Event-based silicon retinas and cochleas

6

Tobi Delbruck, Shih-Chii Liu

Contents

Abstract

This chapter reviews neuromorphic silicon retinas and cochleas that are based on the structure and operation of their biological counterparts. These devices are built using conventional chip fabrication technologies, using transistor circuits that emulate neural computations from biology. In first generation sensors, the analog outputs of every cell were read out serially at fixed sample rates. The new generation of sensors reports only the outputs of active cells through digital events (spikes) that are communicated asynchronously. Such sensors respond more quickly with reduced power consumption. Their digital "address-event" outputs rapidly convey precise timing information about the scene that is only attained from conventional sensors if they are continuously sampled at high rates. The sparseness, low latency, and spatio-temporal structure of this new form of sensor output data can benefit subsequent post-processing algorithms. Tradeoffs in the design of neuromorphic visual and auditory sensors are discussed. Examples are given of vision algorithms that process the address-events, using their spatio-temporal coherence, for low-level feature extraction and for object tracking.

1. Introduction

Biology provides examples of efficient machines which greatly outperform conventional technology. Electronic neuromorphic systems aim to capture the same efficient style of computation by emulating the struc-

Tobi Delbruck
University of Zürich and ETH Zürich
Institute of Neuroinformatics
Winterthurerstr. 190, 8057 Zürich, Switzerland
e-mail: tobi@ini.phys.ethz.ch

F. G. Barth et al., *Frontiers in Sensing*

ture and function of biological counterparts (Mead 1990; Douglas et al. 1995). This chapter discusses recent progress in realizing neuromorphic sensory systems which mimic the biological retina and cochlea, and subsequent sensor processing. The new generation of electronic sensors communicates their activity in much the same way as nervous systems do, by asynchronously outputting streams to digital events representing spikes. Because sensory computation is driven by these events, which originate from the sensory input, such systems are called event-driven systems. The progress in two specific neuromorphic sensors, the retina and the cochlea, is described in this chapter along with examples of how these sensors can help solve challenging problems in vision and audition.

How can asynchronous spike-based communication in neuromorphic systems be implemented? In contrast to biology which uses a slow ionic medium to carry spikes on dedicated axonal connections, event-based systems use the high speed of electrons to transmit source spike addresses from multiple sources (e.g. retina pixels) on a single device (e.g. a silicon retina) onto shared digital buses. Receivers decode these addresses for further use. The name for this asynchronous transmission scheme is the address-event representation, or **AER**. Using AER is a natural match to sensors that reduce redundancy in the input, thus transmitting only sparse useful information. Using AER, outputs can be transmitted with short response latencies and at a low data rate. Both characteristics are useful for practical applications, and they also differentiate AER sensors from conventional sensors that sample redundant information at constant rate.

This chapter begins with the history of silicon retina designs and then describes in detail the design and operation of an example AER retina. Section 3 discusses the history and progress of silicon cochleas. Section 4 discusses how the address events from the retina are processed in vision applications, with a focus on the use of the event timing; and Section 5 summarizes the chapter.

2. Conventional cameras and silicon retinas

The continued shrinking of silicon transistors has enabled rapid development of conventional frame-based electronic camera technology (Fossum 1997). With pixel sizes barely over a micron nowadays, a naïve observer might think that industry can approach a “perfect” image sensor with infinite resolution, infinite dynamic range, infinite frame rate, zero pixel size, and zero power consumption. Of course the output from such an ideal sensor would be impossible to process. As neuromorphic engineers, we observe that biological vision relies on a sensor (the retina) that does local gain control and massive amounts of computation to produce at its output (the optic nerve), an asynchronous stream of digital data (spikes) that represents only relevant visual information. A vision sensor that emulates its biological counterpart could help solve vision problems that seem easy for biology but are difficult using present machine vision techniques.

2.1 History of retina designs

The first integrated silicon retina by Mahowald and Mead (1991) consists of a 3-layered model of the outer retina architecture that included the photoreceptors, bipolar cells, and horizontal cells. The spiking ganglion cells and AER interface were incorpor-

ated in a later design. The key biological features emulated in this retina include the logarithmic, adaptive, phototransduction stage, the horizontal cell spatio-temporal smoothing network, and the rectification of the retina output into complementary ON and OFF channels. This retina was also part of the first AER vision system built by Mahowald and colleagues to compute stereopsis (Mahowald 1992, 1994).

However, this and subsequent AER systems in the following decade were barely useable. Mahowald's retina for instance, would respond reliably only to a flashing LED held up in front of it. Following this retina, Boahen's group (Zaghloul and Boahen 2004) built a more biologically realistic version with four cell types emulating the ON and OFF cell types for the sustained and transient pathways, but it suffered from large pixel response variability. The vision sensor by Ruedi et al. (2003) outputs simultaneous spatial contrast and local orientation using an event-based readout in which the pixel event timing within a frame encodes local contrast information. This design was the first that showed it was possible to overcome the mismatch, poor dynamic range, and otherwise low performance that plagued prior retinas.

Other recent event-based retinas that aim to reduce image redundancy include spatial contrast AER retinas (Barbaro et al. 2002; Ruedi et al. 2009; Lenoro-Bardallo et al. 2010), temporal intensity change retinas (Mallik et al. 2005), temporal-contrast retinas (Kramer 2002; Lichtsteiner et al. 2004), the dynamic vision sensor silicon retina described in Section 2.3 (Lichtsteiner et al. 2008), a prototype temporal-contrast infrared bolometer (Posch et al. 2009), preliminary color retina pixel designs (Fasnacht and Delbruck 2007; Berner et al. 2008; Olsson and Hafliger 2009), and a recent design which outputs both temporal contrast and gray level information (Posch et al. 2010). These designs need to cope with the fact that fabricated transistor characteristics are quite variable.

