Visual sensor with resolution enhancement by mechanical vibrations

Oliver Landolt, Ania Mitros* and Christof Koch
California Institute of Technology, MS 139-74, Pasadena CA 91125

Abstract

The resolution limit of visual sensors due to finite pixel spacing can be overcome by applying continuous low-amplitude vibrations to the image—or taking advantage of existing vibrations in the environment, for instance in a mobile robotics application. Thereby, spatial intensity gradients turn into temporal intensity fluctuations which can be detected and processed by every pixel independently from the others. This approach enhances resolution and virtually eliminates fixed-pattern noise. An integrated circuit is described which implements this visual sensing principle. It incorporates an array of 32 by 32 pixels with local temporal signal processing, and a novel non-arbitrated address-event communication scheme providing timing guarantees on external signals for easy interfacing with off-the-shelf digital components.

1 Introduction

Image sensors can be classified in two broad categories on the basis of their purpose. Cameras are meant to acquire images for replication at another place or time for the benefit of human observers. Visual sensors are meant to extract information about a visual scene for purposes such as robot navigation. In the second category, it is preferrable to incorporate visual data processing as early as possible in the signal flow to reduce the cost of transmitting and processing the tremendous amount of redundant raw image data delivered by an array of photoreceptors. Consistently with this requirement, a number of visual sensing integrated circuits incorporating some amount of processing within each pixel have been described in the literature, many of which are inspired to some degree by biological neural structures [1, 2]. Adding substantial local processing into every pixel unavoidably leads to a steep increase in silicon area compared to the area devoted to photodetection, hence a reduction of the total number of pixels which can be integrated on an affordable chip. The resulting loss in spatial sampling rate is a handicap of existing visual sensors with focal plane processing, in comparison to traditional approaches of machine vision combining a camera—with a fill factor close to 100%—with external processing hardware. Another issue plaguing visual sensors is fixed-pattern noise caused by random spatial fluctuations of device parameters within a pixel array. The level of pixel mismatch is frequently such that only strongly contrasted edges can be detected reliably, whereas dimmer image features are lost in fixed-pattern noise. Signal processing techniques capable of overcoming this problem tend to introduce undesirable side effects, such as temporal sampling or hardware overhead for offset storage.

A novel principle for the acquisition of visual information is emerging, which extends the effective resolution of a pixel array far beyond the limit imposed by pixel spacing. It is also inherently insensitive to fixed-pattern noise. Instead of measuring the distribution of light intensity at fixed locations, continuous small-amplitude oscillatory movements are applied to the imaging system. As a result of such movements, spatial variations of light intensity in the image turn into temporal fluctuations of light intensity at every photoreceptor. For instance, sweeping a photoreceptor over

*To whom correspondence should be addressed at ania@caltech.edu

a thin spatial feature—such as a cable in an outdoor scene—can produce a detectable impulse of photocurrent, even if this feature is much thinner than pixel spacing. The effective spatial resolution of the sensor is limited by the focusing optics and pixel temporal bandwidth. Knowing the pattern of movements applied to the system, local spatial features can be retrieved from the temporal waveform detected by each photoreceptor. These waveforms can be processed locally in such a way that pixels transmit only higher-level feature information off the chip. Each pixel acts as a high resolution local feature detector. Another benefit of scanning is independence from fixed-pattern noise. Since spatial features are analyzed by moving a single photoreceptor over the feature instead of combining the signals delivered by distinct photoreceptors, parameter fluctuations between pixels cannot influence the visual data acquisition process.

It is interesting to point out that visual sensors found in some living organisms appear to rely on a related principle. Jumping spiders (*Salticidae*) acquire visual data by sweeping an essentially linear retina back and forth perpendicularly to its larger dimension, while slowly rotating the retina in its own plane. These spiders are capable of complex visual prey/mate discrimination [3] and route finding tasks [4] using two scanning retinae containing only about 800 photoreceptors each [5]. Scanning has been reported in flies as well [6], which inspired researchers to build a scanning visual sensor for flight control using off-the-shelf components [7]. Humans also rely on tiny periodic vibrations of the retinas (micro-saccades) to prevent the retinal image from fading.

In the field of image sensing for restitution purposes, a number of devices have been proposed (e.g. [8]), which apply subpixel shifts to an image by optical means in order to enhance the intrinsic resolution of a camera. This procedure differs from our scheme in that a camera delivers frames of raw image data at discrete locations, while our sensor exploits temporal waveforms of light intensity resulting from continuous movements to extract spatial image features. A truly continuous two-pixel scanning visual system implemented with discrete components has been described [7]. More recently, the effect of scanning was verified using two linear arrays of p-i-n photodiodes on a millimeter scale mobile robot, and a 1-D microlens array was fabricated for the same purpose [9]. In addition, an elegant implementation of a 2-D pixel array exploiting vibrations has been described [10] and implemented in CMOS [11], incorporating local feature processing based on correlation between the photoreceptor signal and a template temporal waveform. A limitation of this scheme is that only one type of feature can be detected at a time. In this paper, we propose an implementation of a vibrating 2D visual sensor which encodes temporal signal features in the timing of digital pulses transmitted in real time off-chip. Spiking patterns can be processed by external hardware to detected a possibly large number of different features in parallel.

In the first part of this paper, we give a functional description of the vibrating visual sensor. Section 2 presents a quantitative estimate of the resolution enhancement provided by vibrations. Section 3 describes the signal processing required in every pixel for practical use of the concept. In the second part, we describe a hardware implementation of a vibrating visual sensor. Section 5 discusses two different mechanical and optical devices producing image vibrations. Sections 6 and 7 present an analog VLSI chip sensing and processing the resulting image data. The integrated circuit incorporates a novel address-event scheme for transmission of visual data outside of the chip. Some measurement results are presented along with improvements being incorporated into a circuit redesign (in progress).

2 Resolution enhancement by scanning

Let us first consider the case of a 1D image, $I(x)$, of an unchanging visual scene focused onto the surface of a visual sensor. If this image is shifting at a velocity v over the sensor as a consequence of mechanical vibrations occuring at some point in the optical path, a single photoreceptor will detect a light intensity $I_{pix}(t) = I(x_0 + v \cdot t)$, where x_0 depends on the location of the photoreceptor on the sensor. The spatial distribution of light intensity within the image is transformed into a temporal signal. Assuming a constant scanning velocity v, the spectrum of the temporal signal is related to

the spatial spectrum of the image by linear scaling of the frequency axis:

$$f_T = v \cdot f_S \tag{1}$$

where f_T designates temporal frequency and f_S designates spatial frequency in the image plane. If the photoreceptor has a temporal bandwidth of f_{Tmax}, the spatial cutoff frequency for a scanning pixel will be $f_{Smax} = f_{Tmax}/v$. The spatial bandwidth of a non-scanning image sensor is entirely dependent on the spacing Δx of its photoreceptors and equals $1/(2\Delta x)$. Thus, scanning can improve the spatial resolution provided that

$$\frac{f_{Tmax}}{v} > \frac{1}{2\Delta x} \tag{2}$$

In the case of a 2D image subject to mechanical vibrations along both axes, each photoreceptor acquires visual information along a curvilinear scanning path determined by image movements. In the remainder of this paper, we will primarily consider constant velocity scanning along a circular path with a diameter equal to the pixel spacing Δx. Continuous image data is collected along the scanning path with a resolution determined by the same analysis as in the 1D case. The areas of the image inside the circular paths are not scanned. An image feature such as a line segment can be detected if it is long enough to cross the scanning path of at least one photoreceptor, even if it is much thinner than pixel spacing.

