Review Article

ADVANCED PHOTONICS

Review of bio-inspired image sensors for efficient machine vision

Wenhao Tang,[a,†] Qing Yang,[a,b,c,†] Hang Xu,[a] Yiyu Guo,[a] Jiqiang Zhang,[a] Chunfang Ouyang,[d] Leixin Meng,[a,*] and Xu Liu[b,c,*]

[a]Zhejiang Laboratory, Research Center for Frontier Fundamental Studies, Hangzhou, China
[b]Zhejiang University, College of Optical Science and Engineering, State Key Laboratory of Extreme Photonics and Instrumentation, Hangzhou, China
[c]ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, China
[d]Shanghai Jiao Tong University, Chip Hub for Integrated Photonics Xplore (CHIPX), Wuxi, China

Abstract. With the rapid development of sensor networks, machine vision faces the problem of storing and computing massive data. The human visual system has a very efficient information sense and computation ability, which has enlightening significance for solving the above problems in machine vision. This review aims to comprehensively summarize the latest advances in bio-inspired image sensors that can be used to improve machine-vision processing efficiency. After briefly introducing the research background, the relevant mechanisms of visual information processing in human visual systems are briefly discussed, including layer-by-layer processing, sparse coding, and neural adaptation. Subsequently, the cases and performance of image sensors corresponding to various bio-inspired mechanisms are introduced. Finally, the challenges and perspectives of implementing bio-inspired image sensors for efficient machine vision are discussed.

Keywords: bio-inspired image sensor; machine vision; layer-by-layer processing; sparse coding; neural adaptation.

Received Nov. 13, 2023; revised manuscript received Feb. 26, 2024; accepted for publication Mar. 14, 2024; published online Apr. 8, 2024.

[DOI: 10.1117/1.AP.6.2.024001]

1 Introduction

Machine vision is multifunctional and can be applied in manufacturing processes to improve efficiency. In typical machine vision systems, visual perception functions occur in the analog domain, while signals are processed in the digital domain through the von Neumann computing architecture.[1] In this architecture, the sensing, storage, and computing units are separated. To complete an operation, data need to be converted and transmitted among different units. In this process, a large amount of redundant data is generated, resulting in a waste of time and storage space. With the rapid development of sensor networks, it is an urgent task to remove redundant data and improve the processing efficiency of sensor systems. Based on traditional computational frameworks, a great deal of work has been utilizing novel algorithms to compress image data and perform complex processing tasks.[2–5] However, the training and development of these algorithms are often complex and require a lot of time and resources. Optimization from the software side alone seems to be far from enough. To further improve computing efficiency, joint optimization of hardware and software is needed.

The processing of information by the human visual system begins with light signals on the retina and finishes up at the output with the identities and spatial relationships of the objects in the visual scene. The human visual system uses an initial stage of data compression in the retina, which removes much of the redundant data.[6] With the help of neural adaptation, the retina can not only collect information with a high dynamic range but also eliminate noise and redundant data.[7,8] After the retinal conversion and preprocessing, the visual information is encoded into a series of nerve spike signals that can be processed by the subsequent visual centers.[9–11] These spikes are believed to be a key factor in humans' ability to process large amounts of visual information with low power consumption. Spikes from the retina are transmitted to the lateral geniculate nucleus, the first cortical visual area (V1), and other areas, including V2, V3, middle temporal (MT), V4, and inferotemporal cortex.

*Address all correspondence to Leixin Meng, menglx@zhejianglab.com; Xu Liu, liuxu@zju.edu.cn

†These authors contributed equally to this work.

These different layers of visual centers are used to analyze aspects such as motion, stereo, and color. In particular, different levels of visual centers do not process visual information strictly in chronological order. Thanks to the neural network connection of different functional regions, the multilevel visual centers can process and integrate information in parallel and finally let the brain get the visual information in the scene quickly and accurately. The high efficiency of the human visual system is inseparable from the strong ability of the center at all levels to analyze specific information such as motion, stereo, and color. In general, neural adaptation, sparse coding, and layer-by-layer processing mechanisms play a decisive role in the efficient compression and parallel processing of visual information and are also the key learning directions of image sensors for efficient machine vision. By mimicking the human visual system, numerous new bio-inspired image sensors have emerged in recent years. Through the improvement of architecture, information coding mechanism, and neural adaptation, these image sensors have excellent visual information compression capabilities and can be used to solve some high-level visual tasks, improving information processing efficiency (Fig. 1).[12–15]

This review summarizes the recent progress in the field of bio-inspired image sensors. The novel image sensors are analyzed from the perspectives of innovative architectures, sparse coding mechanisms, and neural adaptation. Section 2 briefly introduces the biological structure of the human visual system and related mechanisms of information compression processing. Section 3 summarizes the research progress of bio-inspired image sensors for efficient machine vision, including innovative sensory architecture, sparse coding mechanisms, and neural adaptation in image sensors. Section 4 concludes the review by highlighting the outstanding challenges and perspectives related to the subjects under debate.

2 Information Processing in the Human Visual System

Humans can quickly sense and respond to changes in the environment while consuming very little energy due to the special structure and processing mechanisms of the visual system. Research on human vision has been carried out for many years. Although the complex structure, specific coding mechanisms, and neural properties of the human visual system are not fully understood, the existing biological research results can provide enlightening thinking and design basis for image sensors. As Fig. 2 shows, the components of the human visual pathway are relatively complex. In simple terms, information is mainly received through the eyes and then transmitted to specific areas of the brain through a series of optic nerves. Through neural adaptation mechanism and sparse coding mechanism, visual information will be extracted and converted into low-redundancy spike trains, which will be transmitted to the subsequent visual center for processing, and the brain will finally integrate the information of various parts and reconstruct the original image. Section 2 briefly describes the structure of the human visual system, the process of visual information processing, and internal mechanisms.

2.1 Layer-by-Layer Processing in the Human Visual System

Visual information is not simply transported mechanically from the eyes to the brain; the processing of visual information starts at the sensing end, and it is processed hierarchically when transmitted in the visual pathway, greatly reducing the burden on the visual cortex of the brain. The retina, lateral geniculate nucleus, and visual cortex play a key role in the visual pathway. The human retina is a thin layer of brain tissue in the eye and provides neural processing for photoreceptor signals.[16] The retina can not only convert the external light-intensity information into nerve signals that can be transmitted but also perform certain information preprocessing functions, including the extraction of characteristic information such as color and shape and the filtering of redundant information.[17] Effective feature information is eventually output from the retina in the form of spikes.[18] Then, the lateral geniculate nucleus classifies the different features of the retinal output and sends them to the corresponding functional areas in the visual cortex. Finally, specific visual tasks can be completed through the selective integration of feature information by the visual cortex of the brain.[19] Overall, the human visual system divides complex visual tasks into various levels of the center, and after each level of the center, the information will

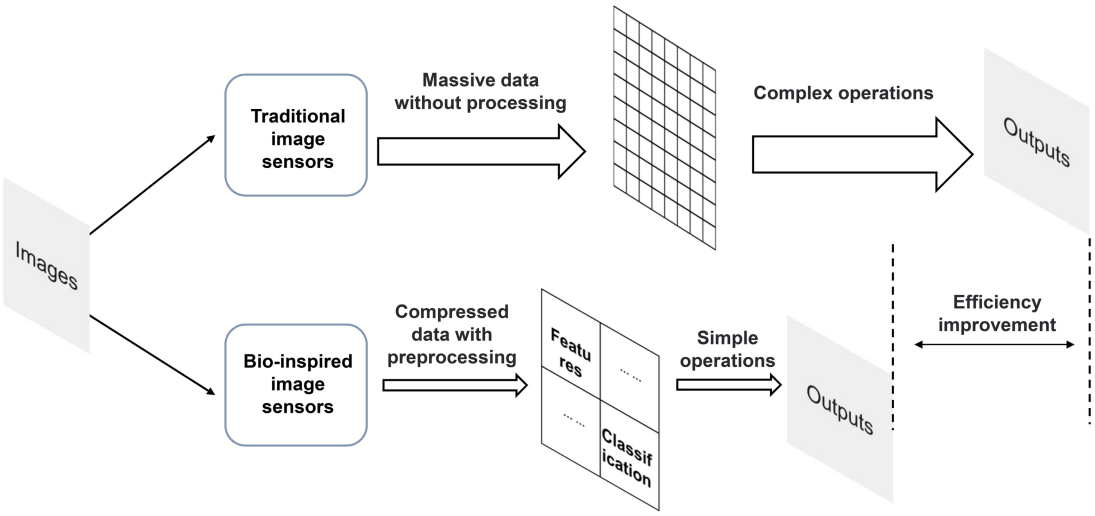

Fig. 1 Comparison between bio-inspired and traditional image sensors.

