KEYWORDS

AI, vision sensors, intelligent systems, image processing, neuromorphic vision sensor

1 Introduction

Vision is the fundamental function of cognitive creatures and agents, responsible for comprehending and seeing their surroundings. The visual system accounts for more than 80% of information in the human perception system, greatly surpassing the total of the auditory, tactile, and other perceptual systems (Picano, 2021). How to create a sophisticated visual perception system for use in computer vision technology and artificial intelligence has long been a research focus in academia and industry (Liu et al., 2019).

A video is a series of still pictures that was created by the advancement of cinema and television technologies. It makes use of the human visual system's visual persistence phenomena (Hafed et al., 2021). Traditional video has made significant advances in the visual viewing angle in recent years (Zare et al., 2019), but there are drawbacks such as high data sampling redundancy, a small photosensitive dynamic range, low resolution in time domain acquisition, and the ability to produce motion blur in high-speed motion scenes (Sharif et al., 2023). Furthermore, computer vision has been moving in the mainstream direction of "video camera + computer + algorithm = machine vision" (Hossain et al., 2019; Wu and Ji, 2019), but few scholars question the logic of using image sequences (videos) to express visual information, and even fewer scholars question whether this computer vision method can be realized.

The human visual system has the advantages of low redundancy, low power consumption, high dynamics and strong robustness. It can efficiently and adaptively process dynamic and static information, and has strong small sample generalization ability and comprehensive complex scene perception ability (Heuer et al., 2020). Explore the mysteries of the human visual system, and learn from the neural network structure of the human visual system and the processing mechanism of visual information sampling and processing (Pramod and Arun, 2022), to establish a new set of visual information perception and processing theories, technical standards, chips and application engineering systems. The ability to better simulate, extend or surpass the human visual perception system (Ham et al., 2021). This is the intersection of neuroscience and information science, called neuromorphic vision (Najaran and Schmuker, 2021; Martini et al., 2022; Chunduri and Perera, 2023). Neuromorphic vision is a visual perception system that includes hardware development, soft support, and biological neural models. One of its ultimate goals is to simulate the structure and mechanism of biological visual perception (Ferrara et al., 2022) in order to achieve real machine vision.

The development of neuromorphic vision sensors is based on scientific and physiologic research on the structure and functional mechanism of biological retinas. A neuron model with computing power was proposed by Alexiadis (Alexiadis, 2019). To describe the neural network, Shatnawi et al. (Shatnawi et al., 2023) established dynamic differential equations for neurons.

The generation and transmission of the action potential is called a spike. A Ph.D. student at Caltech, thought: "The brain is the birthplace of imagination, which makes me very excited. I hope to create a chip for imagining things". The topic of stereo vision is investigated from the standpoints of science and engineering. The author initially introduced the Neuromorphic idea (Han et al., 2022), employing large-scale integrated circuits to replicate the organic nervous system. The writers featured a moving cat on the cover of "Scientific American" (Ferrara et al., 2022), launching the first silicon retina, which models the biology of cone cells, horizontal cells, and bipolar cells on the retina. It was the formal start of the burgeoning area of neuromorphic vision sensors. Marinis et al. (Marinis et al., 2021) presented Address-Event Representation (AER), a novel form of integrated circuit communication protocol, to overcome the challenge of dense three-dimensional integration of integrated circuits, which enabled asynchronous event readout. Purohit et al. (Purohit and Manohar, 2022) created Octopus Retina, an AER-based integral-discharge pulse model that represents pixel light intensity as frequency or pulse interval. Abubakar et al. (Abubakar et al., 2023) created a Dynamic Vision Sensor (DVS) that represented pixel light intensity changes with sparse spatiotemporal asynchronous events, and its commercialization was a watershed moment. DVS, on the other hand, is unable to capture fine texture images of natural scenes. Oliveria et al. (Oliveria et al., 2021) proposed an asynchronous time-based image sensor (ATIS) and an event-triggered light intensity measurement circuit to reconstruct the pixel grey level as the change occurred. Zhang et al. (Zhang et al., 2022a) created a Dynamic and Active Pixel Vision Sensor (DAVIS), a dual-mode technical route that includes an additional independent traditional image sampling circuit to compensate for DVS texture imaging flaws. It was then expanded to colour DAVIS346 (Moeys et al., 2017). To restore the scene texture, Feng et al. (Feng et al., 2020) increased the bit width of the event and let the event carry the pixel light intensity information output. Auge et al. (Auge et al., 2021) adopted the octopus retina's principle of light intensity integral distribution sampling and replaced pulse plane transmission with the AER method to conserve transmission bandwidth. They also confirmed that the integral sampling principle can quickly reconstruct scene texture details. That is, the Fovea-like Sampling Model (FSM), also called Vidar, as shown in Figure 1. Neuromorphic vision sensors simulate biological visual perception systems, which have the advantages of high temporal resolution, less data redundancy, low power consumption and high dynamic range, and are used in autonomous driving (Chen L. et al., 2022), unmanned vehicles machine vision fields such as machine vision navigation (Mueggler et al., 2017; Mitrokhin et al., 2019), industrial inspection (Zhu Z. et al., 2022), and video surveillance (Zhang S. et al., 2022) have huge market potential, especially in scenarios involving high-speed motion and extreme lighting. The sampling, processing, and application of neuromorphic vision is another important field of neuromorphic engineering (Wang T. et al., 2022). This activity validates the computational neuroscience's visual brain model (Baek and Seo, 2022) and is a useful method for examining human intellect. The development of neuromorphic vision sensors is still in its infancy, and more investigation and study are still required to match or even surpass the human visual system's capacity for perception in intricately interconnected settings (Chen J. et al., 2023).

This paper systematically reviews and summarizes the development process of neuromorphic vision, the neural sampling model of biological vision, the sampling mechanism and types of neuromorphic vision sensors, neurovisual signal processing and feature expression, and vision applications, and looks forward to the future of this field. The major challenges and possible development directions of the research are discussed, and its potential impact on the future field of machine vision and artificial intelligence is discussed.

2 Neuromorphic vision model and sampling mechanism

The technical route of neuromorphic vision is generally divided into three levels: the structural level imitates the retina, the device functional level approaches the retina, and the intelligence level surpasses the retina. If the traditional camera is a simulation of the human visual system, then this bionic retina is only a primary simulation of the functional level of the device. In fact, the traditional camera is far inferior to the perception ability of the human retina in various complex environments in terms of structure, function and even intelligence (Chen et al., 2022b).

In recent years, the "Brain Projects" (Fang and Hu, 2021) in various countries have been successively deployed and launched, and the analysis of brain-like vision from the structural level is one of
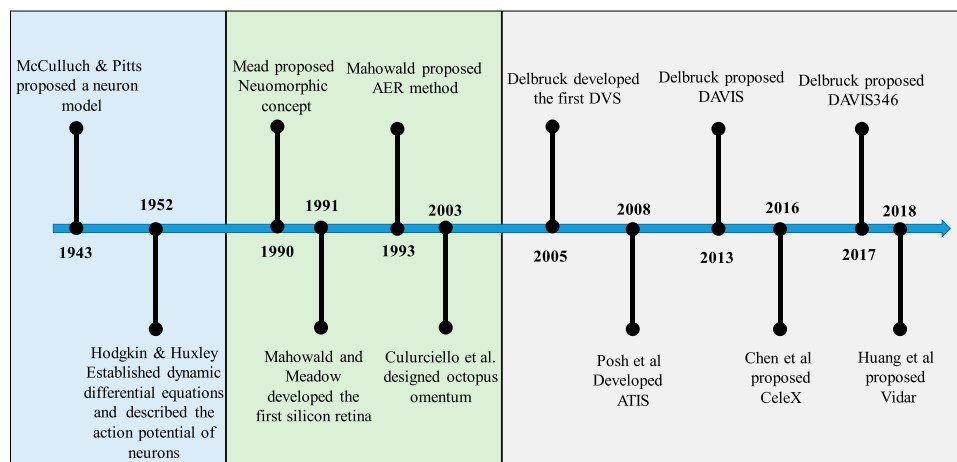
FIGURE 1
The development process of neuromorphic visual sensors.

the important contents to support, mainly through the use of fine analysis and advanced detection technology by neuroscientists to obtain the structure of the basic unit of the retina, functions and their network connections, which provide theoretical support for the device function level approaching the biological visual perception system. The neuromorphic vision sensor starts from the device function level simulation, that is, the optoelectronic nanodevice is used to simulate the biological vision sampling model and information processing function, and the perception system with or beyond the biological vision ability is constructed under the limited physical space and power consumption conditions. In short, neuromorphic vision sensors do not need to wait until they fully understand the analytic structure and mechanism of the retina before simulating. Instead, they learn from the structure-level research mechanism and bypass this more difficult problem, and use simulation engineering techniques such as device function-level approximation. The ability to reach, extend or surpass the human visual perception system.

At present, neuromorphic vision sensors have achieved staged results (Zhou et al., 2019). There are differential visual sampling models that simulate the peripheral sensory motor function of the retina (Jeremie and Perrinet, 2023). There are also integral visual sampling models that simulate the fine texture function of the fovea, such as octopus retina (Purohit and Manohar, 2022), and Vidar (Auge et al., 2021).

Including photoreceptor cells, bipolar cells, horizontal cells, ganglion cells and other main components (Sawant et al., 2023), as shown in Figure 2. Photoreceptor cells are divided into two types: rod cells and cone cells, which are responsible for converting light signals entering the eye into electrical signals, which are transmitted to bipolar cells and horizontal cells. Cone cells are sensitive to color and are mainly responsible for color recognition, and usually work under the condition of strong scene illumination. Rod cells are sensitive to light and can perceive weak light, mainly providing work in night scenes (Soucy et al., 2023), but they have no color discrimination ability. Bipolar cells receive signal input from photoreceptors, and they are divided into ON-type and OFF-type cells according to the different regions of the receptive field (Sun

et al., 2022), which sense the increase in light intensity and the decrease in light intensity, respectively. Horizontal cells are laterally interconnected with photoreceptors and bipolar cells, which adjust the brightness of the signals output by photoreceptors, and are also responsible for enhancing the outline of visual objects. Ganglion cells are responsible for receiving visual signal input from bipolar cells, and respond in the form of spatial-temporal pulse signals, which are then transmitted to the visual cortex through visual fibers. In addition, retinal cells have multiple parallel pathways to transmit and process visual signals, which have great advantages in bandwidth transmission and speed. Among them, the Magnocellular and Parvocellular pathways are the two most important signal pathways (Sawant et al., 2023), which are respectively sensitive to the temporal changes of the scene. and spatial structure sensitive.

The primate biological retina has the following advantages:

(1) Local adaptive gain control of photoreceptors: The change of recorded light intensity replaces the absolute light intensity to eliminate redundancy, and has a high dynamic range (High Dynamic Range, HDR) for light intensity perception;

(2) Spatial bandpass filter of rod cells: filter out the visual information redundancy of low-frequency information and the noise of high-frequency information;

(3) ON and OFF types: both ganglion cells and retinal outputs are encoded by ON and OFF pulse signals, which reduces the pulse firing frequency of a single channel;

(4) Photoreceptor functional area: The fovea has high spatial resolution and can capture fine textures; its peripheral area has high temporal resolution and captures fast motion information.

Additionally, biological vision is represented and encoded by binary pulse information, and the optic nerve only needs to transmit 20 Mb/s data to the vision Cortex, the amount of data is nearly 1,000 times less than what traditional cameras need to transmit in order to match the dynamic range and spatial resolution of human vision. Therefore, the retina efficiently represents and encodes visual information by converting light intensity data into spatiotemporal
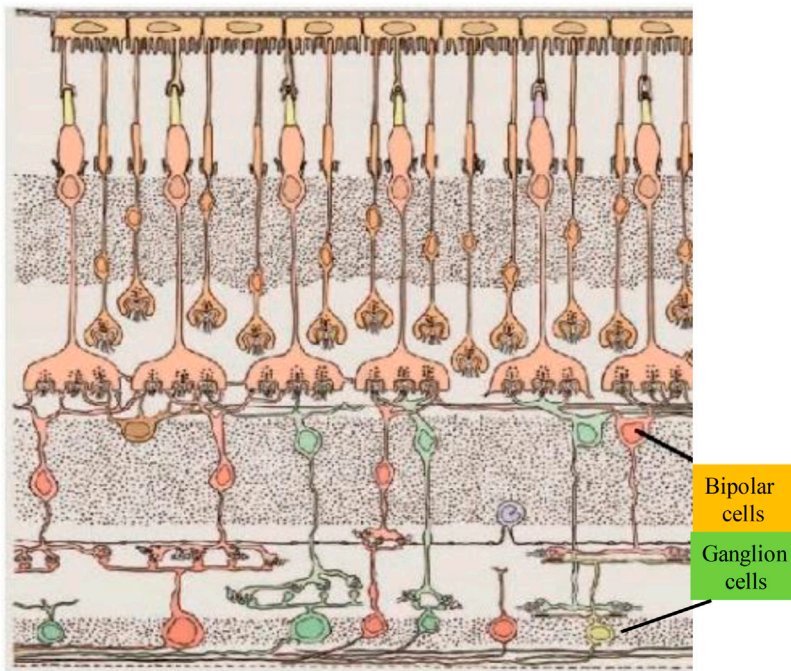
FIGURE 2
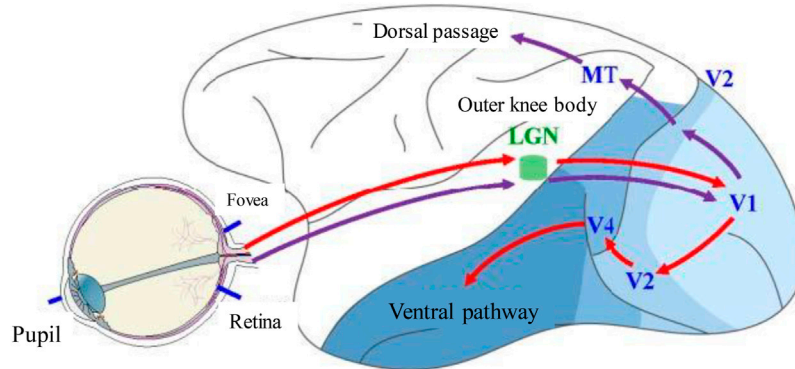Schematic diagram of the cross-section of the primate retina (Sawant et al., 2023).

FIGURE 3
Visual pathway model.

pulse array signals through ganglion cells, which serves as both a theoretical foundation and a source of functional inspiration for neuromorphic visual sensors.

2.1 Vision model of the retina

The information acquisition, processing and processing of the biological visual system mainly occur in the retina, lateral geniculate body and visual cortex (Pramod and Arun, 2022), as shown in Figure 3. The retina is the first station to receive visual information. The lateral geniculate body is the information transfer station that transmits retinal visual signals to the primary visual cortex; the visual cortex is the visual central processor, which is used in advanced visual functions such as learning and memory, thinking language, and perceptual awareness (Jeremie and Perrinet, 2023). The entire process of visual cortex information processing is completed by two parallel pathways: V1, V2, and V4. the ventral pathway mainly deals with the recognition of object shape, color, and other information (Toprak et al., 2018), also known as the what pathway; V1, V2, and MT. The dorsal pathway mainly deals with spatial position, motion and other information (Freud et al., 2016), also known as the where pathway. Therefore, the neural computing model is used to explore the information processing and analysis mechanism of the human visual system, which can provide

FIGURE 4
Schematic diagram of AER model.
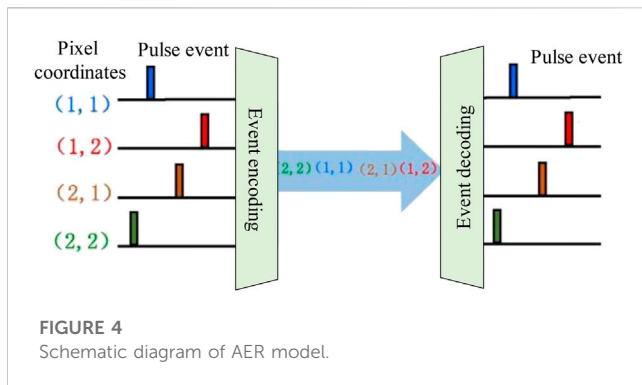
reference ideas and directions for computer vision and artificial intelligence technology, and further inspire the theoretical model and computing method of brain-like vision, so as to better mine visual feature information. It can process dynamic and static information efficiently and adaptively, close to biological vision, and has strong generalization ability of small samples and comprehensive visual analysis ability.

2.2 Differential vision sampling and AER transmission protocol

The complex connections between neurons and the transmission of impulse signals between neurons are asynchronous, so how can neuromorphic engineering systems simulate this feature? It is the Marinis team (Marinis et al., 2021) who proposed a new communication protocol AER method, as shown in Figure 4, which is used for multiple asynchronous transmission of pulse signals, and also solves the three-dimensional dense connection problem of large-scale integrated circuits, that is, the "connection problem".

Like a regular camera, the AER technique interprets each pixel on the sensor as independent rather than sending out a picture at a set frequency. The pulse signal is sent out asynchronously in accordance with the event's temporal sequence and is broadcast as an event. The decoding circuit then parses event attributes according to address and time. The following are the key components of the AER strategy for neuromorphic vision sensors (Buchel et al., 2021):

(1) The output events of silicon retinal pixels simulate the function of neurons in the retina to emit pulse signals;
(2) Light intensity perception, pulse generation and transmission are asynchronous between silicon retina pixels;
(3) When the asynchronous events output by the silicon retina are sparse, the event representation and transmission are more efficient.

As shown in Figure 5, the abstraction of the three-layer structure of cells and ganglion cells approaches or exceeds the ability of high temporal resolution perception of the retina periphery from the functional level of the device. Differential visual sampling is the mainstream of neuromorphic visual sensor perception models (Dai et al., 2023). The DVS series of vision sensors primarily employ the logarithmic differential model, which is to say that the photocurrent and voltage use a logarithmic mapping relationship in order to

increase the dynamic range of light intensity perception (b). The pixel produces a pulse signal when the voltage change exceeds the predetermined threshold due to the relative change in light intensity, as shown in Figure 5. (c). The basic idea is this:

$$\Delta L \doteq \ln L[u,t] - \ln L[u,t-\Delta t] \qquad (1)$$

The AER technique is used by the differential vision sensor, and each pulse signal is represented as an event. Include a quadruple representation that includes the pixel position $u(x,y)$, release time t, and event polarity (x,y,t,p). The last factor, polarity $p = [-1,1]$ represents the weakening of light intensity OFF and the enhancing of light intensity ON, respectively. The first three independently establish the position of the space-time domain. Differential neuromorphic vision sensors provide the following benefits over conventional cameras:

(1) The output of asynchronous sparse pulses, which is no longer constrained by shutter time and frame rate and can perceive changes in light intensity, lacks the concept of "frame" and can therefore eliminate static and invariable visual redundancy;
(2) High-speed motion vision task analysis is suited for sampling due to its high temporal resolution;
(3) The capacity to perceive high and low light levels is improved and the dynamic range is increased because to the logarithmic mapping connection between photocurrent and voltage.

2.3 Integral vision sampling

Integral visual sampling functionally abstracts the three-layer structure of photoreceptors, bipolar cells and ganglion cells in the foveal region of primate retina, such as octopus retina (Purohit and Manohar, 2022), and Vidar (Auge et al., 2021). Integral vision sensors simulate the neuron integral firing model, encode pixel light intensity as frequency or pulse interval, and have the ability to reconstruct the fine texture of visual scenes at high speed (Zhang et al., 2022c), as shown in Figure 6. The photoreceptor converts the light signal into an electrical signal.

The accumulation is carried out under the condition of integrator I(t) to reach the accumulated intensity A(t), when the intensity value exceeds the pulse release threshold φ, the pixel point outputs a pulse signal, and the integrator resets to clear the charge (Gao et al., 2018). The principle is as follows:

$$A(t) = \int_0^t I(t)dt \geq \varphi \qquad (2)$$

The pixels of the integrating vision sensor are independent of each other. The octopus retina (Purohit and Manohar, 2022) uses the AER method to output the pulse signal. Especially when the light intensity is sufficient, the integrating vision sensor emits dense pulses, and the event representation is prone to appear multiple times at the same position and adjacent positions. When requesting pulse output, there will be a huge pressure of data transmission, so the bus arbitration mechanism has to be designed to determine the priority for pulse output, and even the pulse signal will be lost due to bandwidth limitation. Vidar (Auge et al., 2021) explored a high-speed polling method to transmit the pulse release at each sampling moment in the form of a pulse matrix. This method does not require the coordinates and timestamps of the output pulse, and only needs to mark whether the pixel is released as "1" and
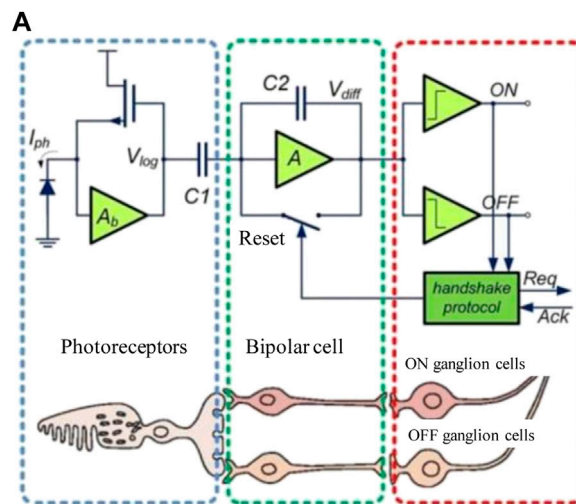
The outer three-layer structure of the primate retina and the DVS circuit.
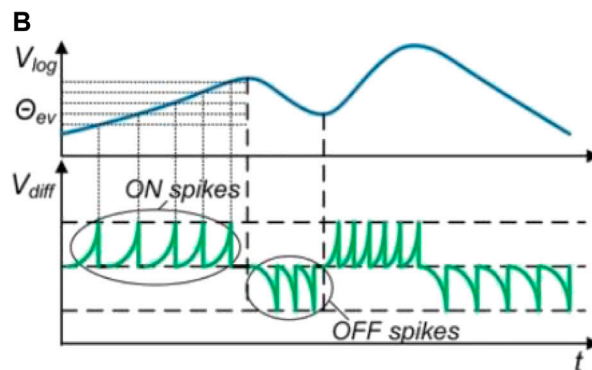
Differential sampling of light intensity

DVS pulse signal accumulation plane

FIGURE 5
Differential visual sampling (Najaran and Schmuker, 2021). **(A)** The outer three-layer structure of the primate retina and the DVS circuit. **(B)** Differential sampling of light intensity. **(C)** DVS pulse signal accumulation plane.