2.2 Achieving precision sensing with imprecise transistors

One of the biggest impediments in bringing the neuromorphic approach to commercially successful products is the problem of transistor mismatch. Transistor mismatch refers to built-in variability in transistor operating characteristics, and leads to the static salt and pepper noise in the output of initial retinas which made the image barely recognizable. Transistor mismatch is inherent to the way in which transistors are fabricated (Pelgrom et al. 1998). It is also one of the two major reasons why the silicon industry revolves around Boolean logic circuits, which reduce the impact of transistor mismatch by restoring the circuit output into one of the two power supply levels of a circuit block. (The other reason is the availability of tools for designing complex synchronous logic circuits.) Because matching precision scales with transistor dimension, precise analog design requires an increase of transistor area devoted to analog processing as process technology scales down. These considerations suggest that analog signals from sensors should be converted quickly to a digital representation so that the subsequent computational precision can be precisely controlled. However, at lower precision, analog computations take less area and consume less power than equivalent digital representations and computations (Sarpeshkar 1998). Analog designers have increasingly focused their efforts on embedding circuit mechanisms that decrease the effect of device imprecision on sensor performance. These mechanisms include dynamic adaptation and gain control that optimally utilizes the internal signal

range, digital calibration, and software post-processing. The most successful neuromorphic sensors have paid particular attention to improvements in their response uniformity. The silicon retina described next uses a combination of techniques to improve its sensitivity.

2.3 The dynamic vision sensor (DVS) silicon retina

The silicon retina discussed in this section is called the dynamic vision sensor or **DVS**, because it emulates retinal processing concerned with dynamic information. The DVS is used as an example of a silicon retina where organizing principles are successfully transferred from biology into silicon. The DVS forms an abstraction of the retina's transient pathway (Fig. 1a), which is blind to static input, monochromatic, and relatively fast (Rodieck 1998). The transient pathway can be contrasted with the sustained pathway, which has properties more like conventional cameras.

The DVS models a simplified 3-layer retina (Fig. 1b) consisting of photoreceptors, bipolar cells, and ganglion cells. In the simplified biological retina in Fig. 1b, the photoreceptors transduce from the input stimulus light to electrochemical signals. Bipolar cells compare the photoreceptor outputs with spatiotemporal lowpass averages computed by the horizontal cell spreading network and rectifiy the differences into ON and OFF activity. ON and OFF ganglion cells generate action potentials (spikes) from the bipolar cell outputs. The DVS simplifies this classical scheme even further by omitting the horizontal spatial averaging computed by the horizontal cells. Each DVS pixel (Fig. 1c) consists of 3 parts: a logarithmic photoreceptor, a temporal-change "differencer" amplifier (bipolar cells), and 2 decision units (ganglion cells). The pixel output is an asynchronous stream of ON and OFF events that signal log intensity changes, which usually correspond with scene reflectance changes. The generation of these events follows the process illustrated in Fig. 1d: The continuous-time photoreceptor output encodes intensity logarithmically as a voltage ($\log I$). The differencer amplifies the change in this output from a memorized value to produce $A \cdot d(\log I) = A \cdot dI/I$. Ganglion cell comparators detect a change in log intensity which exceeds a threshold and emit an ON or OFF event. The emission of an event also causes a new $\log I$ value to be memorized across the capacitor. The ON and OFF thresholds are typically set to about 10 % contrast. AER communication circuits along the periphery of the chip (Fig. 1e) transmit the address events off-chip with latencies of less than a microsecond.

What is the meaning of each DVS event? Each event represents a quantized change in log intensity. Light falling on the retina is a product of the scene illumination and reflectance. In the usual condition where scene illumination is constant over time, the static logarithmic response of the DVS photoreceptor turns this product into a sum. The differencing operation that subsequently produces events thereby encodes scene reflectance changes, which typically represent moving objects. Because this computation is based on a compressive logarithmic transformation in the photoreceptor circuit, the pixel can operate over a wide dynamic range of background intensity (120 dB as compared with 60 dB for a high quality conventional image sensor). This wide dynamic range capability means that the sensor can be used under uncontrolled natural lighting conditions that are typified by wide variations in scene illumination. The events are transmitted off chip in a fraction of a microsecond, so the timing of the pixel event is preserved, leading to a typical effective frame rate of several kHz. The data

