COMPARISON BETWEEN HUMAN AND SILICON RETINA

HAUPTSEMINAR NEUROENGINEERING

TECHNISCHE UNIVERSITÄT MÜNCHEN
WINTER SEMESTER 2014

SUBMITTED BY:
DHINESH KUMAR SUKUMAR

NEUROSCIENTIFIC SYSTEM THEORY
TECHNISCHE UNIVERSITAT MUNCHEN
PROF. DR JORG CONRADT

SUPERVISOR:

MARCELLO MULAS

Contents

Abstract

This paper provides a personal perspective on human vision and event-based vision sensors, algorithms, and applications over the period of last decade. The comparison between the standards of biology and technology has been the major focus throughout. Some recent advancements in the field are also briefly described.When Mahowald and Mead built the first silicon retina with asynchronous digital output around 1992[1], conventional CMOS active pixel sensors (APS) were still research chips. It required the investment by industry of about a billion dollars to bring CMOS APS to high volume production. So it is no surprise that while the imager community has been consumed by the megapixel race to make nice photos, cameras that mimic more closely how the eye works have taken a long time to come to a useful form. These "silicon retinas" are much more complex at the pixel level than APS cameras and they pay the price in terms of fill factor and pixel size; machine vision cameras with capability of synchronous global electronic shutter are about 5um. Silicon retina pixels are roughly 10 times the area of a machine vision camera pixel. So why are silicon retinas still interesting? What was the nessesity to mimic the biological eye? Mostly because of the high cost at the system level of processing the highly redundant data from conventional cameras, and the fixed latencies imposed by the frame intervals. High performance activitydriven event-based sensors could greatly benefit applications in real time robotics, where just as in nature, latency and power are very important[2] [3]. Still is this the closest that technology brought the field towards biology? What important ascpect are missing to make a bionic eye? These are the few aspects that are througly discussed throughout this paper.

[1] Mahowald, M.A.: An Analog VLSI System for Stereoscopic Vision. Kluwer, Boston (1994)
[2] Lichtsteiner, P., Posch, C., Delbruck, T.: A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. IEEE Journal of Solid-State Circuits 43(2), 566–576 (2008)
[3] . Liu, S.C., van Schaik, A., Minch, B.A., Delbruck, T.: Event-based 64-channel binaural silicon cochlea with Q enhancement mechanisms. In: IEEE ISCAS 2009, pp. 2426–2429 (2010)

1 Introduction

In spite of all the noteworthy advancement made amid the most recent decades in the fields of information technology, microelectronics, and computer science, artificial sensory and information processing systems are still substantially less powerful in dealing with real-world tasks than their biological counterparts. Even small insects outperform the most powerful computers in routine functions involving, e.g., real-time sensory data processing, perception tasks, and motor control and are obviously capable of doing all this on an incredibly small energy budget. In stark contrast to human-engineered information processing and computation devices, biological neural systems depend on an extensive number of generally basic, slow, and noisy processing elements and obtain performance and robustness from a massively parallel principle of operation and a high level of redundancy where the failure of single elements usually does not induce any observable system performance degradation. Analyzing and understanding the computational principles of the brain and how they can be used to build intelligent artificial systems are crucial for devising a new generation of neuromorphic systems, that, as the biological systems they model, are adaptive, fault tolerant and scalable, and process information using energy-efficient, asynchronous, event driven strategies.

1.1 Neuromorphic Engineering

Nature has been a source of inspiration for engineers since ancient times. In diverse fields such as aerodynamics, robotics, the engineering of surfaces and structures, or material sciences, approaches developed by nature through extensive evolutionary processes offer dramatic solutions to engineering problems. Additionally the idea of applying computational ideology of biological neural systems to artificial information processing has existed for decades. An early work from the 1940s by McCulloch and Pitts introduced a neuron model and showed that it was able to perform computation [4]

In the late 1980s, Mead at the California Institute of Technology (Caltech, Pasadena, CA, USA) introduced the ‘‘neuromorphic’’ concept to describe systems containing analog and asynchronous digital electronic circuits that mimic neural architectures present in biological nervous systems[5]. This concept revolutionized the frontier of computing and neurobiology to such a degree, to the point that a new engineering discipline emerged, whose goal is to design and build artificial neural systems, like computational arrays of synapse-connected artificial neurons, retinomorphic vision systems or auditory processors, using (micro) electrical components and circuits.

1.2 Retinomorphic vision systems

Test results recommend that adaptive analog systems are 100 times more effective in their utilization of silicon area, consume 10 000 times not as much of power than comparable

[4] A. Hodgkin and A. Huxley, ‘‘A quantitative description of membrane current and its application to conduction and excitation in nerve,’’ J. Physiol., vol. 117, pp. 500–544, 1952.
[5] C. Mead, ‘‘Neuromorphic electronic systems,’’ Proc. IEEE, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.

digital systems, and are much more robust to component degradation and failure than conventional systems[6]. But the actual biology, the human retina is an exquisitely evolved piece of neuronal wetware. It contains about a hundred million black-and-white photoreceptors, complemented by three to four million color receptors. Its output about one million axonal fibers that make up the optic nerve conveys visual information to the rest of brain using an all or none pulse code(binary).

Compared to a state-of-the-art charge-coupled-device (CCD) camera, the retina accomplishes many amazing feats. Parallel processing of visual information begins in the retina with the presence of several channels specialized for such tasks as nocturnal vision, color vision, spatial vision, and motion. Under ideal conditions, these channels allow us to detect reliably the absorption of 10 photons in a pool of 5,000 rods; to perceive color in light wavelengths ranging from 400 to 670 nm; to detect 0.5% contrast; to resolve two lines subtending an angle of 1/60 of a degree; and to tell the onset order of two lines flashed 3 to 5 milliseconds apart[7]. In addition, we can see well both in dim starlight and in bright sunlight a dynamic range of over 10 decades! In contrast, though an 8-bit CCD camera's 0.4% full-scale amplitude resolution comes close to matching the retina's contrast sensitivity, the electronic camera's 1/5-degree angular resolution and its 30-ms temporal resolution are an order of magnitude worse. Its 50-dB dynamic range is six orders of magnitude shore[8].

We can thus move forward the state of the art in vision systems by studying the intensifying body of knowledge gathered by neurobiologists about how the retina works. Retinomorphic vision systems use neurobiological principles to accomplish at the pixel level all four major operations of biological retina such as Continuous sensing for detection, local automatic gain control for amplification, spatiotemporal band-pass filtering for pre-processing, adaptive sampling for quantization.

1.3 Silicon Retina

The output of conventional cameras is organized as a matrix and copies slightly the function of the human eye. Thus, all pixels are addressed by coordinates, and the images are sent to an interface as a whole, e.g., over Camera link. Monochrome cameras deliver grayscale images where each pixel value represents the intensity within a defined range. Color sensors additionally deliver the information of the red, green, and blue spectral range for each pixel of a camera sensor matrix.

A different approach to conventional digital cameras and stereo vision is to use bio-inspired transient sensors. These sensors, called Silicon Retina, are developed to benefit from certain characteristics of the human eye such as reaction on movement and high dynamic range. Instead of digital images, these sensors deliver on and off events which represent the brightness changes of the captured scene. Due to that, new approaches of stereo matching are needed to exploit these sensor data because no conventional images can be used. The silicon

[6] M. A. C. Maher, S. P. Deweerth, M. A. Mahowald, and C. A. Mead, ‘‘Implementing neural architectures using analog VLSI circuits,’’ IEEE Trans. Circuits Syst., vol. 36, no. 5, pp. 643–652, May 1989.
[7] K. A Boahen, 'A retinomorphic vision system', *IEEE Micro,* vol. 16, no. 5, pp. 30-39, 1996.
[8] K. A. Boahen, ‘Neuromorphic microchips’,Sci. Amer., vol. 292, pp. 56–63, May 2005.

retina sensor differs from monochrome/color sensors in the case of chip construction and functionality. These differences of the retina imager can be compared with the principle operation of the human eye.

2 Biological Retina

The retina, initiating some 600 million years ago as an assembly of some light sensitive neural cells and further developed during a long evolutionary process, is the place where acquisition and first stage of processing of the visual information happens. The retina is a filmy piece of tissue, barely half a millimeter thick, that lines the inside of the eyeball. The tissue develops from a pouch of the embryonic forebrain, and the retina is therefore considered part of the brain. This most important part of the eye has a basic structure similar to that of a three-layer cake, with the bodies of nerve cells arrayed in three rows separated by two layers packed with synaptic connections. The retina includes both the sensory neurons that respond to light and intricate neural circuits that perform the first stages of image processing; ultimately, an electrical message travels down the optic nerve into the brain for further processing and visual perception.

2.1 Anatomy & Physiology

Understanding the anatomy of the primate retina is essential to understanding its function. Again, the photoreceptors lie in a layer against the back of the eyeball. In the second of three cell layers, called the inner nuclear layer, lie one to four types of horizontal cells, 11 types of bipolar cells and 22 to 30 types of amacrine cells[9]. The numbers vary depending on species. The surface layer of the retina contains about 20 types of ganglion cells.