"0". Replacing the AER method with the pulse plane polling method can save the transmission bandwidth.

3 Types of vision sensors in neuromorphic process

Neuromorphic vision sensor draws on the neural network structure of biological visual system and the processing mechanism of visual information sampling, and simulates,

extends or surpasses biological visual perception system at the device function level. In recent years, a large number of representative neuromorphic vision sensors have emerged, which are the prototype of human exploration of bionic vision technology. There are differential visual sampling models that simulate the peripheral sensory motor function of the retina, such as DVS (Abubakar et al., 2023), ATIS (Oliveria et al., 2021), DAVIS (Mesa et al., 2019; Sadaf et al., 2023), CeleX (Feng et al., 2020). There are also integral visual sampling models that simulate the fine texture function of the fovea, such as Vidar (Auge et al., 2021).

3.1 DVS

DVS (Abubakar et al., 2023) abstracts the function of the three-layer structure of photoreceptors, bipolar cells and ganglion cells in the periphery of primate retina, which is composed of photoelectric conversion circuit, dynamic detection circuit and comparator output circuit, as shown in Figure 5. The photoelectric conversion circuit adopts the logarithmic light intensity perception model, which improves the light intensity perception range and is closer to the high dynamic adaptation ability of the biological retina. The dynamic detection circuit adopts a differential sampling model, that is, it responds to changes in light intensity, and does not respond if there is no change in light intensity. The comparator outputs ON or OFF events according to the increase or decrease of light intensity.

Traditional cameras have a fixed frame rate sampling technique, which in fast-moving situations is prone to motion blur. As illustrated in Figure 7, DVS employs an asynchronous space-time pulse signal to describe the change in scene light intensity and adopts the differential visual sampling model of the AER asynchronous transmission method. High time resolution (106 Hz), high dynamic range (120 dB), low power consumption, less data redundancy, and low latency are just a few benefits of DVS over conventional cameras.

The first commercial DVS128 (Abubakar et al., 2023) developed by the Delbruck team and IniVation has a spatial resolution of 128 × 128, the sampling frequency in the time domain is 106 Hz, the dynamic range is 120 dB, and it is widely used in high-speed moving object recognition, detection and tracking. In addition, the research and products of neuromorphic vision sensors such as DVS and its derivatives ATIS (Oliveria et al., 2021), DAVIS (Mesa et al., 2019; Sadaf et al., 2023) and CeleX (Feng et al., 2020) have also attracted much attention, and are gradually applied to automatic driving and UAV visual navigation and industrial inspection involving high-speed motion vision tasks. For example, Samsung has developed a spatial resolution of 640 × 480 of DVS-G2 (Xu et al., 2020), and the pixel size is 9 μ m × 9 μ m. The IBM company uses DVS128 as the visual perception system of the brain-like chip TrueNorth for fast gesture recognition (Tchantchane et al., 2023).

The DVS uses the differential visual sampling model to filter the static or weakly changing visual information to reduce data redundancy, and at the same time, it has the ability to perceive high-speed motion. However, this advantage brings the disadvantage of visual reconstruction, that is, the ON or OFF event does not carry the absolute light intensity signal, and no pulse signal is emitted when the light intensity change is weak, so that the refined texture image cannot be reconstructed. In order to solve the visual texture visualization of DVS, neuromorphic visual sensors such as ATIS (Oliveria et al., 2021), DAVIS (Mesa et al., 2019; Sadaf et al., 2023) and CeleX (Feng et al., 2020) are derived.

3.2 ATIS

On the basis of DVS, ATIS (Oliveria et al., 2021) skillfully introduced a light intensity measurement circuit based on time interval to realize image reconstruction. The idea is that every time an event occurs in the DVS circuit, the light intensity measurement circuit is triggered to work. Two different reference voltages are set, by integrating the light intensity, and recording the events that reach the two voltages. Because the time required for the voltage to change by the same amount is different under the conditions of different light intensities, by establishing the light intensity and time mapping can infer that the light intensity is small, so as to output the light intensity information at the pixel where the light intensity changes, which is also called Pulse Width Modulation (PWM) (Holesovsky et al., 2021). In addition, in order to solve the problem that the visual texture information of the static area cannot be obtained without DVS pulse signal issuance, ATIS introduces a set of global emission mechanism, that is, all pixels can be forced to emit a pulse, so that a whole image is used as the background, and then the moving area continuously generates pulses and then continuously triggers the light intensity measurement circuit to obtain the grayscale of the moving area to update the background (Fu et al., 2023).

The commercial ATIS (Holesovsky et al., 2021) developed by the Posch team and Prophesee company has a spatial resolution of 304 × 240, a sampling frequency of 106 Hz in the time domain, and a dynamic range of 143 dB, which is widely used in high-speed vision tasks. In addition, Prophessee has also received a $15 million project from Intel Corporation to apply ATIS to the vision processing system of autonomous vehicles. Subsequently, the team of Benosman (Macireau et al., 2018) further verified the technical scheme of using ATIS sampled pulse signals in the three RGB channels to re-integrate the color (Zhou and Zhang, 2022).

When ATIS is facing high-speed motion, there is still a mismatch between events and grayscale reconstruction updates (Han et al., 2022; Hou et al., 2023). The reasons are as follows: the light intensity measurement circuit is triggered after the pulse is issued, and the measurement result is the average light intensity of a period of time after the pulse is issued, resulting in motion mismatch. Slight changes in the scene do not cause pulses, so the pixels are not updated in time, resulting in obvious texture differences over time.

3.3 DVAIS

DAVIS (Mesa et al., 2019; Sadaf et al., 2023) is the most intuitive and effective fusion technology idea, which combines DVS and traditional cameras, and additionally introduces Active Pixel Sensor (APS) on the basis of DVS for visual scene texture (Zhuo et al., 2022).

The Delbruck team and IniVation further introduced a color scheme based on the DAVIS240 (Zhang K. et al., 2022) with a spatial resolution of 240 × 180.

Color DAVIS346 (Moeys et al., 2017), its spatial resolution reaches 346 260, the time domain sampling frequency is 106Hz, the dynamic range is 120dB, and the spatial position of the event coordinates generated by DVS carries RGB color information, but the sampling speed of APS circuit is far less than DVS (Zhao B. et al., 2022). The frame rate of APS mode is 50FPS, and the dynamic range is 56.7 dB. Particularly in high-speed motion scenes, the images produced by the two sets of sampling circuits cannot be precisely synchronized, and APS images exhibit motion blur.
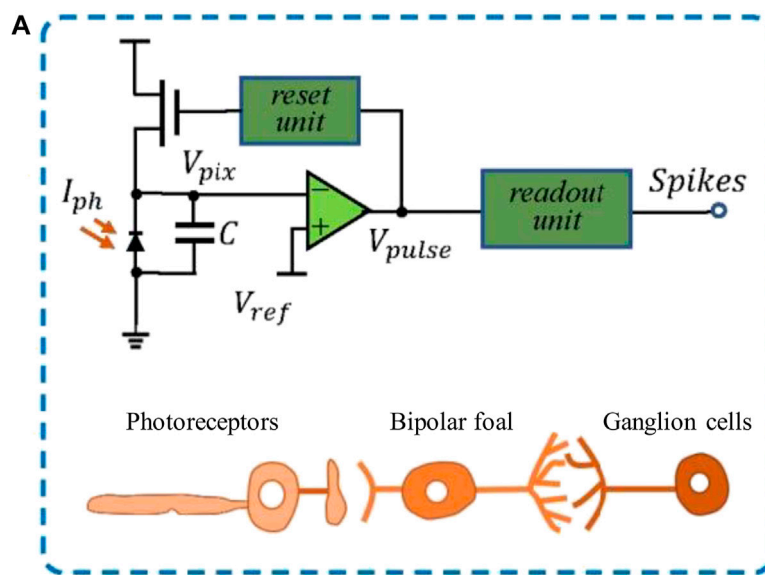
A

Primate retinal foveal three-layer structure and Vidar circuit.

B

Light intensity integral sampling.

C
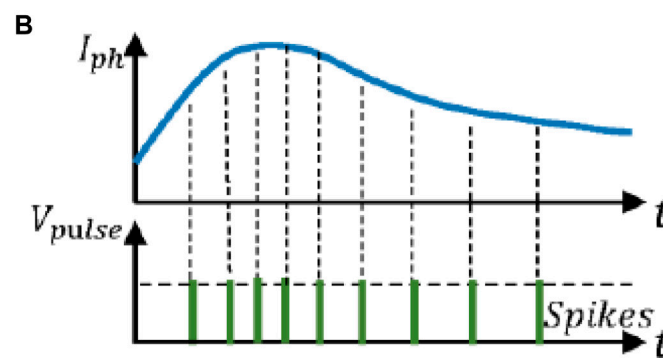
Vidar pulse signal texture imaging.

FIGURE 6
Integral visual sampling. **(A)** Primate retinal foveal three-layer structure and Vidar circuit. **(B)** Light intensity integral sampling. **(C)** Vidar pulse signal texture imaging.

At present, DAVIS is the mainstream of neuromorphic vision sensor commercial products, industrial applications and academic research. It originates from the academic research promotion of DVS series sensors (DVS128, DAVIS240, DAVIS346 and color DAVIS346), and the disclosure, code and a good ecological environment created by open-source software. Therefore, in this paper, the pulse signal processing, feature expression, and vision applications are mainly based on the DVS series sensors of the differential vision sampling model.

FIGURE 7
Schematic diagram of DVS spatiotemporal pulse signal.
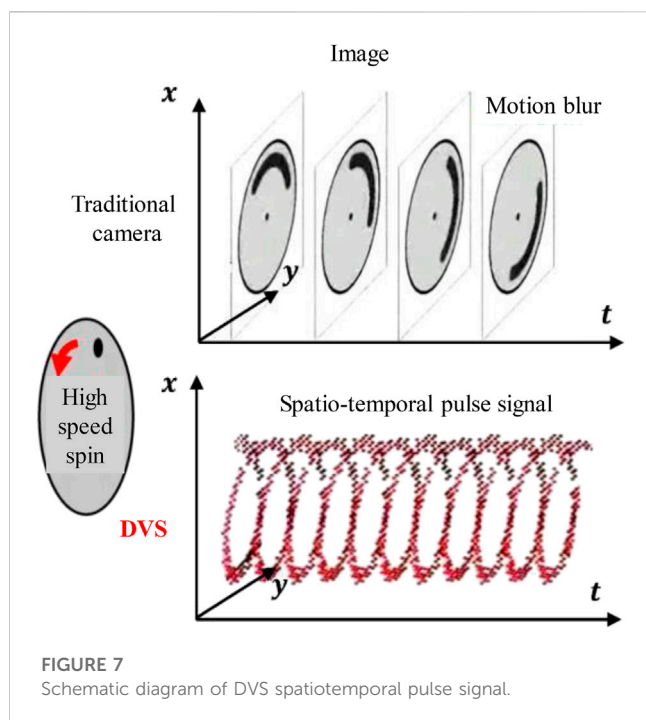
3.4 CeleX

CeleX (Feng et al., 2020) takes into account the hysteresis of the light intensity measurement circuit of ATIS, when the DVS circuit outputs the address (x, y) of the pulse event and the release time t, it also outputs the light intensity information I of the pixel in time, namely Available quads for CeleX output events (x, y, t, I) express. The design idea of CeleX mainly includes three parts (Merchan et al., 2023):

(1) Introduce buffer and readout switch circuits to directly convert the circuit of the logarithmic photoreceptor into light intensity information output;
(2) Use the global control signal to output a whole frame of image, so that the whole image can be obtained as the background and Timely global update;
(3) Specially design the light intensity value of the output buffer of the column analog readout circuit. CeleX cleverly designs the bit width of the pulse event to be 9 bits, which not only ensures the semantic information of the pulse itself, but also carries a certain amount of light intensity information.

The fifth-generation CeleX-V (Tang et al., 2023) recently released by CelePixel has a spatial resolution of 1,280,800, which basically reaches the level of traditional cameras. At the same time, the maximum output sampling frequency in the time domain is 160 MHz and the dynamic range is 120 dB. The "three highs" advantages of high spatial resolution, high temporal resolution and high dynamic range of this product have attracted the attention of the current field of neuromorphic engineering (Ma et al., 2023; Mi et al., 2023). In addition, CelePixel has also received a 40 million project funding from Baidu, using CeleX-V for automatic driving assistance systems in cars, and using its advantages to monitor abnormal driving behaviors in real time (Mao et al., 2022).

The pulse events of CeleX use 9-bit information output. When the scene is in severe motion or high-speed motion, the data volume cannot be transmitted in time, and even part of the pulse data is discarded, so that the sampling signal cannot be fidelity (Zhao J. et al., 2023), and it cannot respond to light in time. Updates and other shortcomings. However, CeleX's "three-high" performance and its advantages in optical flow information output have great application potential in automatic driving, UAV visual navigation (Zhang et al., 2023a), industrial inspection and video surveillance and other tasks involving high-speed motion vision.

3.5 Vidar

Vidar (Auge et al., 2021) abstracted the function of the three-layer structure of photoreceptors, bipolar cells, and ganglion cells in the primate fovea, and adopted an integral visual sampling model to encode pixel light intensity as frequency or pulse interval, with the ability to reconstruct the fine texture of visual scenes at high speed (Zhang H. et al., 2022). Vidar consists of photoelectric conversion circuit, integrator circuit and comparator output circuit, as shown in Figure 6A. The photoreceptor converts the light signal into an electrical signal, the integrator integrates and accumulates the electrical signal, and the comparator compares the accumulated value with the pulse release threshold to determine the output pulse signal (Mi et al., 2022), and the integrator is reset, also known as pulse frequency modulation (PFM) (Purohit and Manohar, 2022).

The pulse signal output between Vidar pixels is independent of each other. The pulse signals of a single pixel are arranged in a "pulse sequence" according to the time sequence, and all pixels form a "pulse array" according to the spatial position relationship. The cross-section of the pulse array at each moment is called "Pulse plane", the pulse signal is represented by "1", and the no pulse signal is represented by "0", as shown in Figure 8.

The first Vidar (Auge et al., 2021) has a spatial resolution of 400×250, a time domain sampling frequency of 4×104 Hz, and an output of 476.3M of data per second. Refined texture reconstruction for static scenes or high-speed motion scenes, such as the use of sliding window accumulation method or pulse interval mapping method (Zhang K. et al., 2022). In addition, Vidar can freely set the duration of the pulse signal for image reconstruction, and has flexibility in the dynamic range of imaging. This integral vision sampling chip can perform refined texture reconstruction for high-speed motion (Zhang H. et al., 2022), and can be used for object detection, tracking and recognition in high-speed motion scenes, and applications in the fields of high-speed vision tasks such as automatic driving, UAV visual navigation, and machine vision.

Vidar uses an integral visual sampling model to encode the light intensity signal by frequency or pulse interval. The essence is to convert the light intensity information into frequency encoding. Compared with the DVS series sensors for motion perception, it is more friendly to visual fine reconstruction. However, Vidar will generate pulses in both static scenes and moving areas, and there is a huge data redundancy in sampling. How to control the pulse emission threshold to adaptively

FIGURE 8
Schematic diagram of Vidar spatiotemporal pulse signal.
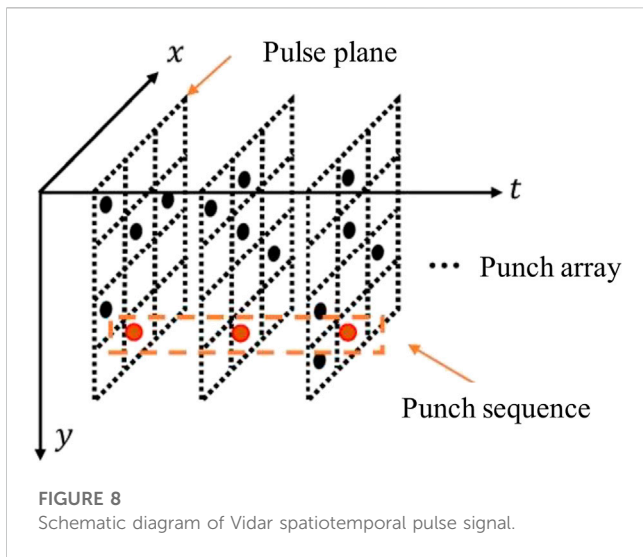
perceive different lighting scenes and control the amount of data is an urgent problem that needs to be solved in integrated visual sampling.

3.6 Performance comparison of imitation retina vision sensors

Recently, a large number of neuromorphic vision sensors have emerged and commercialized, including differential visual sampling models that simulate the peripheral sensory motor function of the retina. There are also integral visual sampling models that simulate the foveal function, such as Vidar (Auge et al., 2021). The comparison of specific performance parameters is shown in Table 1.

Neuromorphic vision sensors have two major advantages:

(1) the ability of high-speed visual sampling, which has great application potential in high-speed motion vision tasks;
(2) low power consumption, which is also the essential advantage of the neuromorphic engineering proposed by Mead (Han et al., 2021) and possible final form in the future. However, how to process the spatiotemporal pulse signals output by neuromorphic vision sensors, feature expression and high-speed visual task analysis is the current research focus of neuromorphic vision. At the same time, how to sample brain-like chips for high-speed processing of pulse signals is used in applications involving high-speed vision tasks are the focus of the neuromorphic engineering industry, such as IBM's TrueNorth (Chen et al., 2022c) chip, Intel's Loihi (Davies and Srinivasa, 2018) chip, and Manchester University's SpiNNaker (Russo et al., 2022) chip.

Currently, the spatial resolution of neuromorphic vision sensors has developed from 128 × 128 of the first commercial DVS128 developed by IniVation (Abubakar et al., 2023) to 640 × 480 of Samsung's DVS-G2 (Xu et al., 2020), CelePixel's CeleX-V (Tang et al., 2023) 1,280 × 800, but compared with traditional HD and UHD cameras, there is a big gap due to: 1) Spatial

resolution and imaging quality; 2) The original intention of dynamic vision sensor design is to perceive high-speed motion instead of high-quality visual viewing. In a word, neuromorphic vision sensors are still in the early stage of exploration, and a lot of exploration and research are needed to achieve the perception ability of the human visual system in complex interactive environments.

4 Asynchronous spatiotemporal pulse signal processing

Neuromorphic vision sensors simulate the pulse firing mechanism of the biological retina. For example, the DVS series sensors using the differential visual sampling model are stimulated by changes in the light intensity of the visual scene to emit pulse signals and record them as address events. The pulse signals present three-dimensionality in the spatial and temporal domains. The sparse discrete lattice of space is shown in Figure 7.

The traditional video signal is represented by the "image frame" paradigm for visual information representation and signal processing, which is also the mainstream direction of existing machine vision. However, "asynchronous spatiotemporal pulse signal" is different from "image frame", and the existing image signal processing mechanism cannot be directly transferred to the application. How to establish a new set of signal processing theory and technology (Zhu R. et al., 2021) is the research difficulty and hotspot in the field of neuromorphic visual signal processing (Qi et al., 2022).

4.1 Analysis of asynchronous space-time pulse signals

In recent years, the analysis of asynchronous spatiotemporal pulse signals (Chen et al., 2020a) mainly focuses on filtering, noise reduction and frequency domain variation analysis.

The filtering analysis of pulse signal is a preprocessing technique from the perspective of signal processing, and it is also the application basis of the visual analysis task of neuromorphic vision sensor. Sajwani et al. (Sajwani et al., 2023) proposed a general filtering method for asynchronous spatiotemporal pulse signals, which consists of hierarchical filtering in the time domain or spatial domain, which can be extended to complex non-linear filters such as edge detection. Linares et al. (Linares et al., 2019) filter noise reduction and horizontal feature extraction of asynchronous spatiotemporal pulse signals on FPGA, which can significantly improve target recognition and tracking performance. Li et al. (Li H. et al., 2019) used an inhomogeneous Poisson generation process to realize the up-sampling of the asynchronous spatio-temporal pulse signal after performing spatio-temporal difference filtering for the pulse signal's firing rate.

Neuromorphic vision sensor outputs an asynchronous spatiotemporal pulse signal that is interfered with by leakage current noise and background noise. Khodamoradi et al. (Khodamoradi and Kastner, 2018) implemented the

TABLE 1 Comparison of performance parameters of neuromorphic vision sensors.

Sensor type	(Abubakar et al., 2023) [DVS128]	(Oliveria et al., 2021) [ATIS]	(Moeys et al., 2017) [DVS346]	(Lenero et al., 2018) [DVS-G2]	(Feng et al., 2020) [CeleX]	(Auge et al., 2021) [Vidar]
Commercial time	2008	2011	2017	2017	2018	2018
Spatial resolution	128×128	304×240	346×260	640×480	1280×800	400×250
Maximum sample rate (Hz)	1×10^6	1×10^6	1.2×10^7	3×10^9	1.6×10^8	4×10^4
Dynamic range (dB)	120	143	120	90	120	70
Power consumption (mW)	23	50–175	10–170	27–50	390–470	370
Chip size (mm^2)	6.3×6	9.9×8.2	8×6	8×5.8	14.3×11.6	10×6
Cell size (μ m^2)	40×40	30×30	18.5×18.5	9×9	9.8×9.8	20×20
Fill factor	8.1%	20%	22%	100%	9%	13.75%
Delay (μ s)	12	3	20	65–410	1	25
Voltage (V)	3.3	1.8 and 3.3	1.8 and 3.3	1.2 and 2.8	1.2 and 3.3	1.5 and 3.3
Texture image	No	Gray scale	Color	No	Gray scale	Gray scale

spatiotemporal correlation filter as hardware on the sensor to minimise the DVS background noise. The spatiotemporal pulse signal generated by ATIS was denoised by the Orchard team (Padala et al., 2018) using the Spiking Neural Network (SNN) on the TrueNorth chip, and the performance of target object identification and recognition was enhanced. Du et al. (Du et al., 2021), presents an events other than moving objects were seen as noise, and optical flow was used to assess the motion consistency in order to denoise the spatiotemporal pulse signal produced by DVS.

