Advances in Technology Innovation, vol. 10, no. 1, 2025, pp. 44-57

A Review of Advances in Bio-Inspired Visual Models Using Event-and Frame-Based Sensors

Aya Zuhair Salim*, Luma Issa Abdul-Kareem

Department of Control and Systems Engineering, University of Technology, Baghdad, Iraq

Received 16 August 2024; received in revised form 29 October 2024; accepted 01 November 2024

DOI: https://doi.org/10.46604/aiti.2024.14121

Abstract

This paper reviews visual system models using event- and frame-based vision sensors. The event-based sensors mimic the retina by recording data only in response to changes in the visual field, thereby optimizing real-time processing and reducing redundancy. In contrast, frame-based sensors capture duplicate data, requiring more processing resources. This research develops a hybrid model that combines both sensor types to enhance efficiency and reduce latency. Through simulations and experiments, this approach addresses limitations in data integration and speed, offering improvements over existing methods. State-of-the-art systems are highlighted, particularly in sensor fusion and real-time processing, where dynamic vision sensor (DVS) technology demonstrates significant potential. The study also discusses current limitations, such as latency and integration challenges, and explores potential solutions that integrate biological and computer vision approaches to improve scene perception. These findings have important implications for vision systems, especially in robotics and autonomous applications that demand real-time processing.

Keywords: bio-inspired model, computer vision, event-based sensor, optic flow, dynamic vision sensor (DVS)

1. Introduction

Technological advancements in robotics and artificial intelligence have propelled the development of computer vision models that emulate human visual capabilities, with DVS and frame-based vision sensors playing critical roles in this advancement. This paper aims to delve into the evolution of computer vision models inspired by visual systems, using these sensors to improve computer systems' capacity to efficiently and effectively perceive and understand their surroundings. Previous research in this sector has highlighted the need to combine continuous data streams from DVS sensors with frame-based data to improve precision and efficiency in computer vision [1-2].

Building a bio-inspired model for visual perception remains a challenge; many researchers have developed models inspired by the dorsal pathway, which is concerned with identifying an object in visual space, and is also called the motion pathway; on the other hand, the ventral pathway, which is focused on form, color, texture, has been studied by other researchers and is referred to as the form pathway. The question here is how data can be collected from two types of sensors, event-based and frame-based sensors, to simulate the visual perception of both pathways by processing the motion data and the form data for an object in the visual scene. Building a bio-inspired model for visual perception remains a challenge. Many researchers have developed models inspired by the dorsal pathway, which identifies where an object is in visual space, also known as the motion pathway. On the other hand, the ventral pathway, which is concerned with form, color, and texture, is the focus of other researchers and is referred to as the form pathway. Merging motion and form data to identify objects in a visual scene could unlock new possibilities for real-world applications, including industrial automation, autonomous robotics, augmented reality,

* Corresponding author. E-mail address: cse.22.06@grad.uotechnology.edu.iq

English language proofreader: Si-Yu Lin

and virtual reality [3]. This approach aims to create more effective and physiologically accurate visual processing. The range of vision tasks has expanded due to recent advances in deep learning approaches for optical flow estimation and reinforcement learning, but incorporating DVS into these models remains a challenge. To complement these gaps, this study aims to highlight the main shortcomings in the existing research while offering a thorough overview of the state-of-the-art in frame-based sensor integration and DVS. It also seeks to provide insights into how bio-inspired systems imitate the processing powers of the brain and can simplify processing while lowering data redundancy.

Deep learning in optical flow estimation applications includes high-level vision tasks such as recognition, reconstruction, and segmentation, as well as low-level vision tasks including feature detection, tracking, and optical flow. This technique provides two variants of ultrasound flow (US flow). One variation is a PWC-like model based on a modification of pyramid, warping, and cost volume (PWC-Net). It incorporates two spike streams that pass through a common dynamic timing representation module before routing to existing backbones in an end-to-end architecture [4].

PWC is an optical flow estimation technique utilizing images at multiple resolutions to capture details, aligning them across different scales, and employing a cost volume to represent pixel-level differences between frames for accurate motion estimation. PWC-Net, on the other hand, enhances this technique by integrating neural networks, significantly improving the accuracy and efficiency of motion estimation at multiple levels of detail.

Computational visual processing mechanisms are categorized into bio-inspired and computer vision models, which focus on tasks such as segmentation and edge detection [5]. For the purpose of enhancing accuracy and efficiency in visual data analysis. Deep reinforcement learning techniques have also been applied to further optimize these processes and achieve higher performance.

This paper is organized as follows. Section 2 covers visual perception and bio-inspired models, focusing on biological systems that inspire computational models. Sections 3 and 4 delve into the evolution and key techniques of computer vision models, respectively. Section 5 reviews visual sensors, detailing both frame-based and event-based sensors, such as DVS. In Section 6, the discussion addresses depth estimation using sensors, exploring how sensors are used to perceive depth. Section 7 presents the data-driven tools employed in the field. Section 8 highlights applications of visual system models across various industries. Finally, Section 9 summarizes the findings and implications of the study, leading to the conclusions presented in Section 10.

2. Visual Perception and Bio-Inspired Models

Examining biological vision systems, particularly the human retinal system, is essential before discussing bio-inspired vision sensors. The human retinal system comprises bipolar cells, ganglion cells, and photoreceptors. These convert light into electrical pulses, which travel through ganglion and bipolar cells before entering optical fibers, which transmit them to the brain for interpretation of these signals. Fig. 1 presents a schematic depiction of the retina network [6].

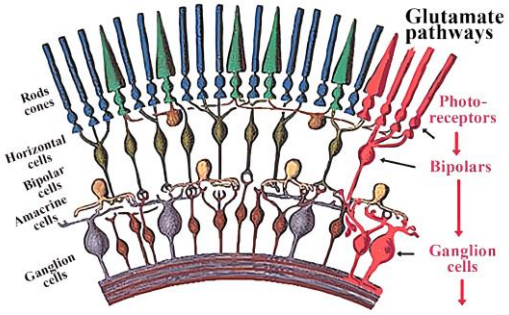


Fig. 1 (Colour online) Retinal network schematic illustration [6]

Ganglion cells, specifically X-cells and Y-cells, play a crucial role in this process. X-cells are located in the parvocellular pathway, which accounts for 80% of nerve fibers and conveys patterns, spatial features, and color information. This system, known as the biological "what" system, is complemented by the magnocellular pathway, which processes changes related to movement, distance, and speed, commonly referred to as the "where" system [6-7].