Impulses from the ganglion cells travel to the brain via more than a million optic nerve fibers. The spaces separating these three layers are also anatomically distinct. The region containing synapses linking the photoreceptors with bipolar and horizontal cell dendrites is known as the outer plexiform layer; the area where the bipolar and amacrine cells connect to the ganglion cells is the inner plexiform layer.

Decades of anatomical studies have shed light on how the retina works. Staining techniques have revealed electrical junctions between cells and the identity and location of neurotransmitter receptors and transporters. The horizontal and amacrine cells send signals using various excitatory and inhibitory amino acids, catecholamine, peptides and nitric oxide. Electrophysiological investigations of the retina started 60 years ago. Researches of the optic nerve fibers showed that they could be stimulated to give conventional depolarizing action potentials, like those observed in other neurons. However, the first recordings of impulses within the retina by Gunnar Svaetichin in the 1950s showed very odd responses to light[10]. Neurons in the outer retina it was not immediately clear which cells he was recording from responded to stimulation not with depolarizing spikes but with slow hyper polarization.

[9] H. Kolb, 'Amacrine cells of the mammalian retina: Neurocircuitry and functional roles', Eye, vol. 11, no. 6, pp. 904-923, 1997.
[10] J. Heckenlively and G. Arden, Principles and practice of clinical electrophysiology of vision. Cambridge, Mass.: MIT Press, 2006, pp. 957-958.

As shown in the Figure 1 The retina has three major functional classes of neurons. Photoreceptors (rods and cones) lie in the outer nuclear layer, interneuron (bipolar, horizontal, and amacrine cells) in the inner nuclear layer, and ganglion cells in the ganglion cell layer. Photoreceptors, bipolar cells, and horizontal cells make synaptic connections with each other in the outer plexiform layer. The bipolar, amacrine, and ganglion cells make contact in the inner plexiform layer. Information flows vertically from photoreceptors to bipolar cells to ganglion cells, as well as laterally via horizontal cells in the outer plexiform layer and amacrine cells in the inner plexiform layer[11].

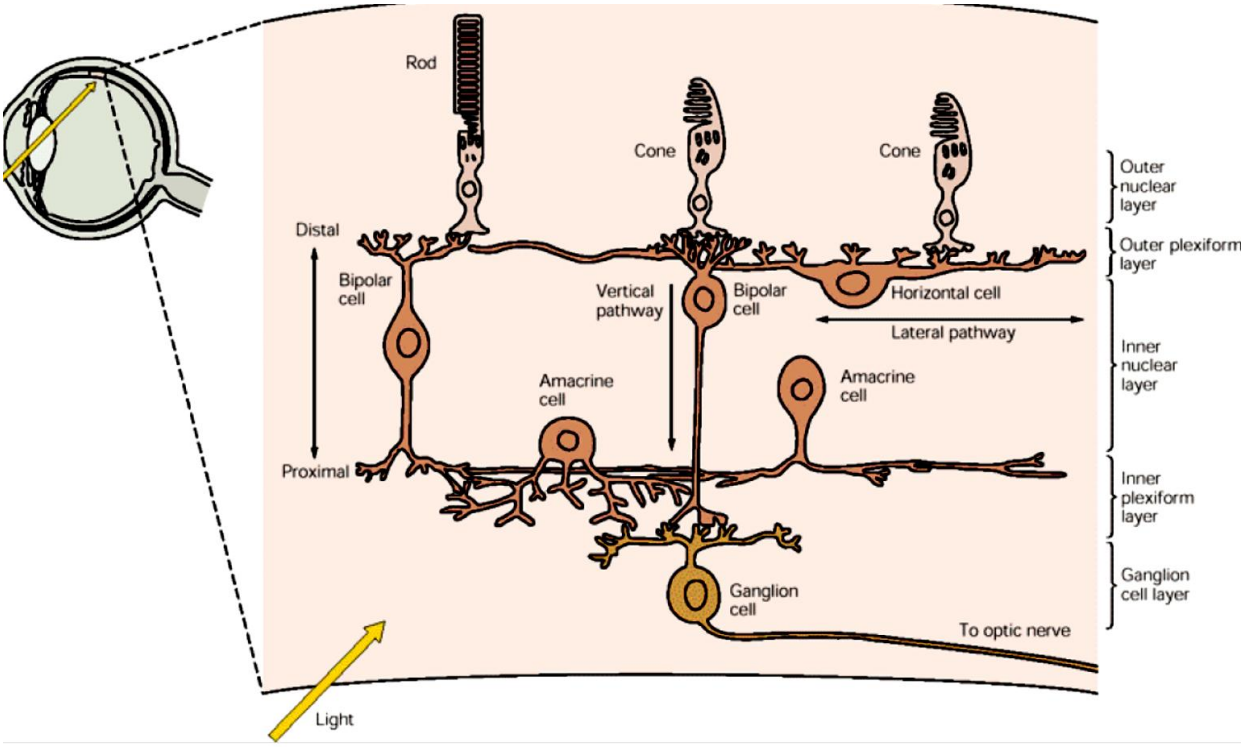

Figure 1 The retina has three major functional classes of neurons.

2.2 Parallel Processing

Our Retina processes diverse 'images' of the outside world to the brain, an picture of contours (line drawing), a RGB image (water color painting) or an picture of moving objects (motion picture). This is frequently coined as parallel processing, and begins from the most basic, the first synapse of the retina, the cone pedicle. At this point, the molecular composition of the transmitter receptors of the postsynaptic neurons characterizes which pictures are exchanged to the inner retina. Within the second synaptic layer the inner plexiform layer circuits that involve complex inhibitory and excitatory furthermore excitatory connections speak to channels that select the information to be passed to the brain[12].

The two noteworthy classes of bipolar cells, the on bipolar cells and the off bipolar cells, independently code for bright contrast and dark contrast changes. They do this by contrasting

[11] Dowling JE. 1979. Information processing by local circuits: the vertebrate retina as a model system. In: FO Schmitt, FG Worden (eds). The Neurosciences: Fourth Study Program, pp. 163-181. Cambridge, MA: MIT Press.
[12] S. W. Kuffler, ‘‘Discharge patterns and functional organization of mammalian retina,’’ J. Neurophysiol., vol. 16, pp. 37–68, 1953.

the photoreceptor signals with spatiotemporal midpoints processed by the horizontally associated layer of horizontal cells, which structure a resistive lattice. The horizontal cells are joined with one another by conductive pores called gap junctions and are associated with bipolar cells and photoreceptors in mind boggling triad synapses. Together with the data current delivered at the photoreceptor synapses, this system figures low-passed duplicate of the photoreceptor yields. The horizontal cells criticism onto the photoreceptors to help set their working focuses furthermore register a smoothed duplicate of the visual input. The bipolar cells are successfully determined by contrasts between the photoreceptor and horizontal cell yields. In the much more unpredictable outer plexiform layer, the on and off bipolar cells synapses onto numerous sorts of amacrine cells and numerous sorts of on and off ganglion cells in the inner plexiform layer. The horizontal and amacrine cells intervene the sign transmission transform between the photoreceptors and the bipolar cells, and the bipolar cells and the ganglion cells, individually.

The bipolar and ganglion cells can be further separated into two separate groups: cells with more sustained reactions and cells with more transient reactions. These cells convey data along no less than two parallel pathways in the retina: the magno cellular pathway, where cells are delicate to worldly changes in the scene, and the parvo cellular pathway where cells are touchy to structures in the scene. This picture of a basic parcel into sustained and transient pathways is excessively straightforward; truly, there are numerous parallel pathways figuring numerous perspectives (presumably no less than 50 in the mammalian retina) of the visual input. In the accompanying, a rearranged perspective of biological vision that is achievable for silicon coordinated circuit focal-plane execution is exhibited.

2.3 Building Images with Amacrine Cells

Amacrine cells arrive in a surprising mixed bag of shapes and utilize a great number of neurotransmitters. There may be well in excess of twenty separate sorts. They all have in like manner, first and foremost, their area, with their cell bodies in the center retinal layer and their courses of action in the synaptic zone between that layer and the ganglion cell layer; second, their associations, connecting bipolar cells and retinal ganglion cells and in this way framing an option, aberrant course in the middle of them; and, at long last, their absence of axons, adjusted for by the capacity of their dendrites to end pre-synaptically on different cells. Amacrine cells appear to have a few diverse capacities, a significant number of them obscure: one sort of amacrine appears to have influence in particular responses to moving articles found in retinas of frogs and rabbits; an alternate sort is mediated in the way that connections ganglion cells to those bipolar cells that get rod input. Amacrines are not known to be included in the center-surround association of ganglion-cell receptive fields, however we can't discount the likelihood. This leaves the vast majority of the shapes unaccounted for, and it is most likely reasonable to say, for amacrine cells as a rule, that our insight into their life structures far exceeds our understanding of their capacity[13].

[13] B. Stephen, 'Contribution of amacrine transmission to fast adaptation of retinal ganglion cells', Frontiers in Neuroscience, vol. 4, 2010.