Transform domain analysis is the basic method of signal processing (Alzubaidi et al., 2021), which transforms the time-space domain into the frequency domain, and then studies the spectral structure and variation law of the signal. Sabatier et al. (Sabatier et al., 2017) suggested an event-based fast Fourier transform for the asynchronous spatiotemporal pulse signal and conducted a cost-benefit analysis of the pulse signal's frequency domain lossy transform.

The analysis and processing of asynchronous spatiotemporal pulse signals has the following exploration directions:

(1) The asynchronous spatiotemporal pulse signal can be described as a spatiotemporal point process in terms of data distribution (Zhu Z. et al., 2022), and the theory of point process signal processing, learning and reasoning can be introduced (Remesh et al., 2019; Xiao et al., 2019; Ru et al., 2022);

(2) Asynchronous spatiotemporal pulse signal is similar to point cloud in spatiotemporal structure, and deep learning can be used in the structure and method of point cloud network (Wang et al., 2021; Lin L. et al., 2023; Valerdi et al., 2023);

(3) The pulse signal is regarded as the node of the graph model, and the graph model signal processing and learning theory can be used (Shen et al., 2022; Bok et al., 2023);

(4) The timing advantage of the high temporal resolution of the asynchronous spatiotemporal pulse signal is to mine the temporal memory model (Zhu N. et al., 2021; Li et al., 2023) and learn from the brain-like visual signal processing mechanism (Wang X. et al., 2022).

4.2 Asynchronous spatiotemporal pulse signal measurement

Asynchronous spatiotemporal impulse signal metric is to measure the similarity between impulse streams, that is, to calculate the distance between impulse streams in metric space (Lin L. et al., 2023). It is one of the key technologies in asynchronous spatiotemporal impulse signal processing. It has a wide range of important applications in fields such as compression and machine vision tasks.

The asynchronous pulse signal appears as a sparse discrete lattice in the space-time domain, lacking the algebraic operation measure in the Euclidean space. Stereo vision research (Steffen et al., 2019) measured the output pulse signal in two-dimensional space projection and extracted the time plane of spatio-temporal relationship, and applied it to depth estimation in three-dimensional vision. Using the perspective of visual features, Lin et al. (Lin X. et al., 2022) methodically quantified the spatiotemporal pulse signal produced by DVS and used it to vision tasks including motion correction, depth estimation, and optical flow estimation. These techniques add up the time-space pulse signal's frequency characteristics, but they do not fully use its time-domain properties. The kernel method metric was discussed in (Lin Z. et al., 2022) from the viewpoint of the signal domain, i.e., converting the discrete time-domain pulse signal into a continuous function and determining the separation of the pulse sequences in the inner product of the Hilbert space. Wu et al. (Wu et al., 2023) utilized the convolutional neural network structure to the test and verification of retinal prosthesis data by mapping discrete pulse signals to the feature space and measuring the distance between the pulse signals. Such techniques do not take into account the labelling characteristics of genuine asynchronous spatiotemporal spikes and are instead tested using neurophysiological or simulation-generated spike data (Kim and Jung, 2023).

Dong et al. (Dong et al., 2018) proposed a pulse sequence measurement method with independent pulse labeling attributes, that is, the pulse signals of ON and OFF labeling attributes output by DVS were measured separately, and the discrete pulse sequence was transformed into a smooth continuous function by using a Gaussian

FIGURE 9
Coding framework of asynchronous spatiotemporal pulse signal.
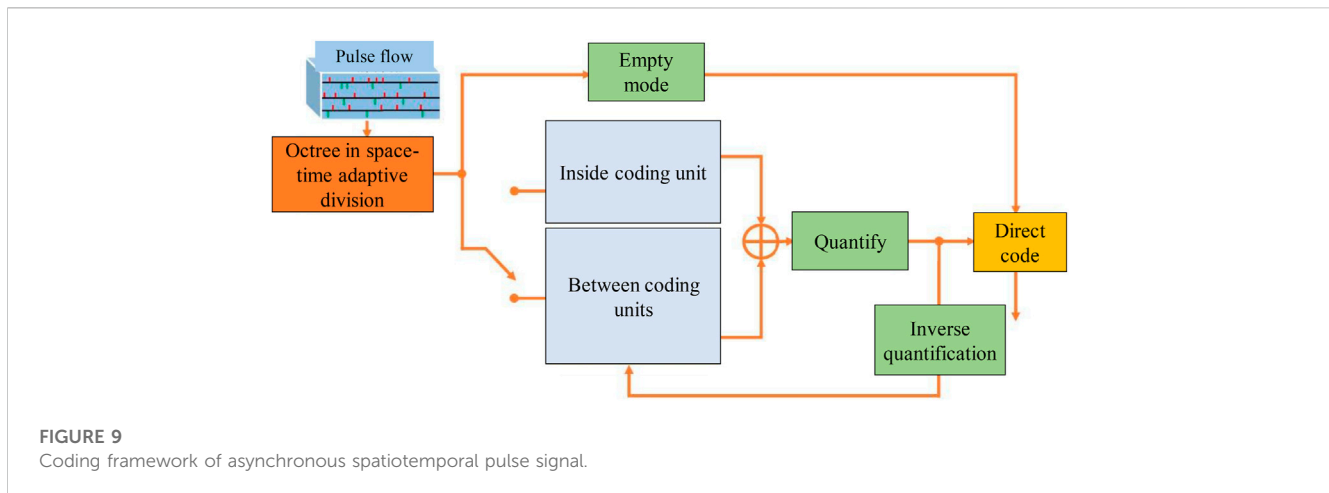
kernel function, using the inner product of the Hilbert space to measure the distance of the pulse sequence. In this method, the pulse sequence is used as the operation unit, and the spatial structure relationship of the pulse signal is not considered. Subsequently, the team (Schiopu and Bilcu, 2023) further modeled the asynchronous spatiotemporal pulse signal as a marked spatiotemporal point process, and used a conditional probability density function to characterize the spatial position and labeling properties of the pulse signal, which was applied to the motion in lossy encoding of the asynchronous pulse signal.

Asynchronous spatiotemporal pulse signals are unstructured data, different from normalizable structured "image frames", and the differences cannot be directly measured in subjective vision. How to orient the measurement of asynchronous pulse signals to visual tasks and the evaluation of normalization is also a difficult problem that needs to be solved urgently.

4.3 Coding of asynchronous space-time pulse signals

With the continuous improvement of the spatial resolution of DVS series sensors, for example, the spatial resolution of Samsung's DVS-G2 (Xu et al., 2020) is 640 × 480, and the spatial resolution of CeleX-V (Tang et al., 2023) is 1,280 × 800. The generated asynchronous spatiotemporal pulse signal is faced with huge challenges of transmission and storage. How to encode and compress the asynchronous spatiotemporal pulse signal is a brand-new spatiotemporal data compression problem (Atluri et al., 2018; Khan and Martini, 2019).

Kim et al. (Kim and Panda, 2021) first proposed a coding compression framework for spatio-temporal asynchronous pulse signals, which took the pulse cuboid as the coding unit and designed address-priority and time-priority predictive coding strategies to achieve effective compression of pulse signals. Subsequently, the team (Dong et al., 2018) further explored more flexible coding strategies such as adaptive division of octrees in the spatiotemporal domain, prediction within coding units, and prediction between coding units, which further improved the compression efficiency of spatiotemporal pulse signals. In addition, the team (Schiopu and Bilcu, 2023) performed a metric analysis of the distortion of pulse

spatial position and temporal firing time, and explores vision-oriented tasks.

Analyzing the lossy coding scheme of (Cohen et al., 2018), the specific coding frame of the asynchronous spatiotemporal pulse signal is shown in Figure 9.

The compression scheme of the asynchronous spatiotemporal pulse signal is a preliminary attempt based on the traditional video coding framework and strategy, but the spatiotemporal characteristics of the asynchronous spatiotemporal pulse signal have not been fully analyzed. Can end-to-end deep learning or adaptive encoders of spiking neural networks inspired by neural computational models be applied to the encoding of asynchronous spatiotemporal spiking signals? At present, it is faced with the problems of network input, distortion measurement and sufficient data labeling of asynchronous pulse signals. Therefore, how to design a robust adaptive encoder for data compression of asynchronous spatiotemporal pulse signals will be a very challenging and valuable research topic, which can be further extended to the field of coding and compression of biological pulse signals.

5 Characteristic expression of asynchronous spatiotemporal pulse signal

Asynchronous spatiotemporal pulse signals are presented as sparse discrete lattices in three-dimensional space in both time and space domains, which are more flexible in signal processing and feature expression than the traditional "image frame" paradigm, especially in the time domain length of the pulse signal processing unit or the choice of the number of pulses also increases the difficulty of inputting the visual analysis algorithm of asynchronous spatiotemporal pulse signals. Therefore, how to express the characteristics of asynchronous spatiotemporal pulse signals (Guo et al., 2023), as shown in Figure 10, mining the spatiotemporal characteristics of asynchronous spatiotemporal pulse signals for the visual analysis tasks of "high precision" and "high speed" is the most important method in the field of neuromorphic vision. Important and core research issues also determine the promotion and application of neuromorphic vision sensors. The overview and distribution of research literature in recent years mainly focus on

FIGURE 10
Image representation of frequency accumulation of asynchronous pulse signal (Guo et al., 2023).
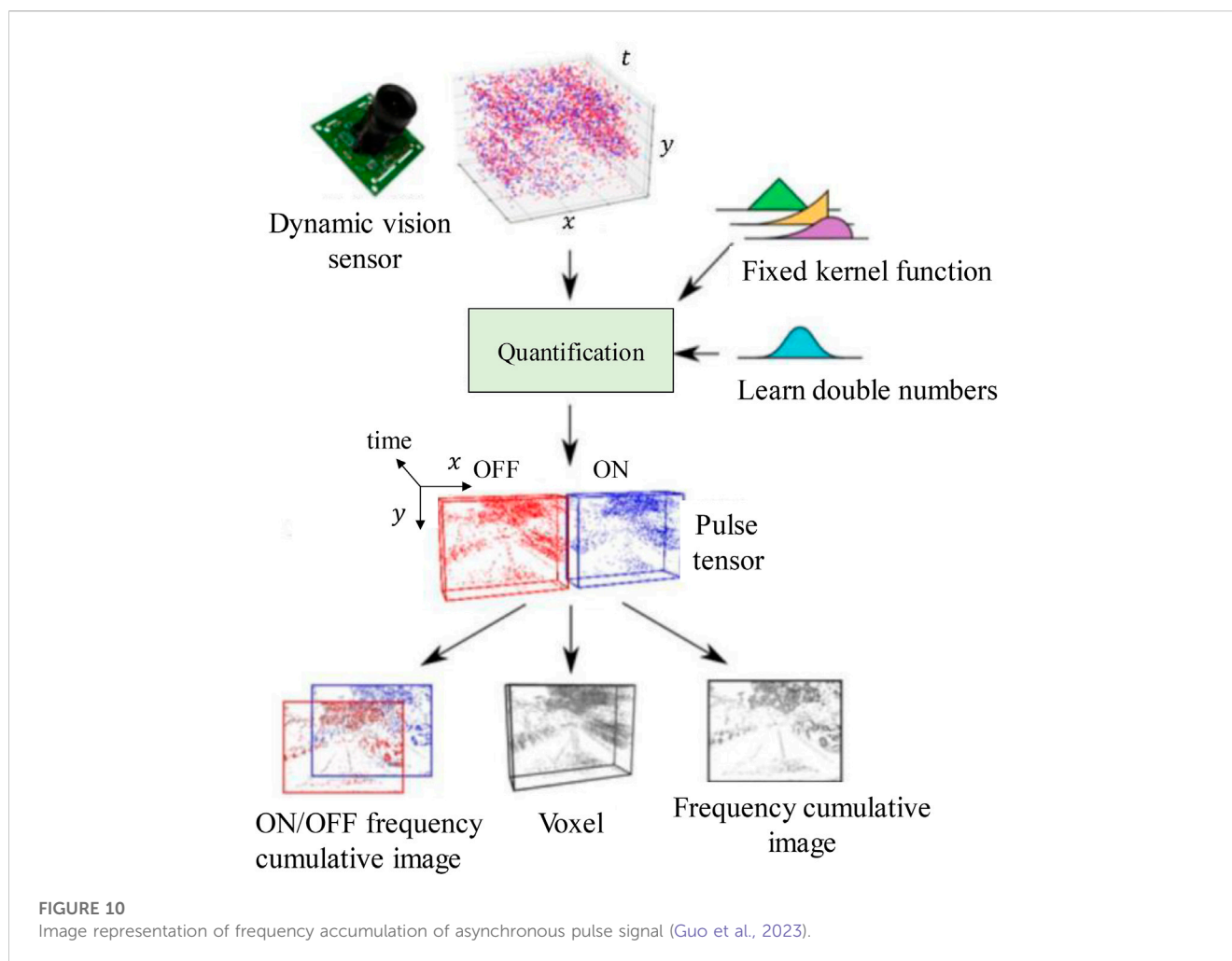
four aspects: rate-based image, hand-crafted feature, end-to-end deep network and spiking neural network.

5.1 Frequency accumulation image

In order to apply the output of neuromorphic vision sensor to the existing visual algorithm based on "image frame", the asynchronous spatiotemporal pulse signal can be projected or accumulated in time domain according to the fixed length or number of time domain, that is, the frequency accumulation image.

Model method: The frequency accumulation image is modeled and feature extracted according to the prior knowledge of the image pattern. Ghosh et al. (Ghosh et al., 2022) projected the pulse signals output by the two DVSs into binary images in the time domain to reconstruct the 360° panoramic visual scene. Huang et al. (Huang et al., 2018) used the pulse signal output by CeleX to interpolate image frames and guide the motion area for high-speed target tracking. The Jiang and Gehrig (Jiang et al., 2020a; Gehrig et al., 2020) accumulated the DVS frequency in the time domain as a grayscale image, and then performed maximum likelihood matching tracking with the APS image.

Deep learning method: The frequency accumulation image is input into the deep learning network based on "image frame". Lai

et al. (Lai and Braunl, 2023) accumulated the ON and OFF pulse streams into grayscale images according to the frequency in the time domain, and then used ResNet to predict the steering wheel angle of the autonomous driving scene. Zeng et al. (Zeng et al., 2023) used the pseudo-label of APS for vehicle detection in autonomous driving scenes after mapping the pulse stream output by DVS into a grayscale image. Ryan et al. (Ryan et al., 2023) combined the APS image for vehicle detection in autonomous driving scenes after extracting the pulse stream output by the DVS as an image using the integral distribution model. Cifuentes et al. (Cifuentes et al., 2022) integrated APS for pedestrian detection with the DVS pulse stream output to create a grayscale picture with a set time domain duration. Shiba et al. (Shiba et al., 2022a) utilized the suggested EV-FlowNet network for optical flow estimation and translated the pulse flow generated by the DVS into grayscale pictures in accordance with the time domain sequence. Lele et al. (Lele et al., 2022) performed high dynamic and high frame rate picture reconstruction after amassing images with a set length or a certain number in the temporal domain in accordance with the pulse flow. Xiang et al. (Xiang et al., 2022) used a fixed time domain duration and toggled ON and OFF. The proposed EV-SegNet network is used to segment autonomous driving scenes using the histogram data, which is separately counted and combined into grayscale images. Yang et al. (Yang F. et al., 2023) uses the attention

mechanism to find the target by mapping the pulse flow to a grayscale image in accordance with the pulse flow's sequence.

These techniques can quickly and successfully apply neuromorphic vision sensors to visual tasks involving high-speed motion because they succinctly convert the asynchronous spatiotemporal pulse stream into time-domain projection or frequency accumulation, which is directly compatible with the "image frame" vision algorithm. However, this approach does not completely take use of the spatiotemporal properties of the impulsive flow. The current mainstream of the feature representation of asynchronous spatiotemporal pulse signals.

5.2 Hand-designed features

Before the dominance of deep learning algorithms, hand-designed features were also widely used in the field of machine vision, such as the well-known SIFT (Tang et al., 2022) operator. How to design compact hand-designed features for asynchronous pulse signals (Ramesh et al., 2019), and have the robust characteristics of scale and rotation invariance, is an important technology for neuromorphic vision sensors to be applied to vision tasks.

How to create hand-designed characteristics for vision tasks is being investigated by certain researchers in the field of neuromorphic vision. Edge and corner features were extracted from asynchronous spatiotemporal pulse signals using their temporal and spatial distribution properties. These features were then employed for tasks including target recognition, depth estimation in stereo vision, and target local feature tracking. Zhang et al. (Zhang et al., 2021) used a sampling-specific convolution kernel chip to extract features from the pulse stream and use it for high-speed target tracking and recognition. Rathi et al. (Rathi et al., 2023) used spatiotemporal filters for unsupervised learning to extract visual receptive field features for target recognition tasks in robot navigation.

A feature of the Bag of Events (BOE) was proposed in (Afshar et al., 2020), which is used for handwritten font recognition data collected by DVS and statistical learning to analyses the probability of events. Ye et al. (Ye et al., 2023) extracted the recognition-oriented HOTS features using hierarchical clustering from the pulse stream produced by ATIS by treating each event as a separate unit and accounting for their temporal relationships within a specific airspace. Dong et al. (Dong et al., 2022) then averaged the event time surfaces in the sample neighborhood and aggregated HATS features in the time domain to remove noise's impact with feature extraction. The edge operator feature of the pulse stream from (Scheerlinck et al., 2019) was extracted using the spatiotemporal filter and then utilized for local feature recognition and target tracking. Wang et al. (Wang Z. et al., 2022) used the time-domain linear sampling kernel to accumulate the weights of the pulse flow as a feature map, and used an unsupervised learning autoencoder network for visual tasks such as optical flow estimation and depth estimation. In addition, also the feature output is used for video reconstruction in the LSTM network. The DART feature operator proposed by the Orchard team (Ramesh et al., 2019) has scale and rotation invariance, and can be applied to the field of target object detection, tracking and recognition.

Hand-designed features have better performance in specific vision task applications, but hand-designed features require a lot of prior knowledge, in-depth understanding of task requirements and data characteristics, and a lot of debugging work. Therefore, using the task-driven cascade method to supervise learning and expressing features (Guo et al., 2023) can better exploit the spatiotemporal characteristics of asynchronous spatiotemporal pulse signals.

5.3 End-to-end deep networks

Deep learning is the current research boom in artificial intelligence, and it has shown obvious performance advantages in image, speech, text and other fields. How to make the asynchronous spatiotemporal pulse signal learn in the end-to-end deep network and fully exploit its spatiotemporal characteristics is the hotspot and difficulty of neuromorphic vision research.

Convolutional Neural Networks: Using 3D Convolution to Process Asynchronous Spatio-Temporal Pulse Signals. Sekikawa et al. (Sekikawa et al., 2018) used an end-to-end deep network to analyze the visual task of asynchronous spatiotemporal pulse signals for the first time. They used 3D spatiotemporal decomposition convolution as the calculation of the input end of the asynchronous spatiotemporal pulse signal, that is, the 3D convolution kernel was decomposed into 2D spatial kernel and 1D motion velocity kernel, and then use recursive operation to process continuous pulse flow in an efficient way, compared with the method of frequency accumulation image, it has a significant improvement in steering wheel angle prediction task in automatic driving scenes, which is a milestone in the field of neuromorphic vision tasks. Uddin et al. (Uddin et al., 2022) proposed an event sequence embedding, which uses spatial aggregation network and temporal aggregation network to extract discrete pulse signals as continuous embedding features, and its performance is compared with frequency accumulation images and hand-designed features in depth. It is estimated that the application has obvious performance advantages.

Point cloud neural network: The asynchronous spatiotemporal pulse signal is treated as a point cloud in three-dimensional space. The pulse signal output by neuromorphic vision sensor is similar to the point cloud in the three-dimensional spatial data structure distribution, but is sparser in the time domain distribution. Valerdi et al. (Valerdi et al., 2023) first used the PointNet (Wang et al., 2021; Lin L. et al., 2023) structure based on point cloud network to process asynchronous spatiotemporal pulse signals, called EventNet, which adopted efficient processing methods of temporal coding and recursive processing, and applied semantic segmentation of autonomous driving scenes and exercise assessment. Chen et al. (Chen X. et al., 2023) regarded asynchronous spatiotemporal impulse signals as event clouds, and adopted the multi-layer hierarchical structure of point cloud network PointNet++ (Lin L. et al., 2023) to extract features for gesture recognition. Jiang et al. (JiangPeng et al., 2022) regarded the asynchronous pulse signal as a point cloud, and proposed an attention mechanism to sample the domain of the pulse signal, which has a significant performance advantage compared to PointNet in gesture recognition.

Graph neural network: The impulse signal is regarded as the node of the graph model, and the processing method of the graph model is adopted. Bok et al. (Bok et al., 2023) proposed to represent the asynchronous spatiotemporal pulse signal by the relevant nodes of the probabilistic graph model, and verified the spatiotemporal representation ability of this method on visual tasks such as pulse signal noise reduction and optical flow estimation. Shen et al. (Shen et al., 2023) modeled asynchronous spatiotemporal pulse signals for the first time in a graph neural network, and achieved a significant performance improvement over frequency cumulative images and hand-designed features in recognition tasks such as gestures (Zhang et al., 2020), alphanumeric, and moving objects (Luo et al., 2022; Lin X. et al., 2023).

The end-to-end deep network can better mine the spatiotemporal characteristics of asynchronous spatiotemporal pulse signals, and its significant performance advantages have also attracted much attention. The performance advantage of deep network supervised learning driven by big data, but the asynchronous pulse signal can hardly be directly subjectively annotated like traditional images, especially in high-level visual tasks such as object detection, tracking and semantic segmentation (Lu et al., 2023a). In addition, the high-speed processing capability and low power consumption of asynchronous spatiotemporal pulse signals are the prerequisites for the wide application of neuromorphic vision sensors, while deep learning currently has no advantages in task processing speed and power consumption. At present, the feature expression of asynchronous pulse signal by end-to-end deep network is still in its infancy, and there are a lot of research points and optimization space.