Concepts from the biological visual system are employed in artificial intelligence models for visual perception and deep learning [8]. These models enhance computational performance, enable rapid response times, maintain high computing efficiency, and support discrimination among elements in images and data [9-10].

DVS operates at a human retina level, capturing scene changes using the event-based sensing method. Such selectivity can be observed in ganglion cells where only movement evokes action, reducing data load and power consumption [11]. This development ensures that AI models are highly efficient in low-latency, low-power applications usage including robotic systems and autonomous vehicles.

Moreover, this represents one of the most efficient temporal pathways for human vision, capable of handling complex visual tasks. Foveal covert receptors located in the rostral pole of the optic tract are triggered by movement and transient ganglion cells, a concept mirrored by DVS, which processes only field changes to improve operational speed in devices such as surveillance and self-driving cars [12-13]. This closely links DVS to biological visual systems, bridging the gap between visual biology and sensor technology [14-15].

In humans, the primary visual cortex (V1) is a foremost component of the visualization process. It analyzes eye signals, converts them into understandable formats, and contributes to binocular interaction. For the correct appreciation of the surroundings, and making agreeable decisions, it is necessary to utilize motion, contours, and borders of visual stimuli, which are the functions of V1 in the brain [8,14,16-17].

Bio-inspired vision systems are expected to improve intelligent structures and self-learning mechanisms while decreasing energy consumption in areas like space imaging, robotics, surveillance, and object identification [18].

In the case of capturing real motion on event-based or frame-based sensors, DVS processes only a small portion of the visual movement taking place, thereby lowering power requirements and processing time. This results in superior performance in high-motion scenarios such as self-driving cars and high-speed surveillance. DVS also improves security and precision via real-time visual processing. According to Gallego et al. (2020), DVS outperforms conventional systems in estimating motion in 2D space via the moving objects and images, enabling bio-inspired systems in practical cases where traditional systems often face high delay and processing power requirements.

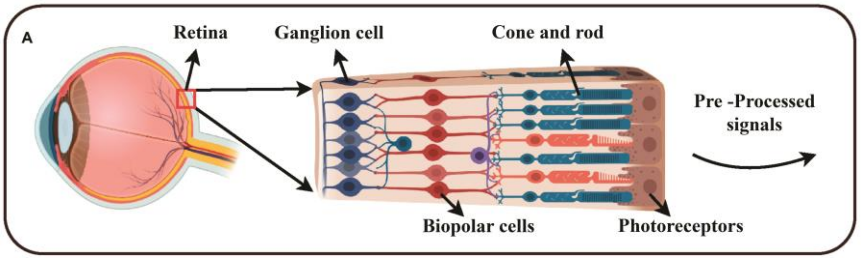
3. Computer Vision

For many years, computer vision has been a prominent research topic, with applications ranging from photogrammetry to medical imaging, car safety, machine inspection, and beyond [3, 19-20].

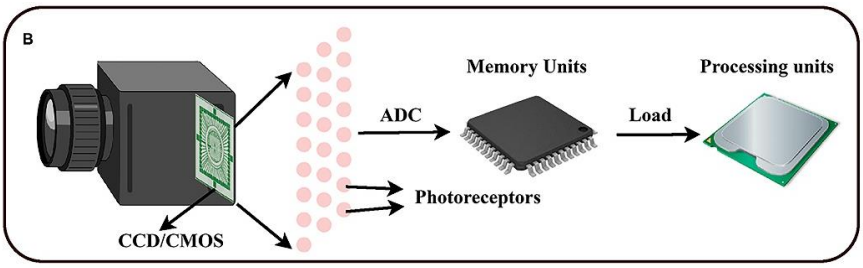
Computer vision increases the accuracy of human action detection by analyzing data from dynamic vision devices and extracting morphological and kinematic information. This provides a useful and effective way of evaluating dynamic observational data [3, 21-23].

Although deep neural networks (DNNs) have revolutionized computer vision, debates continue about their role in vision research. DNNs need to exhibit robust object identification capabilities and recognize objects under changing 3D viewing angles and distortions. Challenges such as manipulated adversarial examples underscore the differences between DNNs and human vision perception, rendering them promising yet insufficient representations vision [8-20].

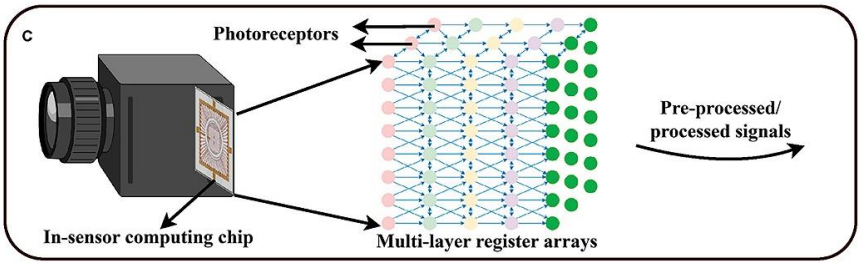
This study demonstrates the proof of concept (POC) of reducing energy consumption by utilizing central processing units (CPUs), graphical processing units (GPUs), video processing units (VPUs), and digital signal processors (DSPs) to save power and mitigate carbon emissions in technology. Techniques from sensor visual perception studies demonstrate non-convolutional methodologies for understanding sensory input, and integrating perceptrons with memory [24]. Fig. 2 illustrates three types of systems performing the same task.



(a) The retina mechanism



(b) The convolution visual system



(c) Hardware design for vision processing

Fig. 2 Systems and mechanisms of visual processing [24]

4. Computer Vision Models

This technique analyzes and tracks the movement of objects in photos and videos, and is considered an important tool in fields such as intelligent robotics, motion analysis [25], and automatic image correction enhancement [11].

Optic flow refers to the global motion during movement, characterized by stimuli with complex speed gradients. Its speed increases with the viewing angle, rendering greater velocities more significant. Neurophysiological data indicates that selectivity for ocular flow originates in area medial superior temporal (MST), influenced by area V5/middle temporal (MT). Imaging and lesion studies reveal a broad cortical network involved in optic flow perception. Age-related effects on motion processing and signal complexity may lead to significant perceptual decline. Limited research provides minimal evidence of perceptual decline, with heading detection thresholds ranging from 1.18 in young to 1.98 in older individuals. Lich and Bremmer's study reveals that older individuals exhibit lower accuracy in determining heading direction using a reference ruler, emphasizing the need for more detailed measures [26-27].