In the discussion so far, it has been assumed that the visual scene and its illumination conditions remain constant. If illumination changes over time, due for instance to AC powered light sources, these fluctuations interfere with the scanning process by modulating the temporal signal generated by scanning. This effect can be cancelled by dividing the photoreceptor current by a baseline signal proportional to illumination intensity. The baseline signal can be acquired by spatial averaging of incoming light over a wide field of view. Beside illumination changes, the visual scene itself may change because of independent object motion. The temporal frequency range of image intensity fluctuations which can be tolerated is limited by the scanning frequency, which is defined as the number of scanning cycles completed per unit time. In this respect, the proposed image sensor is subject to the same temporal aliasing phenomenon as occurs in image sensors operating in discrete frames.

In the current implementation, the visual sensor is designed to operate at a scanning frequency of 300Hz with a photoreceptor bandwidth of 10KHz. Photoreceptor spacing is $68.5\mu m$ (in the first fabricated chip), and the scanning path is circular, with a diameter equal to pixel spacing. Using these parameters and the equations introduced earlier in this section, it can be shown that the effective spatial resolution in the image plane along the scanning path is on the order of $6.5\mu m$. In terms of viewing angle, the resolution would be about 0.08°, for a focal length of 4.5mm. This resolution is close to the diffraction limit of the optics. To achieve the full potential of this scanning scheme, the photoreceptor should be smaller than the spatial resolution calculated in the image plane. However, due to difficulties in achieving a large bandwidth and an acceptable signal-to-noise ratio with extremely low photocurrents, we chose to make the photoreceptor $10\mu m$ by $10\mu m$, thereby reducing the effective resolution.

3 Signal processing

The continuous temporal waveforms delivered by the photoreceptors carry high-resolution visual information, but this information is not readily available in a suitable format for machine vision applications. Besides, it would be impractical to send out of the chip all continuous waveforms as provided by the photoreceptors. For these reasons, signal processing must be performed locally in every pixel for the purpose of detecting key features in the temporal waveform, and encoding them in a format compatible with off-chip communication. In the first stage of signal processing (Fig. 1), the current delivered by the photoreceptor is applied to a logarithmic current-to-voltage converter. The same visual scene under different illumination levels produces images differing only by a scaling factor

in intensity. After logarithmic transform, the temporal waveforms produced by scanning differ only in their DC component, which is ignored by subsequent processing stages. Therefore, the logarithmic operator contributes to make visual data delivered by the sensor independent of illumination level [1]. The signal resulting from logarithmic compression is differentiated with respect to time and half-wave rectified, whereby both the positive and the negative fraction are retained separately. Current signals delivered at both outputs of the rectifier are sent to independent non-leaky integrate-and-fire circuits, where the charge is accumulated over time until the resulting voltage reaches a threshold. At this point, the integrate-and-fire block emits a short pulse, resets its integrator and resumes operation. Pulses from the positive and negative integrate-and-fire blocks are the final output signals delivered by the pixel. They are sent off-chip by means of a communication bus described later (Section 7). Separate encoding of positive and negative transitions in the image is a feature found also in the human retina.

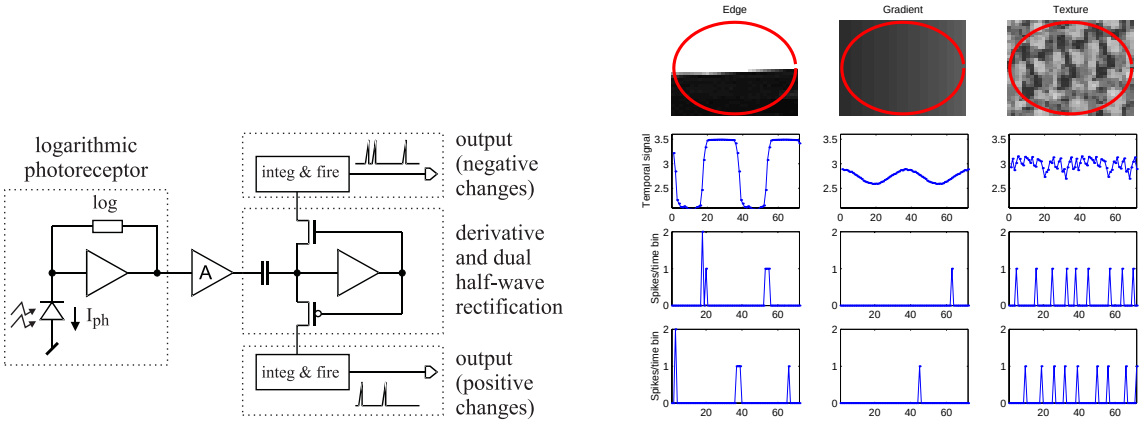

Figure 1. Left: Functional block schematic of a single pixel. Right: Simulation results showing firing rates at the outputs of a pixel during circular scanning for different spatial patterns. The first row shows raw images and the scanning path of the pixel. The second row shows the temporal waveforms of light intensity detected by the photoreceptor over two scanning cycles. The next two rows show the spiking rate at the positive and the negative outputs of the pixel respectively. The duration of a cycle has been split into 36 time bins, which are indicated on the horizontal axis. The vertical axis corresponds to spiking rate, which is measured as the number of spikes per time bin.

