"**Living machines: A handbook of research in biomimetics and biohybrid systems**", T.J Prescott, N.Lepora and P.F.M.J.Verschure (Eds.), ISBN: 9780199674923, Oxford University Press, April 2018

Chapter 14 Vision

Piotr Dudek, The University of Manchester

For many animal species, vision is the primary sense through which they perceive their environment. It is also one, to which they dedicate the largest proportion of their brain resources. The early processing of visual information starts already inside the eye. From insects to vertebrates, eyes are not only sensory organs, but also include neural circuitry providing pre-processing of the visual signal right next to the photoreceptors. Massively parallel near-sensor computation allows rapid extraction of information from the visual signal, compression of redundant or less relevant data, and the creation of signal representations that are more informative, more useful, and easier to process further, than raw light intensity images. This highly efficient scheme has been an inspiration for constructing artificial vision sensors. In this chapter, the principles of operation and key features of the early stages of biological vision systems are reviewed. Several examples, ranging from devices inspired by the compound eyes of insects, through silicon circuits mimicking the structure and operation of vertebrate retinas, to more abstract interpretations of biological principles of by dynamic vision sensors and vision chips with pixel-parallel array processors are presented, and key future research directions in this area identified.

Biological principles

The sense of vision starts with signal transduction and processing of the information inside the eyes. All vertebrates, and many invertebrate species, have complex image-forming eyes, with the basic optical arrangement not dissimilar to that of a camera (Figure 14.1a). The light enters the eye through the cornea, is further focussed by the lens, and falls onto the retina, creating an image there. This image is sensed by a film of light-sensitive cells – photoreceptors - which include proteins that deform in response to incident photons, triggering a sequence of biochemical reactions that result in a change of electrical potential across the cell membrane and the rate at which the photoreceptor cells release a neurotransmitter substance. This electrochemical signal is then processed in several layers of neural cells in the retina, before being sent down the bundle of axon fibres in the optic nerve to other structures in the brain.

In vertebrates, the retinal photoreceptor cells are subdivided into *rods*, which have superior light sensitivity (capable of detecting single photons, and thus providing night-vision), but saturate in bright light, and *cones*, which provide daylight vision, with different spectral sensitivities (e.g. primates have three types of cones, with peak sensitivities in the red, blue and green parts of the light spectrum) that allow the perception of colour. The wide range of responsivity of the photoreceptors, together with adaptive circuitry that moderates the retinal outputs based on the past and surrounding image intensities, result in a very wide dynamic range of image intensities that can be detected.

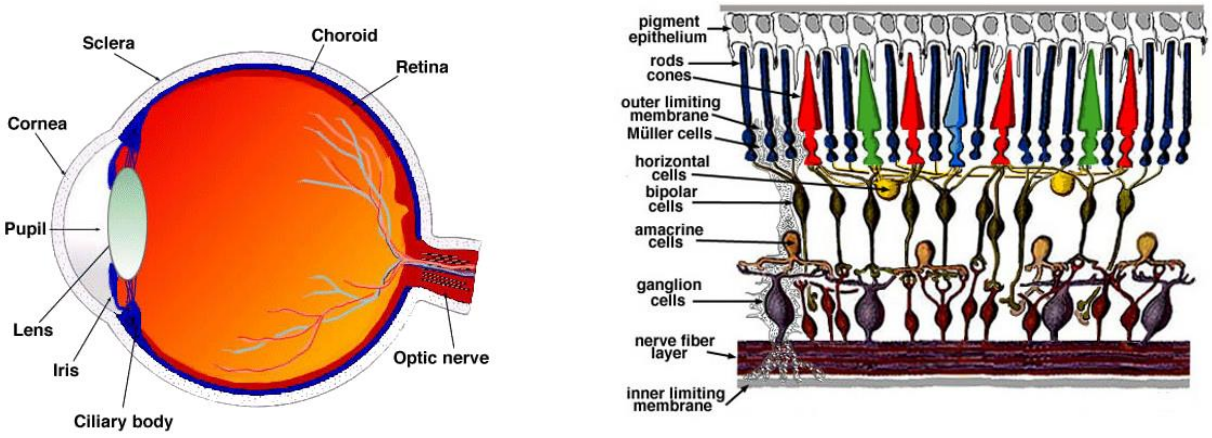
The photoreceptor concentrations are not uniform across the retina. In primates, a small central region called the *fovea* has very tightly packed cones, providing a high resolution image in a very small region of the visual field (a few degrees), while the periphery is populated mostly by rods, with fewer cones, and density of photoreceptors decreasing away from the centre region. Consequently, to resolve the fine detail in a scene, gaze has to be constantly shifting towards the locations of interest in a sequence of rapid eye movements called *saccades*, facilitated by the eye muscles, with peripheral vision identifying the regions of interests for the saccades. Gaze shifts can be also accomplished by movements of the head, and the entire body.