Optical flow estimation presents a challenge in computer vision, as it requires regularization to estimate the object's velocity without prior knowledge of the scene's geometry or motion. Event-based optical flow estimation is challenging due to the novel way visual information is presented as events. Traditional cameras determine optical flow by comparing two

consecutive pictures. Over the past two decades, methods have improved, with new concepts developed to overcome existing limitations. The quadratic regularization in the Horn and Schunck model was replaced by smoothness constraints, facilitating the development of more advanced computer-based applications [13-14, 28-31].

Researchers have developed state-of-the-art deep learning models to address performance challenges in object recognition within dynamic video data. Applications of video analysis and tracking with advanced sensors, include video monitoring, movement detection, optical flow analysis, RGB-D imaging devices, infrared (IR) detectors, radar systems, light detection and ranging (LIDAR) technology, digital photography, augmented reality (AR), and security monitoring. The DEEP-SEE framework employs computer vision and deep convolutional neural networks for real-time object detection, tracking, and recognition in outdoor navigation. The framework identifies both static and moving objects without prior knowledge, employing motion-based monitoring and visual similarity to predict object placements. The DEEP-SEE framework is utilized in an innovative assistive device designed to enhance cognitive abilities and safety for visually impaired individuals navigating urban environments. Validation with a 30-element video dataset demonstrates its high accuracy and robustness [32].

This study focuses on object detection, identification, and classification. Experimental findings indicate that an improved version of the you only look once (YOLO) model surpasses traditional algorithms in categorizing distant scenes with moving objects, achieving a 98.94% accuracy on both public and custom datasets. The YOLO algorithm is employed for various tasks, including classification training pre-trained on ImageNet. Standard data augmentation techniques are applied, with 160 training epochs, a 224 x 224 input image size, and an initial learning rate of 0.1 [33-34].

Furthermore, 33 convolution layers of 11 layers and 1024 filters are added after the final convolution layer for detection tasks. In YOLOv2 and similar models, category probabilities are associated with each box instead of a grid, contributing to increased inference time. This evolution culminates in the development of YOLOv10 [35-36].

Modern vision system models employ an attention mechanism [37]. The authors introduce a "Transformer" network design that relies exclusively on attention mechanisms rather than traditional complex neural networks. This approach surpasses existing sequence-to-sequence models in terms of quality, parallelizability, and training efficiency. Initially applied to natural language understanding (NLU) tasks like machine translation [38], this mechanism was later adapted for visual system tasks [39], including models such as vision transformer (ViT) and data-efficient image transformer (Deit) [40-42].

5. Visual Sensors

Visual sensors are electronic devices applied across various fields, including robotics, surveillance, automotive systems, industrial automation, and healthcare. They detect and record visual data employing technologies, such as charge-coupled devices (CCDs), complementary metal-oxide-semiconductor (CMOS) sensors (see, [43-44] for example), and DVS (see, [7-13] for example). These sensors convert optical signals into electrical signals for computational processing. The data is subsequently analyzed through computer vision and image processing techniques [13, 45-46]. Technological advancements have enabled the development of sophisticated sensors capable of real-time high-resolution photo and film collection. Some notable examples of such sensors include:

5.1. Event-based sensor

Event cameras, also referred to as data-driven sensors operate based on the principle that their reaction is determined by the amount of motion or change in scene brightness they perceive. Each pixel in these cameras adjusts its delta modulator sampling rate based on the rate of change of the log intensity signal it monitors. Consequently, faster motion leads to the creation of more events per second. These sensors respond swiftly to visual stimuli, sending events with sub-millisecond latency and timestamped to microsecond precision [3, 12-13, 23-52].

The DVS represents a departure from conventional cameras, which capture entire pictures at a predetermined pace determined by an external clock, such as 30 frames per second. Instead, event cameras react to scene brightness variations independently and asynchronously for each pixel. Focusing on scene changes rather than static frames makes them a valuable tool in various applications [2-53].

The output of an event camera comprises a changeable data-rate series of digital "events" or "spikes," with each event representing a pixel's change in brightness (log intensity) at a specific time, with a predetermined magnitude. Event cameras, inspired by the spiking properties of biological visual circuits, enable continuous event representation and deliver improved performance [13, 43-54].

When a pixel delivers an event, it memorizes the log intensity and monitors significant changes from this value. The camera transmits an event when a change beyond a threshold is detected, relaying the x, y position, time "t", and 1-bit polarity "p" (indicating an increase or decrease in brightness) from the chip. These events are transmitted from the pixel array to the peripheral and eventually out of the camera using a common digital output bus, typically employing address-event representation (AER) readout. The readout speeds of event cameras can vary from 2 MHz to 1200 MHz, depending on the hardware interface and chip type [11-13, 55-56].

DVS devices show promise for space and scientific applications due to their low power consumption, excellent temporal resolution, and broad dynamic range. Additionally, DVS devices, with higher dynamic range and temporal resolution, are well-suited for applications such as intelligent robotics, autonomous driving, high-speed photography, and intelligent surveillance, owing to their ability to capture images instantaneously [55, 57-60]. Dynamic vision device technology, being cost-effective and impactful in artificial intelligence, human activity recognition, and other specialized applications, has the potential to significantly influence various research fields [21].

Focusing on research into comprehension, event segmentation, episodic memory, and activity planning, this analysis explores the intricate mental representations of everyday experiences and proposes future scientific directions [38].

5.2. Frame-based sensor

Frame-based systems are techniques used for processing or recording data in discrete frames or snapshots at a fixed point in time. In the realm of computer vision, they are employed to sequentially acquire and process visual data, such as photos or films. These systems utilize standard cameras or sensors to periodically capture entire frames of visual data, each comprising a grid of pixels. Frame-based processing involves extracting features, monitoring motion, identifying objects, and performing image comprehension and interpretation.

However, the performance of these sensors is constrained by their mode of operation. Fixed frame rates can give rise to two primary issues: loss of crucial information and significant redundancy, as changes to all pixels unnecessarily increase data transfer and volume. These issues are compounded by the fact that modifications only affect a small section of the image [1-2, 43, 49, 61-62].

Video frame interpolation techniques are crucial in computer vision research as they improve intermediate frames, provide precise motion estimates, boost algorithm performance, reduce memory resource utilization, and improve frame quality [63-65].