Whenever the scanning path of a photoreceptor crosses a sharp edge causing an amplitude change exceeding the built-in threshold, at least one spike is reliably generated at this point at every scanning cycle (Fig. 1). The presence and location of this edge can be inferred off-chip by observing that a spike occurs at the scanning frequency and at an essentially constant phase with respect to the scanning cycle. In another prototypical case where an area of the image contains only a weak intensity gradient instead of a sharp edge, the temporal waveform contains only low amplitude fluctations proportional to the magnitude of the gradient. In this case, it takes several scanning cycles—in inverse proportion to the gradient magnitude—before a spike can be generated, and this spike may occur any time the intensity is changing. Highly textured surfaces are distinguished by a high firing rate and little or no phase-locking. The pixel signal processing chain encodes visual information in spiking patterns in such a way as to devote a high communication bandwidth and low latency to sharply contrasted spatial variations, and a lower bandwidth to weaker gradients. The spike trains sent off-chip are meant to be used directly by external hardware extracting image features from the spiking patterns. Since sharp edges are phase-coded with respect to the scanning cycle, signatures

of specific spatial patterns can be detected using simple delay-and-coincidence detectors. Gradient information is rate-coded, and can be recovered by low-pass filtering or histogramming. We have designed and successfully simulated an algorithm for detecting the presence and orientation of a single edge from the spike train generated by a pixel. We are working on the development of solutions for the detection of more complex spatial features as well.

4 Hardware Implementation Overview

Hardware implementation of a vibrating visual sensor requires building a mechanical and optical device to shift an image along some scanning path while keeping it in focus. Two possible constructions solving this problem are described in Section 5. Another key component of the system is the visual sensing integrated circuit. We have designed a custom VLSI chip incorporating a 32 by 32 array of pixels implementing an analog signal processing chain as described in Section 3, together with a digital communication scheme to transmit visual information outside the chip. The chip has been manufactured in a $0.6\mu m$ double poly, triple metal CMOS process. The design of a second version is in progress. The chip is designed for a supply voltage of 3V. The pixel array occupies 2.2mm by 2.2mm and the entire chip area is about $10mm^2$. Design details of this integrated circuit are given in Section 6 for the pixel and Section 7 for the communication bus.

5 Mechanical Design

A simple way to provide constant-velocity scanning along a circular path consists of spinning a tilted mirror in front of the focusing lens (Fig. 2). The mirror must be mounted on the axis of a motor which should be tilted at an angle of about 45° with respect to the optical axis of the lens. If the mirror is not exactly perpendicular to the motor axis but tilted by a small angle ϵ (measured at 0.56°), rotation of the motor will cause the reflective surface to wobble, thereby causing the image to travel a circular path with a radius of 2ϵ in viewing angle. A position encoder on the motor axis indicates the orientation of the mirror at all times. The signal from this encoder serves as a reference for the interpretation of phase-coded spiking patterns delivered by the visual sensor.

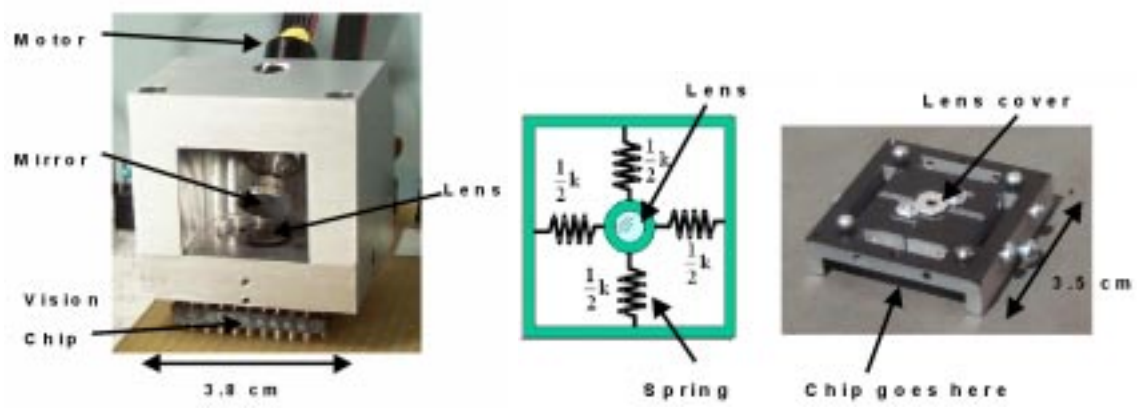

Figure 2. Left: Photograph of mechanical device producing circular scanning. Center and right: Drawing and photograph of mechanical device producing scanning powered by environmental vibrations amplified near a chosen resonant frequency.

The spinning mirror device was easy to build and provides accurate control over the scanning path. Therefore, it is most appropriate for laboratory experiments. For practical applications where

space and power consumption are an issue, we have designed an alternative device where an irregular scanning pattern is caused by displacements of the lens. In this device, the lens is mounted on springs allowing lateral X-Y displacements but maintaining constant spacing between the lens and the chip. If the system is mounted onto a vibrating platform such as a vehicle driving on a rough surface, the mechanical energy available in the vicinity of the resonance frequency of the lens/spring system will cause scanning movements. To be effective, the amplitude of these movements must be on the order of pixel spacing on the chip, e.g. a few tens of microns. The shape of the scanning path will depend on the relative magnitudes and phases of vibrations applied to the X and the Y axes, and on the resonance frequency matching between these axes. As the scanning path will vary over time depending on environmental vibratory conditions, it is necessary to monitor the position of the lens and use this information in the interpretation of the pulse trains generated by the visual sensing chip. The lens position can be monitored by capacitive measurements between the lens socket and surrounding fixed electrodes.

A prototype scanning device operating on the principle described herein has been manufactured (Fig. 2). Its dimensions are 1.36" by 1.36" by 0.32". The calculated mass of the lens and its socket is 760mg. The springs have a measured radially symmetric spring constant $k = 12KN/m$, implying a resonant frequency of $645Hz$. This frequency can be adjusted by modifying the mass of the lens lid. Assuming that the lens/spring resonator has a quality factor of 10, a mechanical power of about $75\mu W$ is required to sustain oscillation with a peak-to-peak amplitude equal to pixel spacing. In applications where this amount of power is not guaranteed to be available in the environment, the system can be mounted on piezoelectric actuators.