Insect eyes have evolved a different optical structure. The compound eyes of these animals consist of a large number of *ommatidia*, direction-sensitive light-sensing units, each consisting of a micro-lens covering a small number of photoreceptor cells; the image is formed across the compound eye as the individual ommatidia are arranged on a curved surface, each one pointing in a slightly different direction. This has the advantage of providing a large field of view in a very compact optical system. Other, simpler optical structures, such as concave mirrors and pinhole

cameras, or even more basic eyes with clusters of photosensitive cells providing rudimentary sensitivity to the intensity and direction of light can be found in less advanced animal species.

The detection of light, however, is only the first stage in the visual pathway. Eyes are not simply "cameras" that capture and transmit the images to be then processed by some distant circuitry in the brain - it is inside the eyes where the actual processing of the visual information starts. Fruit flies have very small brains (about 10^5 neurons), yet perform sophisticated vision processing that allows them to manoeuver at high flight speeds, avoid obstacles, and precisely land on a surface. Much of this ability can be attributed to the image processing carried out right next to the sensors in the eye, using specialised neural circuitry to determine the speed and direction of motion of objects in the fly's visual field.

The human brain is far more complex (about 10^{11} neurons), and of course, capable of a far more sophisticated interpretation of the visual information. Visual perception, ability to recognize objects and their relations in the environment, is a demanding task, which engages large parts of the entire brain. The signals from the retina are relayed in an orderly manner to the other parts of the brain, most notably to the *thalamus* and then onwards to the *primary visual cortex*. While the processing there is organised retinotopically (i.e. adjacent cortical regions correspond to adjacent locations on the retina), the complex networks of cortical neurons start to integrate the various types of available information over larger spatial and temporal scales. It is also there that the signals from both eyes are put together for the first time. From there, information flows to other brain regions, and back, being continually refined and interpreted in the context of other sensory information, past knowledge and experience. The full details of this process are still unknown, but neuroscience is making a steady progress towards unravelling its mysteries. Basic neural pathways and brain regions involved in various tasks that serve visual perception have been identified. Models of neural systems carrying out lower-level vision tasks such as feature, orientation and motion selectivity, contour integration, colour perception, stereopsis, etc., as well as models integrating these competencies towards the accomplishment of higher-level tasks such as object recognition and space awareness, are being continually developed and tested through various physiological, psychological and computational methods. Much of the scientific progress in understanding vision depends on, but also feeds into, our overall ability to understand the information processing principles in the brain.



(a) *(b)*

Figure 14.1. Vertebrate eye: (a) cross section of the eye, (b) structure of the retina; Figures reproduced from www.webvision.med.utah.edu.

