SCIENCE CHINA
Information Sciences

• REVIEW •

June 2018, Vol. 61 060421:1–060421:17
https://doi.org/10.1007/s11432-017-9303-0

Special Focus on Memristive Devices and Neuromorphic Computing

Neuromorphic vision chips

Nanjian WU[1,2,3]

[1]*State Key Laboratory for Superlattices and Microstructures, Institute of Semiconductors, Chinese Academy of Sciences, Beijing,* 100083*, China;*
[2]*Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing,* 100083*, China;*
[3]*University of Chinese Academy of Sciences, Beijing,* 100083*, China*

Received 7 October 2017/Accepted 13 November 2017/Published online 2 February 2018

Abstract The paper reviews the progress of neuromorphic vision chip research in decades. It focuses on two kinds of the neuromorphic vision chips: frame-driven (FD) and event-driven (ED) vision chips. The FD and ED vision chips are very different from each other in system architecture, image sensing, image information coding, image processing algorithm, design methodology. The vision chips can overcome serial data transmission and processing bottlenecks in traditional image processing systems. They can perform the high speed image capture and real-time image processing operations. This paper selects two typical chips from the two kinds of vision chips, respectively, and introduces their architectures, image sensing schemes, image processing processors and system operation. The FD neuromorphic reconfigurable vision chip comprises a high speed image sensor, a processing element array and self-organizing map neural network. The FD vision chip has the advantages in image resolution, static object detection, time-multiplex image processing, and chip area. The ED neuromorphic vision chip system is based on address-event-representation image sensor and event-driven multi-kernel convolution network. The ED vision chip has the advantages in fast sensing, low communication bandwidth, brain-like processing, and high energy efficiency. Finally, this paper discusses the architecture and the challenges of the future neuromorphic vision chip and indicates that the reconfigurable vision chip with left- and right-brain functions integrated in the three dimensional (3D) large-scale integrated circuit (LSI) technology becomes a trend of the research on the vision chip.

Keywords neuromorphic, vison chip, frame-driven, address-event-representation (AER), event-driven, convolution neural network, image sensor, image processing

Citation Wu N J. Neuromorphic vision chips. Sci China Inf Sci, 2018, 61(6): 060421, https://doi.org/10.1007/s11432-017-9303-0

1 Introduction

The visual system is a very important apparatus for human capturing and processing visual information. We obtain about 80% of information from outside world by the visual system. Figure 1 shows schematic diagram of the human visual system that mainly consists of eye, optic nerve and brain visual cortex. The retina in the eye can sense optic image focused through the cornea and lens, and converts the image to electrical signals by photoreceptors. Then the electrical visual image signals are sent to the brain. The brain visual cortex is responsible for processing the visual image information in parallel by a hierarchical neural network. It is attractive to apply the visual system knowledge to realize new information processing systems and new integrated microsystems. Traditional artificial vision systems consist of an imager and

Email: nanjian@red.semi.ac.cn

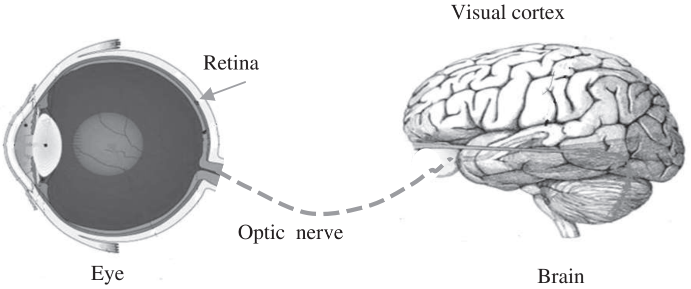

Figure 1 Schematic diagram of the human visual system.

a digital processor. There are several limitations in the systems. First, the serial data transmission scheme results in the image transferring bottleneck and large power dissipation. Second, it is not easy to perform iterative image processing algorithms at high speed by serial processing processor. To overcome these limitations, a neuromorphic vision chip based on Si semiconductor integrated circuit technology was proposed by Mead [1] and Aizawa [2].

The neuromorphic vision chip integrates an imager and brain-inspired parallel-processors on a single Silicon die. It mimics the human visual system in a way that the image sensor and the processors perform functions as the retina and brain visual cortex, respectively: (1) the image sensor captures an image and outputs image data; and (2) the image processors perform intelligent image processing in parallel and output the processing result. The neuromorphic vision chip overcomes serial data transmission and processing bottlenecks in traditional image processing systems. It can speed up an image processing to a speed of higher than 1000 frames per second (fps). Furthermore, it is a typical edge-computing device that performs data processing near the source of the data and reduces the communications bandwidth needed between the sensor and the central data center. It improves the system responding performance and data security remarkably because the transfer of the image data on network is avoided. The single vision chip replaces the combination of an image sensor, transmission line and a powerful computer so that the power consumption, size and cost of the visual system are dramatically reduced. It can be applied in security monitoring, blind navigation, robotic vision, industry automation, visual object tracking, entertainment, and virtual reality/augmented reality.

Researchers have focused on the neuromorphic vision chips for about two decades and made remarkable progress. But, there are many research challenges in chip architecture, photonic/microelectronic hybrid design, high speed brain-inspired neural network, fabrication technology, on-line training algorithm. In this review, we will focus on the recent progresses and research challenges of the neuromorphic vision chips. We will firstly give brief summary of the research progress on both frame-driven (FD) and event-driven (ED) neuromorphic vision chips in Section 2, and then show some typical research results of FD and ED vision chips in Sections 3 and 4, respectively. Finally, Section 5 discusses the architecture of future neuromorphic vision chips and indicates the challenges to realize the vision chip.

2 Neuromorphic vision chips

The neuromorphic vision chips can be categorized into two types of chips: FD vision chip and ED vision chip. The FD vision chip captures the image in a frame form by an image sensor, and then transfers the image data into parallel image processors and processes the image frame by frame by brain-inspired processors, and finally outputs the processing results. On the other hand, the FD vision chip senses temporal or spatial light contrast (event) of the scene that is called as event. And the chip encodes the event in an address-event-representation (AER) scheme [3], and then sends the AER events asynchronously into parallel AER event processors. Finally it processes the AER information in spike mode by the brain-like processors and outputs processing results. The AER event generally may consist of event address, time stamp, event polarity, grey value. The brief summary of the research progress on the FD and ED vision chips will be introduced, respectively, as follows.

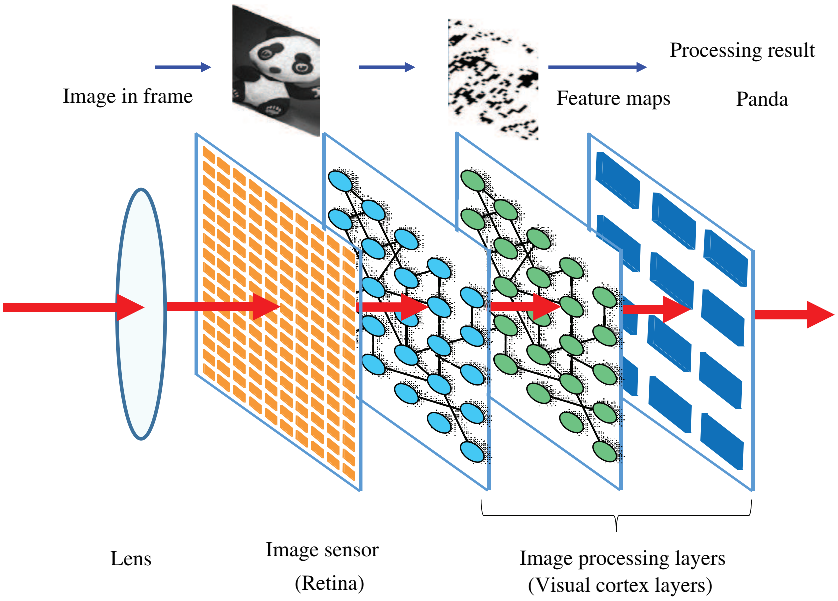

Figure 2 (Color online) Schematic diagram of the FD vision chip.