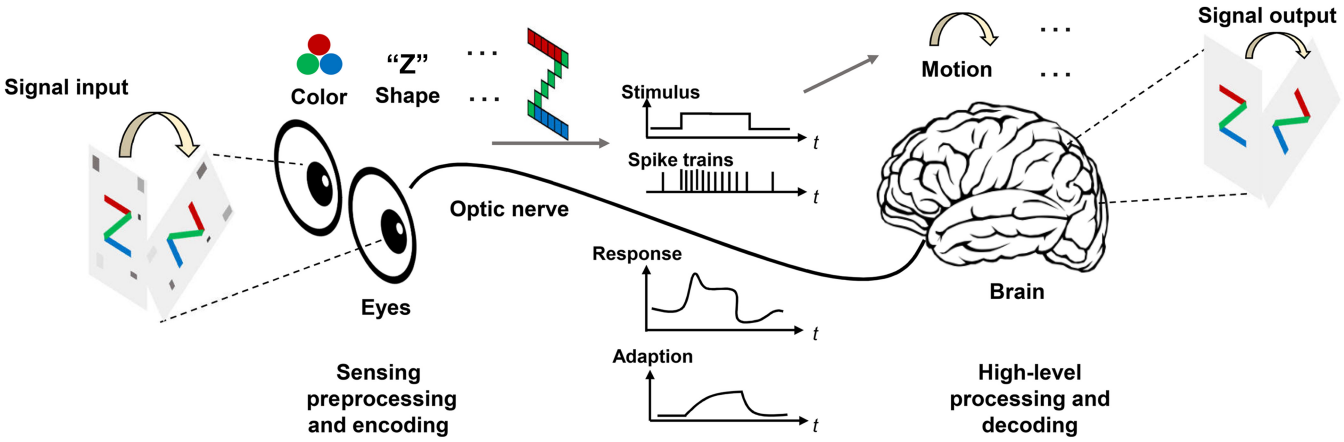

Fig. 2 Diagram of human visual information processing.

be effectively compressed, thus improving the processing efficiency of the whole system.

2.2 Sparse Coding

Theoretical studies suggest that the retina and brain use a sparse code to efficiently represent natural scenes. Retina output is determined by the patterns of spikes produced by retinal ganglion cells. These spike patterns encode all visual information available to the rest of the visual system. The information transmitted by the neuron is contained in the temporal sequence of these spikes, the "spike train." The relationship among spike trains forms the "neural code" (Ref. 9). Over the years, a range of different paradigms for neural code have been developed. Rate encoding and temporal encoding are the two main encoding schemes.[20] The most fundamental formulation of sparse coding is that a single neuron transmits information by the number of spikes produced over an extended temporal period.[21] Sparse coding is computationally efficient for both early vision and advanced visual processing. It allows for increased storage capacity in associative memories and makes the structure of natural signals explicit.[11] By representing complex data in a way that they are easier to read out at subsequent levels of processing, in general, sparse coding reduces the overall neural activity required to represent information.

2.3 Neural Adaptation

Neural adaptation refers to the common phenomenon of a decline in neuronal activity in response to repeated or prolonged stimulation. Neural adaptation is observed along the neuronal pathway from the sensory periphery to the motor output, and adaptation generally becomes stronger at higher levels. Neural adaptation has a typical high-pass filtering property, and the low-frequency stimulus component is gradually weakened by adaptive dynamics. Neural adaptation is also reflected in the adaptive adjustment of stimulus mean and variance. In a natural scene, the photoreceptor conversion light intensity is not constant but fluctuates continuously over different time scales and within certain distributions of intensity levels. The human visual system adjusts to changes in average stimulus levels and higher-order statistics in the environment. The final signal transmitted to the brain is not the intensity of the original image, but the local differences in space and changes in time. This strategy is also known as predictive coding, which can greatly compress visual information.[8] Adaptive mechanisms provide a rich toolkit for the nervous system to perform computations.[22]

3 Bio-Inspired Image Sensors for Efficient Machine Vision

The core mechanism of efficient processing of visual information in the human-vision system can be used in machine vision to guide the design of new image sensors to compress image information and improve computing efficiency. The following sections (Secs. 3.1–3.3) will introduce the applications of bio-inspired layer-by-layer processing mechanisms,[23] sparse coding mechanisms,[24] and neural adaptation in image sensors (Fig. 3).[25] The layer-by-layer processing mechanism can decompose the complex tasks that are only completed by the computing unit, complete some processing at the sensing and storage ends, and reduce the load of the computing unit, improving the overall efficiency. The sparse coding mechanism can achieve dimensionality reduction and compression of data, relieve the storage pressure, and facilitate the subsequent signal processing. Adaptive mechanisms help the vision system capture critical dynamic information and eliminate interference from background noise and static redundant information. In general, the framework of layer-by-layer processing has a clear division of labor, ensuring that the human visual system can efficiently process complex visual tasks. The sparse coding mechanism is a "bridge" connecting all levels in the layer-by-layer processing framework and transmits the low-redundancy visual information to the corresponding functional area. Neural adaptation is an important biological basis for sparse coding, which can effectively filter out a lot of redundant information. To improve processing efficiency, machine vision must not only imitate the computing framework of the human visual system on the macro level but also learn the acquisition and coding mechanism of visual information on the micro level.

To find out how bio-inspired vision can operate as efficiently as human vision, we compared the human-vision system with the bio-inspired vision system, as shown in Table 1. Imaging and signal transmission are not the main factors restricting information processing in bio-inspired vision systems. The development of new neuromorphic image sensors and the construction of artificial neural networks (ANNs) are complicated parts of bio-inspired vision, especially the former. The combination of

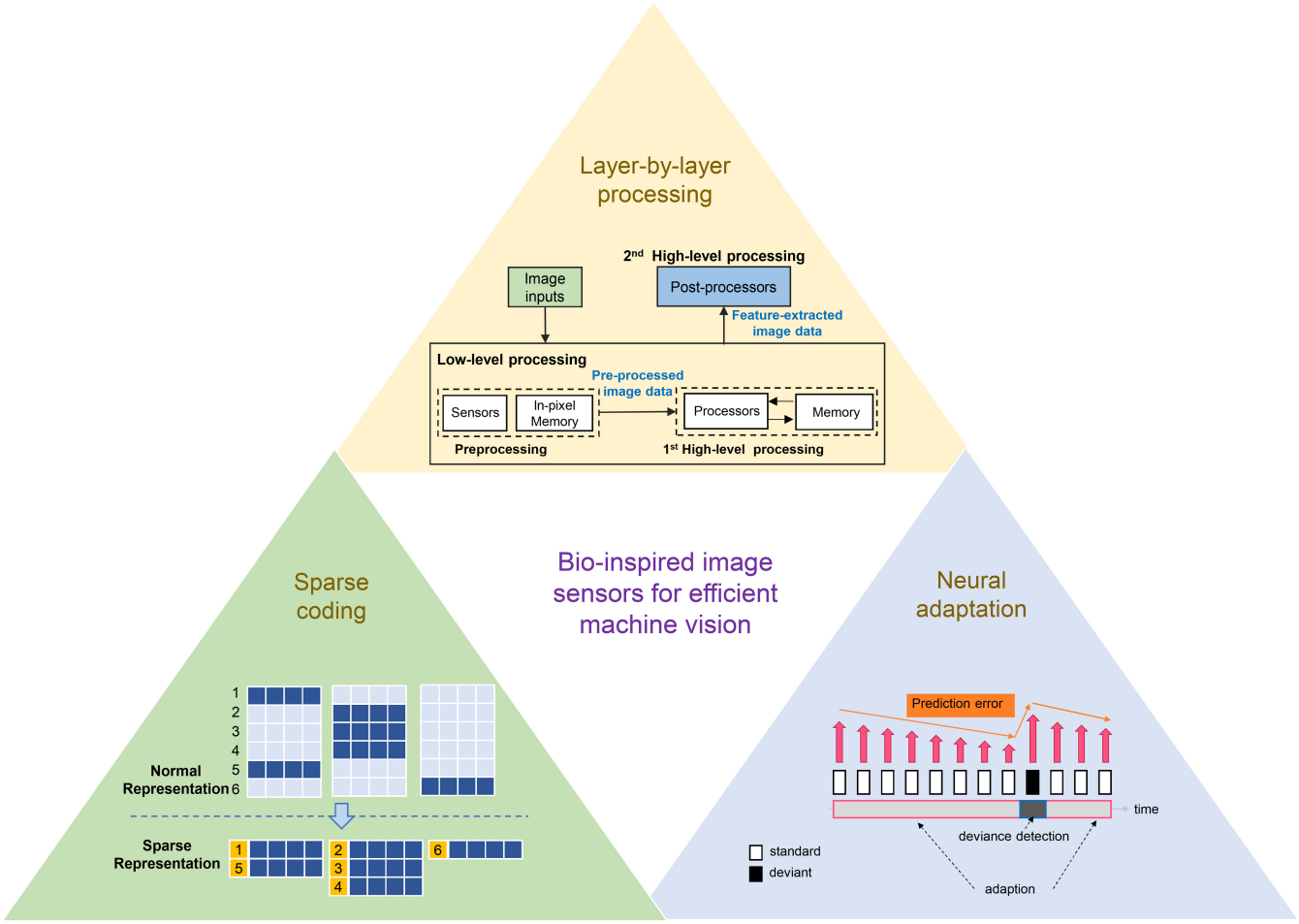

Fig. 3 Bio-inspired image sensors for efficient machine vision. Adapted with permission from Refs. 23–25.