In this chapter, however, our main focus is on how the biological principles have influenced the development of engineered vision systems at the level of sensors and early sensory information processing. The vertebrate retina develops as part of the brain, and contains neural circuitry that performs sophisticated computations. The basic structure of the retina is schematically depicted in Figure 14.1b. The information captured by the photoreceptors is processed in parallel channels, starting with *bipolar* cells that respond with increased or reduced activity, depending on their type,

to changing levels of neurotransmitter released by the photoreceptors. The bipolar cells connect to *ganglion* cells, which then send the action-potentials down their axons, which are bundled to form the optic nerve that transmits the data out of the eye. This direct pathway is modulated by *horizontal* cells, which influence the bipolar cells shaping their receptive fields through the addition of information from neighbouring photoreceptors, and *amacrine* cells providing further spatial and temporal processing of ganglion signals. As a result of this processing, a typical ganglion cell has a centre-surround receptive field that, for instance, emphasises the “edges” in the images. Distinct populations of ganglion cells provide ON signals (increased firing in response to a light patch on a dark background) and OFF signals (detecting dark patches on a light background). Colour information is similarly represented by differences in signals rather than their absolute values. Furthermore, ganglion cells respond on various timescales, with some being activated by an onset or disappearance of a light stimulus (hence detecting rapid appearance of objects and motion in the visual field), and some responding more to a sustained light stimulation. An attempt can be made at describing the operation of these circuits through combinations of spatial and temporal filters, although prevalent nonlinearities and feedback paths, non-uniform physical distribution of rods and various types of cones and their different processing circuits, together with the complex spatio-temporal characteristics of the images impinging on the photoreceptors, caused by saccades and micro-saccades (tiny oscillatory eye movements that are always present between the gaze shifts), make for a very complicated system. And, in addition to several relatively well understood neural circuits in the retina, many more exist, formed by the plethora of amacrine, horizontal and ganglion cell types, communicating through a combination of electrical and biochemical signalling, from direct electrical couplings, through rapid release of neurotransmitters in response to cellular potentials, to slow diffusion of neuromodulators across larger distances. Cells have been identified that, for example, provide a direction-of-motion response (i.e. only fire if the image on the retina moves in a particular direction), and others with a looming response (signalling an expanding object image and thus directly detecting an approaching object) (Münch *et al.*, 2009). Furthermore, the neural code used to represent the information extracted by the retina also appears complex. Ganglion cells have been found that appear to encode both the intensity and edge information using a combination of firing rate and time-to-first spike (Gollisch and Meister, 2010).

While the comprehensive description of the circuits and the function of the retina still eludes scientists, it is clear that retinas perform sophisticated pre-processing on the sensory signals, before sending the information through to the rest of the brain. Undoubtedly, all this helps to relieve the cortex of some of the workload, facilitating the brain’s amazing ability to evaluate complex visual situations in a short time. Retinal computation, together with a region-of-interest processing strategy implied by the saccades, help to reduce the volume of sensory data that needs to be processed by the cortex, and hence to make more efficient use of the limited physical resources available to carry out computations in the brain.