Amacrine cells are about equally divided between those that use glycine and those that use GABA (gamma-amino butyric acid) neurotransmitters. Glycinergic amacrine cells are usually "small field." Their processes can spread vertically across several strata within the inner plexiform layer, but they extend relatively short distances horizontally. Glycinergic amacrine cells receive information from bipolar cells and transmit information to ganglion cells and to other bipolar and amacrine cells. Some glycinergic amacrine cells provide interconnections between ON and OFF systems of bipolar and ganglion cells. The most famous of these is called the AII cell; the AII and a GABA-releasing amacrine cell called A17 are pivotal in the circuitry of rod based, dim-light vision in the mammalian retina[14]. These cells aren't found in mammalian species that are active solely in daylight and have very few rods for example, squirrels.

In the earlier discussion of ON and OFF channels emanating from cones, cones connect in a direct pipeline to bipolar cells to ganglion cells, the bipolar cells that receive input from rods do not synapse with ganglion cells directly. The bipolar cells connected to rods are all of one type, solely transmitting an ON signal, and use the AII and A17 amacrine cells as intermediaries to get signals to ganglion cells. The small-field AII cell collects from about 30 rod-connected bipolar cells and transmits a depolarizing message both to ON (light-detecting) cone bipolar cells and to their ON ganglion cells and to OFF cone bipolar cells and OFF ganglion cells. It is as if the AII cells developed in the rod-dominated parts of the retina as an afterthought to the cone-to-ganglion cell architecture and now takes advantage of the preexisting cone pathway circuitry.

In the meantime, the a17 amacrine cell gathers rod messages from a huge number of rod-connected bipolar cells. It some way or another enhances and tweaks the data from the rod bipolar cells to transmit to the AII cells, however how it does this is not totally caught on. Regardless, the rod pathway with its arrangement of joined and after those dissimilar go-between neurons is obviously intended to gather and enhance scattered vestiges of light for twilight and night vision.

2.4 Evolution of Retina

All vertebrate retinas contain at least two types of photoreceptors the familiar rods and cones. Rods are generally used for low-light vision and cones for daylight, bright-colored vision. The variations among animal eyes reveal adaptations to the different environments in which they live. Most fish, frog, turtle and bird retinas have three to five types of cones and consequently very good color vision. The reptiles and fish are "cold blooded" and need to be active in the warm daytime. Most mammals have retinas in which rods predominate. When the number of mammals started to explode as the dinosaurs died out, the Earth was likely a dark place covered in ash and clouds; the tiny, fur-covered early mammals were able to generate their own body heat and developed visual systems sensitive to dim light. Regardless of their immense criticalness to our vision, cones make up just around 5 percent of the population of photoreceptors. The low thickness of cone photoreceptors in the peripheral

[14] A. Gallego, 'Horizontal and amacrine cells in the mammal's retina', Vision Research, vol. 11, pp. 33-IN24, 1971.

retina is very sufficient for our peripheral vision, even in daytime, as we require just moderately low spatial acuity in the periphery. Despite the fact that the incredible dominant part of peripheral photoreceptors are rods, they bend not truth be told utilized under the majority of the circumstances that we consider vision - rather, they are just utilized at exceedingly low ambient lighting levels. The explanation behind having a high density of rods is to have the capacity to catch each accessible photon under starlight conditions[15].

Most other mammalian retinas also have a preponderance of rods, and the cones are often concentrated in specialized regions. In species such as cats and dogs, images focus to a central specialized area, aptly called the area centralis, where cones predominate. The retinas of mammals such as rabbits and squirrels, as well as those of non-mammals like turtles, have a long, horizontal strip of specialized cells called a visual streak, which can detect the fast movement of predators.

Primates as well as some birds have front projecting eyes allowing binocular vision and thus depth perception; their eyes are specialized for good daylight vision and are able to discriminate color and fine details. Primates and raptors, like eagles and hawks, have a fovea, a tremendously cone-rich spot devoid of rods where images focus. Primates, in fact, have what is called a duplex retina, allowing good visual discrimination in all lighting conditions. The fovea contains most of the cones, packed together as tightly as physically possible, and allows good daylight vision.

More peripheral parts of the retina can detect the slightest glimmer of photons at night. Most mammals have two types of cones, green-sensitive and blue-sensitive, but primates have three types—red-sensitive as well as the other two. With our cone vision, we can see from gray dawn to the dazzling conditions of high noon with the sun burning down on white sand. Initially the cone photoreceptors themselves can adapt to the surrounding brightness, and circuitry through the retina can further modulate the eye’s response. Similarly, the rod photoreceptors and the neural circuitry to which they connect can adapt -to lower and lower intensity of light.

3 Silicon Retina

The first silicon VLSI retina by Mahowald and Mead implemented a model of the photoreceptor cells, horizontal cells, and bipolar cells. Each silicon photoreceptor mimics a cone cell and contains both a continuous-time photo sensor and adaptive circuitry which adjusts its response to cope with changing light level[16]. A network of MOS variable resistors mimics the horizontal cell layer, furnishing feedback based on the average amount of light striking nearby photoreceptors; the bipolar cell circuitry amplifies the difference between the signal from the photoreceptor and the local average and rectifies this amplified signal into on

[15] L. Levin, S. Nilsson, J. Ver Hoeve, S. Wu, P. Kaufman and A. Alm, Adler's Physiology of the Eye. London: Elsevier Health Sciences, 2011, pp. 430-432.
[16] K. A Boahen, 'A retinomorphic vision system', IEEE Micro, vol. 16, no. 5, pp. 30-39, 1996.

and off outputs. The response of the resulting retinal circuit approximates the behavior of the human retina[17].

Zaghloul and Boahen implemented simplified models of all five layers of the retina on a silicon chip starting in 2001[18]. This parvo–magno retina is a very different type of silicon retina that was focused on modeling of both outer and inner retinas including sustained (parvo) and transient (magno) types of cells, based on histological and physiological findings. It is an improvement over the first retina by Mahowald and Mead which models only the outer retina circuitry, that is, the cones, horizontal cells, and bipolar cells[19]. The parvomagno design captures key adaptive features of biological retinas including light and contrast adaptation, and adaptive spatial and temporal filtering. By using small transistor log domain circuits that are tightly coupled spatially by diffuser networks and single-transistor synapses, they were able to achieve 5760 phototransistors at a density of 722 per mm2 and 3600 ganglion cells at a density of 461 per mm2 in a 3.5 _ 3.3-mm2 silicon area, using a 0.25-_m complementary MOS (CMOS) process.

The outer retina's synaptic interactions realize spatiotemporal band pass filtering and local gain control. The model of the inner retina realizes contrast gain control (the control of sensitivity to temporal contrast), through modulatory effects of wide-field amacrine cell activity. As temporal contrast increases, their modulatory activity increases, the net effect of which is to make the transient ganglion cells respond more quickly and more transiently. This silicon retina outputs spike trains that capture the behavior of on- and off-center wide-field transient and narrow-field sustained ganglion cells, which provide 90% of the primate retina's optic nerve fibers. These are the four major types that project, via thalamus, to the primary visual cortex. Clearly, the parvo–magno retina has complex and interesting properties, but the extremely large mismatch between pixel response characteristics makes it quite difficult to use.

3.1 Adapting to Technology: Address–Event Representation

Even though we observe striking parallels between VLSI hardware used to implement neuromorphic devices and neural wetware, some approaches taken by nature cannot be adopted in a feasible way. One prominent challenge posed is often referred to as the ''wiring problem.'' Mainstream VLSI technology does not allow for the dense 3-D wiring observed everywhere in biological neural systems. In vision, the optic nerve connecting the retina to the visual cortex in the brain is formed by the axons of the about one million retinal ganglion cells, the spiking output cells of the retina. Translating this situation to an artificial vision system would imply that each pixel of an image sensor would have its own wire to convey its data out of the array. Given the restrictions posed by chip interconnect and packaging technologies, this is obviously not a feasible approach.

[17] T. Delbruck and C. Mead, ''Adaptive photoreceptor circuit with wide dynamic range,'' in Proc. IEEE Int. Symp. Circuits Syst., 1994, vol. 4, pp. 339–342.
[18] K. Zaghloul and K. Boahen, 'Optic Nerve Signals in a Neuromorphic Chip I: Outer and Inner Retina Models', IEEE Transactions on Biomedical Engineering, vol. 51, no. 4, pp. 657-666, 2004.
[19] M. A.Mahowald and C. A. Mead, ''The silicon retina,'' Sci. Amer., vol. 264, no. 5, pp. 76–82,May 1991

However, VLSI technology does offer a workaround. Leveraging the five orders of magnitude or more of difference in bandwidth between a neuron (typically spiking at rates between 10 and 1000 Hz) and a digital bus enables engineers to replace thousands of dedicated point-to-point connections with a few metal wires and lots of switches, and to time multiplex the traffic over these wires using a packet-based or ‘‘event-based’’ data protocol called address–event representation (AER). AER was originally proposed more than 20 years ago in Mead’s Caltech research lab[20]. For over ten years, AER sensory systems were reported by only a handful of research groups, examples being Lazzaro et al. and The Johns Hopkins University (Baltimore, MD, USA)[21]pioneering work on audition, or Boahen’s early developments on retinas. However, during these years, some basic progress was made.