5.4 Spiking neural networks

The spiking neural network is the third-generation neural network (Bitar et al., 2023), which is a network structure that simulates the biological pulse signal processing mechanism. It considers the precise time information of the pulse signal, and is also one of the important research directions for the feature learning of asynchronous spatiotemporal pulse signals.

The application of spiking neural network in neuromorphic vision mainly focuses on target classification and recognition. Wang et al. (Wang et al., 2022d) proposed a time-domain belief distribution strategy for back-propagation of spiking neural networks, and used GPU for accelerated operations. Fan et al. (Fan et al., 2023) designed a deep spiking neural network for classification tasks, and performed supervised learning and accelerated operations on the deep learning open-source platform. In addition, some researchers have also tried using spiking neural networks in complex visual tasks. Wang and Fan (Wang et al., 2022e; Fan et al., 2023) designed a spiking neural network with multi-layer neuron combination based on empirical information, which was applied to the stereo vision system of binocular parallax and monocular zoom, respectively. Joseph et al. (Joseph and Pakrashi, 2022) proposed a multi-level cascaded spiking neural network to detect candidate regions of fixed scene objects. Quintana (Quintana et al., 2022) designed an end-to-end spiking neural network based on STDP learning rules for robot vision navigation system (Lu et al., 2023b).

At present, the spiking neural network is still in the theoretical research stage, such as the gradient optimization theory of supervised learning of asynchronous spiking spatiotemporal signals (Neftci et al., 2019), the synaptic plasticity mechanism of unsupervised learning (Saunders et al., 2019), the deep learning structure inspired spiking neural network design (Tavanaei et al., 2019). Its performance on neural vision analysis tasks is far less than that of end-to-end deep learning networks. However, the spiking neural network draws on the neural computing model and is closer to the brain visual information processing and analysis mechanism, and has great development potential and application prospects (Jang et al., 2019). Therefore, how to further use the visual cortex information processing and processing mechanism to inspire theoretical models and calculation methods, provide reference ideas and directions for the design and optimization of spiking neural networks, better mine visual feature information and improve computational efficiency, how to solve the spiking neural network supervised learning of neural networks is suitable for complex visual tasks, and how to simulate the efficient calculation of neuron differential equations in hardware circuits or neuromorphic chips is an urgent problem that needs to be solved from theoretical research to practical application of spiking neural networks.

6 Applications

With the rise of cognitive brain science and vision-like computing, machine vision is an important direction to promote the wave of artificial intelligence. Neuromorphic machine vision is an important direction to promote the wave of artificial intelligence. Inspired by the structure and sampling mechanism of biological systems, neuromorphic vision is one of the effective ways to reach or surpass human intelligence, and it has become one of the effective ways for computing gods to reach or surpass human intelligence. The high temporal resolution of the morphological vision sensor, the high temporal resolution of the dynamic range, low power consumption data redundancy, data redundancy and other advantages in automatic driving, low-power data redundancy, in automatic driving, low-power data redundancy and other advantages, in automatic driving, UAV Visual Navigation (Mitrokhin et al., 2019; Galauskis and Ardavs, 2021), Industrial Inspection (Zhu S. et al., 2022), Video Surveillance (Zhang et al., 2022c), and other machine vision fields, especially in the fields of machine vision involving high-speed photography, sports.

6.1 Dataset and software platform

6.1.1 Simulation dataset

The simulation data simulates the neuromorphic vision sensors' sampling mechanism using computational imaging techniques. The simulation data also simulates the optical environment, signal transmission, and circuit sampling in the form of rendering. It generated a pulse flow of DVS, an APS image, and a depth map of the scene while simulating DAVIS moving through a virtual 3D environment. These outputs can be applied to vision tasks like image

reconstruction, depth estimation of stereo vision, and visual navigation. The Event Camera Simulator (ESIM), which simulates and produces data such DVS pulse flow, APS picture, scene optical flow, depth map, camera position and route, was open-sourced by Ozawa et al. (Ozawa et al., 2022). A large-scale multi-sensor interior scene simulation dataset for indoor navigation and localization was supplied by Qu et al. (Qu et al., 2021). In particular, to offer data-driven end-to-end deep learning algorithms, simulation datasets can simulate the effect of actual data gathering in a low-cost manner, which can further enhance the study and development of neuromorphic vision.

6.1.2 Real dataset

At present, the real data sets are mainly classification and recognition tasks: Wunderlich et al. (Wunderlich and Pehle, 2021) recorded motion MNIST and Caltech101 pictures using ATIS and displayed them on an LED monitor. The DVS128 LED was separately reported in (Li et al., 2017; Dong et al., 2022).

MNIST numerical characters and CIFAR-10 image data may be moved around on the screen. IBM (Tchantchane et al., 2023) created the DVS-Gesture gesture recognition dataset by using DVS128 to record 11 gesture motions in a variety of lighting situations. Dong et al. (Dong et al., 2022) created a binary data set of N-CAR vehicles by using ATIS to record cars in real-world road scenes. The largest dataset in the field of recognition, Shen et al. (Shen et al., 2023) assembled the genuine ALS-DVS dataset of 100,000 American letter motions using DAVIS240. Miao et al. (Miao et al., 2019) created the DVS-PAF dataset for pedestrian identification, behavior recognition, and fall detection using color DAVIS346. Zhang et al. (Zhang Z. et al., 2023) recorded the DHP19 dataset of 17 types of 3D pedestrian postures using DAVIS346.

Datasets for scene image reconstruction tasks using neuromorphic vision sensors DDD17 dataset which is also commonly used in neuromorphic vision tasks. Xia et al. (Xia et al., 2023) employed DAVIS346 to record the pulse flow, picture, and steering wheel angle of various illumination conditions. Using DAVIS240 and colour DAVIS346 respectively, Chen et al. (Chen and he, 2020) built the DVS-Intensity and CED datasets for visual scene reconstruction. Zhang et al. (Zhang S. et al., 2022) created the PKU-Spike-High-Speed dataset by using Vidar to record high-speed visual sceneries and moving objects.

Datasets for the Object Detection, Tracking, and Semantic Segmentation Neuromorphic Vision Sensor Vision Task includes (Munir et al., 2022). Ryan et al. (Ryan et al., 2023) created the PKU-DDD17-CAR vehicle detection dataset by marking the vehicles in the driving scene on the DDD17 (Xia et al., 2023) dataset. Cifuentes et al. (Cifuentes et al., 2022) created the DVS-Pedestrian pedestrian detection dataset by using DAVIS240 to capture pedestrians in campus scenarios. Liu et al. (Liu C. et al., 2022) created a DVS-Benchmark data set that may be utilized for target tracking, pedestrian recognition, and target recognition by using DAVIS240 to record the target tracking data set on the LED display. For object detection, tracking, and semantic segmentation, Zhao et al. (Zhao D. et al., 2022) built the EED and EV-IMO datasets, respectively. Munir et al. (Munir et al., 2022) recorded DET data from CeleX-V for lane line detection in scenarios involving autonomous driving.

Datasets for neuromorphic vision sensors include (Zhu et al., 2018; Pfeiffer et al., 2022; Nilsson et al., 2023) for tasks like depth estimation and visual odometry. Using DAVIS346 to record a significant number of autopilot and UAV scenes, Zhu et al. (Zhu et al., 2018) revealed the MVSEC dataset for stereo vision, which is extensively used in the field of neuromorphic vision. A DAVIS240 and an Astra depth camera mounted on a mobile robot were used to record interior scenes in (Nilsson et al., 2023) large-scale multimodal LMED dataset. For the SLAM problem, Pfeiffer et al. (Pfeiffer et al., 2022) created a UZH-FPV dataset with built-in information such as DAVIS346 pulse flow, APS image, optical flow, camera posture, and route (Paredes et al., 2019).

In conclusion, relatively few large-scale datasets are publicly available for neuromorphic vision sensors, especially for complex vision tasks such as object and semantic-level annotation in applications such as object detection, tracking, and semantic segmentation. Developing large-scale datasets for neuromorphic vision applications is the source of data-driven end-to-end supervised learning.

6.2 Visual scene image reconstruction

ATIS and DAVIS can make up for the defect that DVS cannot directly capture the fine texture of the scene, but cannot directly reconstruct the images of high-speed motion and extreme lighting scenes. Some visual scene image reconstruction methods are dedicated to making the DVS series sensors higher. Nagata et al. (Nagata et al., 2021) utilized a sliding spatiotemporal window and minimized optimization functions for scene optical flow and light intensity evaluation. Singh et al. (Singh N. et al., 2022) collected the pulse signal in accordance with the predetermined time domain length before reconstructing the picture using the block sparse dictionary learning technique. A complicated optical flow and manifold construction calculation was presented in (Reinbacher et al., 2018) to rebuild visual scene pictures in real time. Wang et al. (Wang et al., 2020) suggested an asynchronous filtering method that can reconstruct high frame rate and high dynamic video by fusing APS pictures with DVS pulse stream.

Kim et al. (Kim et al., 2023) suggested an event-based double-integral model that can deblur and reconstitute APS pictures using the pulse signal produced by DVS. To survive image-video sequences with high dynamics and high frame rates, Du et al. (Du et al., 2021) employed adversarial generative networks and event-accumulated pictures of defined temporal duration or fixed data. Wang et al. (Wang et al., 2022e) employed the temporal relationship of the LSTM network to reconstruct videos by sampling fixed-length pulse streams in the time domain as feature maps.

Neuromorphic vision sensors are not intended for high-quality visual viewing since they are focused on machine vision perception systems, particularly dynamic vision sensors. Therefore, visual sensors like CeleX-V should directly sample high-quality, high-efficiency, and high-fidelity visual pictures. The visual scene image reconstruction algorithm must also think about how to take use of the benefits of high temporal resolution and high neuromorphism dynamics, as well as how to leverage the timing relationship of pulse signals to further enhance image reconstruction quality.

6.3 Optical flow estimation

Optical flow is the instantaneous speed of pixel motion of space objects on the observation imaging plane. It not only covers the motion information of the measured object, but also has rich three-dimensional structure information. It plays an important role in the research of vision application tasks such as target detection, tracking and recognition. Zhang and Mueggler (Mueggler et al., 2017; Zhang et al., 2023c) proposed optical flow estimation in the pixel domain, which can evaluate targets in high-speed and high-dynamic scenes in real time. In addition, Ieng et al. (Ieng et al., 2017) used 4D spatiotemporal properties to further estimate the optical flow of high-speed moving objects in 3D stereo vision. Wu and Haessig (Haessig et al., 2018; Wu and Guo, 2023) sampled the spiking neural network on the TrueNorth chip for millisecond-level motion optical flow evaluation. Nagata et al. (Nagata et al., 2021) performed scene optical flow on the pulse flow of DVS and used it for visual scene reconstruction. Rueckauer and Hamzah (Rueckauer and Delbruck, 2016; Hamzah et al., 2018) used block matching to evaluate optical flow on FPGA, and further explored the block matching of adaptive time segments (Lee and Kim, 2021), which can evaluate sparse or dense pulse flows in real time. Shiba et al. (Shiba et al., 2022b) accumulated the pulse flow as a feature map by frequency, and then used the supervised learning EV-FlowNet to evaluate the optical flow. A general maximum contrast framework for motion compensation, depth estimation and optical flow estimation is employed. Hordijk et al. (Hordijk et al., 2018) used DVS128 for optical flow estimation, which could keep the UAV landing smoothly. Pardes et al. (Pardes et al., 2019) further used an unsupervised learning hierarchical spiking neural network to perceive global motion. Song et al. (Song et al., 2023) proposed an optical flow estimation method that combines the rendering mode and event gray level in the pixel domain, and samples the optical flow information in CeleX-V.

Due to the difficulty of feature representation in end-to-end supervised learning of asynchronous spatiotemporal pulse signals and the lack of large-scale optical flow datasets, the current optical flow evaluation method is mainly based on the model of prior information, which can directly provide light for the sampling chip of neuromorphic vision sensor. Stream information output. However, the end-to-end supervised learning method can fully exploit the spatiotemporal characteristics of asynchronous spatiotemporal pulse signals, thereby further improving the performance of optical flow motion estimation.

6.4 Object recognition

Neuromorphic vision sensors are widely used in character recognition, object recognition, gesture recognition, gait recognition, and behavior recognition, especially in scenes involving high-speed motion and extreme lighting. Object recognition algorithm is the mainstream of neuromorphic vision task research (Yang H. et al., 2023). From the perspective of processing asynchronous spatiotemporal pulse signals, it is mainly divided into: frequency accumulation image, hand-designed features, end-to-end deep network and spiking neural network.

Frequency cumulative image: Younsi et al. (Younsi et al., 2023) projected the pulse flow into an image according to a fixed time length, and used a feedforward network to recognize the human posture. Morales et al. (Morales et al., 2020) encoded the asynchronous pulse stream into a frequency-accumulated image with a fixed length in the time domain, and then used a convolutional neural network to recognize human poses and high-speed moving characters. Gong (Gong, 2021) accumulated the pulse stream output by DVS as image and speech signal and input it to the deep belief network for character recognition. In addition, Li et al. (Li et al., 2023) used a fixed temporal length to accumulate image sequences and used LSTM to recognize moving characters. Cherskikh (Cherskikh, 2022) accumulated the pulse stream output by ATIS into images according to fixed time domain length or fixed pulse data, and used convolutional network to identify objects. IBM (Tchantchane et al., 2023) accumulated pulse signals into images in time domain, and used convolutional neural network for gesture recognition on the neuromorphic processing chip TrueNorth. Yang et al. (Yang H. et al., 2023) used the attention mechanism to detect and recognize the target from the image of the pulse stream. Lakhshmi et al. (Lakhshmi et al., 2019) accumulated pulse streams as images, and then used convolutional neural networks for action recognition. Du et al. (Du et al., 2021) first preprocessed the pulse signal to denoise, and then input the accumulated images into a convolutional neural network for gait recognition.

Hand-designed features: Mantecon et al. (Mantecon et al., 2019) used an integral distribution model to segment the motion region, and used a hidden Markov model to extract features from the target region for gesture recognition. Afshar et al. (Afshar et al., 2020) carried out handwritten motion font recognition by extracting BOE features. Clady et al. (Clady and Maro, 2017) extracted the motion features of the pulse flow and used it for gesture recognition. In addition, Bartolozzi and Zhang (Bartolozzi et al., 2022; Zhang et al., 2023d) extracted temporal features such as HOTS and HATS respectively for the pulse stream, and used a classifier to recognize handwritten fonts. Shi et al. (Shi et al., 2018) extracted binary features from the pulse stream, and used the framework of statistical learning to recognize characters and gestures. Li et al. (Li et al., 2018) used the time domain coding method to convert the pulse stream into an image, and used the convolutional neural network to carry out features and digital recognition in the classifier.

End-to-end deep network: Chen et al. (Chen J. et al., 2023) treat asynchronous spatiotemporal pulse signals as event clouds, and adopt the hierarchical structure of the end-to-end deep point cloud network PointNet++ (Lin L. et al., 2023) for gesture recognition. Jiang et al. (JiangPeng et al., 2022) adopted an attention mechanism to sample the domain of the impulse signal and used a deep point cloud network structure for gesture recognition. Shen et al. (Shen et al., 2023) modeled asynchronous spatiotemporal pulse signals for the first time in a graph neural network for recognition tasks such as gestures, alphanumeric, and moving objects.

Spiking Neural Network: Xu et al. (Xu et al., 2023) proposed a multi-level cascaded feedforward spiking neural network for handwritten digits.

Character recognition: Zhou et al. (Zhou et al., 2023) constructed a multi-level cascaded spiking neural network model to recognize fast-moving characters. Subsequently, Wang et al.

(Wang et al., 2022e) further adopted an end-to-end supervised learning spiking neural network for alphanumeric recognition. Fan et al. (Fan et al., 2023) designed a deep spiking neural network for classification tasks and performed supervised learning on the deep learning open-source platform.

The end-to-end deep network has obvious advantages in the performance of target classification and recognition tasks, as shown in Table 2. Because the asynchronous spatiotemporal pulse signal is modeled as event points, which can better tap the spatiotemporal characteristics. In addition, large-scale classification and recognition datasets provide data-driven insights for deep network models. How to further utilize the timing advantage of high temporal resolution, mine the temporal memory model, learn from the brain-like visual signal processing mechanism (Schomarker et al., 2023), and perform high-speed recognition on the neuromorphic processing chip is an urgent problem that needs to be studied in current neuromorphic vision tasks.

6.5 Object detection, tracking and segmentation

6.5.1 Object detection

In recent years, object detection methods in neuromorphic vision are divided into two directions from the task perspective: primitive detection and target object detection.

Primitive detection: Khan and Zhao (Khan et al., 2022; Zhao K. et al., 2023) performed corner detection on pulsed streams on the iCub robot platform. In (Khan et al., 2022), a spatial geometric relationship is used to match and detect the corners of targets. In (Zhao K. et al., 2023), the authors extracted the time plane from the pulse signal and used the edge operator to detect the corners of the target. Zheng and Na (Na et al., 2023; Zheng et al., 2023) adopted an asynchronous spatiotemporal filtering method for corner detection, and applied the visual navigation system of high-speed moving robots. Yilmaz et al. (Yilmaz et al., 2021) proposed a velocity-invariant temporal plane feature and used random forest learning for corner detection. Li et al. (Li et al., 2021) proposed a fast computational FA-Harris corner detection operator (Li et al., 2021; Ge et al., 2022; Zhang et al., 2023e; Gomes et al., 2023). Singh et al. (Singh K. et al., 2022) used spiking neural network to implement Hough transform to detect straight lines in pulsating flow. Valeiras et al. (Valeiras and Clady, 2018) used iterative optimization and least squares fitting method to detect the straight line of the pulsed flow. In addition, Lee et al. (Lee and Hwang, 2023) defined a buffer of pulse streams for object edge detection.

Target object detection: Ren et al. (Ren et al., 2022) used the pulse stream output by DVS to accumulate into images according to the frequency and synchronized with APS, and used DVS to cluster the generated target candidate area, and then used convolutional neural network to classify the target area target. Zeng et al. (Zeng et al., 2023) utilized pseudo-labels of APS for vehicle detection in autonomous driving scenarios. Ryan et al. (Ryan et al., 2023) synchronously fused the DVS pulse stream according to the APS frame rate, and then used the convolutional neural network to detect vehicles in autonomous driving scenarios. Cifuentes et al. (Cifuentes et al., 2022) also adopted the strategy of synchronizing DVS and APS

to detect pedestrians jointly. Jiang et al. (Jiang et al., 2020b) used the convolutional neural network to detect the target from the accumulated graph of the ATIS pulse flow, and applied it to the iCub mobile robot platform. Ji et al. (Ji et al., 2023) used convolutional networks to convert the pulse stream into image and alphanumeric detection, respectively.

The research and application of target detection in neuromorphic vision is still in its infancy, mainly for primitive detection, that is, to study the primitive features of vision and detect them to provide basic features for advanced vision tasks such as attitude estimation and visual odometry. At present, some target object detection methods also convert the pulse flow into image features, and do not fully exploit the spatiotemporal characteristics of the pulse flow, especially the advantages of high temporal resolution. DVS has the characteristics of continuous time domain and low spatial resolution, very few large-scale object labeling data sets, and it is difficult to achieve high-precision detection of target objects. Some research methods explore the integration of detection and tracking using time domain information. Therefore, how to further mine the spatiotemporal information of the impulsive flow and use supervised learning and data-driven end-to-end deep networks or bio-inspired brain-like vision methods to achieve high-speed moving object detection is an urgent problem to be solved.

6.5.2 Target tracking

The target tracking algorithm is divided into two directions from the perspective of tracking tasks: primitive feature tracking and target object tracking.

Primitive feature tracking: Zhao and Rui (Rui and Chi, 2018; Zhao K. et al., 2023) established a tracking hypothesis model for the corner motion path of the target (Ryan et al., 2023), which was applied to the visual navigation system of high-speed moving robots. Rakai et al. (Rakai et al., 2022) proposed a probabilistic model of data association, and used optical flow information to optimize the association model to continuously track target feature points under high-speed motion and extreme lighting conditions. Jiang and Gehrig (Jiang et al., 2020a; Gehrig et al., 2020) established a maximum likelihood optimization matching model between the DVS pulse flow integral image and the APS image gradient feature to achieve stable tracking of feature points. In addition, Wu et al. (Wu et al., 2018) proposed a method for fast localization and tracking of straight-line edges for camera pose estimation. Everding et al. (Everding and Conradt, 2018) performed line tracking in a stereo vision system and applied the visual navigation system of mobile robots. Li et al. (Li K. et al., 2019) used DVS pulses, APS images and camera IMU parameters for segment feature tracking.

Target tracking: Lim, Zihao and Berthelon (Berthelon et al., 2017; Zihao et al., 2021; Lim et al., 2022) tracked moving targets such as faces, vehicles in traffic scenes, and high-speed moving particles on optical instruments on ATIS pulse streams. Zhao et al. (Zhao J. et al., 2022) projected the pulse flow as a feature surface in space, and used the methods of motion compensation and Kalman filtering to stably track high-speed moving targets. Ren et al. (Ren et al., 2022) used DVS and APS jointly for moving target detection, which detected target candidate regions from DVS pulse flow and classified them with CNN, and then used particle filter to locate and track the target. Huang et al. (Huang et al., 2018) used CeleX's

TABLE 2 Test performance table of typical target recognition algorithms on public datasets.