6 Pixel Design

6.1 High-Bandwidth Logarithmic Photoreceptor

We selected the n-well/p-substrate diode available in CMOS processes as a photoreceptor because of its low parasitic capacitance compared to other junctions. The dimension of the photodiode is $10\mu m$ by $10\mu m$, resulting from a compromise between spatial resolution and photocurrent intensity constraints (Section 2). Under typical indoor illumination conditions, we expected a DC photocurrent on the order of 30pA. In our dimly lit laboratory and the chip pointing at the ceiling with no focusing optics, we measured a DC photocurrent of 68pA. With a small (3mm diameter) lens, the photocurrent dropped to 4pA, implying the need for a larger lens in this environment. At other illumination levels, the DC photocurrent can differ from this value by several decades. One of the challenges related to the proposed visual sensing scheme is to design a logarithmic current-to-voltage converter achieving a bandwidth of 10KHz even at the low photocurrent levels available indoors. It is critical to prevent any ringing or overshoot, as such ringing would be interpreted as spatial texture by subsequent processing stages. In addition, the converter must be very compact and low-power in order to fit in every pixel. Three possible logarithmic amplifier topologies combining a MOSFET and a transconductance element have been compared (Fig. 3). They differ by the type of the MOSFET and by the control electrode used to apply feedback. The parasitic capacitance on nodes a and

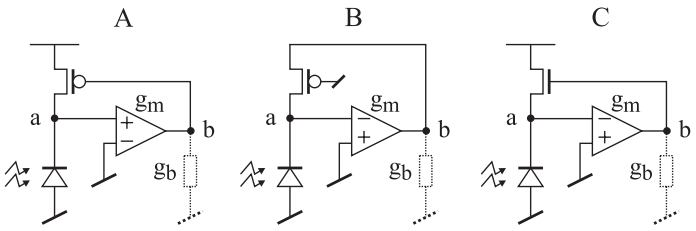

Figure 3. Logarithmic amplifier topologies

Table 1. Analytical expression of parameters determining the frequency response of a logarithmic amplifier

Topology	A	B	C
ω_1	$\frac{g_{mph}}{g_b} \cdot \frac{g_m}{C_a}$	$\frac{1}{1+\frac{g_b}{g_{msph}}} \cdot \frac{g_m}{C_a}$	$\frac{\frac{g_m}{n}+g_b}{C_a \cdot \frac{g_b}{g_{msph}}+C_b}$
ω_2	$\frac{g_b}{C_b}$	$(1+\frac{g_b}{g_{msph}}) \cdot \frac{g_{msph}}{C_b}$	$\frac{g_b}{C_b}+\frac{g_{msph}}{C_a}$

b (not shown in the figure) is denoted C_a and C_b respectively, whereas g_b represents the load conductance on node b. Conductance g_b can be either parasitic—resulting from drain conductances—or intentional. If the current delivered by the photodiode is designated by I_{ph} and the channel current of the MOSFET is called I_{mos}, the small-signal transfer function of the logarithmic amplifier can be written

$$\frac{I_{mos}}{I_{ph}} = H(s) = \frac{1}{1+\frac{s}{\omega_1} \cdot (1+\frac{s}{\omega_2})} \tag{3}$$

Parameter ω_1 is the bandwidth of the amplifier, whereas the ratio ω_2/ω_1 determines damping. Parameters ω_1 and ω_2 are topology-specific and are given in Table 1. In this table, g_m is the transconductance of the amplifier, g_{mph} designates the transconductance of the MOSFET as seen by the gate, g_{msph} is the transconductance seen by the source, and n is the ratio g_{msph}/g_{mph}. The MOSFET transconductance depends on photocurrent, and can therefore vary over several decades. Despite this fact, the circuit must remain close to the point of critical damping, which is reached when

$$\omega_2 = 4 \cdot \omega_1 \tag{4}$$

Examination of Table 1 reveals that this condition can be met at any photocurrent level by making both g_m and g_b proportional to the light-dependent transconductance g_{msph} or g_{mph}. The proper ratios g_m/g_{msph} and g_b/g_{msph} depend on topology, parasitic capacitances C_a and C_b, and on the required bandwidth. In all topologies, it is easy to make g_m proportional to g_{msph} by adapting the bias current of the amplifier. Accurate control over the load conductance g_b is generally more difficult, especially if g_b is determined mostly by drain conductances g_{ds} as in topologies A and C. However, topology B lends itself to a simple and accurate implementation of g_b. In the transistor-level schematic of the proposed logarithmic amplifier (Fig. 4), P1 is the MOSFET depicted in the simplified schematic of Fig. 3B, P2 serves as a load conductance g_b, whereas N1 and P3 implement the transconductance amplifier g_m. A constant bias voltage is applied to terminal g, chosen in such a way to ensure saturation of P1 and P2—typically 200mV above ground. Instead of a classical two-

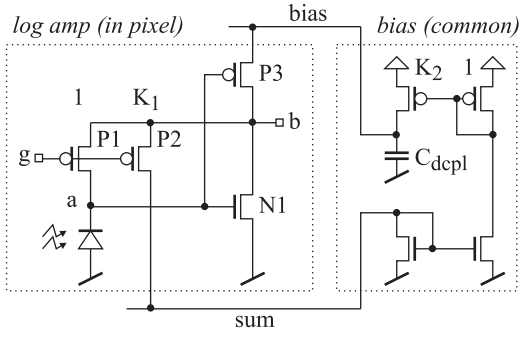

Figure 4. Transistor-level schematic of the proposed high-bandwidth logarithmic amplifier

transistor amplifier with a constant bias voltage at the load, we chose to use a push-pull structure

similar to a CMOS inverter. This circuit has twice the transconductance of a classical amplifier at the same bias current level, which reduces power consumption [12]. To keep accurate control of the bias current independently of supply voltage, process parameters and temperature, the inverter is biased by a current source decoupled by a capacitor, whereby the transconductance doubling feature is available only for AC signals. The main output of the circuit is the voltage on node b, but an auxiliary output at the drain of P2 delivers a current proportional to the photocurrent. Since P1 and P2 have the same gate and source voltage and are both saturated, the ratio of their channel currents is a constant K_1 determined by their geometry[1]. Thereby, the small-signal load conductance g_b depicted in Fig. 3B has a value of $K_1 \cdot g_{msph}$ as required for illumination-independent dynamic behavior. Since the current across P2 is proportional to the photocurrent, it can be used in the bias loop controlling the transconductance g_m of the amplifier. For this purpose, the current is scaled by a factor K_2 by means of two mirrors, the output of which serves as the current source delivering power to P3. For the sake of saving area, the scaling mirrors are shared by all 1024 pixels on the chip. The drain current of P2 from all pixels is collected on global line 'sum' shown in Fig. 4. The bias current delivered by the scaling mirrors is distributed to the sources of P3 in all pixels via global line 'bias'. Thereby, transconductance g_m is adapted to the average illumination level in the image instead of the local intensity in every pixel. The large decoupling capacitor C_{dcpl} is external.