A better understanding of asynchronous design, leading to robust un-arbitrated and arbitrated asynchronous event readout, combined with the availability of user friendly field-programmable gate array (FPGA) external support for interfacing and new sub micrometer technologies allowing complex pixels in reduced areas, heralded a new trend in AER bio-inspired spiking sensor developments. Since 2003, many researchers have embraced this trend and AER has been widely used with retinomorphic vision sensors, in auditory systems and even for systems distributed over wireless networks[22].

3.2 Encoding Images in the Time Domain

First, let us have a closer look at one species of ‘‘event-based’’ cameras that use AER to encode and transmit pixel luminance data. As in biology, these devices encode luminance in the time domain, i.e., in the timing or rate of spike ‘‘events.’’ Yet these devices are not ‘‘event driven’’ in the sense that their pixels autonomously react to visual events in the scene. From an engineering point of view, time-domain encoding of visual information has technical merits as it optimizes the photo transduction individually for each pixel by abstaining from imposing a fixed integration time for all pixels in an array. Exceptionally high DR and improved signal-to-noise ratio (SNR) as compared to conventional imaging techniques are immediate benefits of this approach[23]. In particular, DR is no longer limited by the power-supply rails as in standard CMOS active pixel sensors, thus providing relative immunity to the aggressive supply-voltage scaling of modern CMOS technologies. The so called ‘‘octopus retina’’ encodes and communicates individual pixel intensities in the instantaneous frequency (or inter spike intervals) of AER events emitted concurrently by each pixels. In contrast to conventional, serially scanned arrays that allocate an equal portion of the bandwidth to all pixels independently of activity, this biologically inspired readout method offers activity-driven, pixel-parallel readout. In the octopus sensor, brighter pixels are favored because their

[20] H. Kurino et al., ‘‘Smart vision chip fabricated using three dimensional integration technology,’’ in Advances in Neural Information Processing Systems 13, T. Leen, T. Dietterich, and V. Tresp, Eds. Cambridge, MA, USA: MIT Press, pp. 720–726, 2000.
[21] G. Cauwenberghs, N. Kumar, W. Himmelbauer, and A. G. Andreou, ‘‘An analog VLSI chip with asynchronous interface for auditory feature extraction,’’ IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process., vol. 45, no. 5, pp. 600–606, May 1998.
[22] T. Teixeira, A. G. Andreou, and E. Culurciello, ‘‘Event-based imaging with active illumination in sensor networks,’’ in Proc. IEEE Int. Symp. Circuits Syst., 2005, pp. 644–647.
[23] D. Chen, D. Matolin, A. Bermak, and C. Posch, ‘‘Pulse modulation imagingVReview and performance analysis,’’ IEEE Trans. Biomed. Circuits Syst., vol. 5, no. 1, pp. 64–82, Feb. 2011.

integration threshold is reached faster than darker pixels, thus their AER events are communicated at a higher frequency.

Consequently, brighter pixels request the output bus more often than darker ones and are updated more frequently. The rather large fixed-pattern noise (FPN) makes the sensor hard to sell for conventional imaging. The octopus sensor has the advantage of a smaller pixel size compared to other AER retinas, but has the disadvantage that the bus bandwidth is allocated according to the local scene luminance. Because there is no reset mechanism and the event interval directly encodes intensity, a dark pixel can take a long time to emit an event, and a single highlight in the scene can saturate the AER communication bus. Another drawback of this approach is the complexity of the digital frame grabber required to count the spikes produced by the array. The buffer must either count events over some time interval, or hold the latest spike time and use this to compute the intensity value from the inter spike interval to the current spike time. This, however, leads to a noisier image. The octopus retina would probably be most useful for tracking small and bright light sources.

A biologically inspired approach based on relative spike timing is implemented in the so-called ‘‘time-to-first spike’’ (TTFS) imager[24]. In this encoding method, the system does not require the storage of large number of spikes since every pixel generates only one spike per frame. This coding method was also suggested as a scheme used by neurons in the visual system to code information. The global threshold for generating a spike in each pixel can be reduced over the frame time so that dark pixels will still emit a spike in a reasonable amount of time. A disadvantage of the TTFS sensors is that uniform parts of the scene all try to emit their events at the same time, overwhelming the AER bus. This problem can be mitigated by serial reset of the, e.g., rows of pixels, but, of course, then the problem can arise that a particular image still causes simultaneous emission of events.

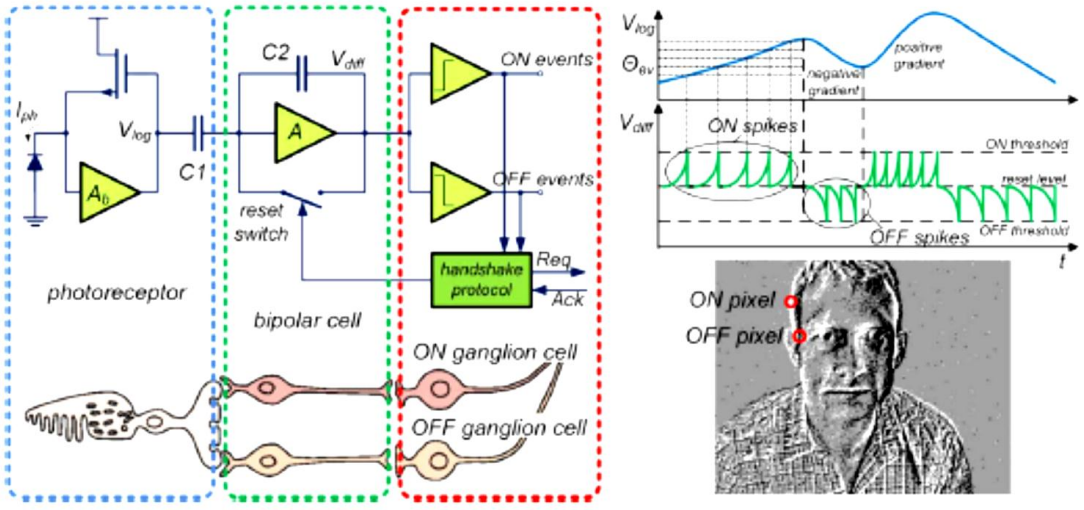

Figure 2 Three-layer model of a human retina and corresponding DVS pixel circuitry (left). Typical signal waveforms of the pixel circuit are shown top right.

[24] Q. Luo and J. Harris, ‘‘A time-based CMOS image sensor,’’ in Proc. IEEE Int. Symp. Circuits Syst., May 2004, vol. IV, pp. 840–843.

3.3 Event Based Vision Sensors

3.3.1 Types of Event Based Vision Sensors

Although gray level image sensors that use the AER protocol have been proposed), these simply use the AER protocol to transmit pixel intensity values, but have the drawback of costing expensive silicon pixel area without providing the benefit of either redundancy or latency reduction[25]. Broadly divided, AER "silicon retinas" which do have these functionalities fall into the following classes[26]:

- **Spatial contrast (SC)** sensors which reduce spatial redundancy based on intensity ratios, vs. **spatial difference (SD)** sensors which use intensity differences. SC is more useful with varying scene illumination while SD is cheaper to implement.
- **Temporal contrast (TC)** sensors which reduce temporal redundancy based on relative intensity change, vs. **temporal difference (TD)** sensors which use absolute intensity change. TC is more useful with non-uniform scene illumination but more expensive to implement than TD, especially with AE sensor. The exposure, readout, and pixel reset mechanisms can be lumped into two classes:
- **Frame Event (FE)** sensors, which use a synchronous exposure of all pixels and then schedule the event readout in order of presumed relevance, e.g. in order of SC or based on detected TD.
- **Asynchronous Event (AE)** sensors, which have autonomous pixels that continuously generate events based on a local decision about relevance, e.g. TC. Contrast here means Weber contrast as opposed to

Raleigh contrast, so that a uniform spatial or unchanging temporal input produces no events. Additional classifications and combinations are possible but are omitted here.

3.3.2 Technology used in Sensors

Different technologies offer advantages and disadvantages for the design of vision chips. The dominant technologies available to date are CMOS, Bi-CMOS, CCD, and GaAs (MESFET and HEMT). CMOS has been exhaustively used in many designs. The additional bipolar transistor in BiCMOS processes, though advantageous in achieving better matching properties and higher speeds, is not easily justifiable when comparing other factors in the design. While commercial grade CMOS processes are accessible through fabrication brokers, such as MOSIS and CMP, the CCD processes available for prototyping are of a low quality. GaAs processes have been used only to a very limited extent because there are no readily available photo detector structures in such processes, and more importantly, analog circuit

[25] J. G. Harris, "The changing roles of analog and digital signal processing in CMOS image sensors," in Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on, 2002, pp. IV-3976-IV-3979 vol.4.
[26] Delbruck, T., et al. "Activity-driven, event-based vision sensors." Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. IEEE, 2010.

design is severely limited by gate leakage in MESFET and HEMT transistors. In the following sections advantages and disadvantages of each process are highlighted[27] [28].