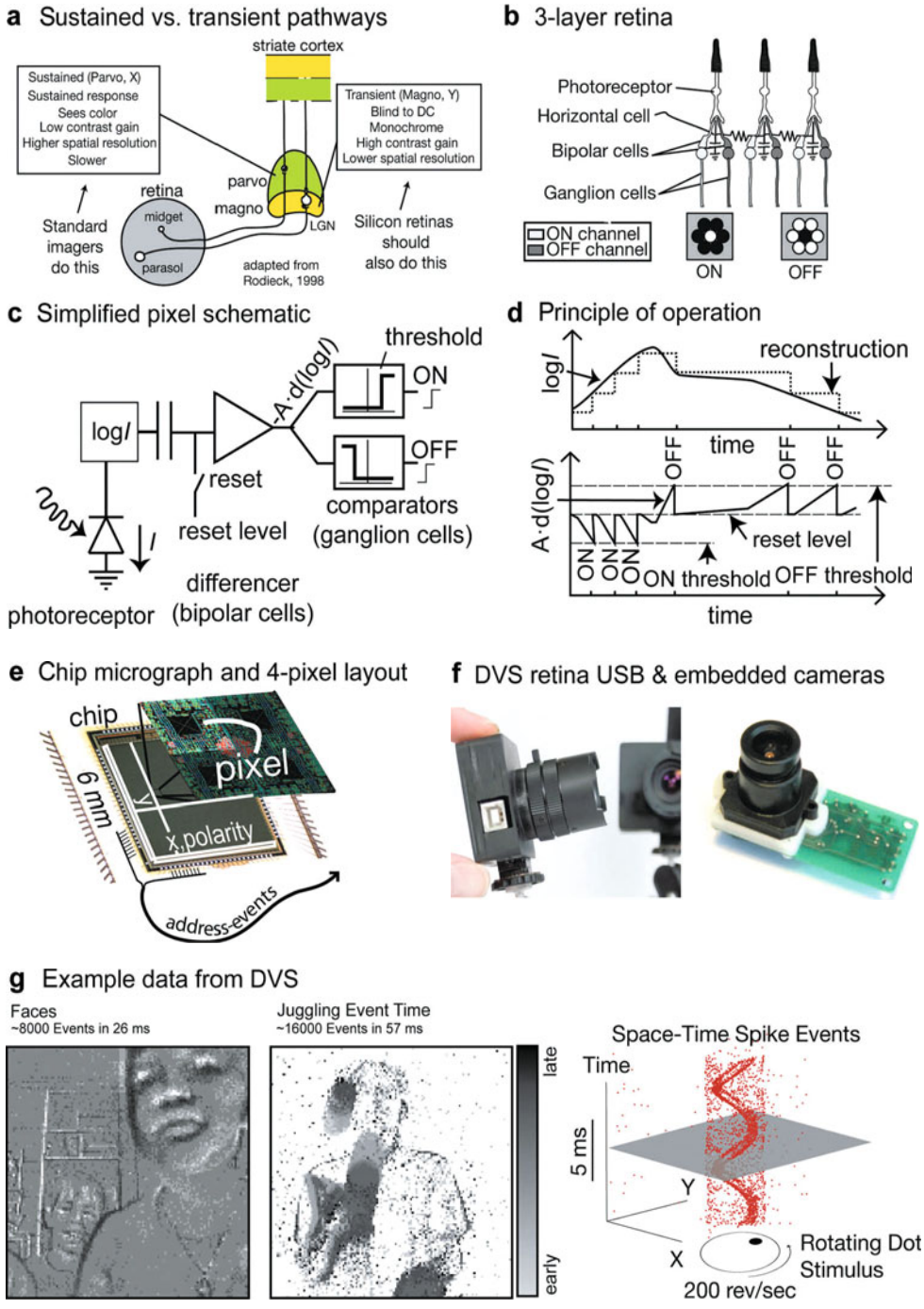

Fig. 1 Dynamic vision sensor (DVS) silicon retina. **a** Sustained and transient pathways from the retina through the lateral geniculate nucleus (LGN) to the striate cortex; **b** 3-layer retina; **c** DVS pixel schematic; **d** principle of operation; **e** DVS die micrograph and mirror symmetric arrangement of 4 pixels; **f** DVS retina USB2.0 camera and embedded microcontroller camera; **g** example DVS output data. "Faces" shows events rendered as a 2 d histogram over a 26 ms time slice. "Juggling Event Time" shows events from a juggler rendered so that lighter events occurred later in the 57 ms time slice. "Space-Time Spike Events" shows events from a spinning dot rendered in 3 d space-time. Adapted from Liu and Delbruck (2010)

rate is often a factor of 100 times lower than that of a frame-based image sensor with an equivalent time resolution. The pixel design provides enough response uniformity such that the pixel contrast threshold can be set to about 10 % contrast, allowing the device to sense real-world contrast signals.

The DVS chip can be connected to a variety of other devices (Fig. 1f). It is only necessary that the receiving device (e. g. custom multi-neuron chips or microcontroller) use the digital AER protocol. The chip has been integrated into a camera unit which interfaces to a computer via a standard USB2.0 interface that delivers time-stamped address-events to a host PC where post-processing of the output events allows us to easily explore event-driven algorithms which we can test in real-world scenarios.

The DVS output data (Fig. 1g) shows interesting characteristics: The "Faces" image shows the 2d histogram of the ON and OFF events over a time slice of 26 ms. The gray level of the "Juggling event time" image represents the event time relative to the start of the time window. The "Space-Time Spike Events" image exposes the spatiotemporal structure of a moving object, in this case a spinning dot, by displaying the events in space-time coordinates.

The DVS retina model is only a loose abstraction of the transient pathway. Even though this simplified silicon retina lacks many classes of retinal cells, it is proving useful for various practical applications as will be discussed in Section 4.

3. Conventional audition and silicon cochleas

Historically, incorporating biological knowledge into machine audition has lead to improved systems. Biologically inspired methods improve the performance of speech recognition systems (Gold and Morgan 2000). However the input stage of these systems is different for that of the biological cochlea. In machines, the acoustic signal is processed by extracting frequency components of the sound within a time window (20 ms for example) using a Discrete Fourier Transform (DFT). The power spectrum of the DFT is transformed logarithmically and again processed through a filter bank which has a log distribution similar to that measured in human cochleas. Systems based on the above methods are dominant because of the efficiency of the Fast Fourier Transform (FFT), and digital signal processing (DSP) chips that support the type of mathematical computation needed for FFTs. Contemporary hearing aids, for example, implement an entire signal processing chain based on a set of customized mostly-digital signal processing chips from microphone to speaker on a power budget of about 1 mW.

However, artificial audition systems still lack the performance of biological systems, for example, in tasks such as understanding speech in noisy backgrounds. Researchers look at replacing the input stage with a continuous-time cochlea processing stage and are also starting to explore the use of spikes for representing signals in speech processing systems, utilizing what is known about speech features in various brain regions (Liu et al. 2010b).

While silicon cochlea designs have been evolving over the last 2 decades; recent cochlea implementations also produce asynchronous spike outputs similar to that of biological cochleas. The analog timing of the silicon cochlea spikes carries precise timing information about the sound, similar to spikes from the spiral ganglion cell outputs of the biological cochlea. Developing methods to process these spikes can guide us on how information is processed from the co-

