

A Biomimetic VLSI Sensor for Visual Tracking of Small Moving Targets

Charles M. Higgins, *Senior Member, IEEE*, and Vivek Pant

Abstract—Taking inspiration from the visual system of the fly, we describe and characterize a monolithic analog very large-scale integration sensor, which produces control signals appropriate for the guidance of an autonomous robot to visually track a small moving target. This sensor is specifically designed to allow such tracking even from a moving imaging platform which experiences complex background optical flow patterns. Based on relative visual motion of the target and background, the computational model implemented by this sensor emphasizes any small-field motion which is inconsistent with the wide-field background motion.

Index Terms—Analog very large-scale integration (VLSI), biomimetic, spatiotemporal frequency, target tracking, visual motion.

I. INTRODUCTION

THE MAJORITY of modern autonomous robots are built using commercial off-the-shelf sensor components, controlled by software running on a microprocessor. This design strategy is general purpose, highly flexible, and grows more power-efficient each year with the introduction of new low-power digital signal processors. However, as the number of sensors on such a robot grows, the sheer number of physical inputs to the microprocessor required and the complexity of the associated software limit the performance and reliability of such systems.

An insect is a sterling example of a biological autonomous system with a vast array of sensors—far beyond the fondest dreams of modern roboticists—which are integrated into a coherent set of behaviors that are highly efficient in real-world situations. This system never experiences anything analogous to a software “crash” or a solder joint failure, is able to robustly recover from sensor limitations and even removal, and despite employing slow computational elements smoothly interprets the inputs from multiple, possibly conflicting, sensor arrays without processing delays sufficient to cause control problems. While the neuronal mechanisms of behavior selection and control in insects are not well understood, clearly the last 400 million years of evolution have led to the design of a system that roboticists would do well to emulate.

Manuscript received March 5, 2004; revised May 17, 2004. This work was supported by the Office of Naval Research under Agreement N68936-00-2-0002. This paper was presented in part at the International Symposium on Circuits and Systems, Vancouver, BC, Canada, May 2004. This paper was recommended by Associate Editor S. Lande.

C. M. Higgins is with the Department of Electrical and Computer Engineering and Arizona Research Laboratories, Division of Neurobiology, The University of Arizona, Tucson, AZ 85721 USA (e-mail: higgins@ece.arizona.edu).

V. Pant is with the Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ 85721 USA.

Digital Object Identifier 10.1109/TCSI.2004.836856

Attempts to use analog and digital integrated circuits to mimic the physical architecture of biological neuronal systems have led to a wide variety of *neuromorphic* very large-scale integration (VLSI) sensors [1]–[3]. However, the amount of processing that ought to be integrated on a VLSI chip is an open area of study. In the area of neuromorphic vision sensors, the majority of designs may be interpreted as highly specialized imagers to reduce the computationally expensive real-time image processing required of the microprocessor. Despite this fact, the largest gains in power efficiency, physical size and weight, and real-time performance are obtained by including in custom VLSI all possible processing steps, reducing or eliminating the need for a powerful microprocessor. These gains come at the cost of flexibility, for once integrated in analog circuit form, a computational model can only be altered in a limited way, through its biases. A number of authors have addressed visual tracking using custom VLSI chips [4]–[10], and although a number of these designs do provide direct control outputs, none were designed to handle tracking in the presence of complex patterns of background motion.

In this paper, we present a biomimetic sensor that processes an image completely into a control signal to be used directly in the guidance of a robot, requiring no microprocessor. This analog VLSI sensor transduces a visual image focused onto it, computes small-field visual motion information, and passes this information through several further stages of computation to produce a control signal in the form of a single output current. This signal responds strongly to small moving targets, but only weakly to large-field background motion. In the following sections, we describe the computational model which underlies this output, give details of a custom VLSI implementation, and show a visual characterization of the VLSI sensor. An earlier version of this work was described in a conference paper [11].

II. COMPUTATIONAL MODEL

The purpose of the present work is to allow a small autonomous robot to visually track a small moving target. In order to intercept a moving target, the tracking system must itself move through the environment. Thus, the patterns of visual motion that it experiences will include in general not only motion generated by the target, but also motion generated by background features due to the self-motion of the tracking system. If visual motion generated by the background is insignificant relative to that generated by the target, this tracking problem can be solved by simply turning in the direction of the strongest motion seen. However, when background motion is strong, detection of the target requires proper accounting for the background motion pattern.

Our working assumption will be that motion generated by the target will cover only a small region of the tracking system's visual field. This is certainly true while the target is relatively far from the tracking system. The background motion, on the other hand, will in general cover a large region of the visual field [12]. By estimating the background motion pattern, the target can be detected when its motion is *inconsistent* with the background [13].