2.1 FD vision chips

Figure 2 shows the schematic diagram of the FD vision chip that consists of a CMOS image sensor and hierarchical parallel processing layers with distributed memories. The CMOS image sensor captures still image and outputs image data frame by frame. The hierarchical parallel processing layers can store the image data and implements image processing algorithms in parallel. Many FD vision chips have been reported [4–19].

The early vision chips consist of a two-dimensional (2D) array of processing elements (PEs) [6–11]. Each PE includes a photodiode pixel and processing circuit. The PEs operate in single-instruction-multiple-data (SIMD) fashion. The vision chips can be categorized into analog and digital ones. In the analog vision chip the analog image processing circuits are integrated with each pixel [4–7]. A 2D Silicon vision chip was implemented in CMOS process [4, 5]. The chips realize smoothing function by using photo-generated carrier redistribution [4] and the pattern matching by using a scanning circuit [5]. They can perform image edge extraction and moving object detection in real time. A general-purpose analog vision chip based on a switched-current analog microprocessor was implemented [6]. It can perform a fine-grain software-programmable SIMD operation at a real-time image processing speed. A vision chip specified for eye tracking was implemented [7]. It integrated a smart pixel array and winner-take-all circuit, and can address the center of the pupil. Advantage of the analog vision chip is high area and energy efficiencies.

The digital vision chip can implement more complicated image processing algorithms and is more flexible [8–15]. The digital vision chips include application-specific and general-purpose chips. The application-specific chips for range finding [9], motion detection [8], image compression [10], and target tracking [11] were reported. The chips show impressive performances but low flexibility. The general-purpose chip includes massively parallel programmable PEs with good flexibility [12–15]. The vision chip can reconfigure its hardware dynamically by chaining PEs and perform edge detection, block matching, image centroid and optical flow calculation through programming [12]. Then it was improved into a multi-SIMD architecture that can carry out more difficult image recognition tasks [12]. To overcome the difficulty in the early general-purpose chips, a novel architecture was proposed [14, 15]. It integrates a pixel-parallel PE array and row-parallel processors that can carry out pixel-parallel and row-parallel operations, respectively. It can implement many algorithms for real-time vision applications.

However, the performances of the early chips are limited for the following reasons. First, the PE array can carry out image enhancement and feature extraction operations efficiently, but it is difficult to implement image recognition operations. Second, the previous reported vision chips were based on a pixel-PE tightly-coupled architecture. Each PE integrates one pixel and one processing circuit unit so that its area is 5–20 times larger than one pixel's area [15]. The architecture suffers from low image resolution and small fill factor (<10%). Moreover, the PE-pixel mapping relationship is fixed as one-to-one so that the image processing flexibly is reduced. To overcome the difficulties, a vision chip based on multiple-level parallel processors was developed [16]. The chip integrates an image sensor, a pixel-parallel PE array, a row-parallel row processor (RP) array and a non-parallel microprocessor unit (MPU). It separates the pixel array and the PE array to beat the sufferings [15] so that the sizes of the pixel array and PE array can be determined independently and the mapping relationship between the image sensor array and the PE array is flexible. The PE array, the RP array and MPU perform low-, mid- and high-level image processing, respectively. But, MPU could not match the PE array and RP array processors so that the high-level processing speed is limited. The time of the high-level operation is much longer than the low- and mid-level image processing and larger than 1 ms. Thus, the system performance would be bottlenecked by the high-level processing according to Amdahl's Law.

A dynamically reconfigurable hybrid architecture of vision chip with PE array processor and self-organizing map (SOM) neural network was proposed and implemented [17]. It consists of a high speed CMOS imager, three von Neumann-type processors, and a non-von Neumann-type SOM neural network. The high speed CMOS imager and three von Neumann-type processors are separated physically. The processors and SOM neural network execute low-level, mid-level and high-level image processing, respectively. The SOM network increases high-level processing speed in image processing tasks and improves the chip performance remarkably. The SOM network and PE array can be dynamically reconfigured from each other within several clock cycles so that the area of the vision chip is saved. The chip can perform various-level image processing at a high speed from several hundreds to 1000 fps and in a programmable fashion. A target tracking system based on the vision processor architecture was realized [18]. The vision processor achieves over 2000 fps processing speed of the tracking algorithm with a 750 × 480 resolution camera. It can track a high-speed moving target under complex conditions. To improve the processing accuracy and efficiency, a heterogeneous parallel processor for high speed vision chip was proposed [19]. It uses patch processing unit (PPU) array processor to replace the RP array processor in [17]. Furthermore, the PE array and PPU array can be reconfigured into SOM neural network [19] and deep convolutional neural network (DCNN) [20]. The neural networks speed up pattern recognition processing and increase the processing accuracy.

Recently a novel vision chip based on pixel-neighborhood-level parallel processing was proposed [21]. The chip integrates a 2D array of neighborhood processors (NPs). Each NP consists of an 8-bit RISC processing core and an 8 × 8 array of pixels. The 8-bit RISC core is physically integrated between the pixels on the same focal plane. The processing algorithms are performed in parallel across the array of the RISC cores. It allows for direct scalability in terms of resolution without reduction in processing speed or frame rate. This chip achieves 1000 fps object tracking. On the other hand, a high resolution 3D-stacked vision chip was reported [22]. The vision chip architecture consists of a 1296 × 976 (1.27 M) pixel array, column-parallel Analog-to-Digital Converters (ADCs) and SIMD column-parallel processing elements (PEs). The 1.27 M pixel array and column-parallel ADCs with the column-parallel PEs are integrated on two separated chips, respectively. Then the two chips were stacked by the TSV process into a 3D complete vision chip. The column-parallel PEs can perform spatiotemporal image processing by 4b and 1b ALUs. The chip can be applied to track multiple targets under different conditions.

2.2 ED vision chips

Figure 3 show the schematic diagram of the ED vision chip. It consists of an AER-based image sensor and hierarchical parallel AER processing layers with distributed memories. AER image sensor is a novel type of sensor device inspired by biologic retinas that do not operate in a frame-by-frame fashion. Each pixel of the imager includes the photodiode and pre-processing circuit to compute temporal or spatial contrast and to generate the events. The sensor encodes the event in the AER information encoding scheme

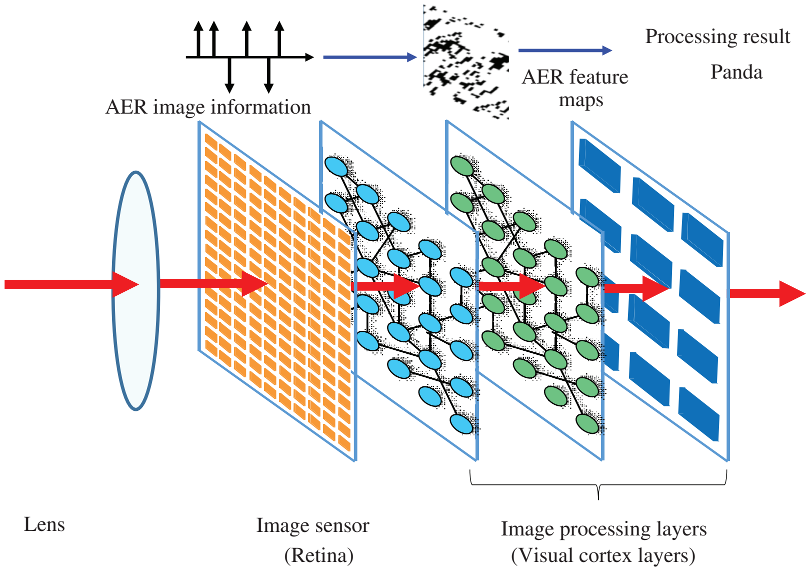

Figure 3 (Color online) Schematic diagram of the ED vision chip.

and outputs a sequence of AER events. The AER information encoding schemes include luminance to frequency transformation [23], time-to-first spike (TFS) coding [24], temporal contrast coding [25]. The sensor has the advantages of high temporal resolution, sparse representation of the scene, and asynchronous fashion. Each pixel monitors the light intensity change and asynchronously outputs AER information to AER processing layers when the change reaches a threshold. The AER information generally consists of event address, time stamp, event polarity, grey value. The hierarchical parallel AER processing layers can perform parallel event processing algorithms.