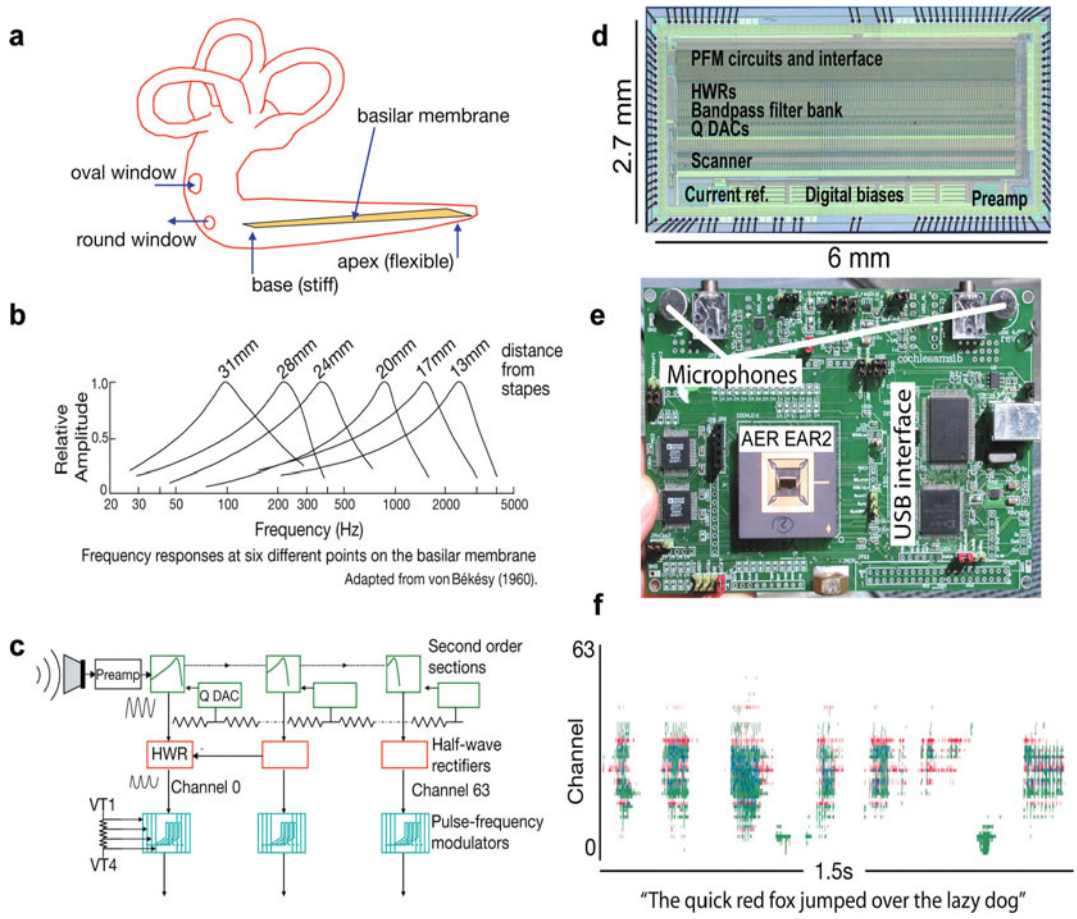

Fig. 2 The AER-EAR2 silicon cochlea. **a** cochlea and its basilar membrane (BM); **b** frequency response of the BM; **c** cascaded second-order-section (SOS) stages of the AER-EAR2 silicon cochlea; **d** integrated binaural 64 channel spike-based cochlea; **e** USB board implementation; **f** response to speech "The quick red fox jumped over the lazy dog". Each dot is an event; lower channel numbers are tuned to higher characteristic frequencies. Adapted from Liu and Delbruck (2010)

chlea to the cochlear nucleus, the midbrain, and other brain structures in the auditory cortex, and might offer insights into how humans are able to process speech and other sounds effectively in adverse environments such as hearing through noise, reverberation and interference from other speakers.

3.1 History of cochlea designs

Silicon cochlea models often implement the basilar membrane (BM) architecture as a cascaded set of second-order-section (SOS) bandpass filters of Lyon and Mead (1988) (see Fig. 2a). Depending on the input frequencies, a frequency component of the pressure wave generated in the fluid leads to a maximum displacement at some place along the basilar membrane (Fig. 2b) depending on its characteristic frequency (CF). The basilar membrane structure has a characteristic frequency (CF) which decreases exponentially from the base to the apex. The cascaded 1-d cochlea captures the sharp roll-off in frequency response of the basilar membrane but has limitations; a failure in one of the stages will lead to a failure in sub-

sequent second-order-section (SOS) stages. Also, the accumulation of noise along the cascaded stages presents a problem. But because the noise outside of the pass band of the filters will be removed as one proceeds along the cascade, the noise accumulation quickly saturates to a level that is tolerable (Sarpeshkar and Lyon 1998). A large number of stages within a frequency range also lead to a larger delay through the cascade than observed in the real cochlea.

Because of these drawbacks, a number of labs have attempted to construct BM models based on a parallel (rather than cascaded) 1-d architecture, or a 2-d architecture that includes the role of the fluid coupling between the stages (Watts et al. 1992; Fragniere 2005; Wen and Boahen 2006; Hamilton et al. 2008). But these approaches have their own drawbacks, such as increased chip area, and sensitivity to mismatches between stages, which cause destructive interference and greatly reduce the theoretically-achievable gain. A recent hot topic in cochlear designs concerns the implementation of local gain control as enabled by the outer hair cells. Developers are considering ways of measuring of the BM signal amplitudes and using these measurements to tune the sharpnesses of the resonances of the stages (Sarpeshkar and Lyon 1998); others include implementation of Hopf-like models of outer hair cell active nonlinear amplification or attenuation of the signal, using discrete chips (Martignoli et al. 2007) or approximations of such models in silicon (Hamilton et al. 2008; Katsiamis et al. 2009).

3.2 Spiking cochlea architectures

The latest silicon cochlea designs include either cascaded or parallel architectures for the BM, inner hair cells, and the spiral ganglion cells with AER outputs (Abdalla and Horiuchi 2005; Fragniere 2005; Sarpeshkar et al. 2005; Wen and Boahen 2006; Chan et al. 2007; Liu et al. 2010a).

Our latest cochlea implementation (AER-EAR2) is a binaural cochlea intended for spatial audition and auditory scene analysis (Liu et al. 2010a). It integrates many features of former designs in a more user-friendly form. It uses cascaded second-order-sections (SOSs) (Fig. 2c) which drive inner hair cells, which in turn drive multiple ganglion cells with different spike thresholds. The resonance of individual sections can be adjusted by a local digital-to-analog converter (QDAC). This chip (Fig. 2d) includes a variety of features including a matched binaural pair of cochleas, on-chip digitally controlled biases, on-chip microphone preamplifiers, and open-sourced host software algorithms (jAER 2007). A bus-powered USB board enables easy interfacing to standard PCs for control and processing (Fig. 2e). An example response of AER-EAR2 to speech is shown in Fig. 2f.