Due to their differing data acquisition methods, integrating event-based and frame-based data remains challenging. The main issue is synchronization, in which event-based systems capture data in response to changes in the scene, while frame-based systems capture the entire scene at fixed intervals. Additionally, frame-based systems face data redundancy and computational overhead challenges, as they record redundant data, while event-based systems capture only significant changes.

A hybrid model (see Fig. 3) could address these challenges by combining event-based sensors for real-time motion detection, with frame-based sensors providing a complete and detailed view of the surroundings at less frequent intervals. This integration could significantly enhance energy efficiency and processing speed while reducing data redundancy, resulting in more robust and accurate systems for real-time applications.

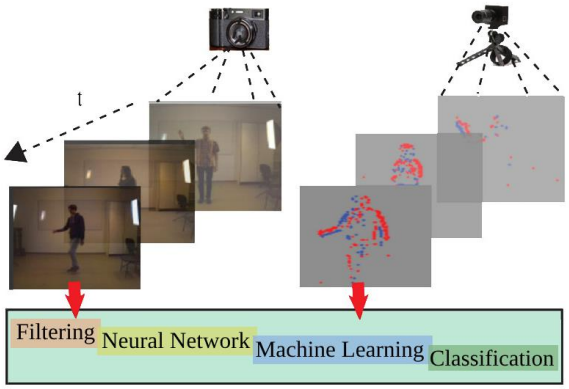


Fig. 3 Bio-inspired visual model utilizing event-based and frame-based sensors

Given these constraints, Table 1 summarizes numerous optical system models that utilize both frame-based and event-based vision sensors, highlighting their key characteristics, advantages, and applications.

Table 1 Summary of optical system models utilizing various sensor technologies

References	Model Type	Sensor Type	Key Characteristics	Advantages	Applications	Summary
[1]	Hybrid Vision	Frame-Based and Event-Based	Combines event- and frame-based data to track photometric features.	Improved feature tracking accuracy, robust to dynamic scenes.	Asynchronous photometric feature tracking.	Uses both frame-based and event-based sensors for tracking features asynchronously.
[2]	Hybrid Vision	Frame-Based and Event-Based	Utilizes combined data from both sensor types to enhance machine vision.	Versatility in various lighting and motion conditions.	Object recognition and machine vision.	Describes the integration of frame-based and event-based sensors to improve machine vision capabilities.
[4]	Event-Based Vision	Dynamic Vision Sensors (DVS)	Dynamic time representation for unsupervised optical flow estimate.	No need for labeled data, suitable for dynamic environments.	Spike camera applications and optical flow estimates.	Demonstrates the use of dynamic timing from spike camera data to estimate optical flow in an unsupervised manner.
[43]	Event-Based Vision	Dynamic Vision Sensors (DVS)	Uses DVS data to estimate the lifespan of occurrences.	Aids in sensor calibration and provides a metric for event dependability.	Calibration of sensors and optimization of performance.	Outlines a technique for calculating an event's lifespan from DVS, which can help with precise sensor calibration and performance enhancement.
[63]	Frame-Based Vision	Frame-Based Sensors	Reviews various techniques for video frame interpolation.	Minimizes motion artifacts and enhances frame quality.	Animation and video processing.	A comprehensive survey of techniques and advancements in video frame interpolation, addressing key challenges.

6. Depth Estimation Using Sensors

Zhao Chen, Vijay Badrinarayanan, Gilad Drozdov, and Andrew Rabinovich introduced a deep model designed to generate dense depth maps with high accuracy from RGB photos with limited depth information, leveraging datasets such as New York University Depth V2 (NYUv2) and Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) [66-67]. This model meets real-time criteria for both outdoor and indoor scenarios, achieving a mean depth error of less than 1% in interior settings [68]. Such sensors could find applications in autonomous vehicles and robotics.

A strategy for dealing with sparse depth data in convolutional neural networks is provided, with the option of using dense RGB. The technique efficiently learns sparse features without the need for extra validity masks, ensuring network durability even at densities as low as 0.8%. The process of transitioning from a sparse to a dense depth map involves a deep neural network that integrates an RGB image and a sparse depth map (see Fig. 4). Mean relative absolute error (MRAE) serves as the evaluation metric to minimize error rates in the generated output. The network architecture includes dense layers along with other components such as pooling layers and rectified linear unit (ReLU) activation functions [69].

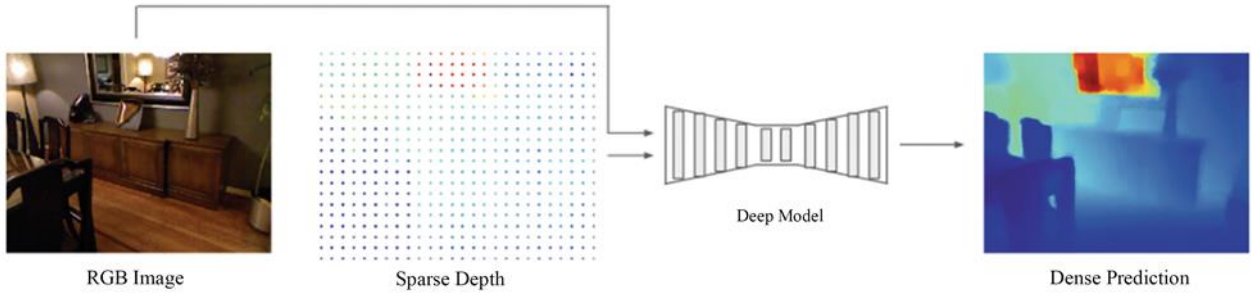


Fig. 4 Artificial neural networks for scaling images from sparse to dense Inference [68]

7. Data-Driven Tools

Event cameras that communicate per-pixel intensity changes have emerged as a feasible choice in a variety of industries, including consumer electronics, industrial automation, and autonomous cars, due to their efficiency and resilience. In event-based algorithms, maintaining these advantages requires balancing accuracy and efficiency [46].

The development of data-driven methodologies has emerged as a thriving research area due to its compatibility with spiking neural networks (SNNs) and deep learning techniques. This study explores the operation and benefits of event cameras, highlighting a recent shift in research focus towards data-driven technologies in event-based vision. Alongside discussions on hardware, datasets, and algorithmic trends, it aims to guide future research towards developing more effective and bio-inspired visual systems.