Table 1 Comparison of human and bio-inspired vision systems.

Functions	Components of the human visual system	Bio-inspired vision system
Imaging	Eye	Optical lens
Signal conversion and coding		Neuromorphic image sensors
Signal transmission	Optic nerve	Transmission units
Interpreting visual information	Visual centers	ANNs

neuromorphic image sensors and artificial neural network has become a common working mode of bio-inspired vision systems. Most of the studies of the three mechanisms described in this section worked in this way.

3.1 Layer-by-Layer Processing Image Sensors

Common visual sensors such as charge-coupled device arrays and complementary metal-oxide-semiconductor (CMOS) arrays need to be combined with a series of storage and computing units to complete complex information processing.[26,27] The transmission and storage of redundant data require a lot of time and space, which also increases the processing difficulty of the end computing units. In recent years, researchers have made great efforts to optimize the structural design of sensing and computing systems, and some computing tasks are transferred from computing units to sensing and storage units. According to the relative spatial location relationship between the processing unit and the sensors, the new sensing system can be divided into processing near- and in-sensor architecture.[28,29]

In near-sensor architecture, the sensor pixel array is physically separated from the processing unit but simultaneously connected in parallel on a chip. The processing units are close to the sensors, and some specific computing tasks can be done near the sensor.[30] At present, the common near-sensor architecture includes omitting analog-to-digital conversion for signal processing in the analog domain or omitting the transmission step between the memory unit and the processing unit and directly

performing operations in the memory unit. Chen et al.[31] proposed processing near-sensor architecture in a mixed-signal domain with a CMOS image sensor (CIS) of the convolutional-kernel-readout method [Fig. 4(a)]. The visual data in this study are collected from intelligent CIS, and the output of the CIS is processed directly by the analog processing unit located near the CIS, unconstrained by the digital clock and analog-to-digital converter (ADC) bottlenecks. The proposed sensing chip achieves an energy efficiency of up to 545.4 GOPS/W with a 20 MHz control clock while consuming 1.8 mW of power. This group also proposed a current-mode computation-in-memory (CIM) architecture enabling near-sensor processing for intelligent internet of thing (IoT) vision nodes. Current mode computing technology was utilized in this work to achieve high energy efficiency while eliminating data conversion overhead [Fig. 4(b)].[32] They fabricated a 2-kbit CIM macro in the proposed architecture, achieving a 60.6-TOPS/W energy efficiency.

In general, the development trend of new sensing systems is more compact, faster, and smarter. Processing in-sensor architecture means that some of the processing tasks are shifted to sensors, reducing data conversion and movement.[35] By essentially eliminating all the parts between the sensor and the processing unit, the processing in-sensor architecture can achieve better computing efficiency than the near-sensor architecture. Based on traditional CMOS sensors, Xu et al.[36] demonstrated a 32×32 processing-in-sensor prototype with a 180-nm CMOS process. Their chip can accomplish MNIST data set (a public database of handwritten digits that is commonly used for training various image processing systems) classification with an accuracy of 93.76%. The energy efficiency of their chip is 13.1 times that of the state-of-the-art work. In addition to superior energy efficiency, the processing in-sensor architecture can also accurately perform high-level processing tasks while greatly reducing the amount of information. Wang et al.[33] developed an optoelectronic in-sensor compression and computing system to mimic the human visual system [Fig. 4(c)]. They used an indium–gallium–zinc–oxide (IGZO) phototransistor to achieve in-sensor compression and computing. The switching characteristics of the phototransistor are the key to forming the compression measurement matrix in the sensor. Figure 4(d) shows the results of single pulse-switching characteristics of the light potentiation and electrical depression. They combined the phototransistor arrays with a reservoir computing (RC) network for signal recognition. The results reveal that even for cases where the signal is compressed by 50%, the recognition accuracy of the reconstructed signal still reaches around 96% [Fig. 4(e)]. In addition to the relatively simple recognition and classification of static images, more complex moving object detection and recognition can also be realized in sensors. Zhang et al.[34] presented a retina-inspired two-dimensional (2D) heterostructure-based hardware device with all-in-one perception, memory, and computing capabilities for the detection and recognition of moving trolleys [Fig. 4(f)]. The device in this work has continuous and progressive adjustable non-volatile positive and negative photoconductivity characteristics, which can truly simulate the signal reception, conversion, and processing in the retina. Through the interframe difference calculation, the device successfully implemented 100% separation detection of moving trichromatic trolleys without ghosting.[34] The way to achieve a layer-by-layer processing mechanism in machine vision is to assign more computing tasks to sensors and storage units. The rapid development of in-memory computing[37,38] and in-sense computing[39,40] neuromorphic devices has enriched the functions of layer-by-layer processing mechanisms and promoted machine vision to approach or even surpass human vision.

3.2 Sparse Coding Image Sensors

Sparse coding extracts relevant information from high-dimensional external stimuli and reduces the data dimension through specific coding rules, thus achieving overall data compression. An important prerequisite for data compression based on sparse coding is to have devices similar to synapses and neurons. Previously, researchers had used traditional CMOS technologies to mimic biological synapses and neurons. Still, the problems were that the circuit needed a lot of transistors and the area of the circuit was very large. To simplify the circuit structure and reduce the circuit scale, it is expected to realize synaptic and neuronal functions at the device level. Many kinds of spike signal artificial neural devices have been studied based on different design principles. Han et al. reviewed the function and principle of different kinds of artificial neural devices, including single transistors, memristors, phase-change memory, magnetic tunnel junction (MTJ), and the leaky ferroelectric field-effect-transistor (FeFET).[41–48] These photoelectric neurons are the basic units of the computing system within the sensor, which can directly perceive and preprocess visual information.

Nonlinear responses to external stimuli are the key to realizing sparse representation. Sun et al.[49] demonstrated a neuromorphic vision system that encodes ambient light intensity and captures optical images by encoding their pixel intensity into spike signals in real time [Fig. 5(a)]. In this work, metal oxide photonic synapses with rich dynamics and nonlinearity are used as neuromorphic image sensors. The photonic synapse responds to the light pulse signal and generates postsynaptic photocurrent. The light intensity information of the input pattern is converted into a series of electrical pulses by a sensor oscillation circuit [Fig. 5(b)]. The electrical output of each photonic synaptic device presents a weighted value proportional to the frequency of the optical input, which is the basis for extracting external light-intensity information from the sensor [Fig. 5(c)]. The array system based on the photonic synapses integrates image perception, storage, and preprocessing and demonstrates dynamic perception and dynamic storage. Sparse coding can be used for the dynamic processing of temporal and sequential information and is essential for advanced applications of machine vision. Sun et al.[50] demonstrated that a 2D memristor based on tin sulfide (SnS) realizes the computation of the sensor memory repository for language learning, using the high-dimensional storage characteristics of sparse coding. Spatiotemporal optoelectronic inputs are applied to the memristors in the array, as schematically illustrated by the pulses (electrical spikes) and discrete optical beam trains (optical spikes) in Fig. 5(d). Such sequential optoelectronic inputs can generate numerous (high-dimensional and dual-mode) reservoir states of RC. By matching the optical input to the current of the memristor, the photoelectric RC system realizes the classification learning of five actual Korean sentences [Figs. 5(e) and 5(f)]. The sparse representation in the vision sensor can effectively transform the complex information of the external environment into electrical pulse signals that are easy to store and process, and greatly reduce the computational network complexity of the subsequent execution of advanced vision tasks.