3.3 Applications of cochleas

Applications of analog silicon cochleas in auditory tasks have been explored by different groups. The first experiments were done on extracting pitch and spatial location (Lazzaro et al. 1993; van Schaik and Shamma 2004). Because of the availability of AER cochleas, the focus has shifted to computations that capitalize on the precise timing of the cochlear events. This precise timing is useful for estimating interaural time difference (ITD) cues which can be used to localize sound sources (Chan et al 2007; Liu et al. 2010a) and for speech recognition tasks (Usyal et al. 2006). Estimating interaural time differences (ITDs) is expensive to compute using conventional approaches because of the high sample rate required on the analog microphone output. Another potential application is the use of these cochlear circuits as

preprocessors for cochlear implant systems (Georgiou and Toumazou 2005; Sarpeshkar et al. 2005).

4. Processing events

AER sensor events can be digitally processed by algorithms running on conventional hardware (jAER 2007; Delbruck 2008). The main characteristics of these algorithms are that they are event-driven and use the precise timing of the events. The focus here is on visual processing because these algorithms are more evolved.

The only prior work on interfacing AER chips with computers was the pioneering work on processing spikes from a silicon cochlea for speech recognition (Lazzaro et al. 1993). The recent availability of AER vision and audition sensors with good performance and user-friendly computer interfaces has given new life to this field. For example, recent applications of the dynamic vision sensor (DVS) include object tracking for commercial products (Litzenberger et al. 2006; Bauer et al. 2007; Gritsch et al. 2008); robotic applications (Delbruck and Lichtsteiner 2007; Conradt et al. 2009), fall detection of the elderly in assisted living (Fu et al. 2008), and in studying the development of topological connections (Boerlin et al. 2009).

Methods for digitally processing events have evolved into the following classes: *filters* that clean up the input to reduce noise or redundancy; *labelers* that assign additional meaning to the events such as contour orientation or direction of motion; and *trackers* that use events to track moving objects. The filters and labelers generally use one or several retinotopic maps of event times. These maps can be thought of as pictures of the most recent event times. The representation of events as software objects allows attachment of arbitrary annotation. As events are processed, some events are discarded, and remaining events are labeled with additional information. Instead of expanding the representation by expanding the number of cell types (as in cortical processing), the digital events are assigned increasing interpretation.

4.1 Low level visual feature extraction

As one example of this style of processing we will consider a low level feature extraction labeler that annotates the events in the input stream with additional interpretation. The software *DirectionSelectiveFilter* (Fig. 3) shows how orientation information is first extracted and then used to compute local optical flow vectors. These vectors are then integrated over space to compute global translational, radial, and tangential optical flows. The input shown in Fig. 3a is generated from a dynamic vision sensor (DVS) camera flying around a room (in this case handheld simulated flying) painted with black and white squares on the walls. The orientation labeler labels the ON and OFF DVS events with an additional 'orientation type' that signals their spatial angle of maximum correlation with past events in the spatial vicinity.
This labeler uses a topographic memory of past event times. Events generated by pixels along a moving edge of a particular orientation will be more correlated in time than events at other orientations. The orientation labeler tags each event using criteria about the maximum allowed correlation time within the receptive field. Events that pass the correlation test are output as orientation events. The correlation time is the average time difference between this event and the past ones stored in the region of the receptive field, with smaller time differences indicating stronger correlations.

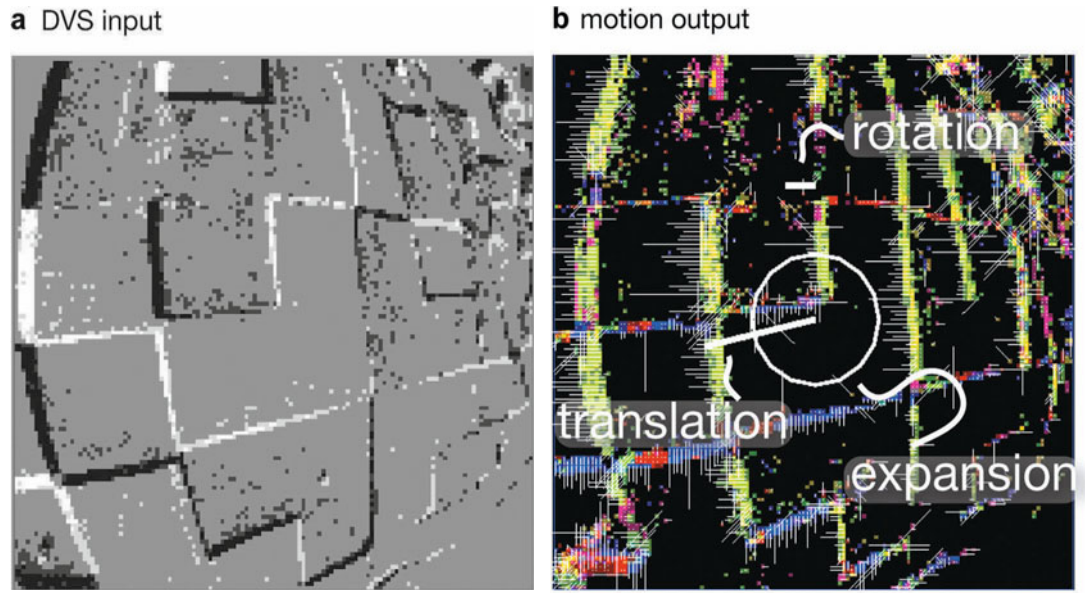

Fig. 3 Example output from the *DirectionSelectiveFilter* event labeler. **a** DVS input to labeler. Grey pixels indicate no events while white/black pixels indicate ON and OFF events caused by moving edges in the scene; **b** output of labeler. Local motion vectors (thin lines) are integrated over space and time to produce the global optical flow components for translation, rotation, and expansion (thick lines). Rotation and expansion vectors are centered on the center of the scene

The orientation events are then used to compute local motion vectors using the time of flight of orientation events. Each orientation event starts a search over the map of past orientation events in a direction normal to the orientation, and the most likely direction of motion is chosen based on a correlation measure similar to the one used for orientation. The speed is computed by the time of flight from the past orientation event. The output of the motion labeler (Fig. 3 b) shows how the local normal flow vectors (thin lines) result in estimates of the global translational flow (thick line from center), which in this case, points to the left. Similar global metrics of rotational and radial flow are also computed.