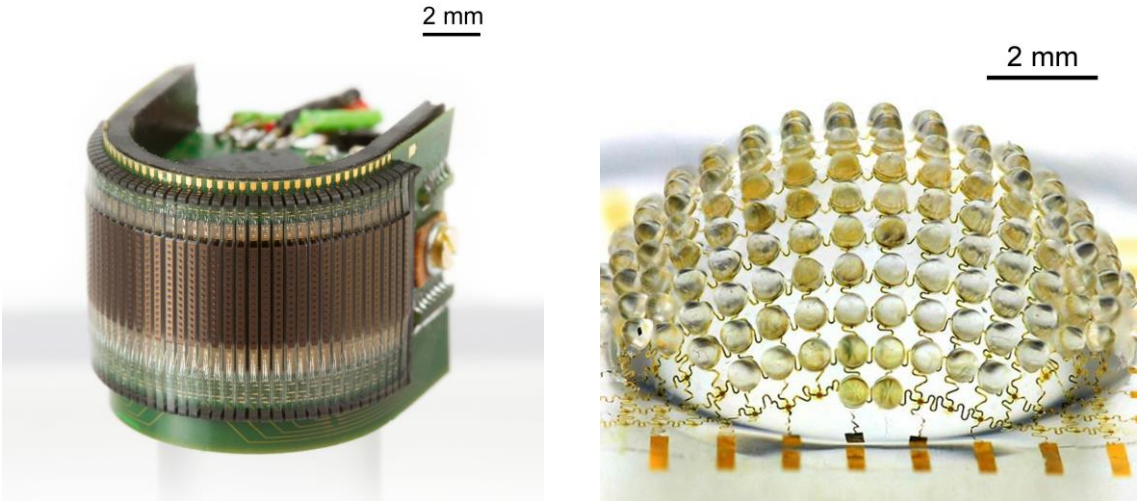
Biomimetic systems

The extraction of features of the visual signal right at the sensor level, using massively parallel networks of locally-interconnected analogue processing units operating in continuous-time and placed right next to the sensors, is a key feature of biological vision systems. The computations are localised, with the neural circuitry processing the information originating from the adjacent photoreceptors, and neurons communicating information to other neurons within a small neighbourhood. Only the results of these computations, providing more informative data than the raw light intensity information obtained by the photosensors, are transmitted onwards from the eye for further processing. This highly efficient scheme has provided cues to several engineered approaches, that try to optimise performance, power efficiency, and size of the vision system beyond what is possible when using a traditional setup of a conventional video camera feeding a central processing unit. The following section presents several examples of such systems.

Mimicking insect eyes

Insect eyes have long been an inspiration for designing vision sensors. The intricate structure of the compound eye has captured the imagination of engineers who have attempted to reproduce

the physical organisation of this system in hardware (Figure 14.2). Recent systems developed by Floreano *et al.* (2013) and Song *et al.* (2013), integrate arrays of polymer microlenses, and silicon photodetector circuits on flexible substrates, which are then mechanically curved to provide the varying direction of the optical axes and thus covering the wide-angle visual field with overlapping direction-sensitive receptive fields of individual sensor elements, much like the insect ommatidia. But it's not just the physical arrangement of the insect eye, with its compact yet providing a wide field of view optics that is intriguing from the engineering perspective. Its processing capabilities are no less remarkable. Researchers working on miniature autonomous flying robots are faced with the challenge of providing enough computational power to carry out visual information processing required for obstacle avoidance and navigation, given extremely stringent weight and power consumption budgets of the micro air vehicles. The simple yet effective way in which insects are able to extract the direction of motion information directly at the sensor level has provided the inspiration for attempting to solve this problem. The work of Barrows (2002) and others has demonstrated that through the implementation of the algorithms mimicking motion processing in the neural circuitry of insect eyes, either in conventional hardware or through specialised electronic circuitry, sufficient information can be obtained to implement basic flight manouvering without the need for a powerful central processor or a high-resolution imaging system.



(a) (b)

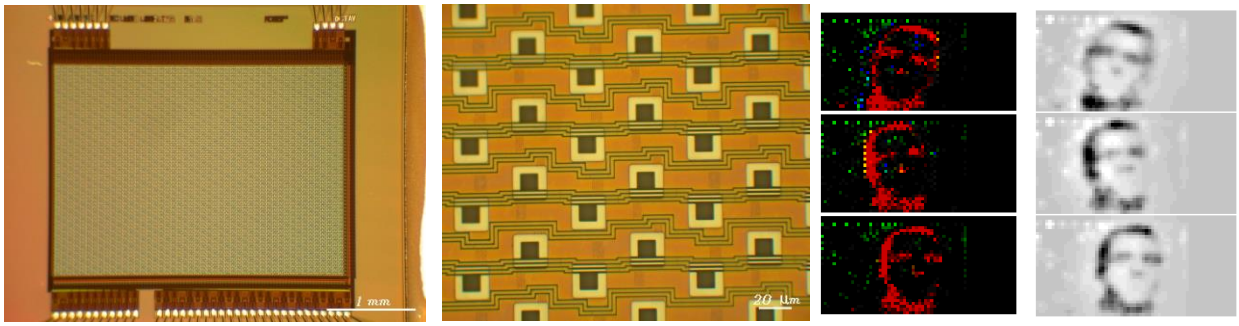
Figure 14.2. Insect-inspired vision sensors: (a) the CURVACE sensor. Photo courtesy of D.Floreano, EPFL; (b) Sensor developed by Song *et al.* (2013). Photo courtesy of J.Rogers, Univ. of Illinois.