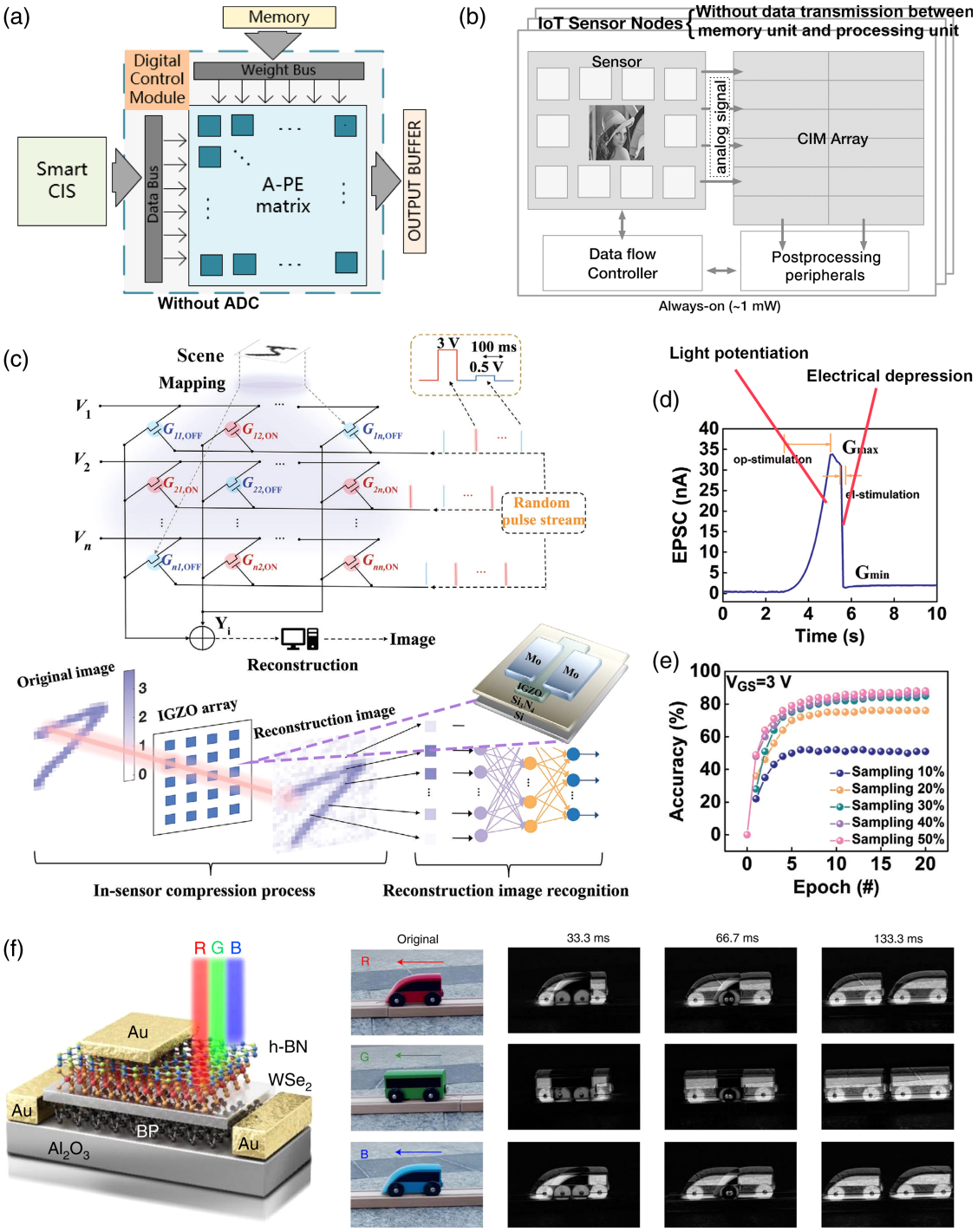

Fig. 4 Bio-inspired sensory architectures. (a) The overview of processing near-sensor architecture system without ADC. Reproduced with permission from Ref. 31. (b) CIM near-sensor architecture. Reproduced with permission from Ref. 32. (c) Schematic of IGZO phototransistor array to realize in-sensor compression simulation. Reproduced with permission from Ref. 33. (d) Optical enhancement and electrical suppression of IGZO phototransistors. Reproduced with permission from Ref. 33. (e) The recognition accuracy of MNIST images reconstructed with different sampling rates. Reproduced with permission from Ref. 33. (f) 2D retinomorphic device structure and motion detection of trichromatic trolleys. Reproduced with permission from Ref. 34.

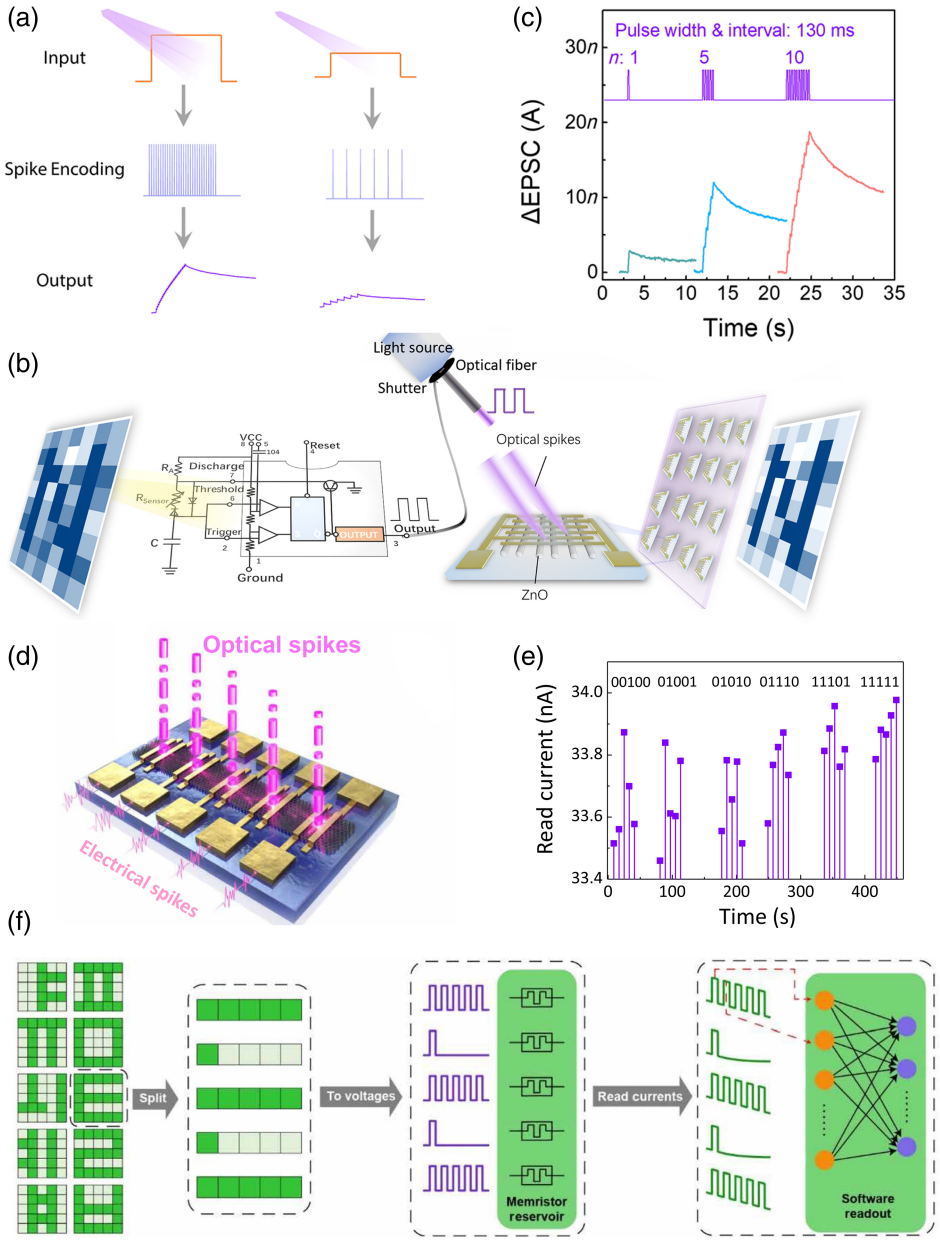

Fig. 5 Novel sensory devices based on sparse coding. (a) Light stimulus-induced spike trains. Reproduced with permission from Ref. 49. (b) Image recognition using photosensor-multivibrator circuit and photonic synapse. Reproduced with permission from Ref. 49. (c) Spike-number-dependent amplitude variation of excitatory postsynaptic current (ΔEPSC) triggered by a train of optical spikes. Reproduced with permission from Ref. 49. (d) Schematic of a multifunctional memristor array stimulated by various electrical and optical inputs. Reproduced with permission from Ref. 50. (e) Read-current responses of a memristor by several optical input signals. Reproduced with permission from Ref. 50. (f) The operation of optoelectronic RC based on 2D SnS memristors for classifying consonants and vowels in the Korean alphabet. Reproduced with permission from Ref. 50.