Flies are highly efficient at tracking other flies at high speed, in complex visual environments [14]. Studies of the fly brain have suggested a neuronally-based model which subserves small-target tracking [13]. We have recently elaborated this model and presented a detailed demonstration and evaluation of its ability to perform small-target tracking in simulation [15]. The computational model implemented by the present sensor is based on a simplified version of this elaborated model. This model utilizes two "eyes" with visual fields to the left and right of the heading direction (refer to Fig. 3), each with an array of one-dimensional (1-D) motion sensors. By differencing the small-target responses from the two eyes, each of which is sensitive to targets which move inconsistently with the background motion pattern, the tracking system can be guided to turn toward and intercept small moving targets.

The first stage of visual processing in the model is small-field motion detection based on the Hassenstein-Reichardt (HR) correlation model [16], described in detail below. Any given small-field motion sensor responds myopically to the motion in its small image region, whether that motion be from a target or from the background. Further processing stages in the model spatially sum small-field motion outputs to create binocular "pool cell" units which respond to particular wide-field patterns of background motion, such as clockwise or counterclockwise rotation. Each small-field motion output is normalized by the response of the pool cell unit to which it contributed. This has the effect that, if the response of a given small-field unit is consistent with the background motion, its response is diminished. However, if a small-field unit is responding inconsistently with the wide-field background motion, its response is enhanced relative to other small-field units. A small-target response synthesized in this way results in a sensor which responds to small moving targets, but not to large-field motion (since this is interpreted as background motion), and which responds most strongly to small moving targets when their motion is in opposition to the background motion pattern. This small-target tracking model is also described below.

A. HR Correlation Model

The HR correlation model [16], [17] was proposed in the 1950s to model the optomotor response of the beetle *Chlorophanus*, and is currently the leading mathematical model of insect elementary motion detection [18]. Several authors have fabricated monolithic VLSI versions of this algorithm [19]–[21]. The HR model (shown in Fig. 1) synthesizes two subunits sensitive to motion in opposite directions by combining the delayed response from one photoreceptor with the undelayed response from a second. These two channels are "correlated" using a multiplication, followed by synthesis of the final output from the two subunits.

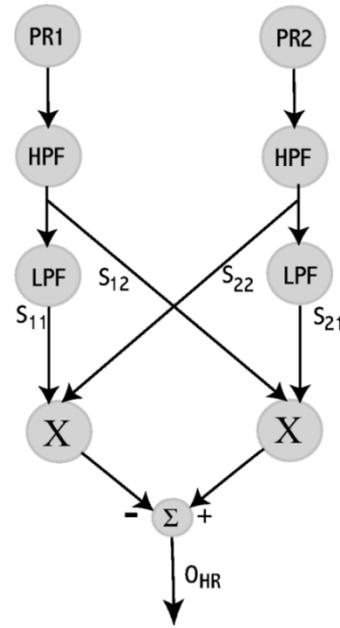


Fig. 1. HR correlation model. PR1 and PR2 are inputs from two neighboring visual sampling points. HPF and LPF denote, respectively, high-pass and low-pass temporal filters. X indicates a product, and Σ indicates a summation. The signals shown are S_{11} and S_{21} (resp. the delayed responses of photoreceptors one and two), S_{12} and S_{22} (resp. the undelayed responses of photoreceptors one and two), and O_{HR} (HR motion output).

In terms of the signals shown in Fig. 1, the output of the HR algorithm is

$$O_{HR} = S_{21} \cdot S_{12} - S_{11} \cdot S_{22}. \quad (1)$$

In the special case of a 1-D moving sinusoidal grating stimulus with contrast C , temporal frequency ω_t , and spatial frequency ω_x ,

$$I(x, t) = \frac{1}{2} \cdot (1 + C \cdot \sin(\omega_t \cdot t + \omega_x \cdot x)). \quad (2)$$

O_{HR} can be shown to be

$$O_{HR} = \frac{C^2}{4} \cdot \sin(\phi_s) \cdot h_1^2 \cdot h_2 \cdot \sin(\phi_2) \quad (3)$$

where ϕ_s is the product of ω_x and the photoreceptor spatial separation Δ , h_1 is the magnitude response of the high-pass filter, and h_2 and ϕ_2 are respectively the magnitude and phase response of the low-pass filter. This output is constant over time, separable in spatial and temporal frequency, and bandpass in both temporal and spatial frequency. Since the sign of the temporal frequency determines the sign of the relative phase ϕ_2 , the sign of the output depends on stimulus direction.

B. Small-Target Tracking Model

The small-target tracking model is diagrammed in Fig. 2; this is a simplified version of the model presented and analyzed in [15]. The first step in the synthesis of a small-target motion unit is to split the individual small-field motion responses into positive responses v_i^+ and negative responses v_i^- . Two monocular

“pool cell” units P^+ and P^- spatially sum these rectified responses

$$P^+ = \sum_{i=1}^N v_i^+ \quad (4)$$

$$P^- = -\sum_{i=1}^N v_i^- \quad (5)$$

where N is the number of small-field motion sensors, and the negative sign makes P^- positive. These operations are performed for both the left and right “eyes.”