The small-signal transfer function parameters of the proposed logarithmic amplifier can be written by substituting K_1 for occurences of g_b/g_{msph} in Table 1. The bandwidth can be written in the following form:

$$\omega_1 = \frac{g_m}{(1+K_1)C_a} \tag{5}$$

Similarly, the condition for critical damping can be written

$$\frac{g_m}{g_{msph}} = \frac{C_a}{4C_b} \cdot (1+K_1)^2 \tag{6}$$

Under the assumption that all devices operate in weak inversion, the current scaling factor K_2 in the bias loop must be

$$K_2 = \frac{C_a}{8C_b} \cdot \frac{(1+K_1)^2}{K_1} \tag{7}$$

Compared to a passive logarithmic photoreceptor, which would have a bandwidth of g_{msph}/C_a, the proposed circuit improves the bandwidth by a factor A given by

$$A = \frac{C_a}{4C_b} \cdot \frac{(1+K_1)^2}{K_1} \tag{8}$$

For $K_1 \gg 1$ and $C_a = 4C_b$, the improvement factor is on the order of K_1. The cost of this increase in bandwidth is a power consumption increase proportional to K_1^2. In our implementation, we chose $K_1 = 20$ as a compromise between power consumption and bandwidth. Substituting K_1 and parasitic capacitance estimates into Equation 7, we determined $K_2 = 25$. Thereby, the current consumption of the photoreceptor front-end is about 500 times the DC photocurrent. To avoid excessive power consumption at high illuminations where bandwidth improvement is unnecessary anyway, a limiting mechanism—not shown in Figure 4—is inserted in the bias loop, which restricts current to about 5nA per pixel. At very high illuminations, this limiting mechanism can cause ringing, but the resonance peak is located well outside the bandwidth of subsequent signal processing stages.

The measured frequency response of the photoreceptor circuit at different DC photocurrent levels is shown in Figure 5. Each trace in this graph represents the amplitude of the output voltage as a function of frequency, at the photocurrent level indicated next to the trace. At low frequencies, the output voltage amplitude is nearly the same at all photocurrent levels, confirming operation as a

[1] Assuming weak inversion operation, a small voltage difference between the gates of P1 and P2 could be applied instead of a geometrical factor. This approach has been used in the actual implementation to save area, but we describe the geometrical approach for conceptual simplicity.

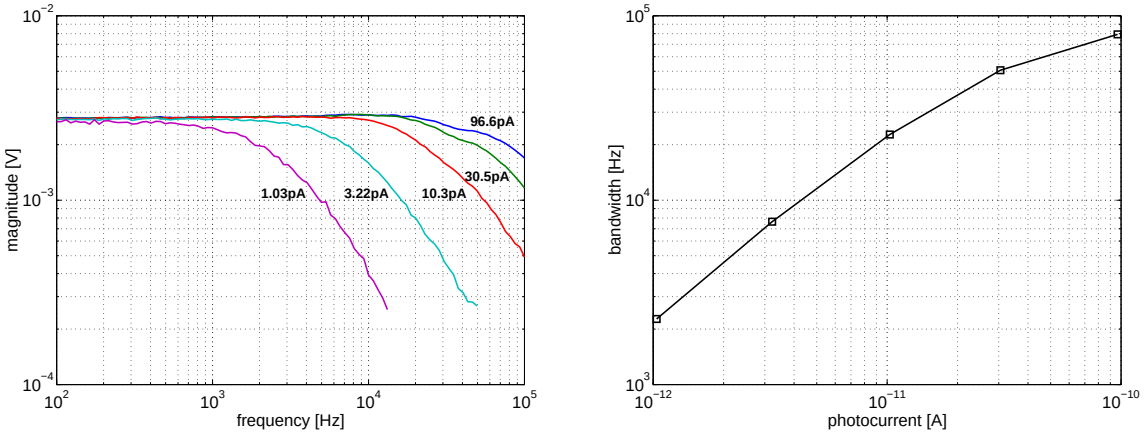

Figure 5. Left: Measured logarithmic amplifier response at different photocurrent levels. Amplitude-modulated current with a modulation factor of 8.5% is applied at the input in lieu of a photocurrent. The voltage is measured at the output of the photoreceptor circuit. Right: Logarithmic amplifier bandwidth as a function of photocurrent level, derived from same data.

logarithmic amplifier. The cutoff frequency is 21.5KHz at a photocurrent of 10pA, higher than the target bandwidth of 20KHz at 30pA. Bandwidth is approximately proportional to photocurrent at low intensities (Fig. 5). It should be noted that subsequent processing stages limit the bandwidth to 20KHz independently from light intensity.

The minimum light intensity contrast detectable by the described photoreceptor front-end is limited by intrinsic noise. Detailed small-signal analysis of the circuit shown in Figure 4 shows that noise current injected onto node b is cancelled by the feedback loop. Thereby, in a first order approximation, intrinsic noise of devices P2, P3 and N1 does not affect the output voltage on node b within the signal bandwidth of the circuit. Dominant noise sources are due to the photodiode itself and feedback transistor P1. The power spectral density of shot noise due to these two devices is given by

$$\frac{dV_N^2}{df} = \frac{4qU_T^2}{I_{ph}} \tag{9}$$

where q is the electron charge, U_T is the thermal voltage, and I_{ph} is the DC photocurrent. Assuming a signal bandwidth of 10KHz limited by a single pole and operation at room temperature, the noise voltage in indoor conditions ($I_{ph} = 30pA$) is about $476\mu V$. It takes a contrast of about 0.7% in the image to produce an output signal of the same amplitude.

6.2 Signal Processing Chain

The voltage delivered by the logarithmic current-to-voltage converter described above must be differentiated and half-wave rectified as described in Section 3. Differentation is implemented by capacitively coupling the voltage delivered by the photoreceptor front-end to a current rectifier with a virtual ground at its input node. It turns out that the amplitude of the photoreceptor signal is relatively low—5mV peak-to-peak for a sinusoidal photocurrent with a modulation factor of 10%. In principle, this signal could be used directly, but a large coupling capacitor would be required to produce charge fluctuations detectable by the rectifier. It turns out to be more economical in space to amplify the signal by a constant factor and reduce the coupling capacitor by the same amount. We elected to add a separate amplification stage after the photoreceptor front-end instead of building gain into the logarithmic amplifier as in other designs [13]. If a voltage gain of A is built into the

logarithmic amplifier—by attenuating the signal in the feedback path by a factor A for instance—then the gain-bandwidth product of the amplifier must be increased by the same factor A, hence the power consumption. It is more economical in terms of power to use an extra amplification stage. The area overhead remains minimal, because the extra transistors require only a modest fraction of the area occupied by the capacitive divider which is required in both approaches. Besides the amplifier and rectifier, the pixel contains two integrate-and-fire circuits followed by digital interfaces for transmission of spikes toward the periphery of the pixel array. Because of space constraints, these blocks are not described in detail.