An AER image sensor has been designed and fabricated [23]. It converts optic image information directly into a sequence of spikes (pulses). The spike frequency encodes the analog signal of the pixel. The pixel signal can be reconstructed simply by counting the number of received spikes during a predetermined time window. Because the time-to-first spike (TFS) coding was considered to be rather than the frequency of the spikes, an AER imager based on TFS coding was demonstrated [24]. The pixel values are converted into TFS information, while the histogram equalization processing is implemented using a compact digital timer. Moreover, an imager based on temporal contrast AER coding was developed [25]. Each pixel independently quantizes its relative intensity changes to output spike events in a sub-millisecond timing precision. The data rate is much lower than the conventional FD imagers. The temporal contrast AER coding imager has an advantage on the binarization, since only the light change on the contours is output by the imager [26,27]. In 2007, an AER imager for multiple transient object detection was reported [28]. It detects moving or light-intensity-changing objects and sends out their locations in AER information coding form. The address-event imagers have an attractive combination of characteristics for low-latency dynamic vision under uncontrolled illumination with low post-processing requirements.

Because the address-event image sensor outputs image information event with high time resolution, the most efficient method of performing address-event image processing would be event by event, such as spike convolution neural networks, and is opposed to conventional FD image processing techniques. The first ED vision chip was reported in 1997 [29]. The cortical layer chip receives the current-to-pulse frequency event information and performs the orientation feature enhancement by the convolutional filtering. But, the AER elliptic convolutional filter in the chip was hardwired. An architecture of the ED vision chip was proposed to perform programmable 2D ED convolution with kernels of more generic shape [30]. Based on the architecture, any 2-D kernel $f(x,y)$ can be decomposed into a horizontal component $H(x)$ and a vertical component $V(y)$ such that the product can be implemented approximatively by a signed minimum operation. Later the architecture was improved and implemented [31,32]. The new architecture does not suffer from this kernel-decomposing restriction and can perform convolutions with arbitrary shape programmable kernels. In 2005, an AER vision system was reported that consists of an AER imager and multi AER Gabor convolution chips [33]. The convolution kernels on the chips are well tuned through analog biases for different orientations and spatial frequencies, respectively. But, these reported chips perform weighted event integration by analog charge packet integration on pixel capacitors and computing circuits with their transistors operating in weak inversion region such that it suffered from serious device mismatch and low computing precision. Secondly, the chips only implement a single-kernel convolution and feature extraction. They do not meet the requirements that the generic convolution system should perform various kernel convolutions for each event. Finally, it is difficult for them to carry out object recognition.

A fully-digital convolutional neural network chip was implemented for the ED vision sensing and processing systems [34]. It integrated a 32 × 32 pixel 2D event convolution processor that can perform the operation of kernels with arbitrary programmable shape and size up to 32 × 32. It can handle the forgetting mechanism that decrements the impact of much earlier events on current convolution processing. Such chips can be combined into chip arrays to carry out larger pixel array and multi-layer image processing. It improves the AER image processing precision and reduces the event flow processing latency. The chip can distinguish two simulated propeller-like shapes rotating simultaneously at a speed of 9400 rps. Later a multi-kernel convolution processing module chip for AER vision system was proposed and implemented [35]. It integrated a 64 × 64 pixel 2D convolution event processor and can handle the leak/forgetting mechanism. The pixel circuit is more compact so that the pixel area is about one quarter of that in the previous design. The module chip has multi-kernel processing capability that it detects the origin of the event and in parallel processes different kernel convolutions for multiple simultaneous input flows. Based on the key property, many of the module chips can be assembled to set up a typical hierarchical feed-forward convolutional neural network for complete AER vision processing [36]. Recently an ED feed-forward classifier for the AER imager was reported [37]. It extracts cortex-like features and classifies different patterns using a leaky integrate-and-fire (LIF) spiking neural network. But, how to train event-driven processing module is still an open research issue. An intermediate solution was proposed [38]. First, it builds an image frame database by collecting events of an AER imager during a fixed time window, and then trains a FD convolutional neural network by this database. Finally the learned parameters of the FD network are applied into an ED network. A novel methodology for training an ED classifier for the AER imager was proposed [39]. An AER image recognition system with a multiple-orientation detector for objects was proposed [40]. It can perform the AER image feature extraction, object tracking and recognition with arbitrary motion orientation.

2.3 Omparison of frame- and event-driven vision chips

The FD and ED vision chips are very different in design methodology, information coding operation mode, image processing algorithm. Table 1 compares the two kinds of the vision chips and shows their advantages and disadvantages. To give more quantitative comparison, the experiment results of the important parameters are listed in Table 1 as well.

The ED vision chip has the advantages of fast sensing, high dynamic range, sparse data, brain-like processing, and high energy efficiency. Its information coding scheme is similar to that in the human visual system. The spike or event is used to represent and transfer the image information. The AER imager that represents the luminance intensity change in the temporal domain allows each pixel to produce a larger dynamic range of outputs [23,25]. The ED vision chip performs the image processing and transfers image information event by event with higher time resolution [35,41]. Its energy efficiency is high. But, the time-multiplexing processing is difficult with ED processors as the AER signal is asynchronous and the neuron keeps its state at each instant [35]. And it is difficult to handle static object without temporal contrast. On the other hand, the FD chip has the advantages in image resolution, static object detection, time-multiplex image processing, and chip area. In FD vision chip, the photo pixel circuit is compact such that the image sensor may have higher fill factor [17,22]. By contrast, the AER imager pixel in ED vision chip integrates one pixel along with one processing circuit unit [25,41]. The area of the processing

Table 1 Comparison of FD and ED vision chips

Functions/parameter	**Frame-driven vision chip**	**Event-driven vision chip**
Information coding	Raw data	AER coding
Resolution	High	Low
	1.27 Mpixles [22]	640 × 480 [41]
Fill factor	High	Low
	$> 60\%$ [17,22]	$< 25\%$ [41]
Time resolution	Low	High
	1 ms [17]	3 ns [41]
Dynamic range	Middle	High
	80 dB [22]	120 dB [25]
Necessary bandwidth	High	Low
Moving object detection	Can handle	Can handle
Static object detection	Can handle	Difficult
Time-multiplex processing	Possible	Not possible
Chip area	Reasonable	Large
Energy efficiency	Middle	High

circuit is much larger than that of the pixel. Such architecture suffers from much smaller fill factor. Therefore, the image resolution of the FD vision chip is higher than that of the ED vision chip under the limited chip area [17]. The FD imager can sense the moving and static objects. The FD vision chip can fetch the intermediate data between the processing layers and memories. The operating speed of the modern digital circuit is much higher than that of the human neural system. Therefore it can perform the processing function of the large-scale neural network by using small-scale high speed processing circuit module in time-multiplexing scheme such that the vision chip area can be reduced significantly.

3 Neuromorphic dynamically reconfigurable vision chips

3.1 Architecture

Figure 4 presents the hybrid architecture of the typical FD vision chip [17,42]. It consists of three main parts: an image sensor with $M \times M$ pixel array, multiple-level parallel processors and SOM neural network. The multiple-level parallel processors consist of an $N \times N$ pixel-parallel PE array, an $N \times 1$ row-parallel RP array, and a dual-core MPU that are the von Neumann-type processors. The chip architecture separates the pixel array from the multiple-level parallel processors and SOM neural network. The sizes of the pixel array and PE array can be designed independently. Furthermore, the image sensor adopts the typical compact four-transistor (4T) pixel structure with high fill factor [43]. Therefore, the image resolution of the FD vision chip is higher than the ED vision chip under the limited chip area [17]. The SOM network is a non-von Neumann-type processor. The network can be trained online by the supervised learning vector quantization (LVQ) method, or trained offline by any feasible method. The image sensor can capture the raw image data and transfer it into the PE array at a high speed. The PE array, the RP array and MPU can implement the low-level, mid-level, and high-level image processing, respectively. The PE array and the RP array increase the low-level and mid-level image processing speed by $O(N \times N)$ and $O(N)$, respectively. The bio-inspired SOM neural network performs image recognition in vector-parallel fashion and can speed up the high-level image processing remarkably. The dual-core MPU is also responsible for managing the overall chip operation.