In the second stage, monocular pool cell units from both eyes are combined to synthesize binocular pool cell units, which respond to background motion patterns in both eyes. To do this, it is necessary to know the type of background motion expected. For the present, we will assume that background motion is purely rotational (see Discussion). A more general case is explained in [15]. If counterclockwise rotation of the tracking system is assumed to result in positive motion responses from both eyes, then the binocular pool cells can be expressed as

$$P_{\text{right}}^{\text{ccw}} = P_{\text{right}}^+ + P_{\text{left}}^+ \quad (6)$$

$$P_{\text{right}}^{\text{cw}} = P_{\text{right}}^- + P_{\text{left}}^- \quad (7)$$

with the two units for the left eye being computed identically.

Directional small-target motion responses are then synthesized by normalizing each small-field motion unit by the pool cell representing the background expected for that sign of motion

$$y_i^+ = \frac{v_i^+}{\beta + P^{\text{ccw}}} \quad (8)$$

$$y_i^- = \frac{v_i^-}{\beta + P^{\text{cw}}} \quad (9)$$

where β is present to limit the small-target motion response. These operations are performed for each small-field motion unit in both “eyes”. If the background motion is of opposite sign from the target motion, the activation of the pool cell unit normalizing the target will be very small and the target will be strongly emphasized. However, if the background is of the same sign as the target, the pool cell unit normalizing the target will be strongly activated and the target signal will be weakened.

The small-target responses for the two directions are passed through a power-law nonlinearity to enhance large values, and then combined without regard to their sign

$$x_{i,\text{right}} = |y_{i,\text{right}}^+|^n + |y_{i,\text{right}}^-|^n \quad (10)$$

$$x_{i,\text{left}} = |y_{i,\text{left}}^+|^n + |y_{i,\text{left}}^-|^n \quad (11)$$

The nonlinearity is required to assure that background responses are more than N times weaker than target responses, since the sensor output will be used in spatial sum. Each x unit responds without regard to direction to small field, but not large-field, motion at its spatial position.

These nondirectional small-target responses are summed across the eyes

$$X_{\text{right}} = \sum_{i=1}^N x_{i,\text{right}} \quad (12)$$

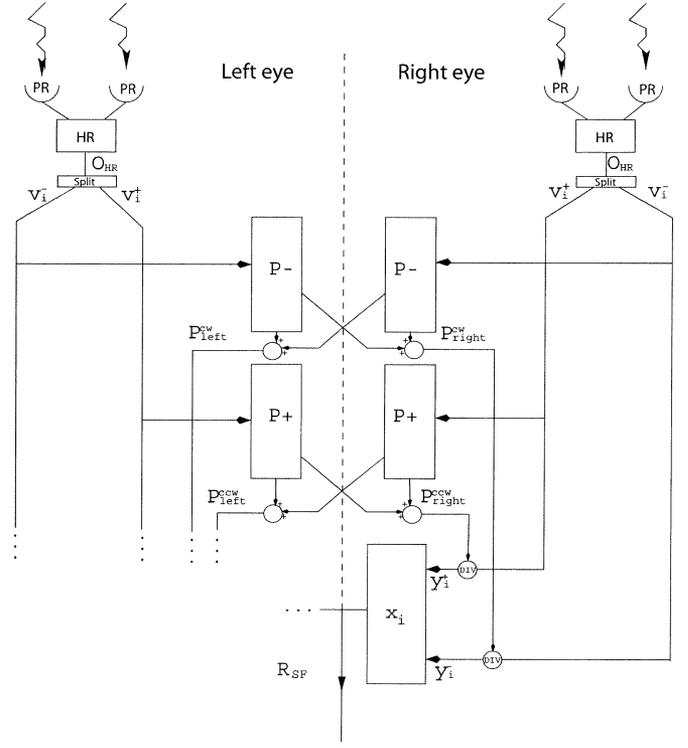


Fig. 2. Computational model of the small-target tracking system. Photoreceptor responses are input to HR motion detectors, the outputs of which are split into positive and negative components. These components are aggregated into directionally selective monocular pool cell units (P^+ , P^-), and then into clockwise and counterclockwise binocular pool cell units ($P_{\text{right}}^{\text{ccw}}$, $P_{\text{right}}^{\text{cw}}$, $P_{\text{left}}^{\text{ccw}}$, and $P_{\text{left}}^{\text{cw}}$). These pool cell units interact divisively (DIV) with the individual motion detector output channels prior to their combination by unit x_i . For simplicity, only the right side computation is diagrammed. The control signal is computed as the difference between the spatially summed x_i from the two sides. Modified from [13].

$$X_{\text{left}} = \sum_{i=1}^N x_{i,\text{left}} \quad (13)$$

where each X unit is sensitive to the motion of small objects anywhere in a wide visual field but not sensitive to wide-field motion.