3.3 Neural Adaptation Image Sensors

In machine-vision application scenarios such as drone detection and security monitoring, most of the static information is redundant, and only a small amount of dynamic information is truly valuable. Traditional cameras or visual sensors only mechanically record the information of all pixels in a period. When the number of cameras is large or the recording time is long, the amount of data generated is very large. In human vision, due to the neural adaptation mechanism described in Sec. 2, a large amount of redundant information is filtered out, and only the useful response changes in the temporal domain or spatial domain are recorded and used for subsequent processing. This way of recording changes is called event-driven. Event-driven sampling offers several advantages over its conventional frame-based counterparts, including lower power requirements, lower data volume, wider dynamic range, and shorter latency times [Fig. 6(a)].[51] Liu et al.[55] showed that using event-based vision

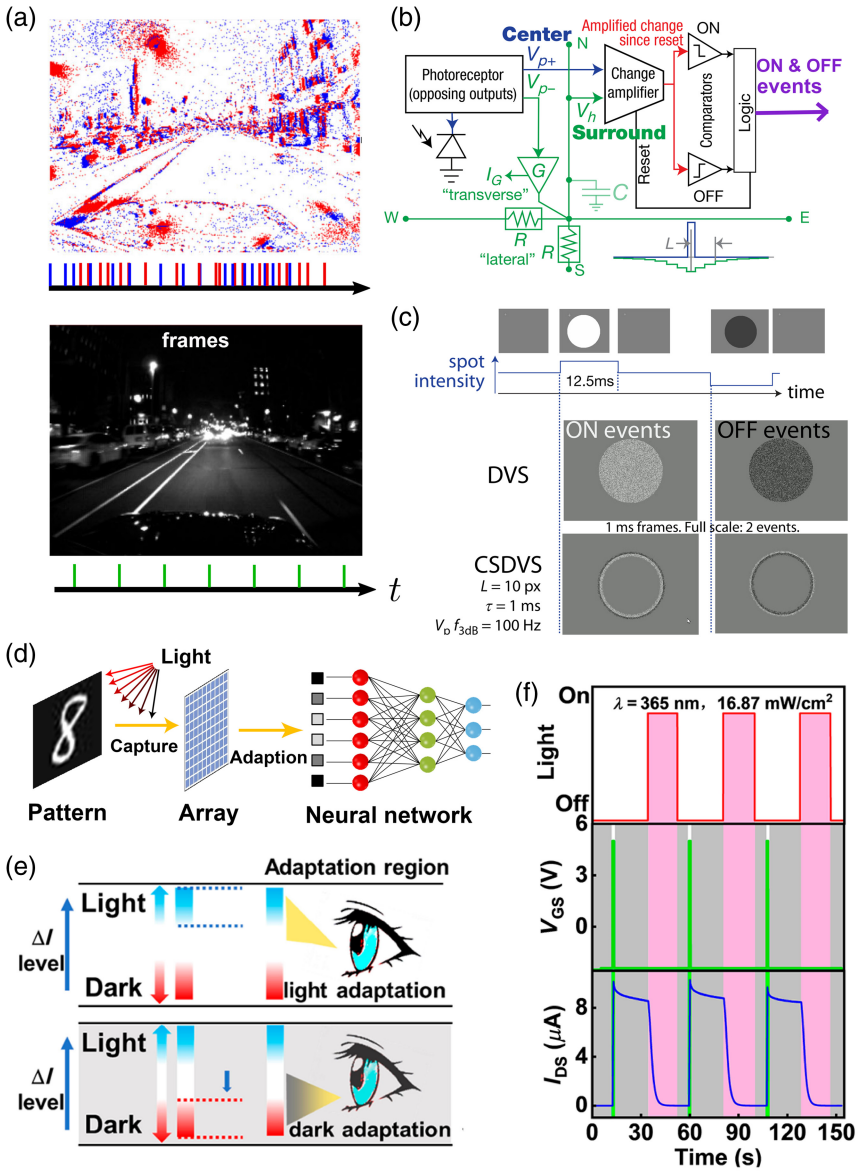

Fig. 6 Neural adaptation sensors for visual compression. (a) Event-driven sampling and frame-based sampling. Reproduced with permission from Ref. 51. (b) CSDVS pixel circuit. Reproduced with permission from Ref. 52. (c) Comparison of simulated normal DVS and CSDVS response to a flashing spot. Reproduced with permission from Ref. 52. (d) Illustration of a machine vision system based on the MoS_2 phototransistor array. Reproduced with permission from Ref. 53. (e) Light- and dark-adapted mechanisms of the In_2O_3 transistor. Reproduced with permission from Ref. 54. (f) Electrical enhancement and light-depression function of an In_2O_3 transistor. Reproduced with permission from Ref. 54.

sensors leads to reduced recording data remarkably. A 1-kfps (fps, frames per second) image sensor would produce about 3×10^8 frames or 5 TB of data from the equivalent resolution 128×128 pixels sensor. However, the dynamic vision sensors (DVSs) in this work only recorded 74 MB of 4-byte event data, which is a factor of 67,000 times fewer data.[55] Vitale et al.[56] first used spiking neural network (SNN) on the chip to solve a high-speed unmanned aerial vehicle (UAV) control task. The event-based vision sensors in this work can achieve up to 3 orders of magnitude better speed versus power consumption trade-off in high-speed control of UAVs compared with conventional image sensors.[56] Event-based sensors differ from conventional imaging systems in that each pixel contains electronics that allow asynchronous operation. However, since the output of the event camera is composed of a series of asynchronous events rather than actual intensity images, it is not possible to apply traditional vision algorithms for image reconstruction, so a paradigm shift is required. Rebecq et al.[57] proposed an event-based multiview stereo image reconstruction scheme that is very computationally efficient and can be run in real time on the central processing unit (CPU). Although DVS has greatly reduced the data volume by converting continuous signal inputs into sparse event outputs, the resulting spike data still face transmission and storage difficulties. Zhu et al.[58] proposed a unified lose-spike coding

framework, which for the first time uses motion patterns hidden in the distribution of spike data to design motion-fidelity coding patterns. The proposed method can further effectively compress the spike data while maintaining visual fidelity.[58] Dong et al.[59] proposed a cube-based spike coding framework for DVSs. Based on representing spatial and temporal information as spike signals to compress sensory signals, they further compressed the spike data. The average data compression achieved by this method is 2.6536 times that of the raw spike data, and the effect is far better than the traditional lossless coding algorithm.

Temporal event-based sensors can greatly reduce the amount of data in dynamic monitoring scenarios and achieve temporal domain compression. However, the spatial resolution of this kind of sensor is insufficient. Each pixel generates a spike response to the external light-intensity change independently, and the lack of position correlation among pixels results in a lot of redundant intensity information or noise in space. The human-vision system can achieve signal compression in the spatial domain through the center-surround structure in the lateral direction of the retina. To realize spatial information preprocessing, researchers first modified and upgraded the relatively mature silicon-based sensors and added pixel circuits with specific functions. Many imaginative silicon vision sensors employ transistor-based spatial and spatiotemporal filtering in the focal plane.[60–62] These devices had complex pixels and lots of transistor mismatch, which produced much fixed-pattern noise (FPN) in the output. As Fig. 6(b) shows, Delbruck et al.[52] proposed a compact and energy-efficient center surround dynamic vision sensor (CSDVS) design. The CSDVS pixel would use ∼10 fewer large analog transistors and provide a surround with a controllable size. Thus, the CSDVS design is feasible with a modest increase in pixel complexity. Combined with switching capacitance DVS change detection, FPN is also expected to be much less than in past center-surround silicon retinas. Ordinary DVS will produce ON and OFF events in the whole point but will not produce ON and OFF events outside the point [Fig. 6(c)]. However, CSDVS only generates events at the edge of the scene. At the center of the spot, the surrounding environment responds almost equally to the photoreceptor, thus suppressing events from this homogeneous region. As a result, CSDVS will amplify high spatial frequencies and significantly reduce DVS activity in uniformly and smoothly changing regions of the scene. Spatial domain and temporal domain compressions of visual information are not discrete, and in the actual application of machine vision, spatial compression and temporal compression are both necessary.

In addition to being event-driven, the human-vision system can also quickly adapt to changes in ambient light intensity, and this adaptive property is also important in the field of machine vision. Adapting to brightness changes helps to improve the perception of the visual system, which is more sensitive to detect faint changes. Liao et al.[53] demonstrated bio-inspired vision sensors that are based on molybdenum disulfide (MoS_2) phototransistors. Their MoS_2 phototransistor arrays exhibit the adaptive capabilities of the human eye, sensing images over a wide range of brightness and achieving contrast enhancement [Fig. 6(d)]. This work is expected to be applied to the field of machine vision, simplifying circuits, and processing algorithms.[53] Jin et al.[54] demonstrated an array of In_2O_3 transistors with negative photoconductivity properties, which provide a new way to create an environmentally adaptive artificial visual perception system [Fig. 6(e)]. Figure 6(f) shows the electrical enhancement and light-depression function of an In_2O_3 transistor, which can be turned on with an electrical pulse and turned off by a light reset. In different external lighting environments, the device self-adapts and adjusts the threshold within a certain range to obtain visual perception.