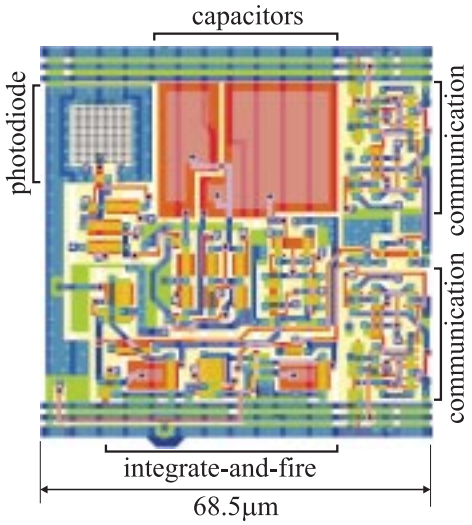

Figure 6. Layout of a single pixel

The layout of a complete pixel is shown in Figure 6. It has a size of $68.5\mu m$ by $68.5\mu m$. Vertical interleaved and overlapping stripes of metal-2 and metal-3 are used for light shielding and routing of power and global signals. The metal-1 layer is used for local interconnects and horizontal lines for the communication of spikes. The area of the photodiode is $10\mu m$ by $10\mu m$, which results in a fill factor of 2.1%.

6.3 Overall Noise in the Pixel

Despite the fact that intrinsic noise within the photoreceptor circuit is close to the minimum level determined by physics, we measured an unacceptably high baseline rate of noise-induced spikes. It turns out that the spiking rate is proportional to the RMS signal after differentiation. Generally, the ratio of noise-induced spikes to signal-induced spikes is substantially larger than the signal-to-noise ratio at the photoreceptor output, because differentiation emphasizes high-frequency components of the signal. To minimize noise-induced spiking at a given illumination level, the pixel bandwidth must be kept as small as resolution enhancement constraints allow (10KHz would suffice). Bandwidth must be limited independently from illumination level by at least a second-order filter. In addition, noise-induced spiking can be reduced or cancelled by introducing an amplitude threshold before rectification. Since this threshold also suppresses very low-amplitude signals, it must be carefully chosen to preserve a useful range of detectable image constrast. These noise reduction techniques are built into a second version of the chip, which is currently under design.

6.4 Measurement Results

We recorded the timing of spikes generated by an individual test pixel in response to a sine wave current input in place of the photocurrent. The spikes are histogrammed against the phase

of the sine wave (Fig. 7). As expected, the envelope of the histogram of the spikes corresponds

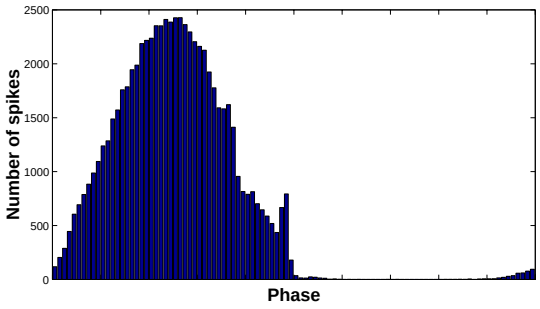

Figure 7. Histogram of spike output of the test pixel to multiple cycles of a sine wave input. Each pixel has two outputs, one for positive intensity transitions and another for negative transitions. Only one of the two outputs has been histogrammed.

to the derivative of the input. That is, the probability of spiking is proportional to the slope of the input. Since the signal is half way rectified and each portion is fed to a separate integrate and fire mechanism, only half the derivative waveform is portrayed. Because the magnitude of the input signal was far greater than that of the noise, the level of the baseline noise-induced spiking is low. The spike baseline curves slightly. When the derivative of the input signal is very large, small-amplitude noise is unlikely to push it towards zero such that charge will be collected by the wrong integrate-and-fire mechanism. When the derivative is near zero, on the other hand, even small noise fluctuations can cause the signal seen by the integrate and fire circuits to fluctuate between positive and negative (Fig. 8). Thereby, fewer noise-induced spikes occur in phases where the signal amplitude is large.

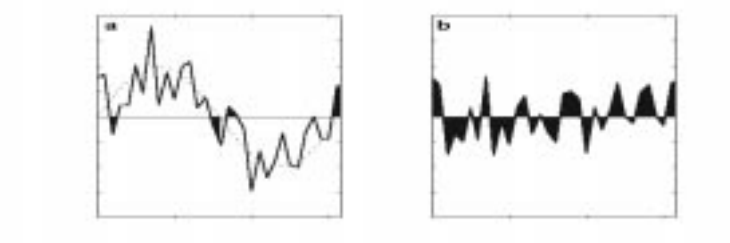

Figure 8. Integration of noise for (a) a large amplitude signal and (b) zero amplitude signal. The times when noise rather than signal is integrated are shown by filling the area under the curve with solid color. This problem is most severe when the signal is close to zero.

7 Off-Chip Communication

The most common technique for transmitting information from an imaging array to other circuitry involves the serial scanning of all pixels and transmission of digital values encoding the level of activity at each pixel. Such a scheme forces the quantization of time into bins of hundreds or, at best, tens of nanoseconds. Because our signal encoding scheme relies on precise temporal localization of asynchronous events we chose an alternate scheme, namely address event representation (AER).

The underlying premise of AER is that the activity of a pixel can be represented in the frequency and timing of digital spikes (similar to neuronal action potentials in biology). Since the pixels are not clocked, the timing of spikes from different pixels is asynchronous and independent. The activity of the pixel array can then be described by a list of events $\{(x_0, t_0), (x_1, t_1), (x_2, t_2), ...\}$ describing the locations x_i and corresponding times t_i when events occurred. Since the events are broadcast off the chip in real-time, t_i is not explicitly encoded. This manner of information transmission allows active pixels more frequent access to the bus (and thus more bandwidth) than quiescent ones.

7.1 Previous Work and Design Overview

Versions of this general idea have previously been implemented by Mortara [14]; Boahen [15]; and Lazzaro, Wawrzynek et al. [16] and used by others [17], [18], [19]. For several reasons, we have chosen to design a custom AER implementation. Mortara's scheme [14] boasts minimal hardware and short latencies but does not deal with collisions (simultaneous spikes) gracefully. All pixels access the output bus simultaneously. No guarantee exists for how long a valid address will remain on the bus before being corrupted by a colliding event. Thus, this AER scheme is difficult to interface with off-the-shelf digital circuits. More sophisticated schemes [15, 16] implement arbitration which allows only one pixel at a time to access the bus. While the address from one pixel is being processed, no other events may appear on the bus. If several other events occur during this time, the active pixels remain active and waiting until they are sequentially read off and reset. No events are lost, but the timing is skewed by waiting for other almost-simultaneous spikes to be read and by the processing delay due to arbitration. Boahen's latest design has latencies of 30-400ns. Since timing is crucial in our system, we have designed our AER to have a latency of about 45ns according to measurements. Error-checking is integrated into the design so that only valid addresses are transmitted off the chip, and a hand-shaking protocol ascertains that all data is cleanly read before a new address is broadcast.