To generate a control signal, these values are then subtracted from each other

$$R_{\text{SF}} = X_{\text{left}} - X_{\text{right}} \quad (14)$$

This final expression indicates by its sign which eye is experiencing a larger small-target response. Turning in this direction yields a system that tracks small moving targets.

III. HARDWARE IMPLEMENTATION

The chip-level hardware system with which we propose to implement the above computational model is shown in Fig. 3. Two identical monolithic VLSI chips, representing the left and right eyes, each contain a two-dimensional array of pixels. Each pixel consists of a 1-D HR motion detector (that is, each motion detector is sensitive to left-right motion only) and local circuitry necessary to produce the small-target response. Together, the pixel array in each chip computes an implementation of (4) through (13) for that eye. The two chips interact using currents

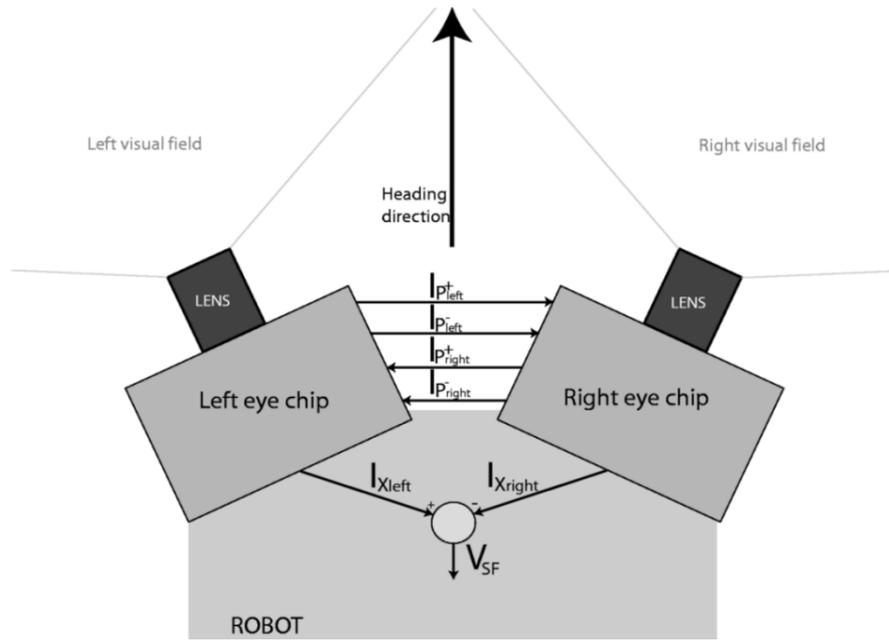


Fig. 3. Chip-level hardware implementation of small target tracking system. Two identical monolithic VLSI vision chips, one with a visual field to the left representing the left eye, and another to the right, interact to produce the control output. The two chips interact using unidirectional currents $I_{P_{left}^+}$, $I_{P_{left}^-}$, $I_{P_{right}^+}$, and $I_{P_{right}^-}$, representing the monocular pool cell activations for each eye. Each chip produces a unidirectional current I_X , the difference of which is used to compute the control output voltage V_{SF} .

to produce the final control output. To compute binocular pool cells on each side, each eye needs from the other the value of its monocular pool cells P^+ and P^- . These values are transferred, in both directions, in the form of unidirectional currents I_{P^+} and I_{P^-} . Each chip produces a unidirectional current I_X [corresponding to (12) and (13)] the difference of which is taken off-chip to produce the control output V_{SF} [corresponding to R_{SF} , (14)].

Details are given below of the hardware implementation of the HR motion detector, and of the circuitry necessary to produce the small-target response. Wherever possible, circuits are operated in the subthreshold regime of the MOSFET to reduce power consumption. The chip also incorporates serial current scanners [22], allowing readout of individual pixel current responses, or aggregation of current signals by Kirchoff's current law (KCL) sum. These scanner circuits incorporate two peripheral shift registers (one vertical and one horizontal) used to store the positions of rows and columns currently selected. The shift registers can select a particular row/column combination in order to view the output of a single pixel, or be set to select all rows and columns simultaneously. If the row and column of a particular pixel is selected, its output current is connected to the common current output line. When all rows and column are selected, the sum of all pixel current outputs is available on the common line.