Mimicking vertebrate retinas

The neural circuitry supporting the near-sensory processing of visual information in insects is relatively simple, and engineered systems may attempt to mimic their functionality in electronic hardware with reasonably high fidelity. Vertebrate retinas present a far more complex picture. Nevertheless, retinas were providing models for some of the very first *neuromorphic* VLSI (Very Large Scale of Integration) silicon integrated circuits developed in the late 80's and early 90's, first in Carver Mead's lab at Caltech and later in other labs worldwide. While the full repertoire of retinal responses could never be captured by these devices, and the technological limitations of the day meant they never provided more than proof-of-concept studies, rather than any practical solution to building artificial vision systems, designs by Mahowald, Indiveri, Delbruck, Liu, Boahen, Andreou, and others have pioneered the implementation of microelectronic circuits mimicking neural circuits, usually focussing on providing an electronic circuit equivalent of one particular aspect of a simplified retina model. Constructed in a manner similar to image sensor arrays, which are nowadays ubiquitous in digital cameras, and implemented in industry-standard CMOS chip technologies, these devices used pixelated photosensor (e.g. photodiode or phototransistor) matrices to transform incoming light into electrical signals. However, rather than directly outputting

thus obtained image data, they placed signal processing transistor circuitry right next to the photosensors, so that image pixels captured by the sensor could be immediately processed in-situ, with the complete silicon retina chip outputting not raw light-intensity images, but instead results of processing the images with corresponding filters, according to the specific circuit design and the implemented model. The work by Zaghloul and Boahen (2006) typifies this approach, with arguably the most comprehensive silicon retina to date (Figure 14.3). Thirteen different neuron cell types have been modelled in electronic circuitry, providing spiking responses resembling those generated by four distinct types of retinal ganglions. The authors speculate that such circuits might find applications in retinal prosthetics. In such application, the biological realism of the produced outputs might indeed be of great importance, although it remains to be seen how much of the retinal complexity needs to be replicated in order to provide a viable artificial replacement for the organ.

From the perspective of constructing artificial vision systems, however, the detailed modelling of the neurophysiology of the retinal tissue in electronic circuits would only make sense if the rest of the system could efficiently use this information to improve the overall performance. Given the complexity, as well as the still incomplete understanding of not only retinal function but also the rest of the biological visual system, this is not the case today. Nevertheless, the basic principle of performing computations right next to the sensors, and outputting not raw images but only data relevant to further processing by the system, has very clear practical advantages, as compared with a more conventional way of sensing, digitising the images, and then processing the data on the inherently sequential (single-core or even many-core) computers. The massive parallelism - thousands of hardware processing units, each processing only the immediate neighbourhood of a pixel - offers huge computational speedups. The reduction in overall data transfer requirements of the system, and efficiency with which individual pixel-level computations can be performed by dedicated circuitry, lead to dramatic power consumption savings.



(a) (b) (c) (d)

Figure 14.3. Silicon retina. (a) microphotograph of the chip; (b) close-up showing individual pixels; (c) output from the chip: three frames of a sequence are shown with colours (red, blue, gree, yellow) corresponding to firing of four distinct retinal ganglion cell types, black cells are not firing; (d) reconstructed gray-scale images. Images courtesy of K.Boahen, Stanford University.