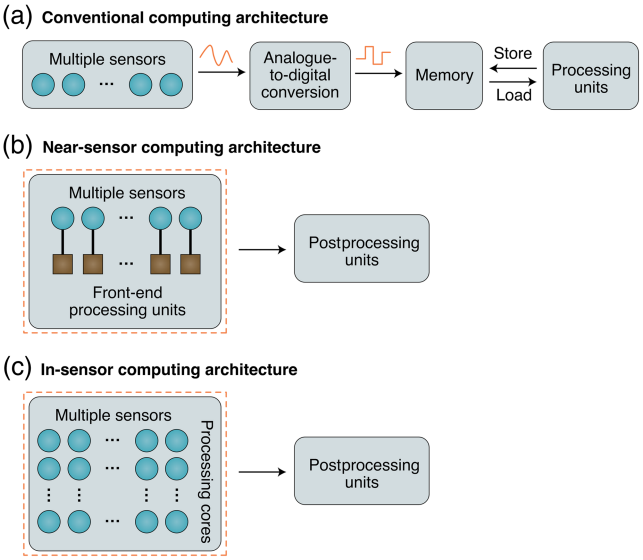

Fig. 7 Basic principles of layer-by-layer processing sensors. (a) Conventional computing architectures. (b) Near-sensor computing architecture. (c) In-sensor computing architecture. Reproduced with permission from Ref. 63.

3.4 Summary of Basic Principles of Bio-Inspired Image Sensor

In conventional architectures [Fig. 7(a)], the analog sensory data are first converted to digital signals using ADC and then temporarily stored in memory before being sent from memory to processing units. This data conversion and transmission-based approach results in inefficient power use and high latency. In a near-sensor computing architecture [Fig. 7(b)], processing units or accelerators reside beside sensors and execute specific computational tasks at sensor endpoints, providing an improved sensor/processor interface and thus minimizing the transfer of redundant data. In the in-sensor computing architecture [Fig. 7(c)], individual self-adaptive or multiple connected sensors can directly process sensory information, eliminating the sensor/processor interface and combining the sensing and computing functions.[63]

Image sensing and processing in conventional linear response and sparse coding are schematically shown in Fig. 8.[64] In a conventional linear response image sensing system (e.g., a digital camera), the light-intensity distribution at the image sensor surface is converted linearly to an electronic signal (e.g., charge or current) and then processed and stored as a digital photo. Usually, this kind of data is redundant. Sparse representation reduces the complexity of the input signals and enables more efficient processing and storage, as well as improved feature extraction and pattern recognition functions. Given a signal x, which may be a vector (e.g., representing the pixel values in an image patch), and a dictionary of features D, the goal of sparse coding is to represent x as a linear combination of features from D using a sparse set of coefficients a, while minimizing the number of features used. The objective of sparse

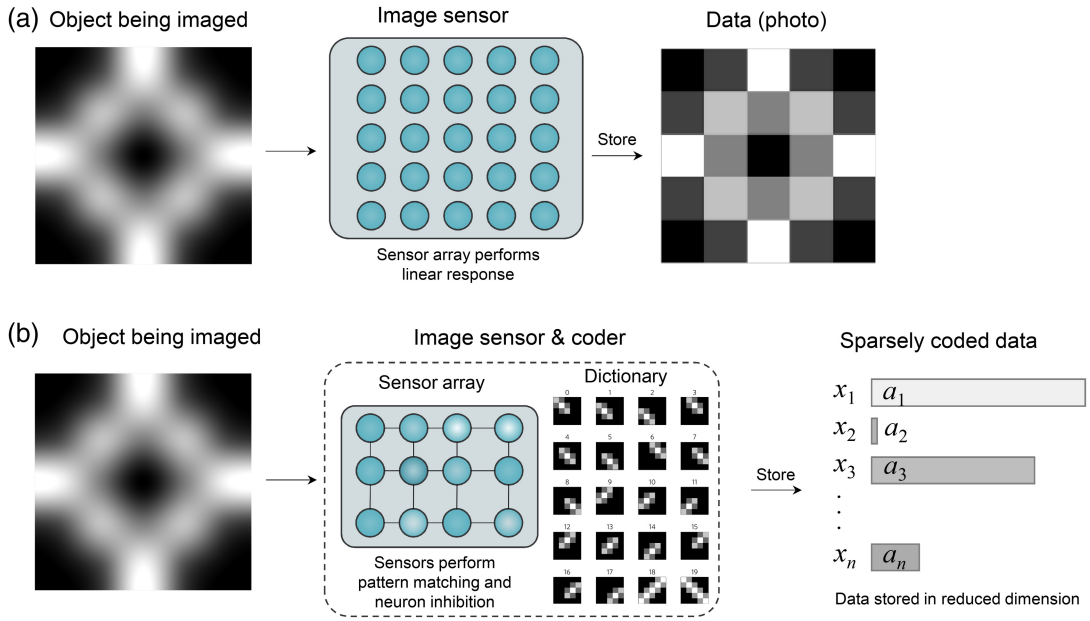

Fig. 8 Basic principles of sparse coding sensors. (a) Conventional linear response image sensing principle. (b) Image sensing and sparse coding principle.

coding can be summarized mathematically as minimizing an energy function, defined as

$$\min_a(|x - Da^T|_2 + \lambda|a|_0), \tag{1}$$

where $|\cdot|_2$ and $|\cdot|_0$ are the L^2- and the L^0-norm, respectively. Here, the first term measures the reconstruction error, which is the difference between the original signal x and the sparse representation Da^T, while the second term measures the sparsity, which is reflected by the number of active elements used to reconstruct the input. Unlike many compression algorithms that focus on reconstruction error only, sparse coding algorithms reduce the complexity by assuming that real signals lie in only a few dimensions (of a high-dimensional space) and attempt to find an optimal representation that also reduces dimensionality. As a result, sparse coding not only enables a more efficient representation of the data but may also be more likely to identify the "hidden" constituent features of the input and thus can lead to improved data analyses such as pattern recognition.

Neural adaptation processing allows the implementation of the Bayesian theories, where prior sensory experiences and current sensory input are used to compute the posterior perceptual estimation, that is, the prediction. At each processing layer, sensors subtract this prediction from the sensory data and register the residual by a "prediction error" signal (Fig. 9).[25]

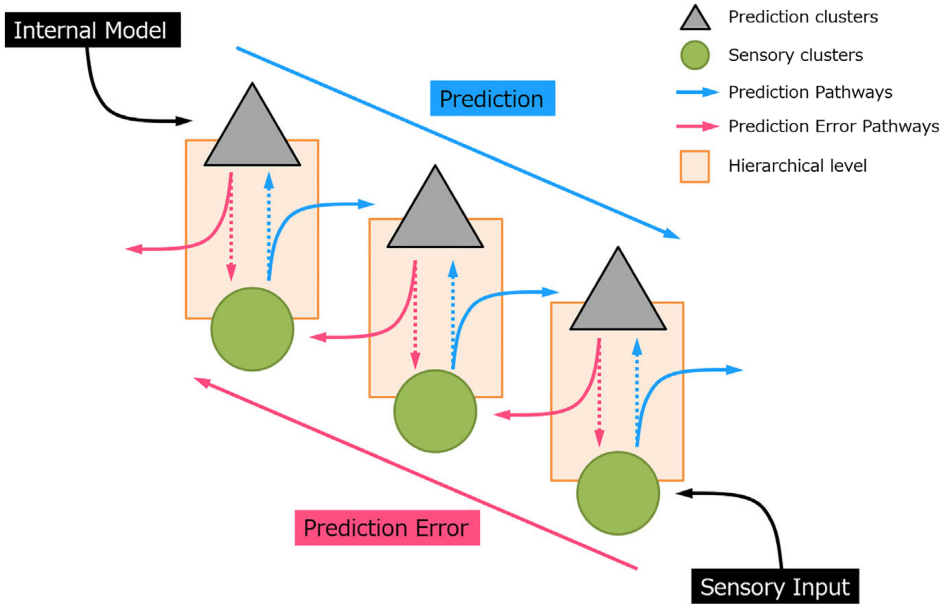

Fig. 9 Basic principles of neural adaptation sensors. Reproduced with permission from Ref. 25.