Way of expression	Recognition methods	Yang et al. (2022)	Dong et al. (2022)	Yang et al. (2022)	Li et al. (2017)	Amosa et al. (2023)	Tchantchane et al. (2023)	Shen et al. (2023)
	VGG_19 (Shen et al., 2023)	0.972	0.983	0.549	0.334	0.728	-	0.806
Frequency accumulation image	RestNet_50 (Shen et al., 2023)	0.984	0.982	0.637	0.558	0.903	-	0.886
	LSTM (Li et al., 2023)	0.837	0.824	0.196	0.245	0.789	0.769	0.621
Hand designed feature	HOTS (Ye et al., 2023)	0.808	0.803	0.210	0.271	0.624	0.785	0.656
	HATS (Dong et al., 2022)	0.991	0.984	0.642	0.524	0.902	0.933	0.871
End-to-end depth network	PATs (JiangPeng et al., 2022)	-	-	-	-	-	**0.974**	-
	RG-CNNs (Shen et al., 2023)	0.990	**0.986**	**0.657**	0.540	**0.914**	0.938	**0.901**
	H-Frist (Zhou et al., 2023)	0.712	0.595	0.054	0.077	0.561	0.529	0.479
Spiking neural network	Direct-SNN (Fan et al., 2023)	**0.995**	-	-	**0.605**	-	-	-
	SLAYER (Wang et al., 2022a)	0.992	0.956	0.598	0.532	0.907	0.936	0.896

The bold values indicates the highest recognition values.

pulse signal to reconstruct and interpolate frames to improve the frame rate of the image sequence, and the pulse flow can guide the moving area for high-speed target tracking. Kong et al. (Kong et al., 2020) converted the pulse stream of DVS into an image, and used rotation evaluation to track the sky star target. Shair et al. (Shair and Rawashdeh, 2022) proposed an adaptive time plane to convert the pulse stream into an image sequence, and used an integrated detection and tracking method to track the target. These methods all convert the pulses into images or time planes, and do not fully exploit the visual spatiotemporal characteristics of asynchronous pulse signals. Some methods used cluster and track pulse signals in space and time, and Wu et al. (Wu et al., 2022) perform online clustering of pulse signals, which are applied to vehicle target tracking in traffic scenes. Makhadmeh et al. (Makhadmeh et al., 2023) proposed a multi-core parallel clustering target tracking method, which is suitable for multi-target high-speed motion with changing direction and scale. In addition, Camunas et al. (Camunas et al., 2017) further adopted the algorithm of stereo vision matching clustering tracking, which can solve the problem of high-speed moving multi-target occlusion. Such methods can effectively track multiple targets in simple visual scenes, but have the disadvantage of poor robustness in overlapping regions of clustering regions.

High temporal resolution, high dynamics for neuromorphic vision sensors range, low redundancy and low power consumption, especially suitable for target tracking under high-speed motion or extreme lighting conditions (Zhao Q. et al., 2022). At present, the target tracking algorithms all convert the pulse signal into an image or feature surface, and the spatiotemporal characteristics of the pulse signal are not fully exploited (Gelen and Atasoy, 2022). In particular, the "image

frame" processing paradigm is difficult to achieve ultra-high-speed processing capabilities. How to learn from the biological visual signal processing mechanism and The computing power of neuromorphic processing chips to achieve "high precision" and "ultra-high speed" target tracking is an urgent problem to be solved.

6.5.3 Object segmentation

The target segmentation of neuromorphic vision is a technique of dividing the pulse flow into several specific and characteristic regions and extracting the precise location of the target of interest. Asad et al. (Asad et al., 2021) performed spatiotemporal clustering of the pulse flow and constructed a stereo vision system, which can perform real-time segmentation and behavior analysis for multiple pedestrians. Chen et al. (Chen et al., 2018) proposed a real-time clustering tracking algorithm, which can segment and track vehicle objects in traffic scenes in real time. Amosa et al. (Amosa et al., 2023) proposed a mean-shift clustering method to segment and track multiple targets in real time on a manipulator robot. Liu et al. (Liu X. et al., 2022) sampled the pulse stream as a feature sequence in the time domain, and used convolutional neural network for object segmentation. Zhang et al. (Zhang et al., 2021) performed spatiotemporal clustering of pulse flow to segment moving objects and backgrounds, which is especially suitable for moving objects segmentation and background modeling in high-speed motion and extreme lighting conditions.

At present, there are relatively few large-scale datasets for target segmentation of pulse flow, and some research works try to use empirical clustering methods. The impulsive flow appears as a three-dimensional sparse lattice in the spatiotemporal domain,

and its essence is the subspace clustering problem of spatiotemporal sparse signals. When DVS is applied to ego-motion scenes with relative motion such as drones, autopilots, and robots, it also triggers the background pulse flow while sensing the moving target. At the same time, the spatial resolution of the pulse signal is low and lacks texture structure information. Therefore, the sparse subspace clustering of asynchronous pulse signals is a difficulty in current neuromorphic vision research, and it is also an urgent problem to be solved for neuromorphic vision sensors to be applied to complex visual tasks.

6.6 3D scene depth estimation

At present, the main stereo vision applications using neuromorphic vision sensors are binocular imaging and measurement systems, and a few researchers have discussed the methods of monocular visual odometry and monocular zoom. From the processing point of asynchronous spatiotemporal pulse signal, it is mainly divided into: frequency accumulation image, hand-designed features, end-to-end deep network and spiking neural network.

Frequency accumulation image: Mantecon et al. (Mantecon et al., 2019) accumulated the DVS output pulse stream as a histogram to calculate the disparity, which was applied to the gesture recognition task in the stereo vision system. Yan and Rebecq (Rebecq et al., 2018; Yan and Zha, 2019) used binocular DVS to simulate lidar 360° panoramic visual scene reconstruction, which projected the pulse flow in the sensor scanning period as an image to calculate the parallax, and the scene depth could be calculated in real time in the autonomous driving scene.

Hand-designed features: Morales et al. (Morales et al., 2019) first developed a stereo vision system for binocular DVS, which uses the output pixel pulse sequence as feature matching to calculate disparity, and Osswal et al. (Osswal et al., 2017) considers the 3D geometric features of the target to match the depth information and perform 3D reconstruction of the target. Ieng et al. (Ieng et al., 2018) aggregated multiple features such as the time plane and other features from the pulse stream output from ATIS to calculate the disparity, and applied it to the binocular and trinocular stereo vision systems. Ghosh et al. (Ghosh and Gallego, 2022) constructed the disparity container feature of the binocular pulse flow respectively, and then constructed the disparity matching optimization function to calculate the disparity. Zhang et al. (Zhang C. et al., 2023) extracted corner features and spatio-temporal text features respectively for the pulse flow for matching and computing depth information. Xie et al. (Xie et al., 2018) used DAVIS240 to build a binocular stereo vision system, and extracted local features from the pulse stream for stereo matching to calculate depth information. In addition, Liu et al. (Liu et al., 2023a) constructed a multi-view 3D reconstruction of a monocular camera for the first time for pulse flow, which used an event-based scan plane method to generate a time difference map to estimate depth information. Wang et al. (Wang et al., 2022e) extracted the temporal surface features of the pulse signal and established an optimization function to calculate the disparity, which was used for multi-view 3D reconstruction of a monocular camera.

End-to-end deep network: Uddin et al. (Uddin et al., 2022) first proposed an end-to-end deep network for depth estimation in stereo vision systems for asynchronous spatiotemporal pulse signals, which

used spatial aggregation network and temporal aggregation network to extract discrete pulse signals as continuous Compared with the frequency accumulated image and hand-designed features, the performance and processing speed of the embedded feature in the visual task of depth estimation are significantly improved.

Spiking Neural Networks: Risi et al. (Risi et al., 2020) proposed a bio-inspired event-driven collaborative network for stand-alone disparity estimation. IBM (Gallego et al., 2022) used DAVIS240 to build a binocular stereo vision system, and performed cascade computing parallax on the neuromorphic processing chip TrueNorth to perform real-time 3D reconstruction of high-speed motion scenes. Osswald et al. (Osswald et al., 2017) used a spiking neural network composed of multi-level cascaded neurons for depth estimation and 3D reconstruction of binocular vision. In addition, Haessig et al. (Haessig et al., 2019) further explored the stereo vision system of monocular zoom by using the spiking neural network.

In recent years, the successful application of deep learning in the field of stereo vision has significantly improved the computational speed and performance of depth estimation and 3D reconstruction, as shown in Table 3, but the application of stereo vision in neuromorphic vision is just in its infancy (Uddin et al., 2022). Large-scale stereo vision datasets, the fusion of ranging sensors such as LiDAR (Huang et al., 2021), the feature expression of asynchronous pulse signals, and the use of neuromorphic processing chips for high-speed processing in stereo vision (Gallego et al., 2022) are all problems that need to be solved urgently.

6.7 Position estimation and visual odometry

Visual Odometry (VO) is to use the output signals of single or multiple visual sensors to estimate the position and attitude of an agent, also known as the problem of obtaining camera attitude through visual information (Huang and Yu, 2022). From the perspective of camera calibration, it can be divided into camera pose estimation (Pose tracking) and Visual-Inertial Odometry (VIO).

Camera attitude estimation: Colonnier et al. (Colonnier et al., 2021) equipped the UAV with DVS, accumulated the pulse flow into images according to the frequency, and tracked the linear features to evaluate the camera attitude. Liu et al. (Liu et al., 2023b) used a monocular DAVIS camera to jointly evaluate 3D scene structure, 6-DOF camera pose and scene light intensity for the first time. Murai et al. (Murai et al., 2023) used APS images in DAVIS240 to detect corners, and used DVS pulse flow for feature tracking to estimate camera pose. Gallego et al. (Gallego et al., 2017) constructed a scene depth map from the pulse stream, and used a Bayesian filtering strategy to estimate the pose of a 6-DOF camera, and could perform real-time estimation under high-speed motion and extreme lighting conditions. In addition, Gao et al. (Gao et al., 2022) further integrated 3D reconstruction method with parallel acceleration for pose tracking. Belter et al. (Belter and Nowicki, 2019) proposed a 3D photometric feature map containing light intensity and depth information, and adopted a most-likelihood optimization model for 6-DOF pose estimation. Jin et al. (Jin et al., 2021) accumulated a stream of pulses as images and employed a spatially stacked LSTM to learn the 6-DOF camera pose.

Visual-Inertial Odometry: Wang et al. (Wang T. et al., 2022) were the first to combine DVS with APS for visual odometry. Li et al.

TABLE 3 Test performance of typical depth estimation algorithms on the MVESC dataset (Zhu et al., 2018).

Way of expression	Estimation method	Average depth estimation error (cm)	
		Sequence1	Sequence3
Frequency accumulation image	Uddin et al. (2022)	99	103
	Uddin et al. (2022)	93	119
	Bi et al. (2023)	36	36
Hand designed feature	Semi-Dense (Liu et al., 2023a)	**13**	33
	CopNet (Bi et al., 2023)	61	64
Deep network	DDES (Uddin et al., 2022)	13.6	**18.4**

The bold value indicates the highest value of depth estimation.

(Li et al., 2020) extracted the feature surface from the pulse flow and used the extended Kalman filter for feature tracking, and then combined with the internal measurement parameters of the IMU to evaluate the 6-DOF camera attitude. Mueggler et al. (Mueggler et al., 2018) combined DVS pulse flow and IMU parameters to propose a continuous visual-inertial odometry.

At present, most of the work of neuromorphic vision sensors in visual odometry is based on the solution of geometric motion constraints, involving feature extraction, feature matching, motion estimation and other processes. Compared with traditional cameras, its processing speed and performance accuracy are significantly improved. Especially in high-speed motion scenes and high-dynamic scenes. However, the model method has shortcomings such as cumbersome design process and re-calibration of camera replacement platform. The end-to-end deep network has the potential to greatly improve the processing speed and performance accuracy under the data drive, but deep learning maps high-dimensional observation vectors to high-dimensional poses. Vector spaces are an extremely difficult problem. In addition, the use of binocular or multi-purpose visual scene depth information to further improve the performance of visual odometry is also an urgent problem to be solved.

6.8 Neuromorphic engineering system applications

The application of neuromorphic vision sensors in engineering systems mainly includes the following aspects:

Consumer electronics: Samsung (Naeini et al., 2020) utilizes the advantages of DVS to quickly perceive dynamic changes, low data redundancy and low power consumption, and apply it to the touch screen wake-up function of mobile phones. In addition, DVS pulse streaming enhances the video capture quality of traditional cameras (Kim et al., 2023), especially in high-speed scenes and extreme lighting conditions, thereby further improving the high-speed photography and dynamic perception capabilities of consumer electronics cameras.

Industrial inspection: Zhang et al. (Zhang H. et al., 2022) used the high temporal resolution of ATIS for tracking and system feedback of the high-speed moving gripper of an industrial micro-machine tool. Lv et al. (Lv et al., 2021) used DVS for real-time identification of industrial components on high-speed running

platforms. As a visual perception system for industrial inspection, the advantages of low power consumption of neuromorphic vision have huge application potential in the Internet of Things era of the Internet of Everything (IoE) (Vanarse et al., 2019).

Mobile robot: Cao et al. (Cao et al., 2019) used DVS as a visual perception system for mobile robots, which can grasp objects in real time under high-speed movement. Rast et al. (Rast et al., 2018) used DVS for the visual perception system of the iCub humanoid robot, which has the ability to perceive low latency, high-speed motion and extreme illumination (Vidal et al., 2018).

Autonomous driving: Feng et al. (Feng et al., 2020) used CeleX for the driver's fatigue detection system. Lai et al. (Lai and Braunl, 2023) used DVS for steering wheel angle prediction in autonomous driving scenarios. Ryan et al. (Ryan et al., 2021) used a combination of DVS and APS for The performance of joint detection of targets for autonomous driving is significantly improved in high-speed motion scenes and extreme lighting conditions.

UAV navigation: Falanga et al. (Falanga et al., 2019; Falanga et al., 2020) used DVS as a visual navigation system for UAVs, which has better tracking ability than traditional cameras in high-speed motion scenes and extreme lighting.

High-speed vision measurement: Hsu et al. (Hsu et al., 2022) designed a 64×64 dynamic vision sampling chip, which integrates an optical flow calculation module and can measure the speed of high-speed moving targets.

7 Research challenges and possible development directions

The research and application of neuromorphic vision sensors have achieved staged development, but there are still many problems and challenges to reach or surpass the perception ability of the human visual system in complex interactive environments. In addition, this chapter further discusses the main research directions for possible development.

7.1 Research challenges

Neuromorphic vision sensors have the advantage of high resolution in the temporal domain. The output pulse signal

presents a three-dimensional sparse lattice in the temporal and spatial domains. The visual signal processing and application have the following challenges:

(1) Large-scale datasets: Supervised learning vision tasks provide data-driven and visual analysis model evaluation. At present, the datasets of neuromorphic vision tasks are mainly classification and recognition tasks. The data scenarios are simple and small in scale. Generally, there are few large-scale datasets publicly available, especially in complex vision tasks such as object detection, tracking and object detection. Segmentation and other applications, the reason is that the pulse flow presents an asynchronous lattice in space and time, which cannot be directly marked and evaluated manually in subjective vision, and needs to be calibrated with the help of other sensors. Developing large-scale datasets for neuromorphic vision applications is an urgent problem, and computational graphics can be used to build simulation datasets at low cost (Falanga et al., 2019; Falanga et al., 2020). Mapping from existing large-scale vision datasets to datasets of spiking signals (Lin et al., 2021). Large-scale collection and annotation of real datasets with the help of other sensors.

(2) Measurement of asynchronous spatiotemporal pulse signals: calculating the distance between pulse streams in metric space (Lin X. et al., 2023) is one of the basic key technologies in asynchronous spatiotemporal pulse signal processing. It has a wide range of applications in many fields such as tasks. Asynchronous spatiotemporal pulse signal is unstructured data, different from normalized structured "image frame", it lacks the direct measurement of Euclidean space, and the difference in subjective vision cannot be directly measured. How to orient the measurement of the asynchronous pulse signal to visual tasks and the evaluation of normalization is also a research problem that needs to be solved urgently. The asynchronous spatiotemporal pulse signal can be modeled as a spatiotemporal point process in the data distribution (Zhu Z. et al., 2022), and the point process signal processing, measurement and learning theory (Xu et al., 2022b).

(3) Characteristic expression of asynchronous spatiotemporal pulse signals: Mining the spatiotemporal characteristics of asynchronous spatiotemporal pulse signals for high-precision visual analysis tasks. The machine vision method of the "image frame" paradigm is the current mainstream direction, but the asynchronous spatiotemporal pulse signal is different from the "image frame", and the existing deep learning algorithms cannot be directly transferred to the application. In addition, compared with the traditional "image frame" paradigm, the pulse signal is more flexible in terms of feature expression, especially in the selection of the time domain length or the number of pulses of the pulse signal processing unit, which also increases the input value of the visual analysis algorithm of the asynchronous spatiotemporal pulse signal. How to efficiently express features of asynchronous spatiotemporal pulse signals is one of the core research issues in the field of neuromorphic vision. Data-driven end-to-end neural networks for specific visual tasks can be established. Mining mechanisms with temporal memory (Zhu R. et al., 2021; Li et al., 2023) spiking neural network (Wang et al., 2022e) for learning and reasoning.

(4) High-speed processing capability: Mining high-resolution characteristics in time domain, oriented to high-speed vision tasks. The high-speed processing capability of asynchronous spatiotemporal pulse signals is the premise for the wide application of neuromorphic vision sensors, while existing feature expression methods such as hand-designed features and end-to-end deep networks have no advantage in task processing speed. Neuromorphic processing chips have high-speed and parallel processing capabilities on pulse signals, such as IBM's TrueNorth chip, Intel's Loihi (Davies and Srinivasa, 2018) chip, and Manchester University's SpiNNaker (Russo et al., 2022) chip. The focus of the neuromorphic engineering industry.

(5) Neuromorphic open source learning framework: Develop an open source framework suitable for asynchronous pulse signal processing and learning. At present, there are few similar deep learning frameworks in the neuromorphic field, and it is urgent to develop frameworks suitable for unstructured spiking data (Hazan et al., 2018), such as spiking neural networks or hybrid frameworks compatible with deep learning and spiking networks, to provide neuromorphic vision researchers open source learning tools and ecological environment (Eshraghian et al., 2020; Yu et al., 2020). In (Zhou et al., 2019), optical resistive random access memory (ORRAM) arrays are utilized as synaptic devices for neuromorphic vision sensors. Lenero et al. (Lenero et al., 2018) used the integral sampling model in the infrared band, which can image in the infrared band and has the advantages of high temporal resolution (Xu C. et al., 2022), high dynamics and low data redundancy (Shu et al., 2022). Galan et al. (Galan et al., 2021) first developed a neuromorphic sensor for speech signals, namely Dynamic Audio Sensor (DAS). Bartolozzi (Bartolozzi, 2018) a neuromorphic tactile sensor and published a paper in Science (Vidar et al., 2018). It also serves as a guide for computer vision and artificial intelligence technological concepts and directions (Roy et al., 2019)

8 Conclusion

This paper provides a detailed comparative study of various state-of-the-art vision sensors in neuromorphic systems. The principles, system models and important characteristics were evaluated. Detailed comparison and contrast of various neuromorphic vision sensors were presented and their key performance indicators were elaborated. The advantages and features were also discussed. The performance of various depth estimation algorithms was compared and evaluated. Finally, the applications and future prospects were described. The coding framework of spatio-temporal visual signal is presented with detailed emphasis on spiking neural network.

Neuromorphic vision sensor draws on the neural network structure of biological visual system and the processing mechanism of visual information sampling, and simulates, extends or surpasses biological visual perception system at the device function level. There are the following possible development directions in the sensor device material process, visual sampling model, visual analysis task, perception system and its application promotion:

(1) Neuromorphic vision sensors for new materials: new materials, new devices and other applications build neuromorphic vision

sensors with neurons, synapses and memory functions. They have preprocessing storage, signal memory, and visual signal perception. It can improve the performance of back-end visual analysis tasks, and has huge application value in edge computing and the Internet of Things (IoT). Therefore, the development of new materials and device processes is one of the most effective ways for neuromorphic vision sensors to simulate, extend or surpass biological visual perception systems from the device functional level.

(2) Neuromorphic sampling model for multi-spectral bands: The differential and integral sampling models of neuromorphic vision sensors are applied to the spectral bands of visible light, such as infrared, ultraviolet and microwave. How to apply the biological vision sampling model to multi-spectral signal sampling to further improve the perception range and capability of existing single-band neuromorphic vision sensors.

(3) Collaboration for multi-vision tasks: Analysis of the same model with processing multiple visual analysis tasks, it shares the underlying features, and the tasks use each other, and has a stronger generalization ability. At present, neuromorphic vision tasks are specific empirical models and learning methods for a single visual task, and the traditional "image frame" multi-task collaborative model can be used for reference, which functionally approximates the human multi-task visual information processing and analysis capabilities.

(4) Neuromorphic engineering system for multi-sensor fusion: the fusion of multiple neuromorphic sensors such as vision, voice and touch is applied to the intelligent perception system of neuromorphic engineering. How to integrate multiple neuromorphic sensors such as vision, voice and touch, and even with traditional cameras, LiDAR and other sensors, to build a perception system in complex interactive environments is the goal of neuromorphic engineering applications.

(5) Coupling of computational neuroscience and neuromorphic engineering: computational neuroscience studies the fine analysis and sampling mechanism of the retina, uses the optic nerve inversion computing theory to study the retinal codec model, and proposes a new type of neuromorphic visual sampling and reconstruction The model provides theoretical support for the device function level approximation to the biological visual perception system. Draws inspiration for the theoretical model and calculation technique of brain-like vision from the biological visual system. At the same time, the application

of neuromorphic vision in signal sampling, processing and analysis provides model verification for computational neuroscience, and even provides bionic medical devices such as retinal prostheses for the field of life medicine, thereby promoting the research and development of computational neuroscience. The two are coupled with each other, providing theoretical and application support for the construction of a new set of visual information perception, processing and analysis.