A. HR Motion Detector

The HR motion detector has been described in detail and fully characterized in [23]. Used for phototransduction in the HR motion pixel (Fig. 4) is the adaptive photoreceptor by Delbrück and Mead [24], as elaborated by Liu [25] to allow control of the adaptation time constant [see Fig. 8(a)]. This circuit provides

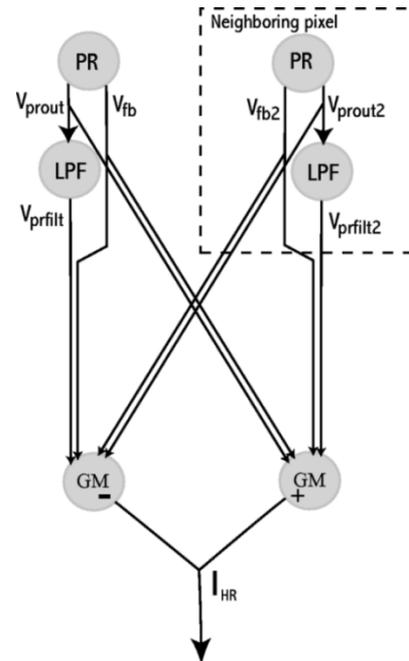


Fig. 4. Diagram of analog circuit implementation of HR motion detector. PR indicates a photoreceptor circuit which transduces local light intensity into a voltage [Fig. 8(a)]. LPF indicates a first-order g_m -C low-pass filter using a 5-transistor transconductance amplifier. GM represents a Gilbert multiplier, used to multiply two differential voltages. Outputs from the two Gilbert multipliers are taken with different signs so the final difference may be done using KCL.

two voltage outputs: V_{prout} , with a high sensitivity to transient changes in contrast and a low sensitivity to ambient illumination levels; and V_{fb} , which is sensitive to the long-term illumination level. The difference of V_{prout} and V_{fb} has very little response

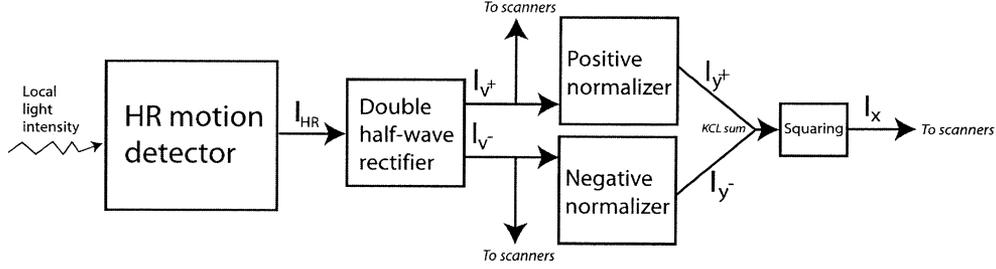


Fig. 5. Diagram of pixel circuitry in the implementation of the small-target response. HR motion detectors get input from a local photodiode, and also communicate with a neighboring photoreceptor to compute I_{HR} . This current is rectified into positive and negative currents I_{v+} and I_{v-} , which go into two normalizer circuits and are copied to the scanners. After normalization (see text), the currents for the two directions I_{y+} and I_{y-} are added, the sum is squared, and the final current is provided to the scanner.

to mean luminance, and thus has effectively been high-pass filtered with the time constant controlled by the adaptation time constant of the photoreceptor. $V_{PR\ out}$ is low-pass filtered using a g_m -C filter [26] to produce $V_{PR\ filt}$. Combining these three local signals with the same signals from a neighboring pixel, Gilbert multipliers [27] are used to perform multiplications of differential voltage signals to produce currents, which are taken with opposite signs from the two multipliers so that KCL can be used for the final subtraction. A bidirectional current I_{HR} [corresponding to O_{HR} , (3)] at the output of this stage represents the motion signal in each local area of the image.

B. Small-Target Response

The circuitry used to implement the small-target response is diagrammed in Fig. 5. See Fig. 6 for a simulation of the pixel circuitry illustrating the function of each stage. To synthesize the response to small moving targets, it is first necessary to split the bidirectional motion detector current output I_{HR} into two unidirectional currents I_{v+} and I_{v-} . This is accomplished using a double half-wave rectifier circuit [Fig. 8(b)]. Each of these currents is copied to the serial scanner network, and is thus available off-chip. With the scanners set to sum all pixels, these output currents represent the sum of the motion responses of each sign, and are used as the I_{P+} and I_{P-} signals (refer to Fig. 3) to communicate with the other eye.

Next, the currents of each sign are aggregated and used to normalize the output of each individual small-field sensor. Equations (4) through (9) are computed using two Gilbert normalizer circuits [Fig. 8(c)], one for positive and one for negative local motion currents. See Fig. 7 for a chip-level diagram of the circuit implementation. The positive normalizer circuit takes three types of inputs (inputs to the negative normalizer are indicated in parentheses):

- From each pixel in the array, the local motion sensor output after rectification I_{v+} (I_{v-}).
- A single external input current representing the sum of local motion signals of the same sign from the other “eye” I_{P+} (I_{P-}).
- A single local bias current I_{β} (I_{β} is independently generated for the negative normalizer).