Table 1 Comparison between AER vision sensor devices.

	Prior work						This session		
Year	2001	2003	2005	2006	2008	2009	2010	2010	2010
Source	Zaghloul, Boahen [30]	Rüedi et al. [16]	Mallik et al.[9, 32]	Lichtsteiner et al. [5, 33]	Massari et al. [27]	Ruedi et al.[31]	Posch et al. 2010 [34]	Linares-Barranco et al. [2]	Culurciello et al. [3]
Functionality	Asynchronous spatial and temporal contrast,	Frame-based spatial contrast and gradient direction, ordered output	Temporal frame-difference intensity change detection APS imager	Asynchronous temporal contrast dynamic vision sensor (DVS)	Binary spatial and temporal contrast	Digital log pixel + RISC proc.	Async. Time-based Image Sensor (ATIS)	Async. Weber Contrast (SC), with either rate or TTFS coding	Temporal intensity change or spatial difference can trigger readout
Type (Sec.3)	SC TD AE	SC FE	TD FE	TC AE	SD TD FE	SC, embedded	TC AE	SC AE	TD SD FE
Gray picture output			●			●	●	●	●
Pixel size um (lambda)	34x40 (170x200)	69x69 (276x276)	25x25 (**100x100**)	40x40 (200x200)	26x26.5 (130x130)	**14x14** (311x311)	30x30 (333x333)	80x80 (400x400)	16x21 (??)
Fill factor (%)	14%	9%	17%	8.1%	20%	20%	10%(TC)/20%(gray)	2.5%	**42%**
Fabrication process	0.35um 4M 2P	0.5um 3M 2P	0.5um 3M 2P	0.35um 4M 2P	0.35um 4M 2P	**180nm** 1P6M	**180nm** 4M 2P MIM	0.35um 4M 2P	**180nm** SiGe BiCMOS 7M
Pixel complexity T=MOS,C=cap	38T	>50T, 1C	**6T (NMOS) 2C**	26T(14 anal), 3C	45T	~80T, 1C	77T, 4C, 2PD	131T, 2C	**11T**
Array size	96x60	128x128	90x90	128x128	128x64	**320x240**	304x240	32x32	128x128
Die size mm²	3.5x3.5	~10x10	3x3	6x6.3	11	5.2x8.4	9.9x8.2	2.5x2.6	??
Power consumption	62.7mW @ 3.3V	300mW @ 3.3V	30mW @ 5V (50 fps)	24mW @ 3.3V	**100uW@2V, 50fps**	**80mW** (11mW sensor)	50-175mW	0.66-6.6mW	<**1.4mW**@3V

- CMOS - CMOS has been and will remain the dominant technology is almost all VLSI design areas, including vision chips.
- BiCMOS - BiCMOS processes provide an additional bipolar device, which has been the workhorse of analog design. The bipolar transistor can be used to increase the speed, reduce the mismatch, and obtain better circuit characteristics when exponential I-V relationship is required.
- CCD and CMOS/CCD - CCD processes have originally been developed for analog signal processing and imaging devices. Although this may have facilitated the design of vision chips, due to their drawbacks there has been limited success in achieving functional and reliable vision chips.
- GaAs MESFET and HEMT - GaAs processes are recognized by their high speed operation for digital and analog circuits. They have also been used in opto-electronic devices. GaAs processes suffer from several problems

3.3.3 Advantages and Disadvantages of Vision Chips

When compared to a vision processing system consisting of a camera and a digital processor, a vision chip provides many system level advantages. These include

[27] Moini, A. "Vision chips or seeing silicon, 1997." URL http://www. eleceng. adelaide. edu. au/Groups/GAAS/Bugeye/visionchips 240.

[28] T.M. Bernard & P.E. Nguyen, \Vision through the power supply of the NCP retina," Proc. SPIE, Charged-Coupled Devices and Solid State Optical Sensors V, Vol. 2415, pp. 159{163, 1995.

- **Speed:** The processing speed achievable using vision chips exceeds that of the camera-processor combination. A main reason is the information transfer bottleneck between the imager and the processor. In vision chips information between various levels of processing is processed and transferred in parallel.

- **Large dynamic range:** Many vision chips use photo detectors and photo circuits which have a large dynamic range over at least 7 decades of light intensity. Many also have local and global adaptation capabilities which further enhance their dynamic range. Conventional cameras are at best able to perform global automatic gain control.

- **Size:** processing algorithms, very compact systems can be realized. The only parts of the system that may not be scalable are the mechanical parts (like the optical interface).

- **Power dissipation:** Vision chips often use analog circuits which operate in sub threshold region. There is also no energy spent for transferring information from one level of processing to another level.

- **System integration:** Vision chips may comprise most modules, such as image acquisition, and low level and high level analog/digital image processing, necessary for designing a vision system. From a system design perspective this is a great advantage over camera-processor option. Although designing single-chip vision systems is an attractive idea, it faces several

Although designing single-chip vision systems is an attractive idea, it faces several dis-advantages:

- **Reliability of processing:** Vision chips are designed based on the concept that analog VLSI systems with low precision are sufficient for implementing many low level vision algorithms. The precision in analog VLSI systems is affected by many factors, which are not usually controllable.
  As a result, if the algorithm does not account for these inaccuracies, the processing reliability may be severely affected. Vision chips also use unconventional analog circuits which may not be well characterized and understood.

- **Resolution:** In vision chips each pixel includes a photocircuit1 which occupies a large proportion of the pixel area. Therefore, vision chips have a low fill-factor and a low resolution. The largest vision chip reported has only 210x230 pixels, for a photo circuit consisting of six transistors only[29].

[29] A.G. Andreou & K.A. Boahen, ``A 48,000 pixel, 590,000 transistor silicon retina in current-mode subthreshold CMOS," *in Proc. 37th Midwest Symposium on Circuits and Systems*, pp. 97-102, 1994.

- **Difficulty of the design:** Vision chips implement a specific algorithm in a limited silicon area. Therefore, often of-the-shelf circuits cannot be used in the implementation. This involves designing many new analog circuits. Vision chips are always full custom designed, and full custom design is known to be time consuming and error-prone.

- **Programming:** None of the vision chips are general purpose. In other words, many vision chips are not programmable to perform different vision tasks. This inflexibility is particularly undesirable during the development of a vision system.

4 Comparison between Human Retina and Artificial silicon retina

4.1 Limitations in Vision Engineering

In order to appreciate how biological approaches and neuromorphic engineering techniques could be beneficial for advancing artificial vision, it is inspiring to look at some shortcomings of conventional machine vision. State-of-the-art image sensors suffer from limitations imposed by their frame-based operation. The sensors acquire the visual information as a series of ‘‘snapshots’’ recorded at discrete points in time, hence time quantized at a predetermined frame rate. Biology does not know the concept of a frame. In fact, comparing the performance of biological vision systems to the best state-of-the-art artificial vision devices, frames do not appear to be a very useful or efficient form of encoding visual information. This is even more obvious if one considers that the world, the source of the visual information, works asynchronously and in continuous time. As a consequence, depending on the time scale of changes in the observed scene, a problem that is very similar to under sampling, known from other engineering fields, arises.
Things happen between frames, and information gets lost. This may be tolerable for the recording of video for a human observer, but artificial vision systems in demanding applications such as, e.g., autonomous robot navigation, high-speed motor control, visual feedback loops, etc., may fail as a consequence of this shortcoming. Nature suggests a different approach: Biological vision systems are driven and controlled by events happening within the scene in view, and not, like image sensors, by artificially created timing and control signals that have no relation whatsoever to the source of the visual information and its dynamics. Translating the frameless paradigm of biological vision to artificial imaging systems implies that control over the acquisition of visual information is no longer being imposed externally to an array of pixels but the decision making is transferred to the single pixel that handles its own information individually. The second problem that is also a direct consequence of the frame-based acquisition of visual information is redundancy. Each recorded frame conveys the information from all pixels, regardless of whether this information, or a part of it, has changed since the last frame had been acquired. This method obviously leads, depending on the dynamic contents of the scene, to a more or less high degree of redundancy in the acquired image data. Acquisition and handling of these dispensable data consume valuable resources and translate into high transmission power dissipation, increased channel bandwidth requirements, increased memory size, and post

processing power demands. Devising an engineering solution that follows the biological pixel-individual, frame-free approach to vision can potentially solve both problems.

4.1.2 Technical Challenges in Vision Chips

Vision chip design is a challenging task. A vision chip obliges photo detecting components, image acquisition circuits, analog conditioning and processing circuits, digital processing and interfacing, and image readout circuitry all on the same chip. A large portion of these parts, for example, low level analog processing elements, ought to exist in number the same as photo detectors. Much of the time these components ought to associate with at any rate their closest neighbors.