Author contributions

ZL: Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Supervision, Writing–original draft, Writing–review and editing. HS: Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing–original draft, Writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abubakar, A., Alharami, A., Yang, Y., and Bermak, A. (2023). Extreme early image recognition using event-based vision. *Sensors* 23 (13), 6195–6218. doi:10.3390/s23136195

Afshar, S., Ralph, N., Xu, Y., Tapson, J., Schaik, A., and Cohen, G. (2020). Event-based feature extraction using adaptive selection thresholds. *Sensors* 20 (6), 1600–1615. doi:10.3390/s20061600

Alexiadis, A. (2019). Deep multiphysics and particle – neuron duality: a computational framework coupling (discrete) multiphysics and deep learning. *Appl. Phys.* 9 (24), 5369–5418. doi:10.3390/app9245369

Alzubaidi, L., Zhang, J., Humaidi, A., Dujaili, A., Duan, Y., Fadhel, M., et al.(2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* 8 (53), 53–18. doi:10.1186/s40537-021-00444-8

Amosa, T., Sebastian, P., Izhar, L., Ibrahim, O., Ayinla, L., Bahashwan, A. A., et al. (2023). Multi-camera multi-object tracking: a review of current trends and future advances. *Neurocomputing* 552, 126558–126569. doi:10.1016/j.neucom.2023.126558

Asad, M., Yang, J., Tu, E., Chen, L., and He, X. (2021). Anomaly3D: video anomaly detection based on 3D-normality clusters. *J. Vis. Commun. Image Represent.* 75, 103047–3047. doi:10.1016/j.jvcir.2021.103047

Atluri, G., Karpatne, A., and Kumar, V. (2018). Spatio-temporal data mining: a survey of problems and methods. *ACM Comput. Surv.* 51 (4), 1–41. doi:10.1145/3161602

Auge, D., Hille, J., Mueller, E., and Knoll, A. (2021). A survey of encoding techniques for signal processing in spiking neural networks. *Neural Process. Lett.* 53, 4693–4710. doi:10.1007/s11063-021-10562-2

Baek, C., and Seo, J. (2022). Encapsulation material for retinal prosthesis with photodetectors or photovoltaics. *IEEE Sensors J.* 22 (2), 1767–1774. doi:10.1109/jsen.2021.3132042

Bartolozzi, C. (2018). Neuromorphic circuits impart a sense of touch. *Science* 360 (6392), 966–967. doi:10.1126/science.aat3125

Bartolozzi, C., Indiveri, G., and Donati, E. (2022). Embodied neuromorphic intelligence. *Nat. Commun.* 13 (1024), 1024–1117. doi:10.1038/s41467-022-28487-2

Belter, D., and Nowicki, M. (2019). Optimization-based legged odometry and sensor fusion for legged robot continuous localization. *Robotics Aut. Syst.* 111, 110–124. doi:10.1016/j.robot.2018.10.013

Berthelon, X., Chenegros, G., Libert, N., Sahel, J. A., Grieve, K., and Benosman, R. (2017). Full-field OCT technique for high speed event-based optical flow and particle tracking. *Opt. Express* 25 (11), 12611–12621. doi:10.1364/oe.25.012611

Bi, Y., Li, C., Tong, X., Wang, G., and Sun, H. (2023). An application of stereo matching algorithm based on transfer learning on robots in multiple scenes. *Sci. Rep.* 13 (12739), 12739–12818. doi:10.1038/s41598-023-39964-z

Bitar, A., Rosales, R., and Paulitsch, (2023). Gradient-based feature-attribution explainability methods for spiking neural networks. *Front. Neurosci.* 17, 1153999–1154014. doi:10.3389/fnins.2023.1153999

Bok, K., Kim, I., Lim, J., and Yoo, J. (2023). Efficient graph-based event detection scheme on social media. *Inf. Sci.* 646 (3), 119415–119431. doi:10.1016/j.ins.2023.119415

Buchel, J., Zendrikov, D., Solina, S., Indiveri, G., and Muir, D. (2021). Supervised training of spiking neural networks for robust deployment on mixed-signal neuromorphic processors. *Sci. Rep.* 11 (23376), 1–13. doi:10.1038/s41598-021-02779-x

Camunas, L., Serrano, T., and Ieng, S. (2017). Event-driven stereo visual tracking algorithm to solve object occlusion. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (9), 4223–4237. doi:10.1109/TNNLS.2017.2759326

Cao, Z., Zhang, D., Hu, B., and Liu, J. (2019). Adaptive path following and locomotion optimization of snake-like robot controlled by the central pattern generator. *Complexity* 19, 1–13. doi:10.1155/2019/8030374

Chen, G., Cao, H., Conradt, J., Tang, H., Rohrbein, F., and Knoll, A. (2020a). Event-based neuromorphic vision for autonomous driving: a paradigm shift for bio-inspired visual sensing and perception. *IEEE Signal Process. Mag.* 37 (4), 34–49. doi:10.1109/msp.2020.2985815

Chen, G., Gao, H., Aafaque, M., Chen, J., Ye, C., Röhrbein, F., et al. (2018). Neuromorphic vision based multivehicle detection and tracking for intelligent transportation system. *J. Adv. Transp.* 18, 1–13. doi:10.1155/2018/4815383

Chen, J., Wang, Q., Cheng, H. H., Peng, W., and Xu, W. (2022b). A review of vision-based traffic semantic understanding in ITSs. *IEEE Trans. Intelligent Transp. Syst.* 23 (11), 19954–19979. doi:10.1109/TITS.2022.3182410

Chen, J., Wang, Q., Peng, W., Xu, H., Li, X., and Xu, W. (2022c). Disparity-based multiscale fusion network for transportation detection. *IEEE Trans. Intelligent Transp. Syst.* 23 (10), 18855–18863. doi:10.1109/TITS.2022.3161977

Chen, J., Xu, M., Xu, W., Li, D., Peng, W., and Xu, H. (2023b). A flow feedback traffic prediction based on visual quantified features. *IEEE Trans. Intelligent Transp. Syst.* 24 (9), 10067–10075. doi:10.1109/TITS.2023.3269794

Chen, L., Xiong, X., and Liu, J. (2022a). A survey of intelligent chip design research based on spiking neural networks. *IEEE Access* 10, 89663–89686. doi:10.1109/access.2022.3200454

Chen, X., Su, L., Zhao, J., Qiu, K., Jiang, N., and Zhai, G. (2023a). Sign language gesture recognition and classification based on event camera with spiking neural networks. *Electronics* 12 (4), 786–818. doi:10.3390/electronics12040786

Chen, Y., and he, Z. (2020). Vehicle identity recovery for automatic number plate recognition data via heterogeneous network embedding. *Sustainability* 2020, 3074–3118. doi:10.3390/su12083074

Cherskikh, E. (2022). A conceptual model of sensor system ontology with an event-based information processing method. *Neurosci. Behav. Physiology* 52, 1310–1317. doi:10.1007/s11055-023-01360-5

Chunduri, R., and Perera, D. (2023). Neuromorphic sentiment analysis using spiking neural networks. *Sensors* 23 (18), 7701–7724. doi:10.3390/s23187701

Cifuentes, A., Vinolo, M., Bescos, J., and Carballeira, (2022). Semantic-driven multi-camera pedestrian detection. *Knowl. Inf. Syst.* 64, 1211–1237. doi:10.1007/s10115-022-01673-w

Clady, X., Maro, J., Barré, S., and Benosman, R. B. (2017). A motion-based feature for event-based pattern recognition. *Front. Neurosci.* 10, 1–14. doi:10.3389/fnins.2016.00594

Cohen, G., Afshar, S., Orchard, G., Tapson, J., Benosman, R., and van Schaik, A. (2018). Spatial and temporal downsampling in event-based visual classification. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (10), 5030–5044. doi:10.1109/tnnls.2017.2785272

Colonnier, F., Vedova, L., and Orchard, G. (2021). ESPEE: event-based sensor pose estimation using an extended kalman filter. *Sensors* 21 (23), 1–16. doi:10.3390/s21237840

Dai, X., Xiao, Z., Jiang, H., and Lui, J. (2023). UAV-assisted task offloading in vehicular edge computing networks. *IEEE Trans. Mob. Comput.* 8 (1), 1–15. doi:10.1109/tmc.2023.3259394

Davies, M., Srinivasa, N., Lin, T. H., Chinya, G., Cao, Y., Choday, S. H., et al. (2018). Loihi: a neuromorphic manycore processor with on-chip learning. *IEEE Micro* 38 (1), 82–99. doi:10.1109/mm.2018.112130359

Dong, J., Jiang, R., Xiao, R., and Tang, H. (2022). Event stream learning using spatio-temporal event surface. *Neural Netw.* 154 (3), 543–559. doi:10.1016/j.neunet.2022.07.010

Dong, S., Bi, Z., Tian, Y., and Huang, T. (2018). Spike coding for dynamic vision sensor in intelligent driving. *IEEE Internet Things J.* 6 (1), 60–71. doi:10.1109/jiot.2018.2872984

Du, B., Li, W., Wang, Z., Xu, M., Gao, T., Li, J., et al. (2021). Event encryption for neuromorphic vision sensors: framework, algorithm, and evaluation. *Sensors* 21 (13), 4320–4417. doi:10.3390/s21134320

Eshraghian, J., Baek, S., Levi, T., Kohno, T., Al-Sarawi, S., Leong, P. H. W., et al. (2020). Nonlinear retinal response modeling for future neuromorphic instrumentation. *IEEE Instrum. Meas. Mag.* 23 (1), 21–29. doi:10.1109/mim.2020.8979519

Everding, L., and Conradt, J. (2018). Low-latency line tracking using event-based dynamic vision sensors. *Front. Neurorobotics* 12, 4–17. doi:10.3389/fnbot.2018.00004

Falanga, D., Kim, S., and Scaramuzza, D. (2019). How fast is too fast? The role of perception latency in high-speed sense and avoid. *IEEE Robotics Automation Lett.* 4 (2), 1884–1891. doi:10.1109/lra.2019.2898117

Falanga, D., Kleber, K., and Scaramuzza, D. (2020). Dynamic obstacle avoidance for quadrotors with event cameras. *Sci. Robotics* 5 (40), eaaz9712–14. doi:10.1126/scirobotics.aaz9712

Fan, X., Zhang, H., and Zhang, Y. (2023). IDSNN: towards high-performance and low-latency SNN training via initialization and distillation. *Biomimetics* 8 (4), 375–418. doi:10.3390/biomimetics8040375

Fang, F., and Hu, H. (2021). Recent progress on mechanisms of human cognition and brain disorders. *Sci. China Life Sci.* 64, 843–846. doi:10.1007/s11427-021-1938-8

Feng, Y., Lv, H., Liu, H., Zhang, Y., Xiao, Y., and Han, C. (2020). Event density based denoising method for dynamic vision sensor. *Appl. Sci.* 10 (6), 2024–2117. doi:10.3390/app10062024

Ferrara, M., Coco, G., Sorrentino, T., Jasani, K., Moussa, G., Morescalchi, F., et al. (2022). Retinal and corneal changes associated with intraocular silicone oil tamponade. *J. Clin. Med.* 11 (17), 5234–5323. doi:10.3390/jcm11175234

Freud, E., Plaut, D., and Behrmann, M. (2016). What's happening in the dorsal visual pathway. *Trends Cognitive Sci.* 20 (10), 773–784. doi:10.1016/j.tics.2016.08.003

Fu, Y., Li, C., Yu, F., Luan, T., and Zhao, P. (2023). An incentive mechanism of incorporating supervision game for federated learning in autonomous driving. *IEEE Trans. Intelligent Transp. Syst.* 7 (5), 1–13. doi:10.1109/TITS.2023.3297996

Galan, D., Morales, J., Fernandez, A., Barranco, A., and Moreno, G. (2021). OpenNAS: open source neuromorphic auditory sensor HDL code generator for FPGA implementations. *Neurocomputing* 436 (3), 35–38. doi:10.1016/j.neucom.2020.12.062

Galauskis, M., and Ardavs, A. (2021). The process of data validation and formatting for an event-based vision dataset in agricultural environments. *Appl. Comput. Syst.* 26 (21), 173–177. doi:10.2478/acss-2021-0021

Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., et al. (2022). Event-based vision: a survey. *IEEE Trans. Pattern Analysis Mach. Intell.* 44 (1), 154–180. doi:10.1109/tpami.2020.3008413

Gallego, G., Lund, J., Mueggler, E., Rebecq, H., Delbruck, T., and Scaramuzza, D. (2017). Event-based, 6-DOF camera tracking from photometric depth maps. *IEEE Trans. Pattern Analysis Mach. Intell.* 40 (10), 2402–2412. doi:10.1109/tpami.2017.2769655

Gao, J., Wang, Y., Nie, K., Gao, Z., and Xu, J. (2018). The analysis and suppressing of non-uniformity in a high-speed spike-based image sensor. *Sensors* 18 (12), 4232–4318. doi:10.3390/s18124232

Gao, X., Xue, H., and Liu, X. (2022). Contrast maximization-based feature tracking for visual odometry with an event camera. *Processes* 10 (10), 2081–2118. doi:10.3390/pr10102081

Ge, Z., Zhang, P., Gao, Y., So, H., and Lam, E. (2022). Lens-free motion analysis via neuromorphic laser speckle imaging. *Opt. Express* 30 (2), 2206–2218. doi:10.1364/oe.444948

Gehrig, D., Rebecq, H., Gallego, G., and Scaramuzza, D. (2020). EKLT: asynchronous photometric feature tracking using events and frames. *Int. J. Comput. Vis.* 128, 601–618. doi:10.1007/s11263-019-01209-w

Gelen, A., and Atasoy, A. (2022). An artificial neural SLAM framework for event-based vision. *IEEE Access* 11, 58436–58450. doi:10.1109/access.2023.3282637

Ghosh, S., Angelo, G., Glover, A., Iacono, M., Niebur, E., and Bartolozzi, C. (2022). Event-driven proto-object based saliency in 3D space to attract a robot's attention. *Sci. Rep.* 12 (7645), 1–15. doi:10.1038/s41598-022-11723-6

Ghosh, S., and Gallego, G. (2022). Multi-event-camera depth estimation and outlier rejection by refocused events fusion. *Adv. Intell. Syst.* 4 (12), 351–364. doi:10.1002/aisy.202200221

Gomes, J., Gasper, J., and Bernardino, A. (2023). Event-based feature tracking in a visual inertial odometry framework. *Front. Robotics AI* 10, 1–18. doi:10.3389/frobt.2023.994488

Gong, T. (2021). Deep belief network-based multifeature fusion music classification algorithm and simulation. *Complexity* 21, 1–10. doi:10.1155/2021/8861896

Guo, G., Feng, Y., Lv, H., Zhao, Y., and Bi, G. (2023). Event-guided image super-resolution reconstruction. *Sensors* 23 (4), 2155–2217. doi:10.3390/s23042155

Haessig, G., Berthelon, X., Ieng, S., and Benosman, R. (2019). A spiking neural network model of depth from defocus for event-based neuromorphic vision. *Sci. Rep.* 9, 3744–3813. doi:10.1038/s41598-019-40064-0

Haessig, G., Cassidy, A., Alvarez, R., Benosman, R., and Orchard, G. (2018). Spiking optical flow for event-based sensors using IBM's TrueNorth neurosynaptic system. *IEEE Trans. Biomed. Circuits Syst.* 12 (4), 860–870. doi:10.1109/tbcas.2018.2834558

Hafed, Z., Yoshida, M., Tian, X., Buonocore, A., and Malevich, T. (2021). Dissociable cortical and subcortical mechanisms for mediating the influences of visual cues on microsaccadic eye movements. *Front. Neural Circuits* 15, 638429–638518. doi:10.3389/fncir.2021.638429

Ham, D., Park, H., Hwang, S., and Kim, K. (2021). Neuromorphic electronics based on copying and pasting the brain. *Nat. Electron.* 4, 635–644. doi:10.1038/s41928-021-00646-1

Hamzah, R., Kadmin, A., Hamid, M., Ghani, S., and Ibrahim, H. (2018). Improvement of stereo matching algorithm for 3D surface reconstruction. *Signal Process. Image Commun.* 65 (4), 165–172. doi:10.1016/j.image.2018.04.001

Han, Y., Wang, B., Guan, T., Tian, D., Yang, G., Wei, W., et al. (2022). Research on road environmental sense method of intelligent vehicle based on tracking check. *IEEE Trans. Intelligent Transp. Syst.* 4 (2), 1–15. doi:10.1109/TITS.2022.3183893

Hazan, H., Saunders, D., Khan, H., Patel, D., Sanghavi, D. T., Siegelmann, H. T., et al. (2018). Bindsnet: a machine learning-oriented spiking neural networks library in python. *Front. Neuroinformatics* 12 (89), 89–18. doi:10.3389/fninf.2018.00089

Heuer, A., Ohl, S., and Rolfs, M. (2020). Memory for action: a functional view of selection in visual working memory. *Vis. Cogn.* 28 (8), 388–400. doi:10.1080/13506285.2020.1764156

Holesovsky, O., Skoviera, R., Hlavac, V., and Vitek, R. (2021). Experimental comparison between event and global shutter cameras. *Sensors* 21 (4), 1–18. doi:10.3390/s21041137

Hordijk, B., Scheper, K., and Croon, G. (2018). Vertical landing for micro air vehicles using event-based optical flow. *J. Field Robotics* 35 (1), 69–90. doi:10.1002/rob.21764

Hossain, M., Sohel, F., Shiratuddin, M., and Laga, H. (2019). A comprehensive survey of deep learning for image captioning. *ACM Comput. Surv.* 51 (6), 1–36. doi:10.1145/3295748

Hou, X., Zhang, L., Su, Y., Gao, G., Liu, Y., Na, Z., et al. (2023). A space crawling robotic bio-paw (SCRBP) enabled by triboelectric sensors for surface identification. *Nano Energy* 105, 108013–108534. doi:10.1016/j.nanoen.2022.108013

Hsu, T., Chen, Y., and Wu, J. (2022). Two ramp reference voltages CDS scheme applied to low noise and high dynamic range PWM pixel. *Microelectron. J.* 135, 5744–5753. doi:10.1016/j.mejo.2023.105744

Huang, J., Wang, S., Guo, M., and Chen, S. (2018). Event-guided structured output tracking of fast-moving objects using a CeleX sensor. *IEEE Trans. Circuits Syst. Video Technol.* 28 (9), 2413–2417. doi:10.1109/tcsvt.2018.2841516

Huang, X., Zhang, Y., and Xiong, Z. (2021). High-speed structured light based 3D scanning using an event camera. *Opt. Express* 29 (25), 35864–35876. doi:10.1364/oe.437944

Huang, Y., and Yu, Z. (2022). Associating latent representations with cognitive maps via hyperspherical space for neural population spikes. *IEEE Trans. Neural Syst. Rehabilitation Eng.* 30 (1), 2886–2895. doi:10.1109/tnsre.2022.3212997

Ieng, S., Carneiro, J., and Benosman, R. (2017). Event-based 3D motion flow estimation using 4D spatio temporal subspaces properties. *Front. Neurosci.* 10, 596–617. doi:10.3389/fnins.2016.00596

Ieng, S., Garneiro, J., Osswald, M., and Benosman, R. (2018). Neuromorphic event-based generalized time-based stereovision. *Front. Neurosci.* 12, 1–16. doi:10.3389/fnins.2018.00442

Jang, H., Simeone, O., Gardner, B., and Gruning, A. (2019). An introduction to probabilistic spiking neural networks: probabilistic models, learning rules, and applications. *IEEE Signal Process. Mag.* 36 (6), 64–77. doi:10.1109/msp.2019.2935234

Jeremie, J., and Perrinet, L. (2023). Ultrafast image categorization in biology and neural models. *Vision* 7 (2), 29–15. doi:10.3390/vision7020029

Ji, M., Wang, Z., Yan, R., Liu, Q., Xu, S., and Tang, H. (2023). SCTN: event-based object tracking with energy-efficient deep convolutional spiking neural networks. *Front. Neurosci.* 17, 1123698–1123717. doi:10.3389/fnins.2023.1123698

Jiang, C., Peng, Y., Tang, X., Li, C., and Li, T. (2022). PointSwin: modeling self-attention shifted window on point cloud. *Appl. Sci.* 12 (24), 1–15. doi:10.3390/app122412616

Jiang, R., Mou, X., Shi, S., Zhou, Y., Wang, Q., Dong, M., et al. (2020b). Object tracking on event cameras with offline-online learning. *CAAI Trans. Intell. Technol.* 5 (3), 165–171. doi:10.1049/trit.2019.0107

Jiang, R., Wang, Q., Shi, S., Mou, X., and Chen, S. (2020a). Flow-assisted visual tracking using event cameras. *CAAI Trans. Intell. Technol.* 6 (2), 192–202. doi:10.1049/cit2.12005

Jin, Y., Yu, L., Li, G., and Fei, S. (2021). A 6-DOFs event-based camera relocalization sstem by CNN-LSTM and image denoising. *Expert Syst. Appl.* 170, 4535–4546. doi:10.1016/j.eswa.2020.114535

Joseph, G., and Pakrashi, V. (2022). Spiking neural networks for structural health monitoring. *Sensors* 22 (23), 9245–9312. doi:10.3390/s22239245

Khan, M., Nazir, D., Pagani, A., Mokayed, H., Liwicki, M., Stricker, D., et al. (2022). A comprehensive survey of depth completion approaches. *Sensors* 22 (18), 6969–7025. doi:10.3390/s22186969

Khan, N., and Martini, M. (2019). Bandwidth modeling of silicon retinas for next generation visual sensor networks. *Sensors* 19 (8), 1751–1818. doi:10.3390/s19081751

Khodamoradi, A., and Kastner, R. (2018). O(N)-space spatiotemporal filter for reducing noise in neuromorphic vision sensors. *IEEE Trans. Emerg. Top. Comput.* 9 (1), 1–23. doi:10.1109/tetc.2017.2788865

Kim, B., Yang, Y., Park, J., and Jang, H. (2023). Machine learning based representative spatio-temporal event documents classification. *Appl. Sci.* 13 (7), 4230–4317. doi:10.3390/app13074230

Kim, J., and Jung, Y. (2023). Multi-stage network for event-based video deblurring with residual hint attention. *Sensors* 23 (6), 2880–2916. doi:10.3390/s23062880

Kim, Y., and Panda, P. (2021). Optimizing deeper spiking neural networks for dynamic vision sensing. *Neural Netw.* 144 (3), 686–698. doi:10.1016/j.neunet.2021.09.022