In this sense, it is not necessary for an expected event to transmit to the final processing area. Instead, only the information that deviates from predictions is further processed and passed upward to the higher-order area in the form of an "error." The greater the prediction error, the larger the sensor response evoked. This prediction error signal propagates through an ascending pathway and updates the subsequent prediction.[25] When the prediction eventually matched with the sensory input, the sensor activity induced by prediction error would be suppressed. Through the above algorithm logic, the sensing system can finally adapt to environmental factors and improve its response to environmental changes.

4 Conclusion and Perspectives

In the field of machine vision, the traditional vision computing architecture suffers from the transmission, storage, and computation of redundant data. Inspired by the efficient processing mechanism of the human-vision system, the new bio-inspired image sensors can effectively compress redundant data through layer-by-layer processing, sparse coding, and neural adaptation mechanisms to improve the computational efficiency of machine vision.

In this review, we discussed the recent advances in bio-inspired sensors for efficient machine vision. First, novel vision sensors based on layer-by-layer processing mechanisms are presented, covering the architecture of such sensors and the visual computing tasks that can be performed. Second, the high-dimensional data compression capability of sparse coding and its hardware implementation cases are exhibited. Third, the principles and functions of event-driven vision sensors and adaptive vision sensors inspired by neural adaptation are introduced in detail.

Despite considerable progress in the bio-inspired image sensors for efficient visual processing, many challenges remain to be addressed. The bio-inspired image sensor is an interdisciplinary project, involving biology, system architecture, integrated circuits, materials, devices, algorithms, and fabrication technologies. This field is in the early stages of development, and there are still many limitations. At the architectural level, process near-sensor and in-sensor architectures are demanding for integration technologies. Sensing, storage, and computing units that were originally separate are now highly integrated into a single micro-device, and it is very challenging to complete the processing and integration of various heterogeneous materials in a tiny area. In terms of coding and information-processing mechanisms, emerging algorithms must be developed to cooperate with hardware systems for high-level information processing. Although current ANNs can solve simple image processing problems, their efficiency decreases, and energy consumption increases when faced with complex tasks.

For the materials, 2D materials have outstanding advantages in spatial-temporal responses and are expected to be applied in the design of vision sensors.[65,66] However, 2D materials are still in the exploration stage, the existing 2D material transfer technology is inefficient, and the technical requirements for operators are very high, which greatly limits the wide application of 2D materials. Perovskite materials are also highly expected due to their excellent photoelectric properties. Perovskite materials have unique advantages compared with 2D materials. Perovskite materials have a direct optical bandgap, which is independent of material thickness, and therefore have a high absorption coefficient and quantum efficiency. In addition, the simple and low-cost preparation of perovskite materials makes the application prospects brighter.[67–72] 2D and perovskite materials with excellent photoelectric properties are undoubtedly promising, but the problems of preparation and preservation need to be solved. In addition, sensor arrays are generally required in practical application scenarios, so how to solve the uniformity problem of emerging materials is also an inevitable challenge.

In the long run, machine vision should approach and surpass the processing efficiency of human vision. However, the current bio-inspired vision sensors can only realize one or two efficient mechanisms of layer-by-layer processing, sparse coding, and neural adaptiveness, and it is difficult to efficiently complete complex vision tasks. Therefore, the future bio-inspired vision system needs to consider how to integrate multiple processing mechanisms, enrich the processing content of visual information, and improve the processing efficiency of visual information. The improved bio-inspired vision systems are expected to be compatible with use in real-time and low-power visual perception applications and with numerous possible applications, such as driverless cars, smart surveillance, and intelligent healthcare.

Disclosures

The authors declare no conflicts of interest.

Acknowledgments

We thank the support from Research Center for Frontier Fundamental Studies, Zhejiang Laboratory. Financial support was provided by the National Natural Science Foundation of China (Grant Nos. 92250304, 62204230, 62020106002, and T2293750), the National Key Research and Development Program of China (Grant No. 2021YFC2401403), and the Department of Science and Technology of Zhejiang Province "Leading Goose" Program (Grant No. 2022C01077).

References

1. Y. Chai, "In-sensor computing for machine vision," *Nature* **579**(7797), 32–33 (2020).
2. S. Dhawan, "A review of image compression and comparison of its algorithms," *Int. J. Electron. Commun. Technol.* **2**(1), 22–26 (2011).
3. F. Mentzer et al., "High-fidelity generative image compression," in *Adv. in Neural Inf. Process. Syst.*, Vol. 33, pp. 11913–11924 (2020).
4. L. C. Ngugi et al., "Recent advances in image processing techniques for automated leaf pest and disease recognition–a review," *Inf. Process. Agric.* **8**(1), 27–51 (2021).
5. M. J. Weinberger et al., "The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS," *IEEE Trans. Image Process.* **9**(8), 1309–1324 (2000).
6. X. Pitkow et al., "Decorrelation and efficient coding by retinal ganglion cells," *Nat. Neurosci.* **15**(4), 628–635 (2012).
7. M. F. Bear et al., "Synaptic plasticity: LTP and LTD," *Curr. Opin. Neurobiol.* **4**(3), 389–399 (1994).
8. T. Hosoya et al., "Dynamic predictive coding by the retina," *Nature* **436**(7047), 71–77 (2005).
9. T. Gollisch, "Throwing a glance at the neural code: rapid information transmission in the visual system," *HFSP J.* **3**(1), 36–46 (2009).
10. T. Gollisch et al., "Eye smarter than scientists believed: neural computations in circuits of the retina," *Neuron* **65**(2), 150–164 (2010).
11. B. A. Olshausen et al., "Sparse coding of sensory inputs," *Curr. Opin. Neurobiol.* **14**(4), 481–487 (2004).