The 2D SOM neural network of $N/4 \times N/4$ neurons can be trained by a supervised LVQ fashion [44]. The neuron stores a K-component reference vector $\mathbf{RV}_{i,j} = (\mathrm{RV}_{i,j}(0), \mathrm{RV}_{i,j}(1), \ldots, \mathrm{RV}_{i,j}(K-1))^{\mathrm{T}}$. The neurons are partitioned into several non-overlapping regions, which represent different pattern classes. For each iteration in the LVQ training procedure, all the neurons simultaneously calculate the value of the sum-of-absolute-difference (SAD) between one feature vector $\mathbf{FV} = (\mathrm{FV}(0), \mathrm{FV}(1), \ldots, \mathrm{FV}(\mathrm{K}-1))^{\mathrm{T}}$

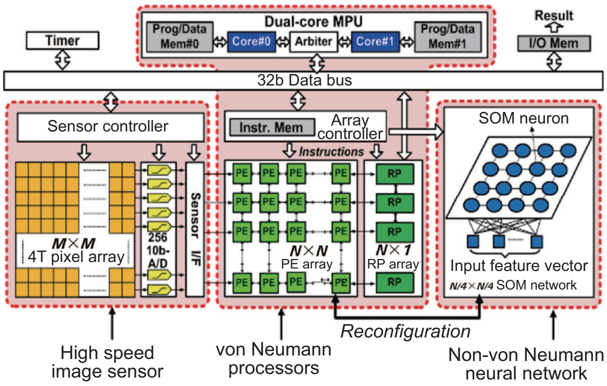

Figure 4 (Color online) Hybrid architecture of the typical FD vision chip [17, 42] @Copyright 2014 IEEE.

and their reference vectors

$$\mathrm{SAD}_{i,j} = \sum_{p=0}^{K-1} |\mathrm{FV}(p) - \mathrm{RV}_{i,j}(p)|. \tag{1}$$

The neuron with the minimum SAD is selected as the winner and its corresponding region stands for the recognized pattern class of the feature vector. If the recognized pattern class is consistent (inconsistent) with its already-known class, the reference vectors of the neurons within the winner neighborhood should be updated towards (against) the training feature vector with a learning rate α $(0 < \alpha < 1)$ by the following equation:

$$\mathrm{RV}_{i,j}(p)(\mathrm{updated}) = \mathrm{RV}_{i,j}(p) \pm \alpha[\mathrm{FV}(p) - \mathrm{RV}_{i,j}(p)], \quad p = 0, 1, \ldots, K-1. \tag{2}$$

The neighborhood radius and the learning rate are decreased gradually. After the network has finished the training procedure, it can be used to recognize new feature vectors extracted from the real-time sensor images.

The architecture has the following advantages: (1) it can effectively eliminate the bottleneck of the high-level processing for a large number of image recognition applications by integrating the 2D SOM network, and largely improve the system performance; (2) the non-von Neumann-type 2D SOM network and the von Neumann-type PE array can be reconfigured from each other dynamically; (3) the image sensor and PE array processor are separated so that the sensitivity and image quality of the image sensor is improved.

3.2 PE and neuron circuits

Figure 5 shows the PE circuit for the PE array and SOM network [17]. The PE circuit consists of a 1-bit ALU, two multiplexers op1_Mux and op2_Mux for ALU operands selection, a 1-bit temporary register T, a carry register C, a PE Memory with 1-bit data width, and eight reconfiguration multiplexers. The ALU can implement the operations of full adder, inverter, AND, and OR logic gates. The multiplexers can switch the topological connection paths between neighboring PEs and can realize different reconfiguration modes. All PEs are topologically connected in a regular 2D mesh and operate in the SIMD fashion.

The $N \times N$ PE array can be reconfigured into the 2D $(N/4) \times (N/4)$ SOM neural network. The $N \times N$ PE array is partitioned into $(N/4) \times (N/4)$ subarrays. Each sub-array consists of 4×4 PEs and becomes one neuron by switching the topological connection paths. Figure 6 shows the dynamic reconfiguration scheme between the PE array and the SOM network. The blue solid paths form a 64×64 pixel-parallel PE array. Otherwise, the red dashed paths reconfigure each sub-array into one SOM neuron. Thus the PE array is reconfigured into the 16×16 neuron SOM network. The dynamical reconfiguration can be finished in 3 clock cycles.

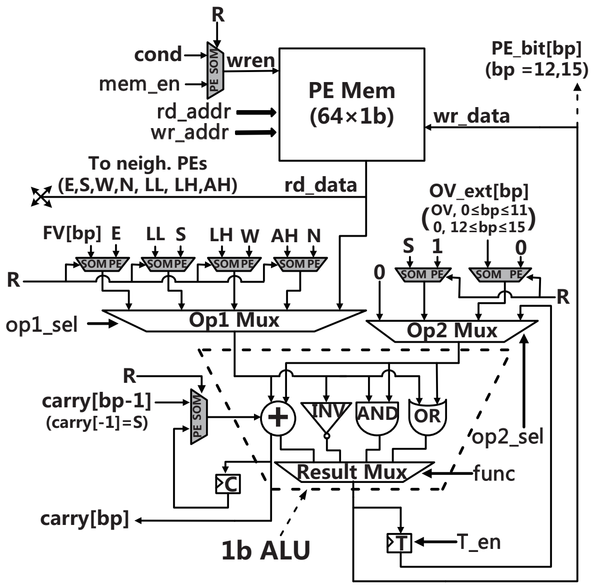

Figure 5 Schematic of PE circuit. The italic inputs are from PE instruction bits and SOM control signals issued by the array controller [17] @Copyright 2014 IEEE.

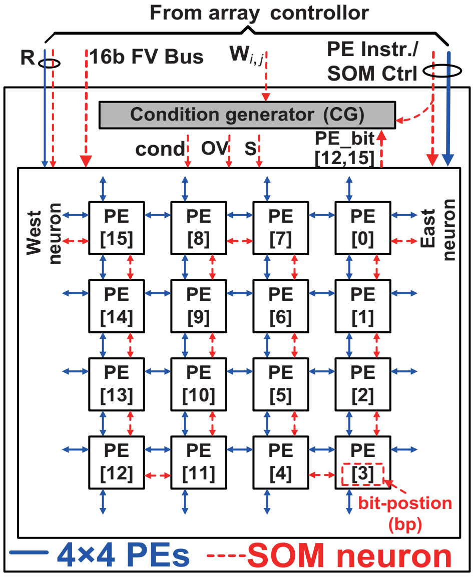

Figure 6 (Color online) The topological structures of one PE sub-array [17] @Copyright 2014 IEEE.

3.3 Implementation and results

A vision chip was implemented in a 1P5M 0.18 μm CMOS technology [17, 42]. The photograph of the chip is shown in Figure 7. The vision chip can capture the image and perform image processing at a high speed of over 1000 fps. Figure 8 shows experimental results of >1000 fps hand gesture recognition: Grasp, Yeah, Up, Fist, Palm, Index, and Down. The vision speed archives a high frame rate of more than 1400 fps and the accuracy is about 90%. The results indicated that the chip has better system performance from the image capture to real-time high-level image processing. To reduce chip area, the image processors based on full-custom distributed memory were implemented also [45].