In contrast to artificial neural networks (ANNs) [70], the emerging generation of neural networks, spiking neural networks (SNNs), which synchronize their firing, can be employed in event-based vision systems to reduce energy consumption [71]. SNNs are particularly suited for processing high temporal resolution data from event-based cameras due to their differences from ANNs in information encoding (spiking neurons encode information via spikes while non-spiking neurons use real-value activations), memory (spiking neurons typically possess memory unlike non-spiking neurons), and time-variance (SNNs exhibit time-varying behavior whereas many ANNs are time-invariant) [72]. Fig. 5 below illustrates a simple SNN neuron.

SNNs employ various spiking neuron models to mimic the information-transmission processes of biological neurons. Three prominent models include the Hodgkin–Huxley model, which accurately represents the electrical properties of neurons, including ion channels; the leaky integrate-and-fire model, which is simpler and more computationally efficient than Hodgkin-Huxley; and the Izhikevich model, which strikes a balance between biological realism and computational efficiency [73-75].

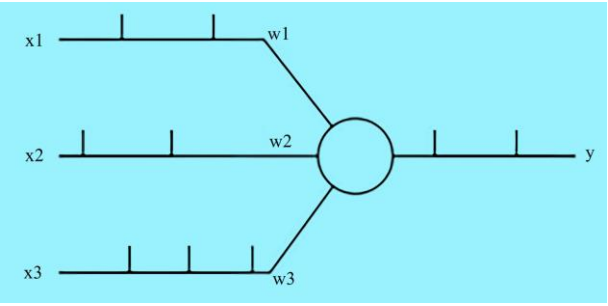


Fig. 5 Simplified SNNs neuron model [72]

8. Applications of Visual System Models

Event- and frame-based sensors are employed across various sectors for visual system models, including autonomous vehicles, robotics, surveillance, and security.

Event-based sensors detect dynamic changes, whereas frame-based sensors deliver comprehensive images for object recognition and lane detection. Event-based sensors improve user experiences in AR and VR, while frame-based sensors contribute to immersive experiences. These sensors enable real-time gesture detection, and facial recognition, and are also applied in biomedical imaging, industrial automation, and sports analytics. Combining both sensor types provides high-speed processing, low latency, and rich visual data [47, 76-77].

9. Discussion

Visual system-inspired vision models (VSVIMs) and traditional computer vision models (CVMs) are widely utilized in the realms of artificial intelligence and machine vision, which are entrusted with the task of interpreting visual data. VSVIMs take their cue from how human vision works biologically; they employ neuromorphic engineering and draw insights from neuroscience to replicate resilience, flexibility, and, above all else, the efficiency exhibited by the human visual system on a computational platform. The traditional models walk a different path altogether; relying on mathematical formulae and algorithms to decipher what the visual data is all about. The strategies used by these two classes of models could not be more dissimilar.

VSVIMs take an event-driven approach towards recording visual information, producing sparse and yet temporally precise data streams that echo what is observed at the level of brain activity. Conventional models with frame-based sensors capture contextual information but fall short when it comes to understanding temporal dynamics.

DVS provides high-resolution pictures and contextual information while maintaining the real-time performance requirements of activities like object recognition and tracking. VSVIM succeeds in both of these resource-intensive domains since their demand exceeds available resources, particularly due to their efficiency. Traditional models perform well in situations requiring limited visual input, sluggish motion, or a low dynamic range. Vision systems based on visual sensorimotor coordination have applications in a variety of fields, including robotics, healthcare, and self-driving automobiles.

Event cameras provide various benefits over traditional cameras, including high temporal resolution, low latency, low power consumption, and a large dynamic range. Using digital readout and rapid analog circuitry, event cameras achieve microsecond precision in time-stamped event detection, allowing quick action capture without motion blur. Their independent pixel feature and immediate propagation of changes ensure low latency, which results in a reaction time of less than a millisecond. Integrated event camera systems consume 100 mW or less power, whereas image-based cameras have a dynamic range of 60 dB. Event cameras can collect data from moonlight to daylight, making them versatile for various applications [23-64].

Sensor technologies are crucial in determining the performance of computer vision systems. Image-based sensors and DVS sensors have different advantages and applications. DVS sensors detect illumination changes with no external synchronization, which is ideal for motion-adaptive real-time systems, while paddle devices take picture snapshots and measure space but cannot cope well when the environment is rapidly changing or objects are in motion. This means DVS sensors can move expeditiously over moving scenes and a variety of objects allowing machines robots and autonomous systems to perceive and track motion efficiently. Image-based sensors have detailed physical motion information, allowing for motion evaluations over time and trajectory analysis, but they are less effective when fast motion is presented with lots of highs and changes over time.

Motion sensor (DVS) techniques allow for energy-efficient data gathering for mobile robotics, wearable technologies as well as battery-powered devices. Their absence of sync and presence in most cases make them cost-effective in terms of computing. Frame-based energy sensors, by contrast, would mostly lead to unwanted battery-powered applications in moving areas or very remote zones where there are no power resources for extended periods [65].

DVS sensors are utilized in various activities including robots, teleportation, and augmented reality due to their high resolution, low latency, and power consumption. Frame-based sensors are applied in scientific research as well as industrial inspection, medical imaging, and surveillance. Future developments aim at improving the parameters of performance, for instance, the working energy efficiency and resolution. Both DVS and frame-based vision sensors are applied in visual data analysis. DVS offers higher spatial and frame resolution while the frame-based sensors offer temporal resolution and power savings [78-80].

10. Conclusions

This paper critiques and analyzes data acquisition from both event-based and frame-based vision sensors. Conventional cameras produce a chain of frames with significant data redundancy, requiring extensive processing. In contrast, event-based sensors respond more effectively to scene modifications, thereby minimizing redundant data. However, DVS faces challenges in detecting objects in static scenes, which limits its effectiveness.

To address these challenges, a hybrid approach integrates the efficient data processing of event-based sensors with the high-resolution spatial data of frame-based systems, thus providing a robust solution for visual computation. This complementary approach enhances real-time processing and computing efficiency, effectively addressing challenges in visual computation across diverse applications, including industrial automation, interactive environments, and driverless vehicles.

Furthermore, future research could explore advanced data fusion techniques and evaluate the hybrid model's performance in various settings, including dimly lit environments or rapidly moving scenes, aiming to enhance real-time performance. Improving the synergy between event-based and frame-based data, could enable AI-driven solutions to enhance visual processing for robotics and augmented reality applications.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, "EKLT: Asynchronous Photometric Feature Tracking Using Events and Frames," International Journal of Computer Vision, vol. 128, no. 3, pp. 601-618, 2020.