Low level methods for processing spikes like the *DirectionSelectiveFilter* could provide a means of bringing a hierarchical style of digital processing to AER events. But since the main success so far in practical applications is in tracking objects, we will now turn to this subject.

4.2 Tracking

Object tracking is widely-used visual capability. The main advantages of using an AER sensor for object tracking include system computational efficiency, low system latency, and the capability of the sensor to operate over a wide dynamic range of background intensities. One example is a tracker called the *RectangularClusterTracker* that tracks multiple moving objects (Litzenberger et al. 2006; Delbruck and Lichtsteiner 2007). This tracker treats an object as a spatially-connected source which generates a *cluster* of temporally correlated events. As objects move, they generate events which drag along the clusters. Clusters are spawned by events and disappear when moving clusters merge

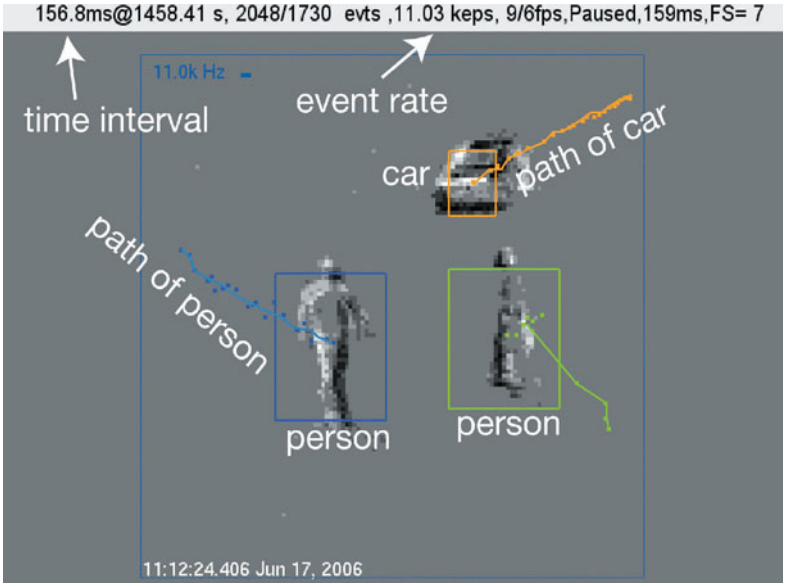

Fig. 4 Object tracking: RectangularClusterTracker tracks two persons and a car. The history of each cluster is shown as the dots

or when they are starved of events. Cluster sizes are determined either by knowledge of the scene (perspective) and object classes (e. g. cars), or they can be dynamic.

Tracking using events is advantageous because clusters do not have to be tracked over frames. Each pixel event updates the information about the size and location of its member cluster and only sensor pixels that generate events need to be processed. The primary processing cost is thus dominated by the search for the nearest existing cluster and the memory cost is limited only to the storage of the list of clusters.

Fig. 4 shows an example of how the *RectangularClusterTracker* can track a car and two people in a scene. Over the 157 ms window shown, the DVS produces 2048 events. 1730 events were left after pre-filtering to remove uncorrelated activity. The past locations of the clusters over 2048-event slices are shown by the dots. The sizes of the clusters dynamically come to equilibrium by balancing the events inside and surrounding the clusters.

The *RectangularClusterTracker* has been used in a robotic goalie that has a reaction latency of 3 ms with a 4 % processor load, using standard USB interfaces (Delbruck and Lichtsteiner 2007). This combination of metrics would be impossible to achieve using conventional frame based vision.

Other types of trackers have also been developed. For instance, the world's first pencil balancing robot uses a pair of dynamic vision sensors (DVSs) with embedded microcontrollers (Fig. 1f) to track the pencil in 3d space. The low memory and processing costs of events allow the pencil balancing robot to process input events at an effective frame rate of more than 10 kHz using only 200 mW embedded fixed-point microcontrollers (Conradt et al. 2009).

Work is ongoing to extend these notions about object tracking to include simultaneous tracking of multiple objects, object classification, and object recognition.

Summary and discussion

AER sensors are starting to make their way into usable high-performance prototypes. The nine papers from 2002 to 2010 on event-based sensors in the very competitive IEEE International Solid State Circuits Conference illustrate that this approach is starting to impact mainstream electronics (Barbaro et al. 2002; Fragniere 2005;

Mallik et al. 2005; Sarpeshkar et al. 2005; Wen and Boahen 2006; Lichtsteiner et al. 2006; Posch et al. 2007; Massari et al. 2008; Posch et al. 2010).

One achievement of this work is the development of event-based digital processing methods that capture the flavor of biological spike-based processing. These methods were only developed after sensors were built in a form that allowed their everyday use away from the lab bench. Although these methods have been developed as software algorithms on standard computers (jAER 2007), one can consider a variety of event-processing platforms ranging from conventional computers to the other extreme of using AER neuromorphic chips (Choi et al. 2005; Chicca et al. 2006; Serrano-Gotarredona et al. 2009). Our industrial partners are using an embedded DSP platform (Belbachir et al. 2007). We ourselves have used tiny microcontrollers (Conradt et al. 2009) and are starting to use field-programmable gate arrays (Linares-Barranco et al. 2007). This field has the potential for realization of small, fast, low power, embedded sensory-motor processing systems that are beyond the reach of traditional approaches under constraints of power, memory, and processor cost. The availability of event-based silicon retinas and cochleas will enable investigations of binding across visual and auditory sensory modalities using the fine temporal structure afforded by spikes. The asynchronous nature of AER data could inspire digital signal processing that breaks away from conventional regular-sample processing and that builds on top of decades of work in understanding computation by the nervous system (Liu and Delbruck 2010).