The range needed for executing the circuits and routing the information over the chip has put upper limits on the realization of dependably practical and high resolution vision chips, and in many usage resolution or functionality has been traded off for the other. The outline of vision chips can clearly profit from the abnormal state integration in present VLSI forms, where more than 10 million transistors can be incorporated on a solitary chip[30].

Unfortunately, developed methodologies for abnormal state combination are generally tuned and portrayed for leading edge digital processors and drams, experiencing sub-micron impacts, for example, short channel impacts, hot-carrier impacts, band-to-band tunneling, gate oxide direct tunneling, gate induced drain leakage (GIDL), drain induced barrier lowering (DIBL), and threshold voltage control[31]. Numerous accessible procedures, then again, don't have any particular photograph discovering photo detecting element, and are not decently tuned for analog circuit design. Device mismatch has likewise seriously influenced the analog circuit design group, and just about no fabrication methodologies have been precisely portrayed and modeled to record for the mismatch.

Design of vision chips has additionally been influenced by the absence of VLSI friendly computer vision algorithms. Most present computational computer vision algorithms are exceptionally perplexing and are even hardly utilizing powerful workstations to run as a part of ongoing. Numerous computer vision algorithms are still not dependable enough for application when all is said and done in uncontrolled environments. Bio inspired algorithms, then again, depend on the way that numerous animals have created extremely proficient visual system. These calculations, be that as it may, are not developed enough and effects of inordinate rearrangements brought about by insufficient understanding of animal's visual system. Regardless of these realities, the configuration of single chip VLSI vision sensors or smart vision sensors is progressively advancing and numerous vision chips focused around biological or computational algorithms have been created in the recent years. The intricacy of

[30] Iee.et.tu-dresden.de, 'Vision Chips or Seeing Silicon', 2014. [Online]. Available: https://www.iee.et.tu-dresden.de/iee/analog/papers/mirror/visionchips/vision_chips/vision_chips.html. [Accessed: 15- Dec- 2014].

[31] C. Fienga*et al.*, ``Scaling the MOS transistor below 0.1 micro m: methodology, device structures, and technology requirements," *IEEE Trans. Electronics Devices*, Vol. 41, No. 6, pp. 941-951, June 1994.

vision chips has likewise essentially expanded, and 2d vision chips with more than 48,000 detectors and processing elements have been composed[32].

4.2 Artificial Vision Chips over Human Vision Chips

The technological advancements of camera chip design led to artificial retinas and manufacturing during the past twenty-five years has made digital imaging more affordable and accessible to the general public. These advancements have become more visible to consumers in eye implant. Although AER vision sensors are easily available today, the state-of-the-art image sensor chips used in these cameras exhibit a performance gap when compared with the capabilities of the human eye. How good these image sensor chips are today when compared to our eyes is a question that will be elaborated on.

The pixels in the DVS also mimic the way an individual eye neuron will calibrate itself to a particular location: that cell and those responsible for another area will respond to incoming data in slightly different ways, so one neuron might be very sensitive to input while another takes more stimulation to fire. Similarly, each pixel of the DVS adjusts its own exposure. This allows the DVS to handle uneven lighting conditions, though it also requires enormous pixels that are 10 times the size of those in a modern cell-phone camera.

It is possible to compare the capabilities of the human visual system, including the eyes, the optic nerve, the visual cortex, etc. with the state-of-the-art image sensor chips which includes optics, event-capture and signal-processing chips etc. The capabilities include ability to see different events, light sensitivity, light-intensity response range, functionality and operation modes, and signal processing capabilities.

4.2.1 Pixel and array size

The size of pixels in today's modern digital cameras is getting closer to the size of the photoreceptors in human eye. The typical human eye contains an average of 130 million photoreceptors. The diameter of the rods and cones varies between 1.0m and 8.0m, depending on their location on the retina.5 Today's state-of-the-art DVS sensor chips contain 0.18um 1P6M MIM CIS CMOS technology which has an array size of 240x180 pixels. It has a pixel size of 18.5 x 18.5 um^2 with 22% fill factor. However, the human being has been equipped with photoreceptors that are as small as 1.0m and has more than 100 million photoreceptors; and this is since the beginning of existence. It is also estimated that the resolution of the human eye is equivalent to an image sensor chip of 576 million pixels with a 120 degree field of view.6 Thus there is still a long way to go in improving the image-sensor pixel and array sizes used in cameras if the goal is to match the human eye.

[32] A.G. Andreou & K.A. Boahen, ``Neural information processing II," . In M. Ismail & T. Fiez, editor, Analog VLSI signal and information processing, chapter 8, pp. 358-413. McGraw-Hill, 1994.

4.2.2 Pixel distribution and formation

In the human eye the photoreceptor size and densities change, depending on their location on the retina. For example, no rods exist on the focus centre of the eye, which is called the fovea. Colour vision photoreceptors, which total only 10% of the eye's photoreceptors, are located mostly on the fovea. There is an irregular distribution of photoreceptors which is unique for every human being, like a fingerprint. Yet,the things that are seen are the same, such as colours (with the exception of people who are colour-blind). In camera chips, however, pixels are arrayed regularly, in two-dimensions. As the image-processing techniques and algorithms used in camera systems are linear and do not closely mimic the signal processing that exists in the human visual system, regularly arrayed pixels are required.

4.2.3 Temporal resolution and latency

Temporal resolutions for human visual features also likely depend on stimulus parameters. Although additional insight into the mechanisms of feature integration will require neuronal recording studies, the present findings do have implications for existing models of feature integration. One model that has been proposed to accomplish integration is that the features of a given object are linked by tagging them with a synchronous 40 Hz oscillation. If this mechanism integrates colour and orientation in our paradigm, synchronization must be able to emerge very quickly, as integration occurred within ~32 ms, ~23 ms of which was required for features to reach perceptual threshold. The events are output asynchronously and nearly instantaneously on an Address-Event bus, so they have much higher timing precision than the frame rate of a frame-based imager. This is shown by these recording from a spinning disk painted with wedges of various contrasts. The disk spins at 17 rev/sec, and the events are painted with coloured-time in the Figure 3 Time resolution of events. The measurements show that it can often achieve a timing precision of 1 us and a latency of 15 us with bright illumination. Because there are no frames, the events can be played back at any desired rate. The low latency is very useful for robotic systems, such as the pencil balancing robot.

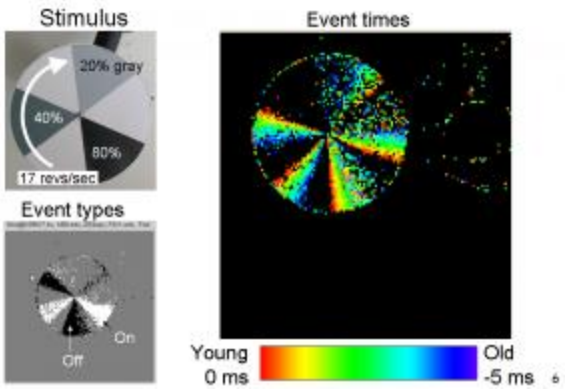

Figure 3 Time resolution of events

4.2.4 Light sensitivity and response range

Although the pixel sizes in image-sensor chips are approaching the size of the photoreceptors in the human eye, camera systems are not yet close to being able to match performance in terms of light sensitivity and response range. The human visual system and photoreceptors can easily adapt to very dim and bright light, with a light-intensity response range of ten billion to one (1010:1).7 This response range goes from light conditions on a bright sunny day to dim night vision. The DVS sensor has a large intra-scene dynamic range because the pixels locally respond to relative change of intensity. This wide dynamic range is demonstrated by the Edmund gray scale chart, which is differentially illuminated by a ratio of 135:1 – a 42dB illumination ratio, which means a normal high-quality CCD based device like the Nikon 995 used below must either expose for the bright or dark part of the image to obtain sensible data. Most of the vision sensor pixels still respond to the 10% contrast steps in both halves of the scene. The rightmost data is captured under 3/4 moon with a high contrast scene. Under these conditions the photocurrent is <20% of the photodiode leakage current, but the low threshold mismatch still allows a good response.

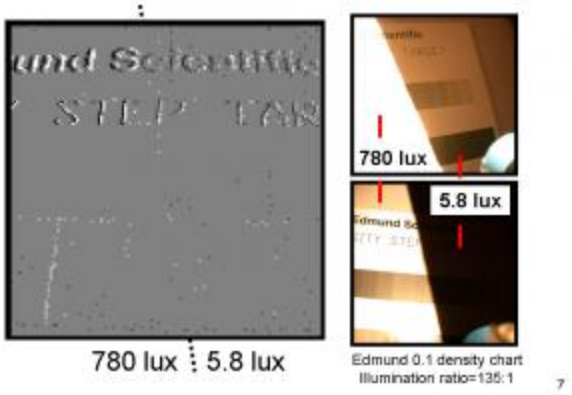

Figure 4 Contrast sensitivity under wide illumination

Typically, a conventional consumer camera pixel has a light intensity response range of one thousand to one (103:1). In a camera system, details of a captured scene are either concealed in the dark regions or washed out by the bright light, depending on the exposure settings of the system. Thus, one could say that the human visual system works ten million times (107) more efficiently than that of consumer cameras in terms of transferring scenes into images.