12. T.-H. Hsu et al., "AI edge devices using computing-in-memory and processing-in-sensor: from system to device," in *IEEE Int. Electron Devices Meet. (IEDM)*, pp. 22.25.1–22.25.4 (2019).
13. K. D. Choo et al., "Energy-efficient motion-triggered IoT CMOS image sensor with capacitor array-assisted charge-injection SAR ADC," *IEEE J. Solid-State Circuit* **54**(11), 2921–2931 (2019).
14. Z. Du et al., "ShiDianNao: shifting vision processing closer to the sensor," in *Proc. 42nd Annu. Int. Symp. on Comput. Archit.*, pp. 92–104 (2015).
15. A. Jimenez-Fernandez et al., "A binaural neuromorphic auditory sensor for FPGA: a spike signal processing approach," *IEEE Trans. Neural Netw. Learn Syst.* **28**(4), 804–818 (2017).
16. P. Sterling, "How retinal circuits optimize the transfer of visual information," in *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, Eds., MIT Press, pp. 234–259 (2004).
17. J. J. O'Brien et al., "Photoreceptor coupling mediated by connexin36 in the primate retina," *J. Neurosci.* **32**(13), 4675–4687 (2012).
18. Z. Yu et al., "Toward the next generation of retinal neuroprosthesis: visual computation with spikes," *Engineering* **6**(4), 449–461 (2020).
19. N. Kruger et al., "Deep hierarchies in the primate visual cortex: what can we learn for computer vision?" *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1847–1871 (2012).
20. M. Kiselev, "Rate coding vs. temporal coding:is optimum between?" in *Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 1355–1359 (2016).
21. M. N. Shadlen et al., "Noise, neural codes and cortical organization," *Curr. Opin. Neurobiol.* **4**(4), 569–579 (1994).
22. J. Benda, "Neural adaptation," *Curr. Opin. Neurobiol.* **31**(3), R110–R116 (2021).
23. D. Lee et al., "In-sensor image memorization and encoding via optical neurons for bio-stimulus domain reduction toward visual cognitive processing," *Nat. Commun.* **13**(1), 5223 (2022).
24. M. Kim et al., "DeepPep: deep proteome inference from peptide profiles," *PLoS Comput. Biol.* **13**(9), e1005661 (2017).
25. C. Y. Fong et al., "Auditory mismatch negativity under predictive coding framework and its role in psychotic disorders," *Front. Psychiatry* **11**, 557932 (2020).
26. N. Waltham, *CCD and CMOS Sensors*, Springer (2013).
27. A. Rodríguez-Vázquez et al., "CMOS vision sensors: embedding computer vision at imaging front-ends," *IEEE Circuits Syst. Mag.* **18**(2), 90–107 (2018).
28. V. Shirmohammadli et al., "A neuromorphic electrothermal processor for near-sensor computing," *Adv. Mater. Technol.* **7**(11), 2200361 (2022).
29. M. Nazhamaiti et al., "NS-MD: near-sensor motion detection with energy harvesting image sensor for always-on visual perception," *IEEE Trans. Circuits Syst. II: Express Briefs* **68**(9), 3078–3082 (2021).
30. R. Forchheimer et al., "Near-sensor image processing: a new paradigm," *IEEE Trans. Image Process.* **3**(6), 736–746 (1994).
31. Z. Chen et al., "Processing near sensor architecture in mixed-signal domain with CMOS image sensor of convolutional-kernel-readout method," *IEEE Trans. Circuits Syst. I: Regular Pap.* **67**(2), 389–400 (2020).
32. Z. Liu et al., "NS-CIM: a current-mode computation-in-memory architecture enabling near-sensor processing for intelligent IoT vision nodes," *IEEE Trans. Circuits Syst. I: Regular Pap.* **67**(9), 2909–2922 (2020).
33. R. Wang et al., "Bio-inspired in-sensor compression and computing based on phototransistors," *Small* **18**(23), e2201111 (2022).
34. Z. Zhang et al., "All-in-one two-dimensional retinomorphic hardware device for motion detection and recognition," *Nat. Nanotechnol.* **17**(1), 27–32 (2022).
35. W. Pan et al., "A future perspective on in-sensor computing," *Engineering* **14**, 19–21 (2022).
36. H. Xu et al., "Senputing: an ultra-low-power always-on vision perception chip featuring the deep fusion of sensing and computing," *IEEE Trans. Circuits Syst. I: Regular Pap.* **69**(1), 232–243 (2022).
37. D. Ielmini et al., "In-memory computing with resistive switching devices," *Nat. Electron.* **1**(6), 333–343 (2018).
38. L. Tong et al., "2D materials-based homogeneous transistor-memory architecture for neuromorphic hardware," *Science* **373**(6561), 1353–1358 (2021).
39. L. Tong et al., "Stable mid-infrared polarization imaging based on quasi-2D tellurium at room temperature," *Nat. Commun.* **11**(1), 2308 (2020).
40. J. Zha et al., "Infrared photodetectors based on 2D materials and nanophotonics," *Adv. Funct. Mater.* **32**(15), 2111970 (2022).
41. J. K. Han et al., "A review of artificial spiking neuron devices for neural processing and sensing," *Adv. Funct. Mater.* **32**(33), 2204102 (2022).
42. J. W. Han et al., "Leaky integrate-and-fire biristor neuron," *IEEE Electron Device Lett.* **39**(9), 1457–1460 (2018).
43. J. K. Han et al., "Mimicry of excitatory and inhibitory artificial neuron with leaky integrate-and-fire function by a single MOSFET," *IEEE Electron Device Lett.* **41**(2), 208–211 (2020).
44. L. Gao et al., "NbO_x based oscillation neuron for neuromorphic computing," *Appl. Phys. Lett.* **111**(10), 103503 (2017).
45. D. Lee et al., "Various threshold switching devices for integrate and fire neuron applications," *Adv. Electron. Mater.* **5**(9), 1800866 (2019).
46. X. Zhang et al., "An artificial neuron based on a threshold switching memristor," *IEEE Electron Device Lett.* **39**(2), 308–311 (2018).
47. T. Tuma et al., "Stochastic phase-change neurons," *Nat. Nanotechnol.* **11**(8), 693–699 (2016).
48. A. Sengupta et al., "Magnetic tunnel junction mimics stochastic cortical spiking neurons," *Sci. Rep.* **6**(1), 30039 (2016).
49. L. Sun et al., "Bio-inspired vision and neuromorphic image processing using printable metal oxide photonic synapses," *ACS Photonics* **10**(1), 242–252 (2022).
50. L. Sun et al., "In-sensor reservoir computing for language learning via two-dimensional memristors," *Sci. Adv.* **7**(20), eabg1455 (2021).
51. D. Gehrig et al., "Combining events and frames using recurrent asynchronous multimodal networks for monocular depth prediction," *IEEE Rob. Autom. Lett.* **6**(2), 2822–2829 (2021).
52. T. Delbruck et al., "Utility and feasibility of a center surround event camera," in *IEEE Int. Conf. Image Process. (ICIP)*, pp. 381–385 (2022).
53. F. Liao et al., "Bioinspired in-sensor visual adaptation for accurate perception," *Nat. Electron.* **5**(2), 84–91 (2022).
54. C. Jin et al., "Artificial vision adaption mimicked by an optoelectrical In_2O_3 transistor array," *Nano Lett.* **22**(8), 3372–3379 (2022).
55. S.-C. Liu et al., "Event-driven sensing for efficient perception: vision and audition algorithms," *IEEE Signal Process. Mag.* **36**(6), 29–37 (2019).
56. A. Vitale et al., "Event-driven vision and control for UAVs on a neuromorphic chip," in *IEEE Int. Conf. Rob. and Autom. (ICRA)*, pp. 103–109 (2021).
57. H. Rebecq et al., "EMVS: event-based multi-view stereo—3D reconstruction with an event camera in real-time," *Int. J. Comput. Vis.* **126**(12), 1394–1414 (2018).
58. L. Zhu et al., "Hybrid coding of spatiotemporal spike data for a bio-inspired camera," *IEEE Trans. Circuits Syst. Video Technol.* **31**(7), 2837–2851 (2020).
59. S. Dong et al., "Spike coding for dynamic vision sensor in intelligent driving," *IEEE Internet Things J.* **6**(1), 60–71 (2019).
60. K. A. Zaghloul et al., "A silicon retina that reproduces signals in the optic nerve," *J. Neural Eng.* **3**(4), 257 (2006).
61. J. Costas-Santos et al., "A spatial contrast retina with on-chip calibration for neuromorphic spike-based AER vision systems," *IEEE Trans. Circuits Syst. I: Regular Pap.* **54**(7), 1444–1458 (2007).

62. J. A. Lenero-Bardallo et al., "A signed spatial contrast event spike retina chip," in *Proc. IEEE Int. Symp. on Circuits and Syst.*, pp. 2438–2441 (2010).
63. F. Zhou et al., "Near-sensor and in-sensor computing," *Nat. Electron.* **3**(11), 664–671 (2020).
64. P. M. Sheridan et al., "Sparse coding with memristor networks," *Nat. Nanotechnol.* **12**(8), 784–789 (2017).
65. V. K. Sangwan et al., "Multi-terminal memtransistors from polycrystalline monolayer molybdenum disulfide," *Nature* **554**(7693), 500–504 (2018).
66. V. K. Sangwan et al., "Neuromorphic nanoelectronic materials," *Nat. Nanotechnol.* **15**(7), 517–528 (2020).
67. Y. Yuan et al., "Ion migration in organometal trihalide perovskite and its impact on photovoltaic efficiency and stability," *Accounts Chem. Res.* **49**(2), 286–293 (2016).
68. A. M. Leguy et al., "The dynamics of methylammonium ions in hybrid organic–inorganic perovskite solar cells," *Nat. Commun.* **6**(1), 7124 (2015).
69. J. You et al., "Low-temperature solution-processed perovskite solar cells with high efficiency and flexibility," *ACS Nano* **8**(2), 1674–1680 (2014).
70. B. J. Kim et al., "Highly efficient and bending durable perovskite solar cells: toward a wearable power source," *Energy Environ. Sci.* **8**(3), 916–921 (2015).
71. S. F. Leung et al., "A self-powered and flexible organometallic halide perovskite photodetector with very high detectivity," *Adv. Mater.* **30**(8), 1704611 (2018).
72. V. K. Hsiao et al., "Photo-carrier extraction by triboelectricity for carrier transport layer-free photodetectors," *Nano Energy* **65**, 103958 (2019).

Wenhao Tang is currently an engineer at Zhejiang Lab. He received his bachelor's degree from the School of Materials Science and Engineering, Shanghai Jiao Tong University in 2020. He received his master's degree from the Department of Materials Science and Engineering, Southern University of Science and Technology in 2022. His research interests include color filter and bionic vision.

Qing Yang received her PhD from the College of Materials Science and Engineering, Zhejiang University in 2006. She was a visiting scientist in the Department of Materials Science and Engineering, Georgia Institute of Technology in 2009–2012. She was a visiting scientist at the University of Cambridge in 2018. Currently, she is a professor at the College of Optical Science and Engineering, Zhejiang University. Her research focuses on micro/nanophotonics, nanomaterials, and endoscopy imaging.

Leixin Meng received his PhD from the School of Physical Science and Technology, Lanzhou University, in 2017. He was a postdoctoral researcher at the School of Nuclear Science and Technology, Lanzhou University in 2018–2020. He is now an associated researcher at Zhejiang Lab. His research focuses on intelligent vision sensing and weak light detection.

Xu Liu received his DSc from L'Ecole Nationale Supérieure de Physiquede Marseille in France. He has been a professor at the College of Optical Science and Engineering, Zhejiang University since 1995. His research interests include optoelectronic display, optics and optoelectronic thin films, optical imaging, and biooptical technologies.

Biographies of the other authors are not available.