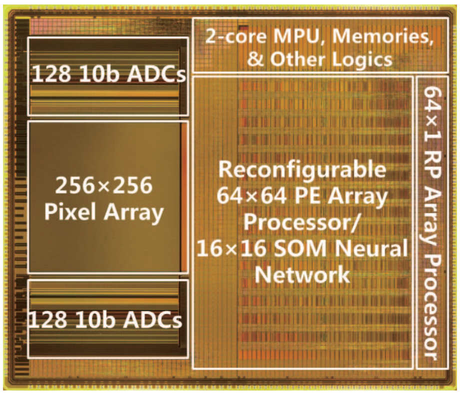
2-core MPU, Memories, & Other Logics
128 10b ADCs
256×256 Pixel Array
128 10b ADCs
Reconfigurable 64×64 PE Array Processor/ 16×16 SOM Neural Network
64×1 RP Array Processor

Figure 7 (Color online) The topological structures of one PE sub-array [17] @Copyright 2014 IEEE.

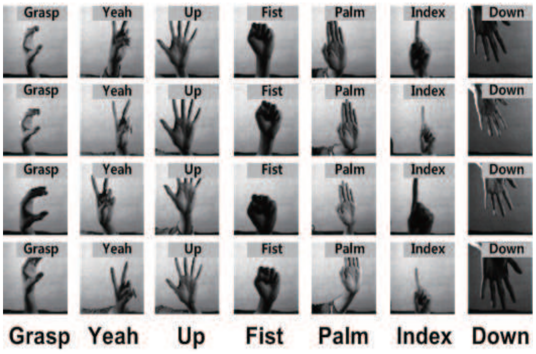

Figure 8 Experimental results of >1000 fps hand gesture recognition [17] @Copyright 2014 IEEE.

4 AER neuromorphic vision chips

4.1 System architecture

Figure 9 shows a typical AER neuromorphic vision system [35, 36]. It contains an AER imager and a sequence of convolutional neural network layers. Each network layer includes an array of convolutional modules. The convolutional module extracts specific features from the AER flow by performing specific kernel convolutions. The first network layer after the AER imager could be a Gabor filter bank that has different scales and angles for oriented segment extractions. The second network layer combines the segments extracted by the first network layer into more complex shapes. The third network layer continues to combine the simpler features into more complex and specialized ones, and so on, until the last network layer implements object recognition. The convolutional modules after the first network layer receive more than one AER event and must use different kernels to implement the convolution operations. Therefore the module should have multi-kernel capability and implement different kernel convolutions in parallel for multiple simultaneous input event flows.

4.2 AER imager

The ARE imager consists of a pixel array, preamplifier biasing cells, a set of programmable current sources, and the address event communication circuit [46]. The AER imager adopts a photodiode-processing circuit tightly-coupled architecture. The pixel consists of one photodiode, three capacitors and fifteen transistors [25, 46]. The number of the capacitors and transistors in the pixel is much more than that in 4T pixel so that the fill factor and the imager resolution are low. The address-event communication circuit generates the output addresses. The AER imager adapts Boahen's row parallel scheme to read out the AER event [23]. The implemented AER imager shows a dynamic range higher

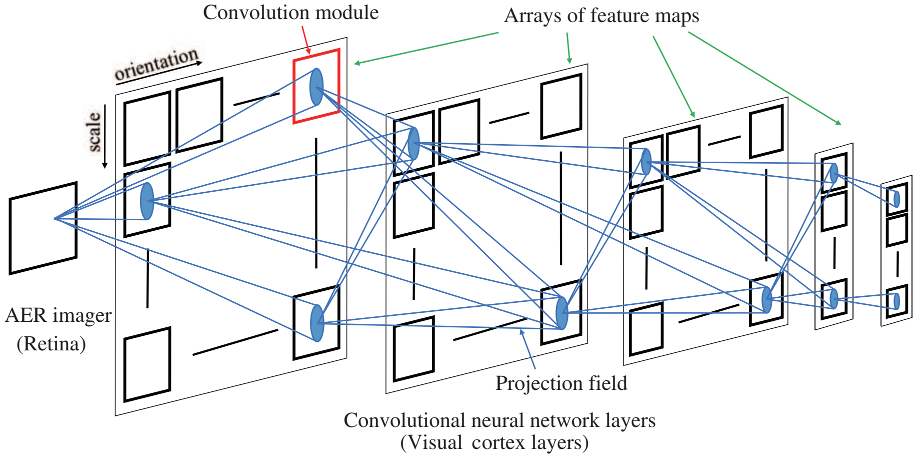

Figure 9 (Color online) A typical AER neuromorphic vision system [35] @Copyright 2012 IEEE.

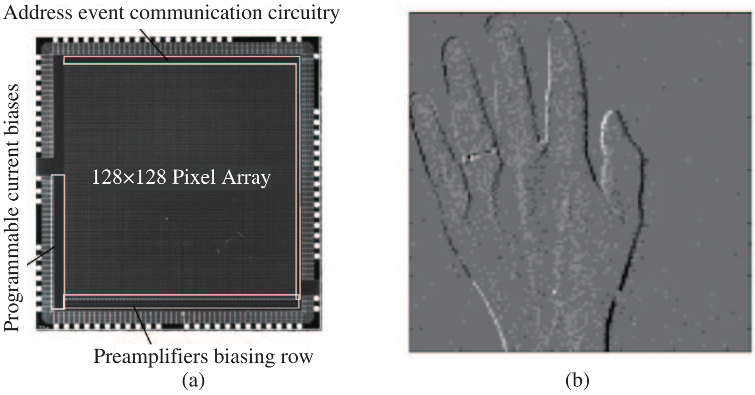
Address event communication circuitry
Programmable current biases
128×128 Pixel Array
Preamplifiers biasing row
(a)
(b)

Figure 10 (a) Microphotograph of the AER imager; (b) moving hand [46] @Copyright 2011 IEEE.

than 100 dB and captures the moving hand, as shown in Figure 10. A methodology was proposed that maps a properly trained neural network for conventional FD image processing to a neural network for ED image processing [38]. The method is demonstrated by applying ED convolutional networks trained to recognize high speed poker cards or rotating human silhouettes. The ED convolutional neural network that consists of a number of ED processing modules is simulated with a dedicated ED simulator.

4.3 Convolutional module with event-driven multi-kernels

Figure 11 shows the architecture of the event-driven multi-kernel convolutional module reported [35]. It contains an array of 64 × 64 pixels, a kernel lookup table (LUT), a kernel-RAM, a controller block, a column blocker, a 2's complement block, a left/right column shifter, and an asynchronous event readout block. The ED convolutions are implemented as follows. When an input event $(x_{\rm in}, y_{\rm in}, s_{\rm in}, k_{\rm in})$ is sent into the module, the kernel $k_{\rm in}$ in the kernel-RAM is added to or subtracted from the "Projection Field" of the pixels around the active address $(x_{\rm in}, y_{\rm in})$. Whether the kernel is added or subtracted depends on the sign $s_{\rm in}$ of the input event. Moreover, there is a constant-rate leak in all pixels. The leak will takes the states to a resting level. If the value in the pixel reaches a threshold, the pixel state is reset to its resting level, and an AER event is output. The sign of the output AER event depends on whether the threshold reached is positive or negative. The module has multi-kernel capability and in parallel implements different convolutions for multiple input events. The convolutional modules are good feature extractors and can produce the output events with the location of these features. After the features are extracted, the object can be recognized and tracked [39, 40].