[2] H. S. Leow and K. Nikolic, "Machine Vision Using Combined Frame-Based and Event-Based Vision Sensor," Proceedings of IEEE International Symposium on Circuits and Systems, pp. 706-709, 2015.

[3] M. Domínguez-Morales, J. P. Domínguez-Morales, Á. Jiménez-Fernández, A. Linares-Barranco, and G. Jiménez-Moreno, "Stereo Matching in Address-Event-Representation (AER) Bio-Inspired Binocular Systems in a Field-Programmable Gate Array (FPGA)," Electronics, vol. 8, no. 4, article no. 410, 2019.

[4] L. Xia, Z. Ding, R. Zhao, J. Zhang, L. Ma, Z. Yu, et al., "Unsupervised Optical Flow Estimation with Dynamic Timing Representation for Spike Camera," Proceedings of 37th International Conference on Neural Information Processing Systems, pp. 48070-48082, 2023.
[5] B. Küçükoğlu, B. Rueckauer, N. Ahmad, J. de Ruyter v. Stevenincks, U. Güçlü, and M. Van Gerven, "Optimization of Neuroprosthetic Vision via End-to-End Deep Reinforcement Learning," International Journal of Neural Systems, vol. 32, no. 11, article no. 2250052, 2022.
[6] D. Kong and Z. Fang, "A Review of Event-Based Vision Sensors and Their Applications," Information and Control, vol. 50, no. 1, pp. 1-19, 2021. (in Chinese)
[7] M. Nowak, A. Beninati, N. Douard, A. Puran, C. Barnes, A. Kerwick, et al., "Polarimetric Dynamic Vision Sensor P(DVS) Neural Network Architecture for Motion Classification," Electronics Letters, vol. 57, no. 16, pp. 624-626, 2021.
[8] F. A. Wichmann and R. Geirhos, "Are Deep Neural Networks Adequate Behavioral Models of Human Visual Perception?," Annual Review of Vision Science, vol. 9, pp. 501-524, 2023.
[9] J. A. Leñero-Bardallo, D. H. Bryn, and P. Häfliger, "Bio-Inspired Asynchronous Pixel Event Tri-Color Vision Sensor," Proceedings of IEEE Biomedical Circuits and Systems Conference, pp. 253-256, 2011.
[10] B. Wei, Y. Zhao, K. Hao, and L. Gao, "Visual Sensation and Perception Computational Models for Deep Learning: State of the Art, Challenges and Prospects," https://arxiv.org/abs/2109.03391v1, 2021.
[11] M. H. Tayarani-Najaran and M. Schmuker, "Event-Based Sensing and Signal Processing in the Visual, Auditory, and Olfactory Domain: A Review," Frontiers in Neural Circuits, vol. 15, article no. 610446, May 2021.
[12] R. Benosman, Sio-Hoi Ieng, C. Clercq, C. Bartolozzi, and M. Srinivasan "Asynchronous Frameless Event-Based Optical Flow," Neural Networks, vol. 27, pp. 32-37, 2012.
[13] G. Gallego T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, et al., "Event-Based Vision: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 1, pp. 154-180, 2022.
[14] S. Tschechne, R. Sailer, and H. Neumann, "Bio-Inspired Optic Flow from Event-Based Neuromorphic Sensor Input," Artificial Neural Networks in Pattern Recognition, pp. 171-182, 2014.
[15] L. I. Abdul-Kreem and H. K. Abdul-Ameer, "Object Tracking Using Motion Flow Projection for Pan-Tilt Configuration," International Journal of Electrical and Computer Engineering, vol. 10, no. 5, pp. 4687-4694, 2020.
[16] R. W. Fleming, "Visual Perception of Materials and Their Properties," Vision Research, vol. 94, pp. 62-75, 2014.
[17] L. I. Abdul-Kreem, "Computational Architecture of a Visual Model for Biological Motions Segregation," Network: Computation in Neural Systems, vol. 30, no. 1-4, pp. 58-78, 2019.
[18] M. J. Dominguez-Morales, A. Jimenez-Fernandez, G. Jimenez-Moreno, C. Conde, E. Cabello, and A. Linares-Barranco, "Bio-Inspired Stereo Vision Calibration for Dynamic Vision Sensors," IEEE Access, vol. 7, pp. 138415-138425, 2019.
[19] J. D. Blair, K. M. Gaynor, M. S. Palmer, and K. E. Marshall, "A Gentle Introduction to Computer Vision-Based Specimen Classification in Ecological Datasets," Journal of Animal Ecology, vol. 93, no. 2, pp. 147-158, 2024.
[20] K. Luma Issa Abdul-Kreem, "Depth Estimation and Shape Reconstruction of a 2D Image Using N.N. and Bézier Surface Interpolation," Iraqi Journal of Computers, Communications, Control, and Systems Engineering, vol. 17, no. 1, pp. 24-32, 2017.
[21] S. A. Baby, B. Vinod, C. Chinni, and K. Mitra, "Dynamic Vision Sensors for Human Activity Recognition," Proceedings of 4th IAPR Asian Conference on Pattern Recognition, pp. 316-321, 2017.
[22] N. V. K. Medathati, H. Neumann, G. S. Masson, and P. Kornprobst, "Bio-Inspired Computer Vision: Towards a Synergistic Approach of Artificial and Biological Vision," Computer Vision and Image Understanding, vol. 150, pp. 1-30, 2016.
[23] M. Yang, S. Chii. Liu, and T. Delbruck, "A Dynamic Vision Sensor with 1% Temporal Contrast Sensitivity and In-Pixel Asynchronous Delta Modulator for Event Encoding," IEEE Journal of Solid-State Circuits, vol. 50, no. 9, pp. 2149-2160, 2015.
[24] Y. Liu, R. Fan, J. Guo, H. Ni, and M. U. M. Bhutta, "In-Sensor Visual Perception and Inference," Intelligent Computing, vol. 2, article no. 0043, 2023.
[25] L. I. Abdul-Kreem, "Motion Estimations of Hand Movement Based on a Leap Motion Controller," IEEE Sensors Journal, vol. 24, no. 11, pp. 17856-17864, 2024.
[26] J. Billino and K. S. Pilz, "Motion Perception as a Model for Perceptual Aging," Journal of Vision, vol. 19, no. 4, pp. 1-28, 2019.
[27] L. I. Abdul-Kreem and H. Neumann, "Estimating Visual Motion Using an Event-Based Artificial Retina," Communications in Computer and Information Science, pp. 396-415, 2016.