7.2 Details of Infomation flow

A diagram of the information flow as well as the time course of a transistor level simulation of the circuit are presented in Fig. 9. First, consider the case of a single active pixel (no collision). Upon reaching threshold it broadcasts a digital spike onto two asynchronous lines, one shared along that column of pixels and the other along that row. Each of these lines is terminated by an SR latch controlled by the input enable signals IEx and IEy applied to the latches of all columns and rows, respectively. If the system is ready to accept a new event, the IE signals are active and allow the event to be latched. The event is acknowledged and the pixel resets. The LTx and LTy signals indicate that an event has been caught, which in turn causes the input enables (IEx, IEy) to go inactive, thereby preventing new events from being accepted. The latched address is presented to the encoders which use a dense binary code to represent the location of the pixel as $log_2 N$ bits for N pixels in the array. The outputs of the encoders are directly connected to external pins as well as fed to validity checking circuitry. Addresses must be encoded in some way that allows verification that no collision has occurred. We have chosen to use a dual rail bit representation on-chip (see Fig. 10). To represent a one, one line is pulled high; to represent a zero, the other line is pulled high. The lines are actively pulled high when an event on the corresponding array row/column occurs, and passively pulled low otherwise by pull-down transistors. Thus, both lines low indicate an idle state. Both lines high indicate an invalid state resulting from one pixel attempting to set that address bit to a zero and another to a one. Only one of each pair of lines is taken off-chip, since error checking is performed on-chip as described herein. If all has gone well and indeed only one pixel had spiked, a valid address is detected and VALx and VALy go high. This in turn results in REQ being raised to signal off-chip circuitry that an address awaits reading. When the address has been read, ACK acknowledges and the reset procedure begins. First, CLR is raised signaling that the event latches should be cleared. When the latches are cleared, the addresses in the address encoder reset, VALx and VALy fall, and finally RESx and RESy go high to indicate that the address lines have finished

resetting. This allows CLR to fall and, in turn, IEx and IEy to be reactivated in anticipation of the next event.

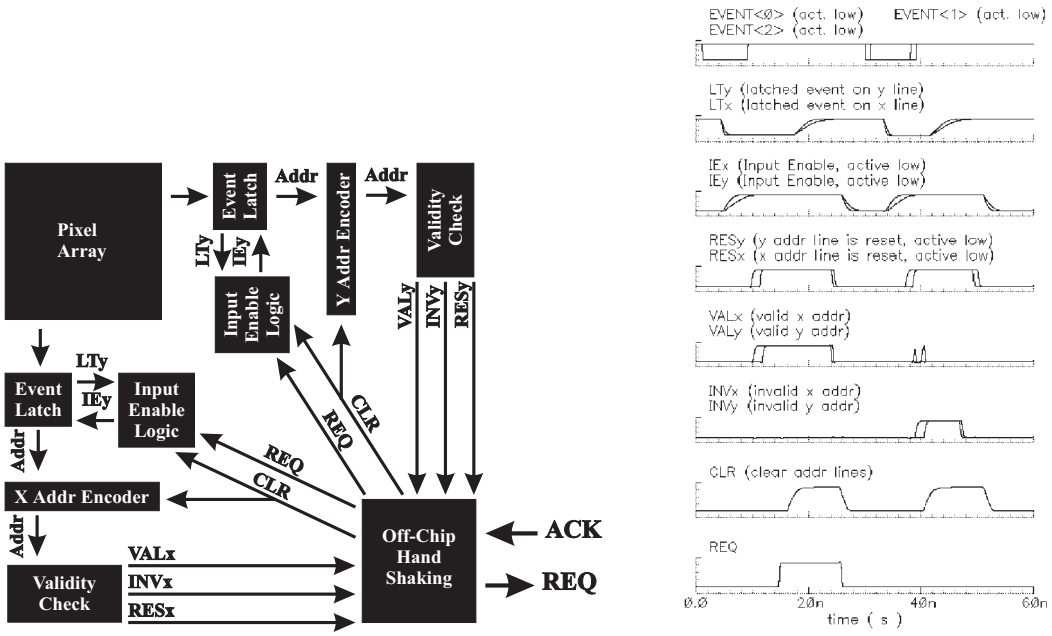

Figure 9. AER signal flow. Left: Schematic of information flow after encoding into digital spikes. ADDR is the address bits; LT indicates an address is latched; IE indicates the system is ready for the next address to be latched; VAL indicates a valid address; INV indicates an invalid address (due to a collision); RES indicates the address lines have been reset; CLR signals that the encoders should be cleared; REQ and ACK are for the hand-shake with the external world, where REQ indicates that the chip ready to have an address read, and ACK signals that off-chip circuitry has completed reading the presented address. Right: Time course of transistor level simulation of digital logic circuitry (most of these signals are not accessible for measurement). The EVENT signals representing pixels spiking were input to the simulation. The REQ and ACK lines were shorted as if the response of the external world was instantanous. The first 30ns show an example of a single pixel cleanly transmitting its information. The second half of the simulation shows a collision between two almost simultaneous events.

Now consider the collision scenario in which two events occur very closely in time. Ideally, we would like the input enables to inactivate immediately when a single event is caught to prohibit simultaneous latching of two events. However, because disabling the latches takes about 1.5ns, two events may be simultaneously latched. As described above, for every address bit in which the pixels differ, both lines representing that bit will be raised. Thus an invalid address will be detected. To avoid the waste of time and error-checking resources necessitated by sending invalid addresses off-chip, CLR is raised and the reset procedure proceeds as above, but REQ never goes high to request that the address be read.

Lastly, it is possible that while an event is being processed and IEx and IEy are inactive, another event or several events occur. The active pixels will continue to attempt to broadcast their spike to the latches until their data is acknowledged by the latches. Once the first event has been fully

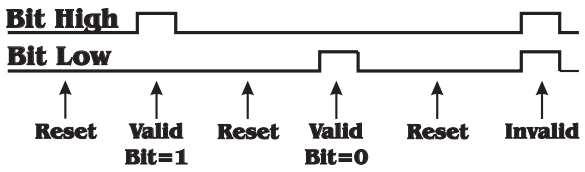

Figure 10. Dual rail bit representation uses two lines to encode each address bit. Using two lines the four possible address states are easily encoded.

processed, the latches cleared out, and the input enables (IEx, IEy) activated, the latches will immediately receive and latch all events which occurred while the first event was being processed. The cycle will proceed as described in the two scenarios above: if only a single event had transpired, a valid address will be transmitted off chip; if multiple events had transpired, an invalid address will be detected and the system reset without off-chip communication.