Dynamic vision sensors

In can be argued, that functional abstractions of retinal function, at a level that well matches the capabilities of the rest of the system, are needed to fully exploit the potential suggested by the basic organisational principles of biological vision. An excellent example of how this can be achieved is provided through the work of Lichtsteiner, Posch and Delbruck (2008), and related developments of Dynamic Vision Sensors (DVS). The retinal function has been abstracted here to a very simple operation – these chips produce instantaneous binary ON and OFF responses (asynchronously firing ganglion cells) in pixels that increase or decrease in light intensity (Fig 4a). Logarithmic sensing ensures that large dynamic range of light intensities are handled internally, and the output is generated by sensing temporal contrast at individual image locations, with no lateral communication between the pixels. Most notably, the output from the chip, in the form of a

stream of addresses (array coordinates) of pixels that change, is entirely input-driven. Conventional vision systems operate with video frames, i.e. discrete images acquired at sampling frequencies dictated by the system's frame rate. For example, when processing a typical video stream of 25 fps (frames per second), a new value is obtained for each pixel every 40 ms. In contrast, the DVS system produces an asynchronous stream of pixel change "events", with latencies in the range of 1 μs, outputting only those pixels that actually change. This reduces the sensor/processor bandwidth, and facilitates, for example, high-speed motion tracking. Several robotic applications of DVS devices have been demonstrated (Fig 4c). In each case, the bulk of the vision processing relies on algorithms running on a conventional microprocessor, but whose computational load is greatly reduced by the pre-processing carried out by the DVS chip.

(a) (b) (c)

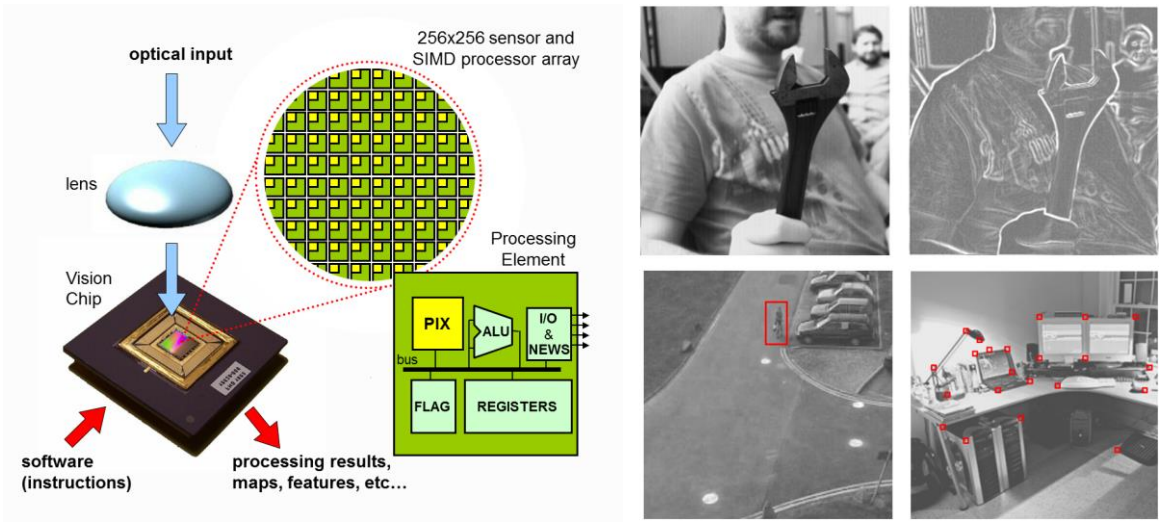
Figure 14.4. Dynamic vision sensor: (a) output of the DAVIS chip (Brandli *et al.* 2013), white pixels represent ON events, black pixels represent OFF events, generated within a short time window as the camera is panned from right to left; (b) corresponding grey-level light intensity image output from the chip; (c) a DVS-based camera (pointed by the arrow) mounted on a line-following robot. Images courtesy of T.Delbruck, INI Zurich.