[28] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping," Lecture Notes in Computer Science, vol. 3024, pp. 25-36, 2004.
[29] G. Farnebäck, "Very High Accuracy Velocity Estimation Using Orientation Tensors, Parametric Motion, and Simultaneous Segmentation of the Motion Field," Proceedings of Eighth IEEE International Conference on Computer Vision, vol. 1, pp. 171-177, 2001.
[30] G. Haessig, A. Cassidy, R. Alvarez, R. Benosman, and G. Orchard, "Spiking Optical Flow for Event-Based Sensors Using IBM's TrueNorth Neurosynaptic System," IEEE Transactions on Biomedical Circuits and Systems, vol. 12, no. 4, pp. 860-870, 2018.
[31] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-Based and Vision-Based Human Activity Recognition: A Comprehensive Survey," Pattern Recognition, vol. 108, article no. 107561, 2020.
[32] R. Tapu, B. Mocanu, and T. Zaharia, "DEEP-SEE: Joint Object Detection, Tracking, and Recognition with Application to Visually Impaired Navigational Assistance," Sensors, vol. 17, no. 11, article no. 2473, 2017.
[33] P. Jiang, D. Ergu, F. Liu, C. Ying, and B. Ma, "A Review of Yolo Algorithm Developments," Procedia Computer Science, vol. 199, pp. 1066-1073, 2022.
[34] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li, "ImageNet: A Large-Scale Hierarchical Image Database," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, 2009.
[35] P. Thangavell and S. Karuppannan, "Dynamic Event Camera Object Detection and Classification Using Enhanced YOLO Deep Learning Architecture," Optica Open Preprint, article no. 110762, 2023.
[36] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, et al., "YOLOv10: Real-Time End-to-End Object Detection," http://arxiv.org/abs/2405.14458, 2024.
[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is All You need," Proceedings of 31st Conference on Neural Information Processing Systems, pp. 1-11, 2017.
[38] L. L. Richmond and J. M. Zacks, "Constructing Experience: Event Models from Perception to Action," Trends in Cognitive Sciences, vol. 21, no. 12, pp. 962-980, 2017.
[39] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," Proceedings of 27th International Conference on Neural Information Processing Systems, vol. 2, pp. 3104-3112, 2014.
[40] W. Weng, Y. Zhang, and Z. Xiong, "Event-Based Video Reconstruction Using Transformer," Proceedings of IEEE /CVF International Conference on Computer Vision, pp. 2543-2552, 2021.
[41] L. Yuan, Y. Chen, T. Wang, W. Yu, Y. Shi, Z. H. Jiang, et al., "Tokens-to-Token ViT: Training Vision Transformers from Scratch on ImageNet," Proceedings of IEEE/CVF International Conference on Computer Vision, pp. 558-567, 2021.
[42] H. Touvron, M. Cord, and H. Jégou, "Deit III: Revenge of the Vit," https://arxiv.org/abs/2204.07118v1, 2022.
[43] E. Mueggler, C. Forster, N. Baumli, G. Gallego, and D. Scaramuzza, "Lifetime Estimation of Events from Dynamic Vision Sensors," Proceedings of IEEE International Conference on Robotics and Automation, pp. 4874-4881, 2015.
[44] L. I. Abdul-Kreem and H. K. Abdul-Ameer, "Shadow Detection and Elimination for Robot and Machine Vision Applications," Scientific Visualization, vol. 16, no. 2, pp. 11-22, 2024.
[45] M. Al-Faris, J. Chiverton, D. Ndzi, and A. I. Ahmed, "A Review on Computer Vision-Based Methods for Human Action Recognition," Journal of Imaging, vol. 6, no. 6, article no. 46, 2020.
[46] R. Sun, D. Shi, Y. Zhang, R. Li, and R. Li, "Data-Driven Technology in Event-Based Vision," Complexity, vol. 21, no. 1, article no.6689337, 2021.
[47] G. Yang, "Application of Indoor Light Sensor Based on Monitoring Image Recognition in Basketball Sports Image Analysis and Simulation," Optical and Quantum Electronics, vol. 56, article no. 315, 2023.
[48] S. Dong, Z. Bi, Y. Tian, and T. Huang, "Spike Coding for Dynamic Vision Sensor in Intelligent Driving," IEEE Internet of Things Journal, vol. 6, no. 1, pp. 60-71, 2019.
[49] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "EV-FlowNet: Self-Supervised Optical Flow Estimation for Event-Based Cameras," Robotics: Science and Systems, vol. 4, 2018.
[50] L. I. Abdul-Kreem and H. Neumann, "Bio-Inspired Model for Motion Estimation Using an Address-Event Representation," Proceedings of 10th International Conference on Computer Vision Theory and Applications, vol. 3, pp. 335-346, 2015.
[51] N. Salvatore and J. Fletcher, "Learned Event-Based Visual Perception for Improved Space Object Detection," Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3301-3310, 2022.
[52] L. A. Camunas-Mesa, T. Serrano-Gotarredona, S. H. Ieng, R. Benosman, and B. Linares-Barranco, "Event-Driven Stereo Visual Tracking Algorithm to Solve Object Occlusion," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 9, pp. 4223-4237, 2018.