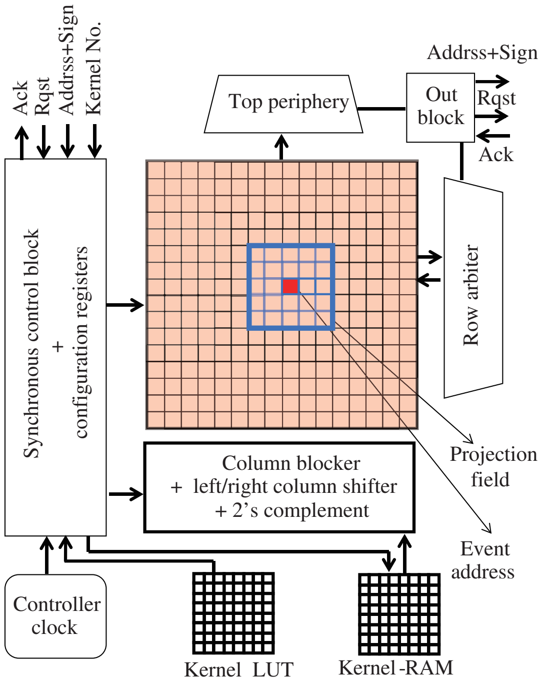

Figure 11 (Color online) The architecture of the convolutional module [35] @Copyright 2012 IEEE.

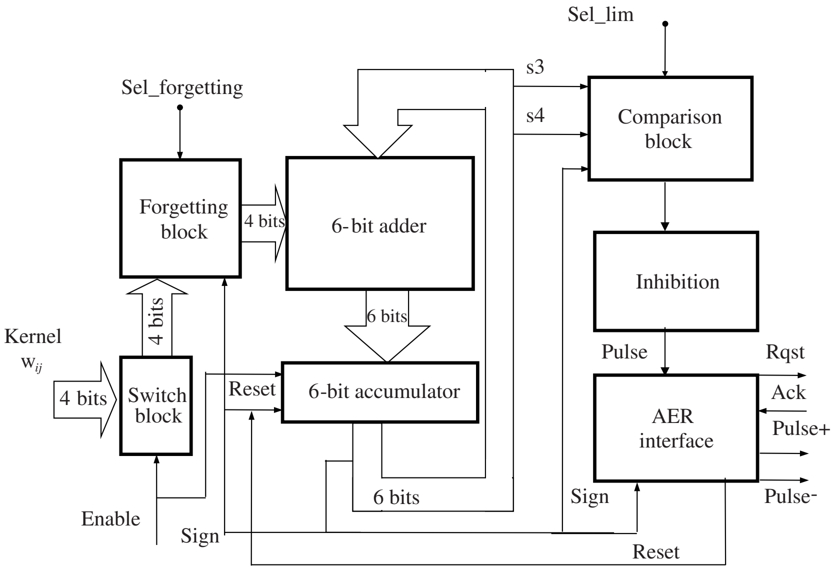

Figure 12 The block diagram of the pixel [35] @Copyright 2012 IEEE.

Figure 12 gives the schematic of the pixel circuit [35]. The kernel parameters are added to or subtracted from the 6-bit accumulator by 2's complement logic circuit. The signal "Enable" enables one-row kernel addition operation in the pixel array. The "Sel_forgetting" signal activates a fixed integer addition/subtraction for all pixels to implement the leak at a constant rate. Leak subtraction or addition depends on the accumulator sign. The comparison block checks whether the value of the 6-bit accumulator reaches the positive or negative threshold. If the value reaches the threshold, the value of the accumulator is reset to zero, and the pixel sends out a signed AER event. Figure 13 shows a chip microphotograph of the convolution module [35]. The module operates at 100 MHz. Its power consumption depends on the event traffic.

Figure 14 shows the high-speed potential of two feedforward modules for pattern recognition by simple template matching. First the AER events are filtered by a center-on convolutional module. Then the

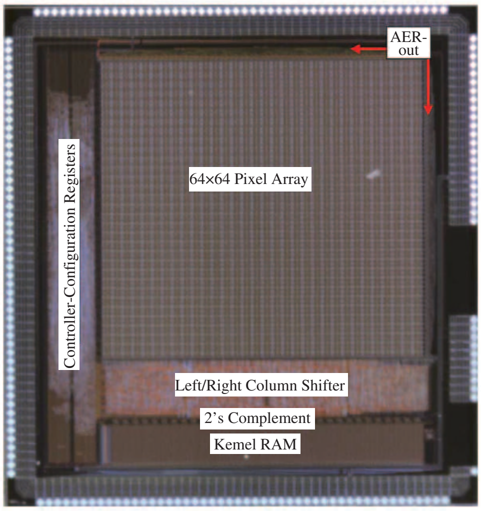

Figure 13 (Color online) A chip microphotograph of the convolution module [35] @Copyright 2012 IEEE.

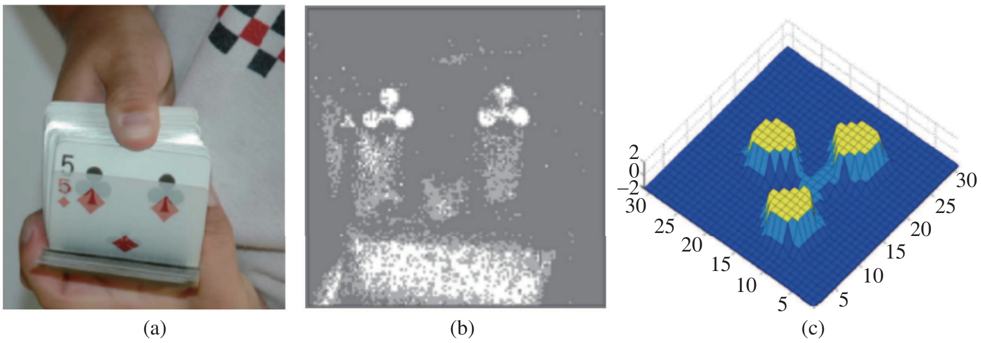

Figure 14 (Color online) Pattern recognition by simple template matching. (a) Browsing card deck; (b) 5 ms event capture image; (c) template matching kernel [35] @Copyright 2012 IEEE.

second convolutional module processes the first module's output AER events for template matching. In the experiment, a deck of 52 cards was browsed at about a rate of 8 ms per card, as shown in Figure 14(a). The AER imager outputs about 460 k AER events during 410 ms time window, while the peak of the event rate reaches 8 Meps [46]. Figure 14(b) shows an event image captured from the events sent to the first module during 5 ms time window. The output AER events of the first convolutional module are fed to the second convolutional module that is programmed with the 31×31 pixel template matching kernel, as shown in Figure 14(c). Finally, the "clover" symbols on the cards are detected.

5 Future neuromorphic vision chips

Comparing with the human binocular visual system, the vision chips rely merely on the 2D imager which loses the object depth information. However, the object depth information plays a very important role in vision processing tasks. The 3D high speed imager has great potential to remarkably improve the performance of the vision chip further in practical applications such as high speed robot vision, high speed object recognition, high speed target tracking and human computer interaction. Although the 3D vision system based on two image sensors has been developed, the system needs two image sensors and is complex. Recently the 2D/3D image sensors have been reported [47, 48]. These sensors are based on the time-of-light (TOF) principle. The sensors have such advantages as high speed, compact size, and low power consumption so that they can be integrated into the future vision chip.

The convolutional neural networks are becoming more popular in vision processing system because

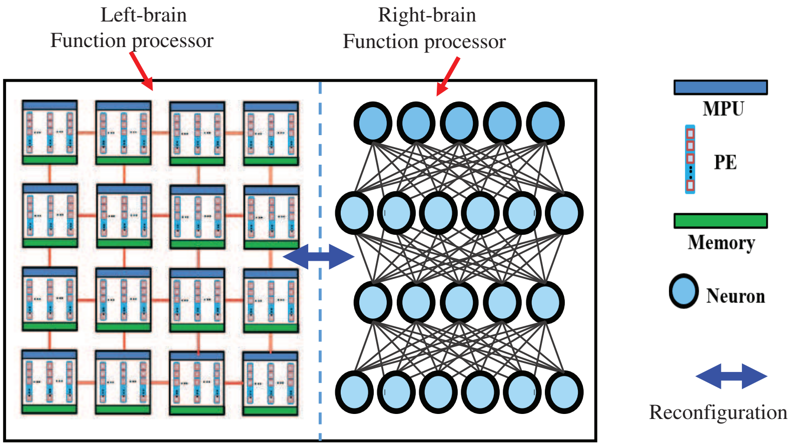

Figure 15 (Color online) Schematic of the left and right brain-function reconfigurable vision chip architecture.

of their excellent accuracy and generalization capability in object recognition, image segmentation, and scene parsing tasks. But, the neuromorphic vision chip is a typical edge-computing device that performs data processing near the source of the data. It improves the system responding performance and data security remarkably because the transfer of the image data on network is avoided. Its application scenes are different from those of the data center-oriented neuromorphic computing chips so that the size, power consumption and cost of the visual system must be dramatically reduced. Although its application scopes are limited as long as it was trained for some specified scenes, its application scopes can be expanded flexibly by retraining it for a new specific scene. One of the challenges in the future neuromorphic vision chip is to develop real-time large scale brain-inspired processor with high processing speed and high energy efficiency.