Vision chips with pixel-parallel processor arrays

Retina-inspired vision chips implementing simplified functional models can provide practical solutions for building artificial vision systems, but applications of devices that carry out only one particular aspect of early visual processing (e.g. temporal contrast, detection of edges, direction of motion, etc.) are limited. The circuits for computing various functions could be integrated on a single chip, but the circuit area in a silicon pixel is limited if practical image resolutions are to be obtained. To relieve this tension between the versatility and space, a multitude of specialised continuous-time circuits (one per function) can be replaced with a programmable processing element, which contains an execution unit capable of a basic set of elementary operations and a memory. Following the Universal Turing Machine concept, the functionality of such processor is then fully described by the code it executes. This does not mean that the desirable features of the biological blueprint must be foregone. Like in retinas, computations can still be carried out in massively parallel fashion, and processing elements can be tightly integrated in physical proximity to the image sensors. Indeed, it is even possible to continue carrying out the computations in the analogue domain – while the software-programmable processor concept has been fundamental to modern digital computing, it can be also applied to a sampled-data analogue system. The vision chip described by Carey *et al.* (2013) demonstrates this approach (Figure 14.5). Each processing element contains a photosensor, arithmetic/logic unit, and memory. A mixture of digital and analogue circuits is used to achieve a balance between the robustness of digital computation and energy efficiency of analogue computing. The system operates as a SIMD (Single Instruction Multiple Data) machine, with an array of 65,536 parallel processors, one per image pixel, executing a sequence of instructions provided by a controller. Lateral connectivity between pixels allows the implementation of spatial neighbourhood filters, while in-pixel memory provides means to execute temporal operations. The chip can be thus programmed to compute spatio-temporal filters akin to retinal pre-processing, as well as more abstract maps (e.g. edge orientation, feature detection,

optic flow), in some sense also emulating early vision processing in parallel networks of the visual cortex. While the computations across the array are synchronised, and pixel data is sampled, the high speed afforded by the massively parallel processing permits very high frame rates (thousands of frames per second, if required). The information is not transmitted out of the chip at this rate; like in a retina the information can be extracted by the near-sensor circuitry, and only relevant data transmitted onwards to further processing stages. High frame rates permit data-driven, event-based output, continually generating neural 'spikes' at detected locations of interest in the visual field. The operating speed can be also traded-off for power consumption, so if lower frame rates are acceptable then ultra-low power operation is possible.

In its ability to extract information on the focal plane, vision chips can even surpass biology, being capable of producing sparse, highly informative 'events', for example describing locations of points of interest computed using elaborate feature extractors. These could be used in vision algorithms carrying out tasks such as object recognition and visual navigation, or as a peripheral system to determine regions-of-interest that should be further examined with a more elaborate, high-resolution (i.e. foveal) vision system.



(a) (b)

Figure 14.5. Vision sensor with a pixel-parallel SIMD processor array by Carey *et al.* (2013): (a) system architecture; (b) illustration of in-pixel processing capabilities, clockwise from top-left: raw image, edge detection, ROI-detection based on motion, extraction of interest points.

Future direction

The systems described in this chapter exploit the advantages of modern microelectronic fabrication technologies that allow the integration of large numbers of transistors in a small silicon area. This enables the placement of sophisticated processing circuitry in each pixel of an image sensor. The space restrictions, however, still limit the achievable pixel densities and today's state-of-the-art vision chips have modest resolutions, in the range of 256x256 pixels. Biological systems place the processing circuitry in a tightly-packed, multi-layered network, adjacent to a layer of photoreceptors. The advances in wafer-stacking silicon integration technologies might allow a more similar arrangement, and indeed some 3D integrated vision chips have been proposed, see for example (Dudek *et al*., 2009). Another advance may come from flexible circuit substrates, to replace the stiff and brittle silicon wafer. This would allow the fabrication of more compact, and more tightly integrated, insect-like compound eyes.

Superior light sensitivities of biological photoreceptor cells can be matched by photodetectors based on Single Photon Avalanche Diodes (SPADs). These, as well as other sensor types (e.g. colour, infra-red, polarisation-sensitive, etc.) have not yet been fully explored in the context of

vision chips. Animals recover depth information from binocular disparity, or estimating depth from motion (in addition to using other senses, e.g. auditory cues or echolocation). While brain-inspired systems pursuing these strategies have been investigated, engineered systems can also use methods relying on active illumination such as time of flight measurements or using structured light. These capabilities could be integrated into the vision sensors. Development of methods to fuse information from various sensor types is also a challenge.