Acknowledgments

The authors gratefully acknowledge the constructive comments by the anonymous reviewer, F. Barth, M. Srinivasan, J. Harris, M. Rostogi, and N. da Costa.

References

Abdalla H, Horiuchi T (2005) An ultrasonic filterbank with spiking neurons, IEEE Intl. Symp. on Circuits and Systems (ISCAS 2005), pp. 4201–4204

Barbaro M, Burgi PY, Mortara A, Nussbaum P, Heitger F (2002) A 100 × 100 pixel silicon retina for gradient extraction with steering filter capabilities and temporal output coding. IEEE J. of Solid-State Circuits 37: 160–172

Bauer D, Belbachir AN, Donath N, Gritsch G, Kohn B, et al. (2007) Embedded vehicle speed estimation system using an asynchronous temporal contrast vision sensor. EURASIP J Embedded Syst 2007 (art. ID 82 174): 1–12

Belbachir AN, Litzenberger M, Posch C, Schon P (2007) Real-time vision using a smart sensor system. IEEE Intl Symp Industrial Electronics 2007. ISIE 2007, pp. 1968–1973.

Berner R, Lichtsteiner P, Delbruck T (2008) Self-timed vertacolor dichromatic vision sensor for low power face detection. IEEE Intl Symp Circuits and Systems (ISCAS 2008), pp. 1032–1035

Boerlin M, Delbruck T, Eng K (2009) Getting to know your neighbors: Unsupervised learning of topography from real-world, event-based input. Neural Computation 21: 216–238

Chan V, Liu S-C, van Schaik A (2007) AER EAR: A matched silicon cochlea pair with address event representation interface. IEEE Trans Circuits and Systems I: Regular Papers 54: 48–59

Chicca E, Whatley AM, Lichtsteiner P, Dante V, Delbruck T, et al. (2006) A multi-chip pulse-based neuromorphic infrastructure and its application to a model of orientation selectivity. IEEE Trans Circuits and Systems I: Regular Papers 54: 981–993

Choi TYW, Merolla PA, Arthur JV, Boahen KA, Shi BE (2005) Neuromorphic implementation of orientation hypercolumns. IEEE Trans Circuits and Systems I: Regular Papers 52: 1049–1060

Conradt J, Berner C, M., Delbruck T (2009) An embedded AER dynamic vision sensor for low-latency pole balancing. 5th IEEE Workshop on Embedded Computer Vision (in conjunction with ICCV 2009), Kyoto, Japan

Delbruck T, Lichtsteiner P (2007) Fast sensory motor control based on event-based hybrid neuromorphic-procedural system. IEEE Intl Symp Circuits and Systems (ISCAS 2007), pp. 845–848

Delbruck T (2008) Frame-free dynamic digital vision. Proc Intl Symp Secure-Life Electronics, Advanced Electronics for Quality Life and Society, pp. 21–26. Tokyo: University of Tokyo

Douglas R, Mahowald M, Mead C (1995) Neuromorphic Analog VLSI. Ann Rev Neurosci 18: 255–281

Fasnacht D, Delbruck T (2007) Dichromatic spectral measurement circuit in vanilla CMOS. IEEE Intl

Symp Circuits and Systems (ISCAS 2007), pp. 3091–3094

Fossum ER (1997) CMOS image sensors: electronic camera-on-a-chip. IEEE Trans Electron Devices 44: 1689–98

Fragniere E (2005) A 100-channel analog CMOS auditory filter bank for speech recognition. IEEE ISSCC Dig of Tech Papers, pp. 140–589

Fu Z, Delbruck T, Lichtsteiner P, Culurciello E (2008) An address-event fall detector for assisted living applications. IEEE Trans Biomed Circuits and Systems 2: 88–96.

Georgiou J, Toumazou C (2005) A 126-uW cochlear chip for a totally implantable system. IEEE J. Solid-State Circuits 40: 430–443

Gold B, Morgan N (2000) Speech and audio signal processing: John Wiley and Sons, Inc. New York, NY

Gritsch G, Litzenberger M, Donath N, Kohn B (2008) Real-time vehicle classification using a smart embedded device with a 'silicon retina' optical sensor. ITSC08, pp. 534–538. Bejing, China

Hamilton T, Tapson J, Jin CT, van Schaik A (2008) An active 2-D silicon cochlea. IEEE Trans Biomed Circuits and Systems 2: 30–43

Indiveri G, Liu S-C, Delbruck T, Douglas R (2009) Neuromorphic systems. In: LSquire (ed) Encyclopedia of neuroscience, pp. 521–528: Academic Press

jAER (2007) jAER Real time sensory-motor processing for spike based address-event representation (AER) sensors and systems available: http://jaer.wiki.sourceforge.net

Katsiamis A, Drakakis E, Lyon R (2009) A biomimetic, 4.5uW, 120+ dB, log-domain cochlea channel with AGC. IEEE J Solid-State Circuits 44: 1006–1022

Kramer J (2002) An ON/OFF transient imager with event-driven, asynchronous read-out. IEEE Intl Symp Circuits and Systems (ISCAS 2002), pp. 165–168

Lazzaro J, Wawrzynek J, Mahowald M, Sivilotti M, Gillespie D (1993) Silicon auditory processors as computer peripherals. IEEE Trans Neural Networks 4: 523–528

Lennie P (2003) The cost of cortical computation. Current Biology 13: 493–497

Lenoro-Bardallo JA, Serrano-Gotarredona T, Linares-Barranco B (2010) A spatial calibrated contrast AER vision sensor with adjustable contrast threshold. IEEE Intl Symp Circuits and Systems (ISCAS 2010), pp. 2426–2429

Lichtsteiner P, Posch C, Delbruck T (2006) A 128×128 120 dB 30 mW asynchronous vision sensor that responds to relative intensity change. ISSCC Dig Tech. Papers, pp. 508–509 (27.9). San Francisco