Kong, F., Lambert, A., Joubert, D., and Cohen, G. (2020). Shack-hartmann wavefront sensing using spatial-temporal data from an event-based image sensor. *Opt. Express* 28 (24), 36159–36175. doi:10.1364/oe.409682

Lai, Z., and Braunl, T. (2023). End-to-end learning with memory models for complex autonomous driving tasks in indoor environments. *J. Intelligent Robotic Syst.* 37, 37–17. doi:10.1007/s10846-022-01801-2

Lakshmi, A., Chakraborty, A., and Thakur, C. (2019). Neuromorphic vision: from sensors to event-based algorithms. *WIREs Data Min. Knowl. Discov.* 9 (4), 614–625. doi:10.1002/widm.1310

Lee, H., and Hwang, H. (2023). Ev-ReconNet: visual place recognition using event camera with spiking neural networks. *IEEE Sensors J.* 23 (17), 20390–20399. doi:10.1109/jsen.2023.3298828

Lee, S., and Kim, H. (2021). Low-latency and scene-robust optical flow stream and angular velocity estimation. *IEEE Access* 9, 155988–155997. doi:10.1109/access.2021.3129256

Lele, A., Fang, Y., Anwar, A., and Raychowdhury, A. (2022). Bio-mimetic high-speed target localization with fused frame and event vision for edge application. *Front. Neurosci.* 16, 1010302–1010315. doi:10.3389/fnins.2022.1010302

Lenero, J., Guerrero, J., Carmona, R., and Rodriguez-Vazquez, A. (2018). On the analysis and detection of flames with an asynchronous spiking image sensor. *IEEE Sensors J.* 18 (16), 6588–6595. doi:10.1109/jsen.2018.2851063

Li, G., Yu, L., and Fei, S. (2020). A binocular MSCKF-based visual inertial odometry system using LK optical flow. *J. Intelligent Robotic Syst.* 100, 1179–1194. doi:10.1007/s10846-020-01222-z

Li, H., Li, G., Ji, X., and Shi, L. (2018). Deep representation via convolutional neural network for classification of spatiotemporal event streams. *Neurocomputing* 299, 1–9. doi:10.1016/j.neucom.2018.02.019

Li, H., Li, G., and Shi, L. (2019a). Super-resolution of spatiotemporal event-stream image. *Neurocomputing* 335, 206–214. doi:10.1016/j.neucom.2018.12.048

Li, H., Liu, H., Ji, X., Li, G., and Shi, L. (2017). CIFAR10-DVS: an event-stream dataset for object classification. *Front. Neurosci.* 11, 309–318. doi:10.3389/fnins.2017.00309

Li, K., Shi, D., Zhang, Y., Li, R., Qin, W., and Li, R. (2019b). Feature tracking based on line segments with the dynamic and active-pixel vision sensor (DAVIS). *IEEE Access* 7, 110874–110883. doi:10.1109/access.2019.2933594

Li, R., Dhi, D., and Wang, M. (2021). Asynchronous event feature generation and tracking based on gradient descriptor for event cameras. *Int. J. Adv. Robotic Syst.* 18 (4), 873–884. doi:10.1177/17298814211027028

Li, Z., Wang, Q., Wang, J., Qu, H., Dong, J., and Dong, Z. (2023). ELSTM: an improved long short-term memory network language model for sequence learning. *Expert Syst.* 4 (1), 3882–3890. doi:10.1111/exsy.13211

Lim, S., Huh, J., and Kim, J. (2022). Deep feature based Siamese network for visual object tracking. *Energies* 15 (17), 6388–6418. doi:10.3390/en15176388

Lin, L., Huang, P., Fu, C., Xu, K., Zhang, H., and Huang, H. (2023a). On learning the right attention point for feature enhancement. *Sci. China Inf. Sci.* 66 (112207), 112107–112467. doi:10.1007/s11432-021-3431-9

Lin, X., Yang, C., Bian, X., Liu, W., and Wang, C. (2022a). EAGAN: event-based attention generative adversarial networks for optical flow and depth estimation. *IET Comput. Vis.* 16 (7), 581–595. doi:10.1049/cvi2.12115

Lin, X., Zhang, Z., and Zheng, D. (2023b). Supervised learning algorithm based on spike train inner product for deep spiking neural networks. *Brain Sci.* 13 (2), 168–216. doi:10.3390/brainsci13020168

Lin, Y., Ding, W., Qiang, S., Deng, L., and Li, G. (2021). ES-ImageNet: a million event-stream classification dataset for spiking neural networks. *Front. Neurosci.* 15, 726582–726616. doi:10.3389/fnins.2021.726582

Lin, Z., Wang, H., and Li, S. (2022b). Pavement anomaly detection based on transformer and self-supervised learning. *Automation Constr.* 143, 104544–111145. doi:10.1016/j.autcon.2022.104544

Linares, A., Perez, F., Moeys, D., Gomez-Rodriguez, F., Jimenez-Moreno, G., Liu, S. C., et al. (2019). Low latency event-based filtering and feature extraction for dynamic vision sensors in real-time FPGA applications. *IEEE Access* 7, 134926–134942. doi:10.1109/access.2019.2941282

Liu, C., Qi, X., Lam, E., and Wong, N. (2022a). Fast classification and action recognition with event-based imaging. *IEEE Access* 10, 55638–55649. doi:10.1109/access.2022.3177744

Liu, L., Chen, J., Fieguth, P., Zhao, G., Chellappa, R., and Pietikainen, M. (2019). From BoW to CNN: two decades of texture representation for texture classification. *Int. J. Comput. Vis.* 127 (1), 74–109. doi:10.1007/s11263-018-1125-z

Liu, X., Zhao, Y., Yang, L., and Ge, S. (2022b). A spatial-motion-segmentation algorithm by fusing EDPA and motion compensation. *Sensors* 22 (18), 6732–6815. doi:10.3390/s22186732

Liu, Z., Shi, D., Li, R., and Yang, S. (2023a). ESVIO: event-based stereo visual-inertial odometry. *Sensors* 23 (4), 1998–2014. doi:10.3390/s23041998

Liu, Z., Shi, D., Li, R., Zhang, L., and Yang, S. (2023b). T-ESVO: improved event-based stereo visual odometry via adaptive time-surface and truncated signed distance function. *Adv. Intell. Syst.* 5 (9), 1571–1583. doi:10.1002/aisy.202300027

Lu, S., Ding, Y., Liu, M., Yin, Z., Yin, L., and Zheng, W. (2023a). Multiscale feature extraction and fusion of image and text in VQA. *Int. J. Comput. Intell. Syst.* 16 (1), 54–782. doi:10.1007/s44196-023-00233-6

Lu, S., Liu, M., Yin, L., Yin, Z., Liu, X., Zheng, W., et al. (2023b). The multi-modal fusion in visual question answering: a review of attention mechanisms. *PeerJ Comput. Sci.*, 9, 1–14. doi:10.7717/peerj-cs.1400

Luo, Z., Wang, H., and Li, S. (2022). Prediction of international roughness index based on stacking fusion mode. *Sustainability* 14 (12), 1–17. doi:10.3390/su14126949

Lv, H., Feng, Y., Zhang, Y., and Zhao, Y. (2021). Dynamic vision sensor tracking method based on event correlation index. *Complexity* 21, 1–11. doi:10.1155/2021/8973482

Ma, X., Dong, Z., Quan, W., Dong, Y., and Tan, Y. (2023). Real-time assessment of asphalt pavement moduli and traffic loads using monitoring data from Built-in Sensors: optimal sensor placement and identification algorithm. *Mech. Syst. Signal Process.* 187, 109930. doi:10.1016/j.ymssp.2022.109930

Macireau, A., Ieng, S., Simon, C., and Benosman, R. B. (2018). Event-based color segmentation with a high dynamic range sensor. *Front. Neurosci.* 12, 1–14. doi:10.3389/fnins.2018.00135

Makhadmeh, S., Betar, M., Doush, I., Awadallah, M., Kassaymeh, S., Mirjalili, S., et al. (2023). Recent advances in grey wolf optimizer, its versions and applications: review. *IEEE Access* 11, 1–38. doi:10.1109/access.2023.3304889

Mantecon, T., Blanco, C., Jaureguizar, F., and Garcia, N. (2019). A real-time gesture recognition system using near-infrared imagery. *Plos One* 14 (10), 1–22. doi:10.1371/journal.pone.0223320

Mao, Y., Zhu, Y., Tang, Z., and Chen, Z. (2022). A novel airspace planning algorithm for cooperative target localization. *Electronics* 11 (18), 2950–3013. doi:10.3390/electronics11182950

Marinis, L., Cococcioni, M., Landouceur, O., Contestabile, G., Castoldi, P., and Andriolli, N. (2021). Photonic integrated reconfigurable linear processors as neural network accelerators. *Appl. Sci.* 11 (3), 1–20. doi:10.3390/app11136232

Martini, M., Adhuran, J., and Khan, N. (2022). Lossless compression of neuromorphic vision sensor data based on point cloud representation. *IEEE Access* 10, 121352–121364. doi:10.1109/access.2022.3222330

Merchan, R., Antonio, J., and Vazquez, A. (2023). A self-powered asynchronous image sensor with TFS operation. *IEEE Sensors* 23 (7), 6779–6790. doi:10.1109/jsen.2023.3248177

Mesa, L., Barranco, B., and Gotarredona, (2019). Neoromorphic spiking neural networks and their memristor-CMOS hardware implementations. *Materials* 12 (17), 1–18. doi:10.3390/ma12172745

Mi, C., Huang, S., Zhang, Y., Zhang, Z., and Postolache, O. (2022). Design and implementation of 3-D measurement method for container handling target. *J. Mar. Sci. Eng.* 10 (12), 1961–1972. doi:10.3390/jmse10121961

Mi, C., Liu, Y., Zhang, Y., Wang, J., Feng, Y., and Zhang, Z. (2023). A vision-based displacement measurement system for foundation Pit. *IEEE Trans. Instrum. Meas.* 8 (3), 1–15. doi:10.1109/TIM.2023.3311069

Miao, S., Chen, G., Ning, X., Zi, Y., Ren, K., Bing, Z., et al. (2019). Neuromorphic vision datasets for pedestrian detection, action recognition, and fall detection. *Front. Neurosci.* 13, 38–14. doi:10.3389/fnbot.2019.00038

Mitrokhin, A., Sutor, P., Fermüller, C., and Aloimonos, Y. (2019). Learning sensorimotor control with neuromorphic sensors: toward hyperdimensional active perception. *Sci. Robotics* 4 (30), 1711–1725. doi:10.1126/scirobotics.aaw6736

Moeys, D., Corradi, F., Li, C., Bamford, S., Longinotti, L., Voigt, F. F., et al. (2017). A sensitive dynamic and active pixel vision sensor for color or neural imaging applications. *IEEE Trans. Biomed. Circuits Syst.* 12 (1), 123–136. doi:10.1109/tbcas.2017.2759783

Morales, M., Morales, J., Fernandez, A., Braranco, A., and Moreno, G. (2019). Stereo matching in address-event-representation (AER) bio-inipred binocular systems in field-programmable grate array (FPGA). *Electronics* 8 (4), 1–15. doi:10.3390/electronics8040410

Morales, R., Maro, J., Fernandez, A., Moreno, G., Benosman, R., and Barranco, A. (2020). Event-based gesture recognition through a hierarchy of time-surfaces for FPGA. *Sensors* 20 (12), 1–17. doi:10.3390/s20123404

Mueggler, E., Gallego, G., Rebecq, H., and Scaramuzza, D. (2018). Continuous-time visual-inertial odometry for event cameras. *IEEE Trans. Robotics* 34 (6), 1425–1440. doi:10.1109/tro.2018.2858287

Mueggler, E., Rebeeq, H., Gallego, G., Delbruck, T., and Scaramuzza, D. (2017). The event-camera dataset and simulator: event-based data for pose estimation, visual odometry, and SLAM. *Int. J. Robotics Res.* 36 (2), 142–149. doi:10.1177/0278364917691115

Munir, F., Azam, S., Jeon, M., Lee, B., and Pedrycz, W. (2022). LDNet: end-to-End lane marking detection approach using a dynamic vision sensor. *IEEE Trans. Intelligent Transp. Syst.* 23 (7), 9318–9334. doi:10.1109/tits.2021.3102479

Murai, R., Saeedi, S., and Kelly, P. (2023). High-frame rate homography and visual odometry by tracking binary features from the focal plane. *Aut. Robots* 8 (3), 1–19. doi:10.1007/s10514-023-10122-8

Na, W., Sun, K., Jeon, B., Lee, J., and Shin, Y. (2023). Event-based micro vibration measurement using phase correlation template matching with event filter optimization. *Measurement* 215, 112867–112875. doi:10.1016/j.measurement.2023.112867

Naeini, F., Makris, D., Gan, D., and Zweiri, Y. (2020). Dynamic-vision-based force measurements using convolutional recurrent neural networks. *Sensors* 20 (16), 1–18. doi:10.3390/s20164469

Nagata, J., Sekikawa, Y., and Aoki, Y. (2021). Optimal flow estimation by matching time surface with event-based cameras. *Sensors* 21 (4), 1–17. doi:10.3390/s21041150

Najaran, M., and Schmuker, M. (2021). Event-based sensing and signal processing in the visual, auditory, and olfactory domain: a review. *Front. Neural Circuits* 15, 1–19. doi:10.3389/fncir.2021.610446

Neftci, E., Mostafa, H., and Zenke, F. (2019). Surrogate gradient learning in spiking neural networks: bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Process. Mag.* 36 (6), 51–63. doi:10.1109/msp.2019.2931595

Nilsson, M., Schelen, O., Lindgren, A., Paniaga, C., Delsing, J., Sandin, F., et al. (2023). Integration of neuromorphic AI in event-driven distributed digitized systems: concepts and research directions. *Front. Neurosci.* 17, 1–16. doi:10.3389/fnins.2023.1074439

Oliveria, V., Lopes, T., Reis, F., Gomes, J., and Santos, G. (2021). Asynchronous time-based imager with DVS sharing. *Analog Integr. Circuits Signal Process.* 108, 539–554. doi:10.1007/s10470-021-01893-0

Osswald, M., Ieng, S., and Benosman, R. (2017). A spiking neural network model of 3D perception for event-based neuromorphic vision systems. *Sci. Rep.* 7, 1–18. doi:10.1038/srep40703

Ozawa, T., Sekikawa, Y., and Saito, H. (2022). Accuracy and speed improvement of event camera motion estimation using a bird's-eye view transformation. *Sensors* 22 (3), 773–817. doi:10.3390/s22030773

Padala, V., Basu, A., and Orchard, G. (2018). A noise filtering algorithm for event-based asynchronous change detection image sensors on TrueNorth and its implementation on TrueNorth. *Front. Neurosci.* 12, 118–212. doi:10.3389/fnins.2018.00118

Paredes, F., Scheper, K., and Croon, G. (2019). Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: from events to global motion perception. *IEEE Trans. Pattern Analysis Mach. Intell.* 42 (8), 2051–2064. doi:10.1109/tpami.2019.2903179

Pfeiffer, C., Wengeler, S., Loquercio, A., and Scaramuzza, D. (2022). Visual attention prediction improves performance of autonomous drone racing agents. *Plos One* 17 (3), 02644711–e264516. doi:10.1371/journal.pone.0264471

Picano, B. (2021). Multi-sensorial human perceptual experience model identifier for haptics virtual reality services in tactful networking. *IEEE Access* 9, 147549–147558. doi:10.1109/access.2021.3124607

Pramod, R., and Arun, S. (2022). Improving machine vision using human perceptual representations: the case of planar reflection symmetry for object classification. *IEEE Trans. Pattern Analysis Mach. Intell.* 44 (1), 228–241. doi:10.1109/tpami.2020.3008107

Purohit, P., and Manohar, R. (2022). Field-programmable encoding for address-event representation. *Front. Neurosci.* 16, 1018166–1018216. doi:10.3389/fnins.2022.1018166

Qi, M., Cui, S., Chang, X., Xu, Y., Meng, H., Wang, , et al. (2022). Multi-region nonuniform brightness correction algorithm based on L-channel gamma transform. *Secur. Commun. Netw.* 22 (1), 1–9. doi:10.1155/2022/2675950

Qu, Y., Yang, M., Zhang, J., Xie, W., Qiang, B., and Chen, J. (2021). An outline of multi-sensor fusion methods for mobile agents indoor navigation. *Sensors* 21 (5), 1605–1618. doi:10.3390/s21051605

Quintana, F., Pena, F., and Galindo, P. (2022). Bio-plausible digital implementation of a reward modulated STDP synapse. *Neural Comput. Appl.* 34, 15649–15660. doi:10.1007/s00521-022-07220-6

Rakai, L., Song, H., Sun, S., Zhang, W., and Yang, Y. (2022). Data association in multiple object tracking: a survey of recent techniques. *Expert Syst. Appl.* 192, 116300–116319. doi:10.1016/j.eswa.2021.116300

Rast, A., Adams, S., Davidson, S., Davies, S., Hopkins, M., Rowley, A., et al. (2018). Behavioral learning in a cognitive neuromorphic robot: an integrative approach. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (12), 6132–6144. doi:10.1109/tnnls.2018.2816518

Rathi, N., Chakraborty, I., Kosta, A., Sengupta, A., Ankit, A., Panda, P., et al. (2023). Exploring neuromorphic computing based on spiking neural networks: algorithms to hardware. *ACM Comput. Surv.* 55 (12), 1–49. doi:10.1145/3571155

Rebecq, H., Gallego, G., Mueggler, E., and Scaramuzza, D. (2018). EMVS: event-based multi-view stereo-3D reconstruction with an event camera in real-time. *Int. J. Comput. Vis.* 126 (12), 1394–1414. doi:10.1007/s11263-017-1050-6

Reinbacher, C., Graber, G., and Pock, T. (2018). Real-time intensity-image reconstruction for event cameras using manifold regularisation. *Int. J. Comput. Vis.* 126 (12), 1381–1393. doi:10.1007/s11263-018-1106-2

Remesh, B., Yang, H., Orchard, G., Thi, N., Zhang, S., and Xiang, C. (2019). DART: distribution aware retinal transform for event-based cameras. *IEEE Trans. Pattern Analysis Mach. Intell.* 42 (11), 2767–2780. doi:10.1109/TPAMI.2019.2919301

Ren, Y., Benedetto, E., Borrill, H., Savchuk, Y., Message, M., Flynn, K., et al. (2022). Event-based imaging of levitated microparticles. *Appl. Phys. Lett.* 121, 423–434. doi:10.1063/5.0106111

Risi, N., Aimar, A., Donati, E., Solinas, S., and Indiveri, G. (2020). A spike-based neuromorphic architecture of stereo vision. *Front. Neurorobotics* 14, 568283–568316. doi:10.3389/fnbot.2020.568283

Roy, K., Jaiswal, A., and Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature* 575 (7784), 607–617. doi:10.1038/s41586-019-1677-2

Ru, Y., Qiu, X., Tan, X., Gao, Y., and Jin, Y. (2022). Sparse-attentive meta temporal point process for clinical decision support. *Neurocomputing* 485 (1), 114–123. doi:10.1016/j.neucom.2022.02.028

Rueckauer, B., and Delbruck, T. (2016). Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor. *Front. Neurosci.* 10, 176–218. doi:10.3389/fnins.2016.00176

Rui, I., and Chi, M. (2018). Asynchronous corner detection and tracking for event cameras in real time. *IEEE Robotics Automation Lett.* 3 (4), 3177–3184. doi:10.1109/lra.2018.2849882

Russo, N., Huang, H., Donati, E., Madsen, T., and Nikolic, (2022). An interface platform for robotic neuromorphic systems. *Chips* 2 (1), 20–30. doi:10.3390/chips2010002

Ryan, C., Elrasad, A., Shariff, W., Lemley, J., Kielty, P., Hurney, P., et al. (2023). Real-time multi-task facial analytics with event cameras. *IEEE Access* 11, 76964–76976. doi:10.1109/access.2023.3297500

Ryan, C., Sullivan, B., Elrasad, A., Cahill, A., Lemley, J., Kielty, P., et al. (2021). Real-time face & eye tracking and blink detection using event cameras. *Neural Netw.* 141, 87–97. doi:10.1016/j.neunet.2021.03.019

Sabatier, Q., Ieng, S., and Benosman, R. (2017). Asynchronous event-based Fourier analysis. *IEEE Trans. Image Process.* 26 (5), 2192–2202. doi:10.1109/tip.2017.2661702

Sadaf, M., Sakib, N., Pannone, A., Ravichandran, H., and Das, S. (2023). A bio-inspired visuotactile neuron for multisensory integration. *Nat. Commun.* 14 (5729), 5729–5813. doi:10.1038/s41467-023-40686-z

Sajwani, H., Ayyad, A., Alkendi, Y., Halwani, M., Abdulrahman, Y., Abusafieh, A., et al. (2023). TactiGraph: an asynchronous graph neural network for contact angle prediction using neuromorphic vision-based tactile sensing. *Sensing* 23 (14), 6451–6518. doi:10.3390/s23146451

Saunders, D., Patel, D., Hazan, H., Siegelmann, H. T., and Kozma, R. (2019). Locally connected spiking neural networks for unsupervised feature learning. *Neural Netw.* 119, 332–340. doi:10.1016/j.neunet.2019.08.016

Sawant, A., Saha, A., Khoussine, J., Sinha, R., and Hoon, M. (2023). New insights into retinal circuits through EM connectomics: what we have learnt and what remains to be learned. *Annu. Rev. Neurosci.* 3, 1–13. doi:10.3389/fopht.2023.1168548

Scheerlinck, C., Barnes, N., and Mahony, R. (2019). Asynchronous spatial image convolutions for event cameras. *IEEE Robotics Automation Lett.* 4 (2), 816–822. doi:10.1109/lra.2019.2893427

Schiopu, I., and Bilcu, R. (2023). Memory-efficient fixed-length representation of synchronous event frames for very-low-power chip integration. *Electronics* 12 (10), 2302–2318. doi:10.3390/electronics12102302

Schomaker, L., Timmermans, J., and Banerjee, T. (2023). Non-linear adaptive control inspired by neuromuscular systems. *Bioinspiration Biomimetics* 18, 375–387. doi:10.1088/1748-3190/acd896