Traditional computer vision algorithms are designed to execute efficiently on conventional sequential CPUs, or GPU-type parallel hardware, and assume the availability of a sequence of high-resolution video frames. New algorithms and approaches, and better understanding of both the advantages and the limitations of carrying out some of the processing at the sensor level, and the desirable balance between the on-sensor and off-sensor computation, are needed to fully realise the potential of bio-inspired vision sensors. This can only be achieved by future research into the application of these devices in the context of complete vision systems.

Learning more

For more in-depth information on the biology of the eye and the retina, the 'Webvision' website (http://webvision.med.utah.edu/) is an excellent resource. Gollish and Meister (2010) provide a good review of recent findings on neural computations in retinal circuits. A book on "Flying Insects and Robots" edited by Floreano *et al.* (2010) contains many contributions on state-of-the-art in insect-inspired vision. A book on "Vision Chips" by Moini (1999) charts many early attempts at building artificial retinas and bio-inspired vision chips, while several more recent vision chips with pixel-parallel processing capabilities are presented in "Focal-plane processor arrays" book, edited by Zarandy (2012).

REFERENCES

Barrows, G. L., Chahl, J. S., & Srinivasan, M. V. (2002). Biomimetic visual sensing and flight control. In Proc. Bristol UAV Conf (pp. 159-168).

Brandli C., Berner R., Yang M., Liu S.-C., and Delbruck T. (2013), A 240x180 130dB 3us Latency Global Shutter Spatiotemporal Vision Sensor, IEEE J. Solid State Circuits, submitted 2013

Carey S.J., Barr D.R.W., Lopich A. and Dudek P. (2013), A 100,000 fps Vision Sensor with Embedded 535 GOPS/W 256x256 SIMD Processor Array, VLSI Circuits Symposium 2013, Kyoto, June 2013

Dudek P., Lopich A. and Gruev V. (2009), A pixel-parallel cellular processor array in a stacked three-layer 3D silicon-on-insulator technology, European Conference on Circuit Theory and Design, ECCTD 2009, pp.193-197, August 2009

Floreano D., Zufferey J-C., Srinivasan M.V. and Ellington C. (Eds.), (2010), Flying Insects and Robots, Springer

Floreano D., et al. (2013), Miniature curved artificial compound eyes. Proceedings of the National Academy of Sciences, vol. 110, no. 23, pp 9267-9272,

Gollisch T. and Meister M. (2010), Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina, Neuron, vol. 65, pp.150-164, 28 January 2010

Lichtsteiner P., Posh C. and Delbruck T. (2008), A 128×128 120dB 15 us Asynchronous Temporal Contrast Vision Sensor. IEEE Journal of Solid-State Circuits, vol. 43, No 2, pp. 566-576

Moini A. (1999), Vision Chips, Kluwer Academic Publishers

Münch T.A., Azeredo da Silveira R., Siegert S., Viney T.J., Awatramani G.B. and Roska B. (2009), Approach sensitivity in the retina processed by a multifunctional neural circuit, Nature Neuroscience, 12 (10), pp.1308-1316, October 2009

Song Y.M., Xie Y., Malyarchuk V., Xiao J., Jung I., Choi K-J., Liu Z., Park H., Lu C., Kim R-H., Li R., Crozier K.B., Huang Y. and Rogers J.A. (2013), Digital cameras with designs inspired by the arthropod eye, Nature, vol. 497, pp. 95-99, 2 May 2013

Webvision (2013), http://webvision.med.utah.edu/ [Internet] “The Organization of the Retina and Visual System”, Kolb H, R Nelson R, E Fernandez E, Jones B editors.

Zaghloul K.A. and K Boahen K. (2006), A silicon retina that reproduces signals in the optic nerve, Journal of Neural Engineering., vol 3, no 4, pp 257-267, December 2006

Zarandy A. (2011), Focal Plane Sensor-Processor Chips, Springer