Lichtsteiner P, Posch C, Delbruck T (2008) A 128×128 120 dB 15us latency asynchronous temporal contrast vision sensor. IEEE J Solid State Circuits 43: 566–576

Linares-Barranco A, Gómez-Rodríguez F, Jiménez A, Delbruck T, Lichtsteiner P (2007) Using FPGA for visuo-motor control with a silicon retina and a humanoid robot. IEEE Intl Symp Circuits and Systems (ISCAS 2007), pp. 1192–1195

Litzenberger M, Posch C, Bauer D, Schön P, Kohn B, et al. (2006) Embedded vision system for real-time object tracking using an asynchronous transient vision sensor. IEEE Digital Signal Proc Workshop 2006, pp. 173–178. Grand Teton, Wyoming

Liu SC, Kramer J, Indiveri G, Delbruck T, Douglas R (2002) Analog VLSI: circuits and principles: MIT Press, Cambridge, MA

Liu SC and Delbruck T (2010) Neuromorphic sensory systems. Curr. Opin. in Neurobiol 20: 288–295

Liu SC, van Schaik A, Minch BA, Delbruck T (2010a) Event-based 64-channel binaural silicon cochlea with Q enhancement mechanisms. IEEE Intl Symp Circuits and Systems 2010 (ISCAS 2010), pp. 2027–2030

Liu SC, Mesgarani N, Harris, J, Hermansky, H (2010b) The use of spike-based representations for hardware audition systems. IEEE Intl Symp Circuits and Systems 2010 (ISCAS 2010), pp. 505–508

Lyon RF, Mead C (1988) An analog electronic cochlea. IEEE Trans Acoustics Speech and Signal Processing 36: 1119–1134

Mahowald MA (1992) VLSI analogs of neuronal visual processing: a synthesis of form and function. Computation and neural systems, Caltech, Pasadena, California

Mahowald MA (1994) An analog VLSI system for stereoscopic vision: Kluwer, Boston, MA

Mahowald MA, Mead C (1991) The silicon retina. Sci Am 264: 76–82

Mallik U, Clapp M, Choi E, Cauwenberghs G, Etienne-Cummings R (2005) Temporal change threshold detection imager. IEEE ISSCC Dig Tech. Papers, pp. 362–363

Martignoli S, van der Vyver J-J, Kern A, Uwate Y, Stoop R (2007) Analog electronic cochlea with

mammalian hearing characteristics. Applied Physics Letters 91 (064 108)

Massari N, Gottardi M, Jawed S (2008) A 100uW 64 × 128-pixel contrast-based asynchronous binary vision sensor for wireless sensor networks. IEEE ISSCC Dig Tech Papers, pp. 588–638

Mead C (1990) Neuromorphic electronic systems. Proc IEEE 78: 1629–1636

Olsson JAM, Hafliger P (2009) Live demonstration of an asynchronous integrate-and-fire pixel-event vision sensor. IEEE Intl Symp Circuits and Systems (ISCAS 2009), pp. 774–774

Pelgrom M, Tuinhout H, Vertregt M (1998) Transistor matching in analog CMOS applications. IEDM Tech Dig: 915–918

Posch C, Hofstatter M, Matolin D, Vanstraelen G, Schon P, et al. (2007) A dual-line optical transient sensor with on-chip precision time-stamp generation. IEEE ISSCC Dig Tech Papers, pp. 500–618

Posch C, Matolin D, Wohlgenannt R (2010) A QVGA 143 dB DR asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression. IEEE ISSCC Dig Tech. Papers, pp. 400–401

Posch C, Matolin D, Wohlgenannt R, Maier T, Litzenberger M (2009) A microbolometer asynchronous dynamic vision sensor for LWIR. IEEE Sensors Journal 9: 654–664

Rodieck R (1998) The first steps in seeing: Sinauer Associates, Sunderland, MA

Ruedi PF, Heim P, Gyger S, Kaess F, Arm C, et al. (2009) An SoC combining a 132 dB QVGA pixel array and a 32 b DSP/MCU processor for vision applications. IEEE ISSCC Dig Tech Papers, pp. 46–47

Ruedi PF, Heim P, Kaess F, Grenet E, Heitger F, et al. (2003) A 128 × 128, pixel 120-dB dynamic-range vision-sensor chip for image contrast and orientation extraction. IEEE J. Solid-State Circuits 38: 2325–2333

Sarpeshkar R (1998) Analog versus digital: Extrapolating from electronics to neurobiology. Neural Computation 10: 1601–38

Sarpeshkar R, Baker MS C., Sit JJ, Turicchia L, Zhak S (2005) An analog bionic ear processor with zero-crossing detection. IEEE ISSCC Dig Tech Papers, pp. 78–79

Sarpeshkar R, Lyon RF (1998) A low-power wide-dynamic-range analog VLSI cochlea. Analog Integrated Circuits and Signal Processing 16: 245–274

Serrano-Gotarredona R, Oster M, Lichtsteiner P, Linares-Barranco A, Paz-Vicente R, et al. (2009) CAVIAR: A 45 k neuron, 5 M synapse, 12 G connects/s AER hardware sensory–processing–learning–actuating system for high-speed visual object recognition and tracking. IEEE Trans Neural Networks 20: 1417–1438

Uysal I, Sathyendra H, Harris JG (2006) A biologically plausible system approach for noise robust vowel recognition. IEEE Proc Midwest Symp Circuits and Systems, pp. 245–249

van Schaik A, Shamma S (2004) A neuromorphic sound localizer. Analog Integrated Circuits and Signal Processing 39: 267–273

Watts L, Kerns DA, Lyon RF, Mead CA (1992) Improved implementation of the silicon cochlea. IEEE J. of Solid State Circuits 27: 692–700

Wen B, Boahen K (2006) A 360-channel speech preprocessor that emulates the cochlear amplifier. IEEE ISSCC Dig Tech Papers, pp. 556–557

Zaghloul KA, Boahen K (2004) Optic nerve signals in a neuromorphic chip II: Testing and results. IEEE Trans Biomed Engineering 51: 667–675