Thus, the positive normalizer circuit in pixel i in the left eye computes

$$I_{y_i^+} = \frac{I_{v_i^+}}{I_{\beta} + I_{P_{\text{right}}^+} + \sum_{j=1}^N I_{v_j^+}} \quad (15)$$

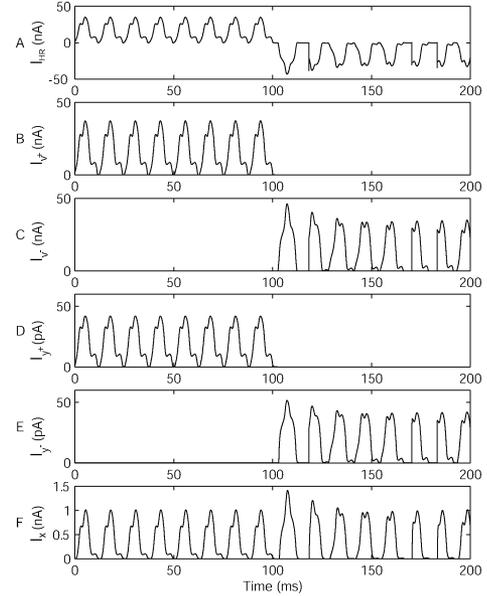


Fig. 6. Simulation of small-target response pixel circuitry. Panel A shows I_{HR} , the HR motion sensor response, as a visual stimulus moves to the right for 100 ms (eliciting a positive response), and then to the left for 100 ms (eliciting a negative response). Panels B and C show the two half-wave rectified currents I_{v+} and I_{v-} , each representing a single direction of motion only, which are input to separate normalizer circuits. Panels D and E show the normalizer circuit outputs I_{y+} and I_{y-} . The magnitude of these outputs is dependent on the other inputs to the chip-wide normalizer circuits. In the simulation, other normalizer inputs were constant. Finally, panel F shows the small-target output I_x , which results from the square of the sum of I_{y+} and I_{y-} , and which responds to motion without regard to direction.

implementing (8) in the current mode. The negative normalizer circuit simultaneously implements (9). The normalizer circuit shown in Fig. 8(c) is actually distributed throughout the chip, with the circuit of transistors $M1a$ and $M1b$ in each pixel, the node V_c common to each pixel, and the currents I_{bias} and I_{β} generated at the periphery. The current outputs of the two normalizer circuits corresponding to the I_{β} inputs (shown in Fig. 7 as $I_{out,\beta+}$ and $I_{out,\beta-}$) are also available off chip.

The two normalizer circuit current outputs are summed before being passed through a translinear squaring circuit [Fig. 8(d)] to implement the necessary nonlinearity ($n = 2$). This circuit exactly implements (10) and (11), despite the reversal of order of the sum and nonlinearity, due to the fact that only one of I_{y+} and I_{y-} can be nonzero at any given time (refer to Fig. 6). The reversal of ordering saves a squaring circuit. Finally, the scanners may be set to sum the entire chip such that the small-target current output implements (12) or (13).