4.2.5 Operation principle

In terms of operation principles, the photoreceptors in the human eye convert light rays into electrical signals with extremely rapid electro-chemical reactions which can detect a single photon. Typically, in the image sensor pixel of a digital camera the photoelectric effect is used to convert impinging photons into electrical charges. Electrical charges are collected and stored in each pixel during the exposure period. Collected electric charges in each pixel are amplified and converted into digital ones (logic-1) and zeros (logic-0).

In a biological vision system the information is acquired and processed continuously in time. Biological retina send the image information to the visual cortex coded as spikes (also called events). When the activity level of a certain pixel reaches a threshold the pixel sends a spike.

Very active pixels send more spikes, and spikes propagate through the processing chain as soon as they are generated without waiting for the whole image to be processed. Consider, for example, the vision processing system shown in Figure 5 Multilayer vision processing system where the image from the retina is processed by two layers of convolutions. If the system is frame-based, each layer of convolutions will have to wait until the previous layer finishes processing the whole image to start its operation.

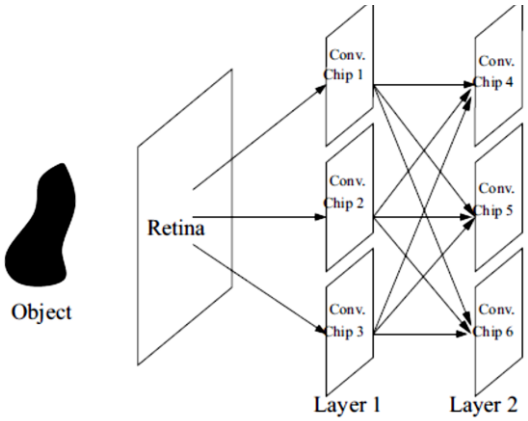

Figure 5 Multilayer vision processing system

However, if information between chips is sent as asynchronous spikes, spikes are communicated and processed by the next layer as soon as they are generated. In fact, the system in Figure 5 Multilayer vision processing system resembles the architecture of biological vision systems where the image from the retina is sent to the visual cortex. The visual cortex is organized as layers of neurons. Neurons from each layer project their output to a population of neurons of the following layer, thus performing a projection field or, equivalently, a convolution operation[33].

4.2.6 Signal processing capabilities

The captured image in the human eye is pre-processed before it is sent to the visual cortex of the brain. This pre-processing consists of a data reduction operation in which nothing is lost, with a compression ratio of 130 to 1, as only 1 million optic nerves leave each eye carrying the information from 130 million photoreceptors. This compression allows the brain to process information at a rate of 25 to 150 scenes or frames per second. Typically, every pixel in an image sensor chip is first transferred to higher processing units. A data compression method is either carried out with some loss of details in the image or the compression is never used. It does not perform frame-based image processing, but event-based. It receives input spikes from a previous stage (which can be a retina, another convolution chip or any other AER based processing module) and applies the programmed[34]. Synchronous controller performs the sequencing of all operations for each input event, which basically consists of

[33] G. M. Shepherd, *The synaptic organization of the brain*, Oxford University Press, 3rd Edition, 1990
[34] R. Serrano-Gotarredona, T. Serrano-Gotarredona, A.Acosta-Jiménez and B. Linares-Barranco, "A Neuromorphic Cortical-Layer Microchip for Spike-Based Processing Vision Systems", in *IEEETransactions on Circuits and Systems, Part I*, vol.53, No. 12, pp. 2548-2564, December 2006

adding row by row the kernel onto the array of pixels. Finally, uses an asynchronous circuitry for arbitrating and sending out the output address events generated by the array of pixels.
The aim of the overall design is to reduce the computational and communication load without any loss of efficiency. The inherent inefficiencies of image-capture in today's image-sensor chips are hidden by employing the limitations of the human eye. For example, solid-state image sensors have always been produced with row or column-vice uncanny stripes which are easily picked up by the human eye. However, psycho-visual experiments have shown that the human eye can only detect contrasts between two adjacent gray lines when the difference is greater than 0.5%. Thus, if a camera chip is designed to have a column to column or row to row contrast of less than 0.5%, these odd stripes would not be visible.

4.3 Future of Silicon Retina Pixels

Bernabe Linares-Barranco and Teresa Serrano-Gotarredona at the Inst. of Microelectronics in Sevilla and Christoph Posch, now at the Vision Instititute in Paris, have been particularly creative in devising interesting retina pixels with good performance.

4.3.1 The ATIS

Posch designed the ATIS1 pixel with colleagues while at the Austrian Inst. of Technology[35]. This pixel consists of two sub pixels. The first sub pixel is a DVS temporal contrast pixel. Events from the DVS pixel trigger time-based intensity readings in the second sub pixel. The intensity is measured by the time is takes the photodiode voltage to integrate between two levels. The beautiful thing about this mechanism is the way it avoids both mismatch and kTC noise, by integrating not from a reset voltage to a threshold, but rather between two thresholds, which are multiplexed to a common comparator. This way, the kTC reset level variation and the comparator offset are both suppressed[36]. The main advantage of the ATIS pixel is the event triggered and wide dynamic range intensity readout; however the price of this is a large pixel size and small fill factor (the ATIS is effectively about twice the area of the DVS pixel and must use a separate photodiode for each measurement), and intensity capture time that can be up to several hundred ms at low intensities.

4.3.2 Faster and More Sensitive DVS Pixels

The latest DVS pixels from Linares-Barranco and Serrano-Gotarredona are also very interesting. They addressed the need in some applications of higher speed and sensitivity by realizing that the best improvement in performance results from adding moregain and bandwidth to the photoreceptor that precedes the differencing amplifier. They have taken two approaches to this improvement but only the first is published[37]. In their pixel, they interposed two non-inverting voltage gain amplifiers between the logarithmic photoreceptor

[35] Posch, C., Matolin, D., Wohlgenannt, R.: A QVGA 143dB dynamic range asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression. In: 2010 IEEE Solid-State Circuits Conference Digest of Technical Papers, ISSCC (2010)
[36] Matolin, D., Posch, C., Wohlgenannt, R.: True correlated double sampling and comparator design for time-based image sensors. In: ISCAS 2009 (2009)
[37] Camunas-Mesa, L., Zamarreno-Ramos, C., Linares-Barranco, A., Acosta-Jimenez, A.J.,Serrano-Gotarredona, T., Linares-Barranco, B.: An Event-Driven Multi-Kernel Convolution Processor Module for Event-Driven Vision Sensors. IEEE Journal of Solid-State Circuits47(2), 504–517 (2012)

and the capacitive differencing amplifier. The voltage amplifiers are formed by current mirror stages using strong inversion operation with transistor geometry and operating current determining the voltage gain. This photoreceptor requires global gain control to keep the circuits in range over the entire intensity range of natural lighting. The time constant for this global gain control must be carefully chosen to provide sufficiently fast response to changes in lighting while not being so fast that it by itself generates oscillations or "gain control events". By using this circuit, they increase the gain of the photoreceptor by a factor of about 6, to result in an overall gain increase from 20 to 125. This increase allows them to set a lower nominal event threshold of about 2% contrast, compared with about 10% for our original DVS. They also use a different feedback arrangement for the photoreceptor. Instead of supplying photocurrent from the source of an nfet with feedback to the gate of the nfet, they use the photoreceptor from Oliver Landolt, where the feedback photocurrent is supplied from the drain of a pfet, with feedback applied to the source of the pfet. The gate of the pfet is tied to a fixed voltage, which determines the clamped photodiode voltage. The main advantage of this circuit is the reduced Miller capacitance, which allows lower latency responses. The main disadvantages are that the photocurrent cannot be read from the drain of the transistor, and the requirement that the feedback amplifier bias must be larger than the largest photocurrent. This requirement means that bias current cannot be arbitrarily reduced to control and width. However this is not a severe constraint for the high speed applications of this photoreceptor.

4.3.3 The apsDVS Pixel

The apsDVS tries to address some of the drawbacks of the ATIS in our newest pixel, which is called the apsDVS pixel. Here "aps" stands for "active pixel sensor" and is used to describe any kind of conventional CMOS image sensor pixel with inpixel active buffering of the integrated photodiode voltage. The current versions have a corrected 240x160 design with 18.5um pixels in fabrication. The apsDVS chip marries the advantages of simple small synchronous pixels with the low latency, wide dynamic range detection capabilities of the DVS pixels. The main disadvantage of the apsDVS will be the small dynamic range of the aps pixels. In hope, to take advantage of this combination in future application areas that extend on the obvious advantage of simply having a DC view of the scene in front of the sensor. In particular, the aps frames can be extrapolated using the DVS events to complete a richer and more powerful retinal output.

5 Conclusion

Humans are visually oriented and without a doubt, our eyes are considered to be our primary source of information. It is obvious that the human visual system is extremely complex and this complexity has fascinated human beings throughout history. Yet, the underlining principles and basic functions of human vision and the eye have only been discovered during the last two centuries.