To realize the future neuromorphic vision chip, we can develop a kind of novel vision chip architecture with right and left brain-inspired functions. Figure 15 shows a schematic of the left and right brain-function reconfigurable vision chip architecture. It consists of right-brain function and left-brain function processors. The left-brain function processor consists of an array of the von Neumann-type processing cores, while the right-brain function processor is the neural network processor. The energy-efficient DCNN [49], RNN [50], SDCNN [51] or SNN [52] network, memristive neuroprocessor [53, 54] can be adopted as the right brain-inspired processor. The two kinds of the processors can be reconfigured between each other by reconfigurable circuit techniques. The novel vision chip can perform visual feature recognition and accurate visual feature information estimation by right- and left-brain function processors for different applications, respectively. How the networks can be integrated in compact area and how their energy consumption can be reduced will become the important issues. Moreover, the research on the on-line training technique of the vision chip will attract the attention of researchers [55].

Because the imager and the image processor are suitable for adopting different-node CMOS processes and the process node for the processor is more advanced than that for the imager, it is difficult to balance image processing performances and sensing abilities for diverse applications by integrating the imager and imaging processors on a single Silicon die in conventional 2D LSI technology. Figure 16 shows a concept of the future neuromorphic vision chip. It is a 3D integrated circuit chip that consists of the retina layer, brain-inspired processor layer, memory layer, power and I/O layer. The different layers can be implemented by different semiconductor processes, respectively, and then be integrated into one 3D stacked chip. Recently the examples of the 3D neuromorphic vision chips have been reported [22, 56]. This will become a trend of the research on the neuromorphic vision chip.

6 Conclusion

The paper focused on two kinds of the neuromorphic vision chips: FD and ED vision chips, and reviewed

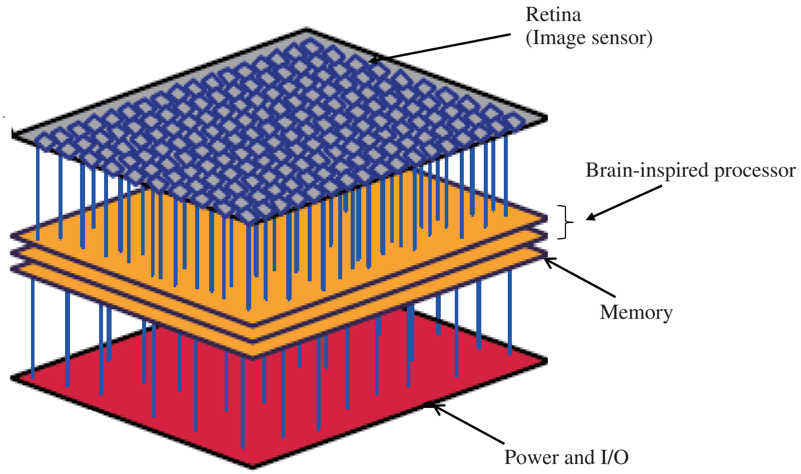

Figure 16 (Color online) Schematic of future 3D neuromorphic vision chip.

their research progress in decades. It introduced their architectures, image sensing schemes, image processing processors and system operation. The FD and ED vision chips are very different in chip architecture, image sensing scheme, information coding operation mode, design methodology, image processing algorithm. The paper showed some typical research results of FD and ED vision chips and gave the comparison of their advantages and disadvantages. The ED vision chip has the advantages in fast sensing, low communication bandwidth, brain-like processing, and high energy efficiency. On the other hand, the FD vision chip has the advantages in image resolution, static object detection, time-multiplex image processing, and compact chip area. Finally, it discussed the architecture of the future vision chip and indicated that the left and right brain-function reconfigurable vision chip integrated in 3D LSI technology becomes a trend of the research on the neuromorphic vision chip in future.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61234003, 61434004, 61504141), Brain Project of Beijing (Grant No. Z161100000216129), and CAS Interdisciplinary Project (Grant No. KJZD-EW-L11-04). The author would like to thank all members in the research group for their collaborations.

References

1 Mead C. Neuromorphic electronic systems. Proc IEEE, 1990, 78: 1629–1636
2 Aizawa K. Computational sensors — vision VLSI. IEICE Trans Inf Syst, 1999, 82: 580–588
3 Boahen K A. Communicating neuronal ensembles between neuromorphic chips. In: Neuromorphic Systems Engineering. Berlin: Springer, 1998. 229–259
4 Wu C Y, Chiu C F. A new structure of the 2-D silicon retina. IEEE J Solid-State Circ, 1995, 30: 890–897
5 Funatsu E, Nitta Y, Miyake Y, et al. An artificial retina chip with current-mode focal plane image processing functions. IEEE Trans Electron Dev, 1997, 44: 1777–1782
6 Dudek P, Hicks P J. A general-purpose processor-per-pixel analog SIMD vision chip. IEEE Trans Circ Syst I Regul Pap, 2005, 52: 13–20
7 Kim D, Cho J, Lim S, et al. A 5000S/s single-chip smart eye-tracking sensor. In: Proceedings of IEEE International Solid-State Circuits Conference — Digest of Technical Papers, San Francisco, 2008
8 Moini A, Bouzerdoum A, Eshraghian K, et al. An insect vision-based motion detection chip. IEEE J Solid-State Circ, 1997, 32: 279–284
9 Oike Y, Ikeda M, Asada K. A 375/spl times/365 high-speed 3-D range-finding image sensor using row-parallel search architecture and multisampling technique. IEEE J Solid-State Circ, 2005, 40: 444–453
10 Leon-Salas W D, Balkir S, Sayood K, et al. A CMOS imager with focal plane compression using predictive coding. IEEE J Solid-State Circ, 2007, 42: 2555–2572
11 Miao W, Lin Q Y, Wu N J. A novel vision chip for high-speed target tracking. Jpn J Appl Phys, 2007, 46: 2220–2225
12 Komuro T, Kagami S, Ishikawa M. A dynamically reconfigurable SIMD processor for a vision chip. IEEE J Solid-State Circ, 2004, 39: 265–268
13 Yamaguchi K, Watanabe Y, Komuro T, et al. Design of a massively parallel vision processor based on multi-SIMD architecture. In: Proceedings of IEEE International Symposium on Circuits and Systems, New Orleans, 2007.