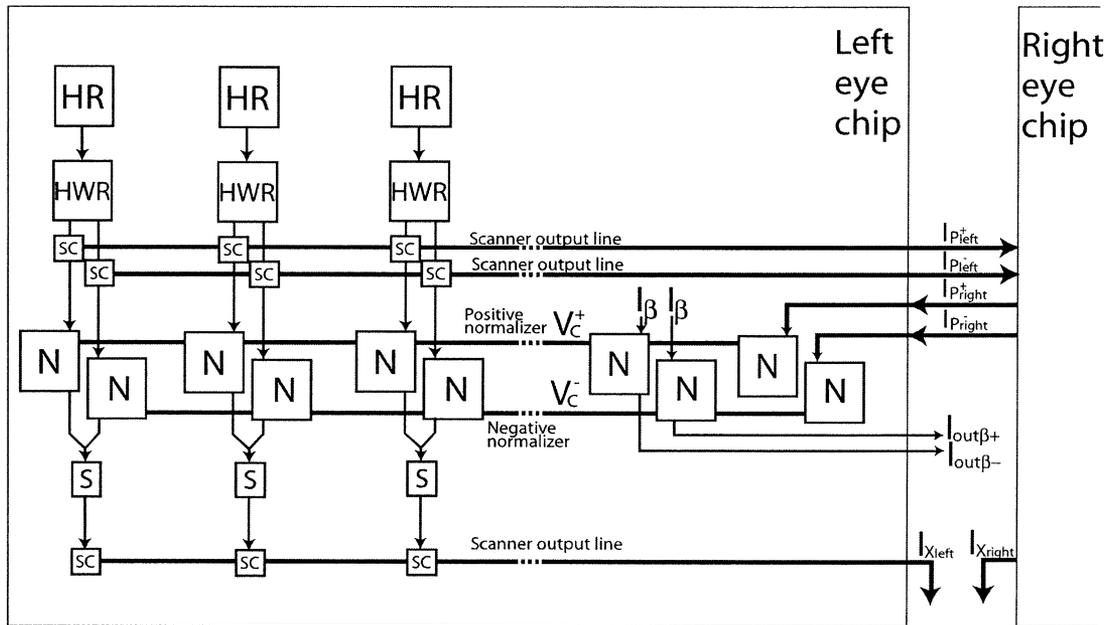


Fig. 7. Chip-level diagram of circuitry implementing the small-target response. See Fig. 5 for detail of the pixel circuitry. HR indicates a motion detector, HWR a double half-wave rectifier, N a normalizer subcircuit (consisting of two transistors like $M1a$ and $M1b$ in Fig. 8(c)), and S a squaring circuit. Scanner circuits (SC) connect their current input signals to the common horizontal bus when that column is selected. Multiple columns may be simultaneously selected resulting in a spatial KCL sum. Three scanner outputs are provided by each pixel and in spatial sum these correspond to signals I_{P+} , I_{P-} , and I_X . There are two chip-wide normalizer circuits, one for positive and one for negative motion directions. Subcircuits of each normalizer are joined by chip-wide common nodes V_c^+ and V_c^- . Each normalizer circuit has two peripheral inputs (at right), one for the current I_β , and one for the input from the other chip I_{P+} or I_{P-} . The outputs $I_{out,\beta+}$ and $I_{out,\beta-}$ corresponding to the two I_β inputs are made available off-chip. Although the diagram is 1-D for simplicity, in the implementation both scanner and normalizer networks are two-dimensional.

Each pixel thus has three outputs which can be viewed through the scanners individually: I_{v+} and I_{v-} , the HR motion detection outputs after rectification; and I_x [(10) or (11)], the small-target output. When the scanner is used to spatially sum these signals over multiple pixels, they correspond, respectively, to I_{P+} , I_{P-} , and I_X .

IV. CHARACTERIZATION RESULTS

The small-target tracking implementation was fabricated through MOSIS in a $1.5 \mu\text{m}$ CMOS process. Details of the implementation are provided in Table I. A photomicrograph of the fabricated chip is shown in Fig. 9.

The response of a single central pixel of the fabricated VLSI sensor was characterized by presenting computer-generated moving visual stimuli on an liquid crystal display (LCD) screen [used for its low flicker relative to cathode ray tube (CRT) displays]. An image of this stimulus was focused onto each chip using a camera lens. The on-chip bias current I_β was set to zero to maximize the response. The external current inputs I_{P+} and I_{P-} were also zero. Outputs of the chip were in the form of currents in the nanoampere range, and were converted to voltages using off-chip current sense amplifiers.

The response of the HR motion detector before rectification (I_{HR}) is shown in Fig. 10, in response to a moving sinusoidal grating. The sensor has a positive mean response to motion in the “preferred” direction (right along a row), and a negative mean response to motion in the “null” direction (left along a row). Note that in no case is the signal entirely positive or negative, but overlaps to the opposite sign to some extent. Due to

transistor mismatch, the motion detector response of each pixel has a small random dc offset. The contrast and spatio-temporal frequency response of this detector were discussed in detail in [23].

In order to characterize the motion response of the small-target output I_x [comparable to (10) or (11)], we used compound visual stimuli (illustrated in Fig. 11) to perform several experiments.

First, on top of a gray background covering the entire field of view of the chip, we presented a preferred-direction sinusoidal grating stimulus centered on the pixel the output of which was being measured [Fig. 11(a)]. The spatial frequency of this grating was set at 0.33 cycles per pixel to simultaneously stimulate as many motion sensors as possible, and the temporal frequency at the optimum for the tuning of the motion sensor ($f_t = -3 \text{ Hz}$). The size of this stimulus was varied from zero until it covered the entire field of view of the chip, while the mean response of the small-target output was recorded. The results of this experiment are shown in Fig. 12 (circles). When the stimulus is very small, the pixel responds very weakly. As the stimulus size grows, the pixel’s mean response also grows as long as only the local motion sensor in the pixel being measured is stimulated. As soon as the stimulus grows beyond the current pixel, the small-target output begins to drop off as the normalizer network divides down its response.