[53] I. A. Lungu, F. Corradi, and T. Delbruck, "Live Demonstration: Convolutional Neural Network Driven by Dynamic Vision Sensor Playing Roshambo," Proceedings of IEEE International Symposium on Circuits and Systems, p. 1, 2017.
[54] L. Camuñas-Mesa, C. Zamarreño-Ramos, A. Linares-Barranco, A. J. Acosta-Jiménez, T. Serrano-Gotarredona, and B. Linares-Barranco, "An Event-Driven Multi-Kernel Convolution Processor Module for Event-Driven Vision Sensors," IEEE Journal of Solid-State Circuits, vol. 47, no. 2, pp. 504-517, 2012.
[55] D. P. Moeys, F. Corradi, C. Li, S. A. Bamford, L. Longinotti, and F. F. Voigt, "A Sensitive Dynamic and Active Pixel Vision Sensor for Color or Neural Imaging Applications," IEEE Transactions on Biomedical Circuits and Systems, vol. 12, no. 1, pp. 123-136, 2018.
[56] R. Benosman, S. H. S. H. Ieng, P. Rogister and C. Posch, "Asynchronous Event-Based Hebbian Epipolar Geometry," IEEE Transactions on Neural Networks, vol. 22, no. 11, pp. 1723-1734, 2011.
[57] S. Dong, Z. Bi, Y. Tian, and T. Huang, "Spike Coding for Dynamic Vision Sensor in Intelligent Driving," IEEE Internet of Things Journal, vol. 6, no. 1, pp. 60-71, 2019.
[58] C. Brandli, R. Berner, M. Yang, S. C. Liu, and T. Delbruck, "A 240 × 180 130 dB 3 μs Latency Global Shutter Spatiotemporal Vision Sensor," IEEE Journal of Solid-State Circuits, vol. 49, no. 10, pp. 2333-2341, 2014.
[59] B. J. McReynolds, R. P. Graca, and T. Delbruck, "Experimental Methods to Predict Dynamic Vision Sensor Event Camera Performance," Optical Engineering, vol. 61, no. 7, article no. 074103, 2022.
[60] A. Yousefzadeh, G. Orchard, T. Serrano-Gotarredona, and B. Linares-Barranco, "Active Perception with Dynamic Vision Sensors. Minimum Saccades with Optimum Recognition," IEEE Transactions on Biomedical Circuits and Systems, vol. 12, no. 4, pp. 927-939, 2018.
[61] T. Delbruck, "Frame-Free Dynamic Digital Vision," Proceedings of International Symposium on Secure-Life Electronics, Advanced Electronics for Quality Life and Society, pp. 21-26, 2008.
[62] L. Steffen, D. Reichard, J. Weinland, J. Kaiser, A. Roennau, and R. Dillmann, "Neuromorphic Stereo Vision: A Survey of Bio-Inspired Sensors and Algorithms," Frontiers in Neurorobotics, vol. 13, no. 28, pp. 1-20, 2019.
[63] A. S. Parihar, D. Varshney, K. Pandya, and A. Aggarwal, "A Comprehensive Survey on Video Frame Interpolation Techniques," The Visual Computer, vol. 38, pp. 295-319, 2022.
[64] C. Posch, D. Matolin, and R. Wohlgenannt, "A Qvga 143 dB Dynamic Range Frame-Free PWM Image Sensor with Lossless Pixel-Level Video Compression and Time-Domain CDS," IEEE Journal of Solid-State Circuits, vol. 46, no. 1, pp. 259-275, 2011.
[65] A. Lakshmi, A. Chakraborty, and C. S. Thakur, "Neuromorphic Vision: From Sensors to Event-Based Algorithms," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 4, article no. 1310, 2019.
[66] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, et al., "A Dataset for Semantic Scene Understanding of LiDAR Sequences," Proceedings of IEEE/CVF International Conference on Computer Vision, pp. 9296-9306, 2019.
[67] Y. Cabon, N. Murray, and M. Humenberger, "Virtual KITTI 2," http://arxiv.org/abs/2001.10773, 2020.
[68] Z. Chen, V. Badrinarayanan, G. Drozdov, and A. Rabinovich, "Estimating Depth from RGB and Sparse Sensing," Proceedings of 15th European Conference, pp. 176-192, 2018.
[69] M. Jaritz, R. D.e Charette, E. Wirbel, X. Perrotton, and F. Nashashibi, "Sparse and Dense Data with Cnns: Depth Completion and Semantic Segmentation," Proceedings of International Conference on 3D Vision, pp. 52-60, 2018.
[70] J. A. Pérez-Carrasco, B. Zhao, C. Serrano, B. Acha, T. Serrano-Gotarredona, S. Chen, "Mapping from Frame-Driven to Frame-Free Event-Driven Vision Systems by Low-Rate Rate Coding and Coincidence Processing-Application to Feed Forward ConvNets," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 11, pp. 2706-2719, 2013.
[71] J. Wu, Y. Wang, Z. Li, L. Lu, and Q. Li, "A Review of Computing with Spiking Neural Networks," Computers, Materials & Continua, vol. 78, no. 3, pp. 2909-2939, 2024
[72] J. Furmonas, J. Liobe, and V. Barzdenas, "Analytical Review of Event-Based Camera Depth Estimation Methods and Systems," Sensors, vol. 22, no. 3, article no. 1201, 2022.
[73] Y. Xu, S. Gao, Z. Li, R. Yang, and X. Miao, "Adaptive Hodgkin–Huxley Neuron for Retina-Inspired Perception," Advanced Intelligent Systems, vol. 4, no. 12, article no. 202200210, 2022.
[74] J. Kim, Y. I. Choi, J. W. Sohn, S. P. Kim, and S. J. Jung, "Modeling Long-Term Spike Frequency Adaptation in SA-I Afferent Neurons Using an Izhikevich-Based Biological Neuron Model," Experimental Neurobiology, vol. 32, no. 3, pp. 157-169, 2023.
[75] M. Botvinick, J. X. Wang, W. Dabney, K. J. Miller, and Z. Kurth-Nelson, "Deep Reinforcement Learning and Its Neuroscientific Implications," Neuron, vol. 107, no. 4, pp. 603-616, 2020.
[76] H. Qiao, Y. X. Wu, S. L. Zhong, P. J. Yin, and J. H. Chen, "Brain-Inspired Intelligent Robotics: Theoretical Analysis and Systematic Application," Machine Intelligence Research, vol. 20, no. 1, pp. 1-18, 2023.

[77] Y. Li, J. Moreau, and J. Ibanez-Guzman, “Emergent Visual Sensors for Autonomous Vehicles,” IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 5, pp. 4716-4737, 2023.

[78] F. B. Naeini, A. M. AlAli, R. Al-Husari, A. Rigi, M. K. Al-Sharman, and D. Makris, “A Novel Dynamic-Vision-Based Approach for Tactile Sensing Applications,” IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 5, pp. 1881-1893, 2020.

[79] A. Censi, J. Strubel, C. Brandli, T. Delbruck, and D. Scaramuzza, “Low-Latency Localization by Active LED Markers Tracking Using a Dynamic Vision Sensor,” Proceedings of IEEE International Conference on Intelligent Robotics and Systems, pp. 891-898, 2013.

[80] A. Andreopoulos, H. J. Kashyap, T. K. Nayak, A. Amir, and M. D. Flickner, “A Low Power, High Throughput, Fully Event-Based Stereo System,” Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7532-7542, 2018.