Sekikawa, Y., Ishikawa, K., and Saito, H. (2018). Constant velocity 3D convolution. *IEEE Access* 6, 76490–76501. doi:10.1109/access.2018.2883340

Shair, Z., and Rawashdeh, S. (2022). High-temporal-resolution object detection and tracking using images and events. *J. Imaging* 8 (8), 1–17. doi:10.3390/jimaging8080210

Sharif, S., Naqvi, R., Ali, F., and Biswas, (2023). DarkDeblur: learning single-shot image deblurring in low-light condition. *Expert Syst. Appl.* 222 (2), 119739–119754. doi:10.1016/j.eswa.2023.119739

Shatnawi, M., Khennaoui, A., Ouannas, A., Grassi, G., Radogna, A., Bataihah, A., et al. (2023). A multistable discrete memristor and its application to discrete-time FitzHugh-Nagumo model. *Electronics* 12 (13), 2929–3019. doi:10.3390/electronics12132929

Shen, G., Zhao, D., and Zeng, Y. (2022). Backpropagation with biologically plausible spatiotemporal adjustment for training deep spiking neural networks. *Patterns* 3 (6), 100522–102289. doi:10.1016/j.patter.2022.100522

Shen, G., Zhao, D., and Zeng, Y. (2023). EventMix: an efficient data augmentation strategy for event-based learning. *Inf. Sci.* 644 (3), 119170–120158. doi:10.1016/j.ins.2023.119170

Shi, C., Li, J., Wang, Y., and Luo, G. (2018). Exploiting lightweight statistical learning for event-based vision processing. *IEEE Access* 6, 19396–19406. doi:10.1109/access.2018.2823260

Shiba, S., Aoki, Y., and Gallego, G. (2022a). Fast event-based optical flow estimation by triplet matching. *IEEE Signal Process. Lett.* 29, 2712–2716. doi:10.1109/lsp.2023.3234800

Shiba, S., Aoki, Y., and Gallego, Go. (2022b). Event collapse in contrast maximization frameworks. *Sensors* 22 (14), 5190–5216. doi:10.3390/s22145190

Shu, J., Yang, T., Liao, X., Chen, F., Xiao, Y., Yang, K., et al. (2022). Clustered federated multitask learning on Non-IID data with enhanced privacy. *IEEE Internet Things J.* 10 (4), 3453–3467. doi:10.1109/jiot.2022.3228893

Singh, K., Kapoor, D., Thakur, K., Sharma, A., and Gao, X. (2022b). Computer-vision based object detection and recognition for service robot in indoor environment. *Comput. Mater. Continua* 72 (1), 197–213. doi:10.32604/cmc.2022.022989

Singh, N., Rathore, S., and Kumar, (2022a). Towards a super-resolution based approach for improved face recognition in low resolution environment. *Multimedia Tools Appl.* 81, 38887–38919. doi:10.1007/s11042-022-13160-z

Song, B., Gao, R., Wang, Y., and Yu, G. (2023). Enhanced LDR detail rendering for HDR fusion by TransU-Fusion network. *Symmetry* 15 (7), 1463–1515. doi:10.3390/sym15071463

Soucy, J., Aguzzi, E., Cho, J., Gilhooley, M., Keuthan, C., Luo, Z., et al. (2023). Retinal ganglion cell repopulation for vision restoration in optic neuropathy: a roadmap from the RReSTORe consortium. *Mol. Neurodegener.* 18 (64), 64–151. doi:10.1186/s13024-023-00655-y

Steffen, L., Reichard, D., Weinland, J., Kaiser, J., Roennau, A., and Dillmann, R. (2019). Neuromorphic stereo vision: a survey of bio-Inspired sensors and algorithms. *Front. Neurobotics* 13, 28–19. doi:10.3389/fnbot.2019.00028

Sun, Y., Zeng, Y., and Li, Y. (2022). Solving the spike feature information vanishing problem in spiking deep Q network with potential based normalization. *Front. Neurosci.* 16, 953368–953415. doi:10.3389/fnins.2022.953368

Tang, L., Ma, S., Ma, X., and You, H. (2022). Research on image matching of improved SIFT algorithm based on stability factor and feature descriptor simplification. *Appl. Sci.* 12 (17), 8448–8517. doi:10.3390/app12178448

Tang, S., Lv, H., Zhao, Y., Feng, Y., Liu, H., and Bi, G. (2023). Denoising method based on salient region recognition for the spatiotemporal event stream. *Sensors* 23 (15), 6655–6718. doi:10.3390/s23156655

Tavanaei, A., Ghodrati, M., Kheradpishep, S., Masquelier, T., and Maida, A. (2019). Deep learning in spiking neural networks. *Neural Netw.* 111, 47–63. doi:10.1016/j.neunet.2018.12.002

Tchantchane, R., Zhou, H., Zhang, S., and Alici, G. (2023). A review of hand gesture recognition systems based on noninvasive wearable sensors. *Adv. Intell. Syst.* 8 (1), 142–153. doi:10.1002/aisy.202300207

Toprak, S., Guerro, N., and Wermter, S. (2018). Evaluating integration strategies for visuo-haptic object recognition. *Cogn. Comput.* 10, 408–425. doi:10.1007/s12559-017-9536-7

Uddin, S., Ahmed, S., and Jung, Y. (2022). Unsupervised deep event stereo for depth estimation. *IEEE Trans. Circuits Syst. Video Technol.* 32 (11), 7489–7504. doi:10.1109/tcsvt.2022.3189480

Valeiras, D., Clady, X., Ieng, S. H., and Benosman, R. (2018). Event-based line fitting and segment detection using a neuromorphic visual sensor. *IEEE Trans. Neural Netw. Learn. Syst.* 30 (4), 1218–1230. doi:10.1109/TNNLS.2018.2807983

Valerdi, J., Bartolozzi, C., and Glover, A. (2023). Insights into batch selection for event-camera motion estimation. *Sensors* 23 (7), 3699–3719. doi:10.3390/s23073699

Vanarse, A., Osseiran, A., and Rassau, A. (2019). Neuromorphic engineering — a paradigm shift for future IM technologies. *IEEE Instrum. Meas. Mag.* 22 (2), 4–9. doi:10.1109/mim.2019.8674627

Vidal, A., Rebecq, H., Horstschaefer, T., and Scaramuzza, D. (2018). Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios. *IEEE Robotics Automation Lett.* 3 (2), 994–1001. doi:10.1109/lra.2018.2793357

Wang, T., Wang, X., Wen, J., Shao, Z., Huang, H., and Guo, X. (2022a). A Bio-inspired neuromorphic sensory system. *Adv. Intell. Syst.* 4 (7), 3611–3624. doi:10.1002/aisy.202200047

Wang, X., Zhong, M., Cheng, H., Xie, J., Zhou, Y., Ren, J., et al. (2022b). SpikeGoogle: spiking neural networks with GoogLeNet-like inception module. *CAAI Trans. Intell. Technol.* 7 (3), 492–502. doi:10.1049/cit2.12082

Wang, Y., Shao, B., Zhang, C., Zhao, J., and Cai, Z. (2022e). REVIO: range-and event-based visual-inertial bio-inspired sensors. *Biomimetics* 7 (4), 1–16. doi:10.3390/biomimetics7040169

Wang, Y., Yang, J., Peng, X., Wu, P., Gao, L., Huang, K., et al. (2022d). Visual odometry with an event camera using continuous ray wrapping and volumetric contrast maximization. *Sensors* 22 (15), 1–17. doi:10.3390/s22155687

Wang, Y., Yue, C., and Tang, X. (2021). A geometry feature aggregation method for point cloud classification and segmentation. *IEEE Access* 9, 140504–140511. doi:10.1109/access.2021.3119622

Wang, Z., Dai, Y., and Mao, Y. (2020). Learning dense and continuous optical flow from an event camera. *IEEE Trans. Image Process.* 31, 7237–7251. doi:10.1109/tip.2022.3220938

Wang, Z., Ren, J., Zhang, J., and Luo, P. (2022c). Image deblurring aided by low-resolution events. *Electronics* 11 (4), 631–716. doi:10.3390/electronics11040631

Wu, K., Mina, M., Sahyoun, J., Kalevar, A., and Tran, S. (2023). Retinal prostheses: engineering and clinical perspectives for vision restoration. *Sensors* 23 (13), 5782–5815. doi:10.3390/s23135782

Wu, Y., and Ji, Q. (2019). Facial landmark detection: a literature survey. *Int. J. Comput. Vis.* 127 (2), 115–142. doi:10.1007/s11263-018-1097-z

Wu, Y., Sheng, H., Zhang, Y., Wang, S., Xiong, Z., and Ke, W. (2022). Hybrid motion model for multiple object tracking in mobile devices. *IEEE Internet Things J.* 10 (6), 4735–4748. doi:10.1109/jiot.2022.3219627

Wu, Y., Tang, F., and Li, H. (2018). Image-based camera localization: an overview. *Vis. Comput. Industry, Biomed. Art* 1 (8), 8–324. doi:10.1186/s42492-018-0008-z

Wu, Z., and Guo, A. (2023). Bioinspired figure-ground discrimination via visual motion smoothing. *Plos Comput. Biol.* 19 (4), e1011077–e1011088. doi:10.1371/journal.pcbi.1011077

Wunderlich, T., and Pehle, (2021). Event-based back propagation can compute exact gradients for spiking neural networks. *Sci. Rep.* 11 (12829), 12829–12918. doi:10.1038/s41598-021-91786-z

Xia, R., Zhao, C., Sun, Q., Cao, S., and Tang, Y. (2023). Modality translation and fusion for event-based semantic segmentation. *Control Eng. Pract.* 136, 105530–105541. doi:10.1016/j.conengprac.2023.105530

Xiang, K., Yang, K., and Wang, K. (2022). Polarization-driven semantic segmentation via efficient attention-bridged fusion. *Opt. Express* 29 (4), 4802–4820. doi:10.1364/oe.416130

Xiao, S., Yan, J., Farajtabar, M., Song, L., Yang, X., and Zha, H. (2019). Learning time series associated event sequences with recurrent point process networks. *IEEE Trans. Neural Netw. Learn. Syst.* 30 (10), 3124–3136. doi:10.1109/tnnls.2018.2889776

Xie, Z., Zhang, J., and Wang, P. (2018). Event-based stereo matching using semiglobal matching. *Int. J. Adv. Robotic Syst.* 15 (1), 172988141775275–15. doi:10.1177/1729881417752759

Xu, C., Liu, Y., Chen, D., and Yang, Y. (2022a). Direct training via back propagation for ultra-low-latency spiking neural networks with multi-threshold. *Symmetry* 14 (9), 1933–17. doi:10.3390/sym14091933

Xu, C., Liu, Y., and Yang, Y. (2023). Ultra-low latency spiking neural networks with spatio-temporal compression and synaptic convolutional block. *Neurocomputing* 550 (3), 126485–493. doi:10.1016/j.neucom.2023.126485

Xu, J., Miao, J., Gao, Z., Nie, K., and Shi, X. (2020). Analysis and modeling of quntization error in spike-frequency-based image sensor. *Microelectron. Reliab.* 111 (3), 3705–3716. doi:10.1016/j.microrel.2020.113705

Xu, J., Park, S., Zhang, X., and Hu, J. (2022c). The improvement of road driving safety guided by visual inattentional blindness. *IEEE Trans. Intelligent Transp. Syst.* 23 (6), 4972–4981. doi:10.1109/TITS.2020.3044927

Xu, J., Zhang, X., Park, S., and Guo, K. (2022b). The alleviation of perceptual blindness during driving in urban areas guided by saccades recommendation. *IEEE Trans. Intelligent Transp. Syst.* 23 (9), 16386–16396. doi:10.1109/TITS.2022.3149994

Yan, Z., and Zha, H. (2019). Flow-based SLAM: from geometry computation to learning. *Virtual Real. Intelligent Hardw.* 1 (5), 435–460. doi:10.1016/j.vrih.2019.09.001

Yang, F., Zhao, J., Chen, X., Wang, X., Jiang, N., Hu, Q., et al. (2023a). SA-FlowNet: event-based self-attention optical flow estimation with spiking-analogue neural networks. *IET Comput. Vis.* 5 (3), 576–584. doi:10.1049/cvi2.12206

Yang, H., Huang, H., Liu, X., Li, Z., Li, J., Zhang, D., et al. (2023b). Sensing mechanism of an Au-TiO2-Ag nanograting based on Fano resonance effects. *Appl. Opt.* 62 (17), 4431–4438. doi:10.1364/AO.491732

Yang, L., Dong, Y., Zhao, D., and Zeng, Y. (2022). N-omniglot, a large-scale neuromorphic dataset for spatio-temporal sparse few-shot learning. *Sci. Data* 9, 351–363. doi:10.1038/s41597-022-01851-z

Ye, F., Kiani, F., Huang, Y., and Xia, Q. (2023). Diffusive memristors with uniform and tunable relaxation time for spike generation in event-based pattern recognition. *Adv. Mater.* 35 (37), e2204778–1763. doi:10.1002/adma.202204778

Yilmaz, O., Chane, C., and Histace, A. (2021). Evaluation of event-based corner detectors. *J. Imaging* 7 (2), 1–15.

Younsi, M., Diaf, M., and Siarry, P. (2023). Comparative study of orthogonal moments for human postures recognition. *Eng. Appl. Artif. Intell.* 120 (3), 105855–867. doi:10.1016/j.engappai.2023.105855

Yu, Z., Liu, J., Jia, S., Zhang, Y., Zheng, Y., Tian, Y., et al. (2020). Toward the next generation of retinal neuroprosthesis: visual computation with spikes. *Engineering* 6 (4), 449–461. doi:10.1016/j.eng.2020.02.004

Zare, A., Homayouni, M., Aminlou, A., Hannuksela, M., and Gabbouj, M. (2019). 6K and 8K effective resolution with 4K HEVC decoding capability for 360 video streaming. *ACM Trans. Multimedia Comput. Commun. Appl.* 15 (2), 661–22. doi:10.1145/3335053

Zeng, C., Wang, W., Nguyen, A., Xiao, J., and Yue, Y. (2023). Self-supervised learning for point cloud data: a survey. *Expert Syst. Appl.* 237 (2), 121354–1367. doi:10.1016/j.eswa.2023.121354

Zhang, C., Xiao, P., Zhao, Z., Liu, Z., Yu, J., Hu, X., et al. (2023f). A wearable localized surface plasmons antenna sensor for communication and sweat sensing. *IEEE Sensors J.* 23 (11), 11591–11599. doi:10.1109/JSEN.2023.3266262

Zhang, H., Luo, G., Li, J., and Wang, F. (2022e). C2FDA: coarse-to-fine domain adaptation for traffic object detection. *IEEE Trans. Intelligent Transp. Syst.* 23 (8), 12633–12647. doi:10.1109/TITS.2021.3115823

Zhang, K., Zhao, Y., Chu, Z., and Zhou, Y. (2022d). Event-based vision in magneto-optic Kerr effect microscopy. *AIP Adv.* 12, 1–16. doi:10.1063/5.0090714

Zhang, S., Huangfu, L., Zhang, Z., Huang, S., Li, P., and Wang, H. (2020). Corner detection using the point-to-centroid distance technique. *IET Image Process.* 14 (14), 3385–3392. doi:10.1049/iet-ipr.2020.0164

Zhang, S., Wang, W., Li, H., and Zhang, S. (2022b). EVTracker: an event-driven spatiotemporal method for dynamic object tracking. *Sensors* 22 (16), 6090–18. doi:10.3390/s22166090

Zhang, S., Wang, W., Li, H., and Zhang, S. (2023d). E-Detector: asynchronous spatio-temporal for event-based object detection in intelligent transportation system. *ACM Trans. Multimedia Comput. Commun. Appl.* 20 (2), 1–20. doi:10.1145/3584361

Zhang, S., Wei, Z., Xu, W., Zhang, L., Wang, Y., Zhou, X., et al. (2023e). DSC-MVSNet: attention aware cost volume regularization based on depthwise separable convolution for multi-view stereo. *Complex & Intelligent Syst.* 8, 6953–6969. doi:10.1007/s40747-023-01106-3

Zhang, Y., Li, W., and Yang, P. (2021). Surveillance video motion segmentation based on the progressive spatio-temporal tunnel flow model. *Electron. Lett.* 57 (13), 505–507. doi:10.1049/ell2.12186

Zhang, Y., Lv, H., Zhao, Y., Feng, Y., Liu, H., and Bi, G. (2023c). Event-based optical flow estimation with spatio-temporal backpropagation trained spiking neural network. *Micromachines* 14 (1), 203–14. doi:10.3390/mi14010203

Zhang, Y., Peng, L., Chang, Y., Huang, K., and Chen, G. (2022c). An ultra-high-speed hardware accelerator for image reconstruction and stereo rectification on event-based camera. *Microelectron. J.* 119 (3), 813–824. doi:10.1016/j.mejo.2021.105312

Zhang, Y., Xiang, S., Han, Ya., Guo, X., Zhang, Y., Shi, Y., et al. (2023a). Supervised learning and pattern recognition in photonic spiking neural networks based on MRR and phase-change materials. *Opt. Commun.* 549 (3), 129870–883. doi:10.1016/j.optcom.2023.129870

Zhang, Y., Zhao, Y., Lv, H., Feng, Y., Liu, H., and Han, C. (2022a). Adaptive slicing method of the spatiotemporal event stream obtained from a dynamic vision sensor. *Sensors* 22 (7), 2614–19. doi:10.3390/s22072614

Zhang, Z., Chai, K., Yu, H., Majaj, R., Walsh, R., Wang, E., et al. (2023b). Neuromorphic high-frequency 3D dancing pose estimation in dynamic environment. *Neurocomputing* 547 (2), 126388–397. doi:10.1016/j.neucom.2023.126388

Zhao, B., Huang, Y., Ci, W., and Hu, X. (2022a). Unsupervised learning of monocular depth and ego-motion with optical flow features and multiple constraints. *Sensors* 22 (4), 1383–16. doi:10.3390/s22041383

Zhao, D., Zeng, Y., and Li, Y. (2022b). BackEISNN: a deep spiking neural network with adaptive self-feedback and balanced excitatory – inhibitory neurons. *Neural Netw.* 154 (2), 68–77. doi:10.1016/j.neunet.2022.06.036

Zhao, J., Ji, S., Cai, Z., Zeng, Y., and Wang, Y. (2022c). Moving object detection and tracking by event frame from neuromorphic vision sensors. *Biomimetics* 7 (1), 31–15. doi:10.3390/biomimetics7010031

Zhao, J., Su, L., Wang, X., Li, J., Yang, F., Jiang, N., et al. (2023a). DTFS-eHarris: a high accuracy asynchronous corner detector for event cameras in complex scenes. *Appl. Sci.* 13 (9), 5761–14. doi:10.3390/app13095761

Zhao, K., Jia, Z., Jia, F., and Shao, H. (2023b). Multi-scale integrated deep self-attention network for predicting remaining useful life of aero-engine. *Eng. Appl. Artif. Intell.* 120, 105860–621. doi:10.1016/j.engappai.2023.105860

Zhao, Q., Liu, J., Yang, H., Liu, H., Zeng, G., and Huang, B. (2022d). High birefringence D-shaped germanium-doped photonic crystal fiber sensor. *Micromachines* 13 (6), 826–722. doi:10.3390/mi13060826

Zheng, J., Dong, L., Wang, F., and Zhang, Z. (2023). Semantic plane-structure based motion detection with a nonstationary camera. *Displays* 79, 102454–463. doi:10.1016/j.displa.2023.102454

Zhou, F., Zhou, Z., Chen, J., Choy, T. H., Wang, J., Zhang, N., et al. (2019). Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol.* 14 (8), 776–782. doi:10.1038/s41565-019-0501-3

Zhou, X., and Zhang, L. (2022). SA-FPN: an effective feature pyramid network for crowded human detection. *Appl. Intell.* 52 (11), 12556–12568. doi:10.1007/s10489-021-03121-8

Zhou, Z., Yu, X., and Chen, X. (2023). Object detection in drone video with temporal attention gated recurrent unit based on transformer. *Drones* 7 (7), 466–20. doi:10.3390/drones7070466

Zhu, A., Thakur, D., Ozaslan, T., Pfrommer, B., Kumar, V., and Daniilidis, K. (2018). The multivehicle stereo event camera dataset: an event camera dataset for 3D perception. *IEEE Robotics Automation Lett.* 3 (3), 2032–2039. doi:10.1109/lra.2018.2800793

Zhu, N., Zhao, G., Zhang, X., and Jin, Z. (2021b). Falling motion detection algorithm based on deep learning. *IET Image Process.* 16 (11), 2845–2853. doi:10.1049/ipr2.12208

Zhu, R., Hochstetter, J., Loeffler, A., Alvarex, A., Nakayama, T., Lizier, J. T., et al. (2021a). Information dynamics in neuromorphic nanowire networks. *Sci. Rep.* 11 (13047), 13047–15. doi:10.1038/s41598-021-92170-7

Zhu, S., Ding, R., Zhang, M., Hentenryck, P., and Xie, Y. (2022b). Spatio-temporal point processes with attention for traffic congestion event modeling. *IEEE Trans. Intelligent Transp. Syst.* 23 (7), 7298–7309. doi:10.1109/tits.2021.3068139

Zhu, Z., Li, M., Xie, Y., Zhou, F., Liu, Y., and Wang, W. (2022a). The optimal projection intensities determination strategy for robust strip-edge detection in adaptive fringe pattern measurement. *Optik* 257, 168771–879. doi:10.1016/j.ijleo.2022.168771

Zhuo, Z., Du, L., Lu, X., Chen, J., and Cao, Z. (2022). Smoothed Lv distribution based three-dimensional imaging for spinning space debris. *IEEE Trans. Geoscience Remote Sens.* 60, 1–13. doi:10.1109/TGRS.2022.3174677

Zihao, Ji., Hu, W., Wang, Z., Yang, K., and Wang, K. (2021). Seeing through events: real-time moving object sonification for visually impaired people using event-based camera. *Sensors* 21 (10), 1–17. doi:10.3390/s21103558