3498–3501
14 Miao W, Lin Q Y, Zhang W C, et al. A programmable SIMD vision chip for real-time vision applications. IEEE J Solid-State Circ, 2008, 43: 1470–1479
15 Lin Q Y, Miao W, Zhang W C, et al. A 1000 frames/s programmable vision chip with variable resolution and row-pixel-mixed parallel image processors. Sensors, 2009, 9: 5933–5951
16 Zhang W C, Fu Q Y, Wu N J. A programmable vision chip based on multiple levels of parallel processors. IEEE J Solid-State Circ, 2011, 46: 2132–2147
17 Shi C, Yang J, Han Y, et al. A 1000 fps vision chip based on a dynamically reconfigurable hybrid architecture comprising a PE array processor and self-organizing map neural network. IEEE J Solid-State Circ, 2014, 49: 2067–2082
18 Yang Y X, Yang J, Liu L Y, et al. High-speed target tracking system based on a hierarchical parallel vision processor and gray-level LBP algorithm. IEEE Trans Syst Man Cybern Syst, 2017, 47: 950–964
19 Yang J, Yang Y X, Chen Z, et al. A heterogeneous parallel processor for high-speed vision chip. IEEE Trans Circ Syst Video Technol, 2016. doi: 10.1109/TCSVT.2016.2618753
20 Li H L, Zhang Z X, Yang J, et al. A novel vision chip architecture for image recognition based on convolutional neural network. In: Proceedings of the 11th International Conference on ASIC, Chengdu, 2015
21 Schmitz J A, Gharzai M K, Balkir S, et al. A 1000 frames/s vision chip using scalable pixel-neighborhood-level parallel processing. IEEE J Solid-State Circ, 2017, 52: 556–568
22 Yamazaki T, Katayama H, Uehara S, et al. 4.9 A 1ms high-speed vision chip with 3D-stacked 140GOPS column-parallel PEs for spatio-temporal image processing. In: Proceedings of IEEE International Solid-State Circuits Conference, San Francisco, 2017. 82–83
23 Culurciello E, Etienne-Cummings R, Boahen K A. A biomorphic digital image sensor. IEEE J Solid-State Circ, 2003, 38: 281–294
24 Chen S S, Bermak A. Arbitrated time-to-first spike CMOS image sensor with on-chip histogram equalization. IEEE Trans VLSI Syst, 2007, 15: 346–357
25 Lichtsteiner P, Posch C, Delbruck T. A 128×128 120 dB 15 µs latency asynchronous temporal contrast vision sensor. IEEE J Solid-State Circ, 2008, 43: 566–576
26 Xu J T, Zhang M X, Yan S, et al. A method to solve the side effects of dual-line timed address event vision system. J Circ Syst Comput, 2015, 24: 1550028
27 Xu J T, Zou J W, Yan S, et al. Effective target binarization method for linear timed address-event vision system. Opt Eng, 2016, 55: 063103
28 Chan V, Jin C, van Schaik A. An address-event vision sensor for multiple transient object detection. IEEE Trans Biome Circ Syst, 2007, 1: 278–288
29 Venier P, Mortara A, Arreguit X, et al. An integrated cortical layer for orientation enhancement. IEEE J Solid-State Circuits, 1997, 32: 177–186
30 Serrano-Gotarredona T, Andreou A G, Linares-Barranco B. AER image filtering architecture for vision-processing systems. IEEE Trans Circ Syst I Fund Theory Appl, 1999, 46: 1064–1071
31 Serrano-Gotarredona R, Serrano-Gotarredona T, Acosta-Jimenez A, et al. A neuromorphic cortical-layer microchip for spike-based event processing vision systems. IEEE Trans Circ Syst I Regul Pap, 2006, 53: 2548–2566
32 Serrano-Gotarredona R, Serrano-Gotarredona T, Acosta-Jiménez A, et al. On real-time AER 2-D convolutions hardware for neuromorphic spike-based cortical processing. IEEE Trans Neural Netw, 2008, 19: 1196–1219
33 Choi T Y W, Merolla P A, Arthur J V, et al. Neuromorphic implementation of orientation hypercolumns. IEEE Trans Circ Syst I Regul Pap, 2005, 52: 1049–1060
34 Camunas-Mesa L, Acosta-Jimenez A, Zamarreno-Ramos C, et al. A 32×32 pixel convolution processor chip for address event vision sensors with 155 ns event latency and 20 Meps throughput. IEEE Trans Circ Syst I Regul Pap, 2011, 58: 777–790
35 Camunas-Mesa L, Zamarreno-Ramos C, Linares-Barranco A, et al. An event-driven multi-kernel convolution processor module for event-driven vision sensors. IEEE J Solid-State Circ, 2012, 47: 504–517
36 Serrano-Gotarredona R, Oster M, Lichtsteiner P, et al. CAVIAR: a 45 k neuron, 5 M synapse, 12 G connects/s AER hardware sensory processing learning actuating system for high-speed visual object recognition and tracking. IEEE Trans Neural Netw, 2009, 20: 1417–1438
37 Zhao B, Ding R X, Chen S S, et al. Feedforward categorization on AER motion events using cortex-like features in a spiking neural network. IEEE Trans Neural Netw Learn Syst, 2015, 26: 1963–1978
38 Pérez-Carrasco J A, Zhao B, Serrano C, et al. Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing—application to feedforward convNets. IEEE Trans Pattern Anal Mach Intell, 2013, 35: 2706–2719
39 Stromatias E, Soto M, Serrano-Gotarredona T, et al. An event-driven classifier for spiking neural networks fed with synthetic or dynamic vision sensor data. Front Neuros, 2017, 11: 350
40 Wang H Y, Xu J T, Gao Z Y, et al. An event-based neurobiological recognition system with orientation detector for objects in multiple orientations. Front Neuros, 2016, 10: 498
41 Son B, Suh Y, Kim S, et al. 4.1 A 640×480 dynamic vision sensor with a 9 µm pixel and 300 Meps address-event representation. In: Proceedings of IEEE International Solid-State Circuits Conference, San Francisco, 2017. 66–67
42 Shi C, Yang J, Han Y, et al. 7.3 A 1000fps vision chip based on a dynamically reconfigurable hybrid architecture comprising a PE array and self-organizing map neural network. In: Proceedings of IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, 2014. 128–129

43 Cao Z X, Zhou Y F, Li Q L, et al. Design of pixel for high speed CMOS image sensors. In: Proceedings International Image Sensor Workshop, Snowbird, 2013, 229–232

44 Kohonen T. Self-organizing Maps. Berlin: Springer, 2001

45 Chen Z, Yang J, Shi C, et al. High speed vision processor with reconfigurable processing element array based on full-custom distributed memory. Jpn J Appl Phys, 2016, 55: 04EF08

46 Lenero-Bardallo J A, Serrano-Gotarredona T, Linares-Barranco B. A 3.6 µs latency asynchronous frame-free event-driven dynamic-vision-sensor. IEEE J Solid-State Circ, 2011, 46: 1443–1455

47 Kim S J, Kang B, Kim J D K, et al. A 1920×1080 3.65 µm-pixel 2D/3D image sensor with split and binning pixel structure in 0.11 pm standard CMOS. In: Proceedings of IEEE International Solid-State Circuits Conference, San Francisco, 2012. 396–398

48 Chen Z, Di S, Cao Z X, et al. A 256×256 time-of-flight image sensor based on center-tap demodulation pixel structure. Sci China Inf Sci, 2016, 59: 042409

49 Chen Y H, Krishna T, Emer J S, et al. Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks. IEEE J Solid-State Circ, 2017, 52: 127–138

50 Shin D, Lee J, Lee J, et al. 14.2 DNPU: an 8.1 TOPS/W reconfigurable CNN-RNN processor for general-purpose deep neural networks. In: Proceedings of IEEE International Solid-State Circuits Conference, San Francisco, 2017. 240–241

51 Cao Y Q, Chen Y, Khosla D. Spiking deep convolutional neural networks for energy-efficient object recognition. Int J Comput Vision, 2015, 113: 54–66

52 Merolla P A, Arthur J V, Alvarez-Icaza R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. Science, 2014, 345: 668–673

53 Wu H Q, Wang X H, Gao B, et al. Resistive random access memory for future information processing system. Proc IEEE, 2017, 105: 1770–1789

54 Zheng Z J, Weng J Y. Mobile device based outdoor navigation with on-line learning neural network: a comparison with convolutional neural network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, 2016. 11–18

55 Fan D L, Shim Y, Raghunathan A, et al. STT-SNN: a spin-transfer-torque based soft-limiting non-linear neuron for low-power artificial neural networks. IEEE Trans Nanotechnol, 2015, 14: 1013–1023

56 Koyanagi M, Nakagawa Y, Lee K W, et al. Neuromorphic vision chip fabricated using three-dimensional integration technology. In: Proceedings of IEEE International Solid-State Circuits Conference, San Francisco, 2001. 270–271