7.3 Races

One race situation is possible and has been addressed in the design: when two pixels spike very closely in time, it is possible that the first one will cause a valid address to be detected and signalled. Before the input enables (IEx, IEy) can be deactivated to prevent new events from being caught, a second event may be latched. This results in VALx or VALy transiently going high before the INVx or INVy signal rises. We have designed a slow logic gate to process the VALx and VALy signals, such that REQ will be raised more slowly than the possible delay in raising IEx and IEy (about 2ns).

7.4 Performance

The measured transmission latency is about 45ns. This discrepancy with simulation is due to an underestimate of parasitic capacitances and can be improved in the redesign. A 45ns latency implies a 22MHz peak bus capacity. Because pixel spiking is stochastic, pixel spiking rates resulting in 22 million spikes per second would result in many collisions and unacceptable information loss. Measurements of the observed collision rate, as well as simulations based on measured latencies and on latencies derived from circuit simulations, suggest that a data transmission rate of up to about 4MHz results in an acceptable collision rate (see Fig. 11). For a 32 by 32 pixel array with a scanning frequency of 300Hz, an average firing rate of 13 spikes per pixel per scanning cycle can be supported, which we believe is enough in typical visual scenes. This calculation assumes that the receiver can handle this data rate and indeed First-In First-Out (FIFO) digital chips meeting this requirement are commercially available.

8 Conclusion

A new approach to visual sensing for machine vision purposes has been described, which relies on mechanical vibrations in the optical path to turn image features into temporal signals. A system-level analysis, behavioral simulations, and preliminary measurements show that an implementation of this principle is feasible. The potential enhancement in effective resolution is such that pixel spacing is no longer the limiting factor of resolution, but rather optical phenomena. This allows us to approach the resolution of standard cameras while incorporating visual data processing in the focal plane. An integrated circuit implementing a visual sensor taking advantage of this principle has been designed and fabricated, and a redesign to address noise issues is in progress. Two mechanical systems for inducing the mechanical vibrations have also been designed and built. The sensor is intended to be used in robotics navigation applications.

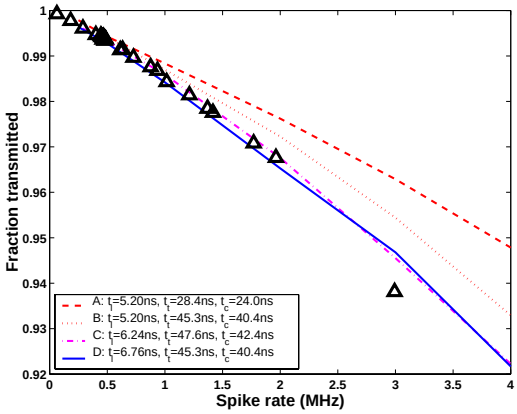

Figure 11. Measurement and simulation of fraction of spikes transmitted as a function of overall array activity. The triangles represent measured data. The lines correspond to simulation outputs given different parameters. The three parameters are the t_l, the time necessary to latch a signal (to inactivate IE); t_t, the time to transmit a spike and reset; and t_c, the time to process and clear a collision. Line A corresponds to values derived purely from a transistor-level simulation and an estimate of parasitic capacitances. Line B combines measurement of the REQ and INV signals from the chip with simulated values. Lines C and D are good fits to the data obtained by only slightly modifying the values of t_l, t_t, and t_c.

Acknowledgements

This work was funded by the Office of Naval Research, DARPA, and the Center for Neuromorphic Systems Engineering, as part of the National Science Foundation Engineering Research Center Program.

References

[1] C. Mead, *Analog VLSI and Neural Systems*, Addison Wesley, 1989.

[2] T.S. Lande, Ed., *Neuromorphic Systems Engineering - Neural Networks in Silicon*, Kluwer Academic Publishers, Dordrecht, 1998.

[3] M.F. Land, "Movements of the retinae of jumping spiders in response to visual stimuli," *J. Exp. Biol.*, vol. 51, pp. 471–493, 1969.

[4] M.S. Tarsitano and R. Andrew, "Scanning and route selection in the jumping spider portia labiata," *Animal Behaviour*, vol. 58, no. 2, pp. 255–265, 1999.

[5] M.F. Land, "Mechanisms of orientation and pattern recognition by jumping spiders," in *Information Processing in the Visual Systems of Arthropods*, R. Wehner, Ed. 1972, pp. 231–247, Springer.

[6] N. Franceschini and R. Chagneux, "Repetitive scanning in the fly compound eye," in *Proc Göttingen Neurobiology Conf*, Wässle and Elsner, Eds., 1997.

[7] S. Viollet and N. Franceschini, "Visual servo system based on a biologically-inspired scanning sensor," in *Sensor Fusion and Decentralized Control in Robotic Systems II*, Bellingham, 1999, SPIE, vol. 3839, pp. 144–155.

[8] N. Sztanko, S.P. Smith, and W.B. Jones, "Cam actuated optical offset image sampling system," Nov. 1994, US Patent #5,363,136.

[9] K. Hoshino, F. Mura, and I. Shimoyama, "Design and performance of a micro-sized biomorphic compound eye with a scanning retina," *Journal of Microelectromechanical Systems*, vol. 9, no. 1, pp. 32–37, 2000.

[10] A. Kimachi, R. Imaizumi, and S. Ando, "Intelligent image sensor with a vibratory mirror mimicking involuntary eye movement," in *Technical Digest of the 16th Sensor Symposium*, 1998, pp. 171–176.

[11] S. Ando and A. Kimachi, "Time-domain correlation image sensor: First cmos realization of demodulator pixels array," in *Proc. 1999 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Japan, 1999, pp. 33–36.

[12] E.A. Vittoz, "Micropower techniques," in *Design of VLSI Circuits for Telecommunication and Signal Processing*, J.E. Franca and Y.P. Tsividis, Eds. 1994, Prentice Hall.

[13] T. Delbruck, *Investigations of Analog VLSI Visual Transduction and Motion Processing*, Ph.D. thesis, California Institute of Technology, Pasadena CA, 1993.

[14] A. Mortara, E.A. Vittoz, and P. Venier, "A communication scheme for analog vlsi perceptive systems," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 6, pp. 660–669, 1995.

[15] K.A. Boahen, "Point-to-point connectivity between neuromorphic chips using address-events," *IEEE Transactions on Circuits and Systems*, 2000, (In press).

[16] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti, and D. Gillespie, "Silicon auditory processors as computer peripherals," *IEEE Transactions on Neural Networks*, vol. 4, no. 3, pp. 523–528, 1993.

[17] M. Mahowald, *An analog VLSI system for stereoscopic vision*, Kluwer Academic Pub., Boston, MA, 1994.

[18] C.M. Higgins and C. Koch, "Multi-chip neuromorphic motion processing," *1999 Conference on Advanced Research in VLSI*, 1999.

[19] T. Horiuchi, Personal communication.