These discoveries have led research in how to mimic these functions, which has resulted in moving and still-photographic and camera equipment, and the image sensors chips used in digital cameras today. Even though human beings are only taking baby steps in fully mimicking the human eye, curiosity and scientific inquiry allows us to discover functions and features of the eye and the visual pathways that will increase our knowledge and help us to build better pixels and image sensor chips.

Although object tracking is natural and easy with the DVS, it is somehow limited by the lack of a full cortically-inspired hierarchy of computation. However even object tracking already takes advantage of spatio-temporal event occurrence: Moving objects emit events like the familiar sparklers waved around on holiday occasions. It is the spatio-temporal coincidences of these events that drive the tracker models. Vision is often considered to be the process of object recognition. Now we observe from biology that there exists an impressive amount of cortical tissue that expands the visual representation of the dynamic visual input to a high dimensional representation. How can we bring these ideas into algorithmic processing of the retina output, while somehow taking advantage of the event-based output which affords us information about spatio-temporal coincidences in the sensor input?

The detection of obstacles in the environment should be roboust on a power budget more competitive with that of flying insects. This target has long been an aim of neuromorphic engineering and although it is not there yet, but getting closer. Together with developments of new sensors, new hardware for sensor processing, and inventive new algorithms, the aim is sure to have a grand time over the next few years.

List of figures

Bibliography

[1] Mahowald, M.A.: An Analog VLSI System for Stereoscopic Vision. Kluwer, Boston (1994)

[2] Lichtsteiner, P., Posch, C., Delbruck, T.: A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. IEEE Journal of Solid-State Circuits 43(2), 566–576 (2008)

[3] Liu, S.C., van Schaik, A., Minch, B.A., Delbruck, T.: Event-based 64-channel binaural silicon cochlea with Q enhancement mechanisms. In: IEEE ISCAS 2009, pp. 2426–2429 (2010)

[4] Hodgkin and A. Huxley, ‘‘A quantitative description of membrane current and its application to conduction and excitation in nerve,’’ J. Physiol., vol. 117, pp. 500–544, 1952.

[5] Mead, ‘‘Neuromorphic electronic systems,’’ Proc. IEEE, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.

[6] M. A. C. Maher, S. P. Deweerth, M. A. Mahowald, and C. A. Mead, ‘‘Implementing neural architectures using analog VLSI circuits,’’ IEEE
[7] Trans. Circuits Syst., vol. 36, no. 5, pp. 643–652, May 1989.

[8] K. A Boahen, 'A retinomorphic vision system', IEEE Micro, vol. 16, no. 5, pp. 30-39, 1996.

[9] K. A. Boahen, ‘Neuromorphic microchips’,Sci. Amer., vol. 292, pp. 56–63, May 2005.

[10] H. Kolb, 'Amacrine cells of the mammalian retina: Neurocircuitry and functional roles', Eye, vol. 11, no. 6, pp. 904-923, 1997.

[11] J. Heckenlively and G. Arden, Principles and practice of clinical electrophysiology of vision. Cambridge, Mass.: MIT Press, 2006, pp. 957-958.

[12] Dowling JE. 1979. Information processing by local circuits: the vertebrate retina as a model system. In: FO Schmitt, FG Worden (eds). The Neurosciences: Fourth Study Program, pp. 163-181. Cambridge, MA: MIT Press.

[13] S. W. Kuffler, ‘‘Discharge patterns and functional organization of mammalian retina,’’ J. Neurophysiol., vol. 16, pp. 37–68, 1953.

[14] B. Stephen, 'Contribution of amacrine transmission to fast adaptation of retinal ganglion cells', Frontiers in Neuroscience, vol. 4, 2010.

[15] Gallego, 'Horizontal and amacrine cells in the mammal's retina', Vision Research, vol. 11, pp. 33-IN24, 1971.

[16] L. Levin, S. Nilsson, J. Ver Hoeve, S. Wu, P. Kaufman and A. Alm, Adler's Physiology of the Eye. London: Elsevier Health Sciences, 2011, pp. 430-432.

[17] K. A Boahen, 'A retinomorphic vision system', IEEE Micro, vol. 16, no. 5, pp. 30-39, 1996.

[18] T. Delbruck and C. Mead, ‘‘Adaptive photoreceptor circuit with wide dynamic range,’’ in Proc. IEEE Int. Symp. Circuits Syst., 1994, vol. 4, pp. 339–342.

[19] K. Zaghloul and K. Boahen, 'Optic Nerve Signals in a Neuromorphic Chip I: Outer and Inner Retina Models', IEEE Transactions on
[20] Biomedical Engineering, vol. 51, no. 4, pp. 657-666, 2004.

[21] M. A.Mahowald and C. A. Mead, ‘‘The silicon retina,’’ Sci. Amer., vol. 264, no. 5, pp. 76–82,May 1991

[22] H. Kurino et al., ‘‘Smart vision chip fabricated using three dimensional integration technology,’’ in Advances in Neural Information Processing Systems 13, T. Leen, T. Dietterich, and V. Tresp, Eds. Cambridge, MA, USA: MIT Press, pp. 720–726, 2000.

[23] G. Cauwenberghs, N. Kumar, W. Himmelbauer, and A. G. Andreou, ‘‘An analog VLSI chip with asynchronous interface for auditory feature extraction,’’ IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process., vol. 45, no. 5, pp. 600–606, May 1998.

[24] T. Teixeira, A. G. Andreou, and E. Culurciello, ‘‘Event-based imaging with active illumination in sensor networks,’’ in Proc. IEEE Int. Symp. Circuits Syst., 2005, pp. 644–647.

[25] D. Chen, D. Matolin, A. Bermak, and C. Posch, ‘‘Pulse modulation imagingVReview and performance analysis,’’ IEEE Trans. Biomed. Circuits Syst., vol. 5, no. 1, pp. 64–82, Feb. 2011.

[26] Q. Luo and J. Harris, ‘‘A time-based CMOS image sensor,’’ in Proc. IEEE Int. Symp. Circuits Syst., May 2004, vol. IV, pp. 840–843.

[27] J. G. Harris, "The changing roles of analog and digital signal processing in CMOS image sensors," in Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on, 2002, pp. IV-3976-IV-3979 vol.4.

[28] Delbruck, T., et al. "Activity-driven, event-based vision sensors." Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. IEEE, 2010.

[29] Moini, A. "Vision chips or seeing silicon, 1997." URL http://www. eleceng. adelaide. edu. au/Groups/GAAS/Bugeye/visionchips 240.

[30] T.M. Bernard & P.E. Nguyen, \Vision through the power supply of the NCP retina," Proc. SPIE, Charged-Coupled Devices and Solid State Optical Sensors V, Vol. 2415, pp. 159{163, 1995.

[31] A.G. Andreou & K.A. Boahen, ``A 48,000 pixel, 590,000 transistor silicon retina in current-mode subthreshold CMOS," in Proc. 37th Midwest Symposium on Circuits and Systems, pp. 97-102, 1994.

[32] Iee.et.tu-dresden.de, 'Vision Chips or Seeing Silicon', 2014. [Online]. Available: https://www.iee.et.tu-dresden.de/iee/analog/papers/mirror/visionchips/vision_chips/vision_chips.html. [Accessed: 15- Dec- 2014].

[33] C. Fiengaet al., ``Scaling the MOS transistor below 0.1 micro m: methodology, device structures, and technology requirements," IEEE Trans. Electronics Devices, Vol. 41, No. 6, pp. 941-951, June 1994.

[34] A.G. Andreou & K.A. Boahen, ``Neural information processing II," . In M. Ismail & T. Fiez, editor, Analog VLSI signal and information processing, chapter 8, pp. 358-413. McGraw-Hill, 1994.

[35] G. M. Shepherd, The synaptic organization of the brain, Oxford University Press, 3rd Edition, 1990

[36] R. Serrano-Gotarredona, T. Serrano-Gotarredona, A.Acosta-Jiménez and B. Linares-Barranco, “A Neuromorphic Cortical-Layer Microchip for Spike-Based Processing Vision Systems”, in IEEETransactions on Circuits and Systems, Part I, vol.53, No. 12, pp. 2548-2564, December 2006

[37] Posch, C., Matolin, D., Wohlgenannt, R.: A QVGA 143dB dynamic range asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression. In: 2010 IEEE Solid-State Circuits Conference Digest of Technical Papers, ISSCC (2010)

[38] Matolin, D., Posch, C., Wohlgenannt, R.: True correlated double sampling and comparator design for time-based image sensors. In: ISCAS 2009 (2009)

[39] Camunas-Mesa, L., Zamarreno-Ramos, C., Linares-Barranco, A., Acosta-Jimenez, A.J.,Serrano-Gotarredona, T., Linares-Barranco, B.: An Event-Driven Multi-Kernel Convolution Processor Module for Event-Driven Vision Sensors. IEEE Journal of Solid-State Circuits47(2), 504–517 (2012)