A model equation [based on (8) and (10)] was fit to this data

$$R_{st}(n) = g \cdot \left(\frac{r_{sf}(n)}{B + n} \right)^2 \quad (16)$$

$$r_{sf}(n) = \min(\sqrt{n}, 1) \quad (17)$$

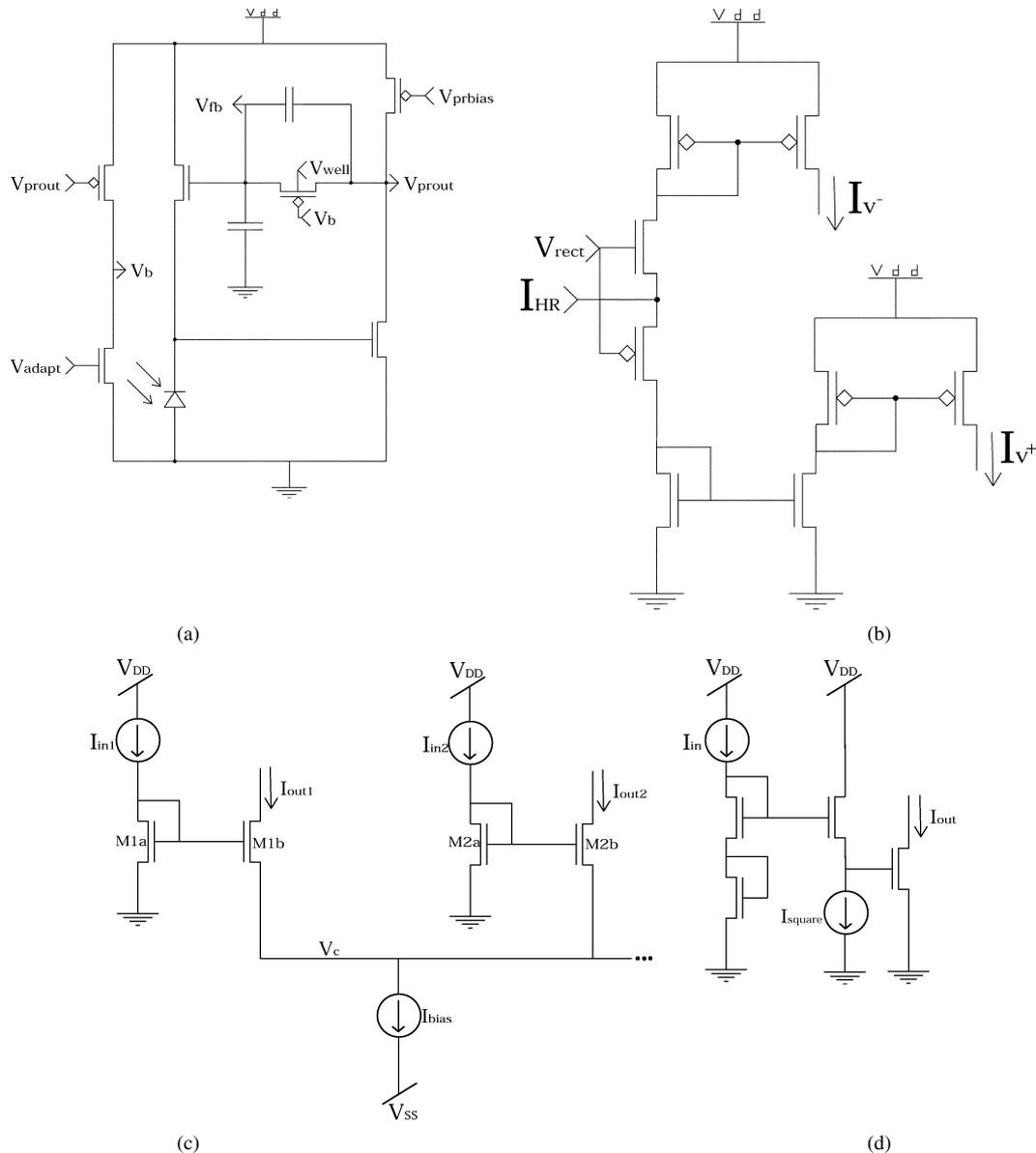


Fig. 8. Analog circuits used in small-target tracking sensor implementation. (a) Adaptive photoreceptor [25]. This circuit converts photocurrent into a voltage signal, amplifying transients while adapting to the mean light intensity. The bias V_{adapt} allows electronic variation of the adaptation rate. (b) Double half-wave rectifier circuit [32] used in small-target model. The outputs represent positive and negative half-wave rectified versions of the HR motion detector current. Note that I_{v+} is mirrored so as to provide the same sign of current as I_{v-} . (c) Translinear normalizer circuit [33]. The circuit of transistors $M1a$ and $M1b$ is replicated for each of N inputs, with each circuit connected by the common source node V_c ; only two inputs are shown. Operating in subthreshold, this circuit can be shown to compute $I_{\text{out},j} = I_{\text{bias}} \cdot I_{\text{in},j} / \sum_{i=1}^N I_{\text{in},i}$. Thus each input is normalized by the sum of all the other inputs. (d) Translinear squaring circuit. Operating in subthreshold, this circuit computes $I_{\text{out}} = I_{\text{in}}^2 / I_{\text{square}}$.

where n is the number of motion sensors being stimulated (proportional to the area of the visual motion stimulus), g is an empirical scale factor used to match the magnitude of the mean sensor response, and B is a parameter indicating the strength (in units of “motion sensors stimulated”) of the fixed input to the normalizer. Equation (16) matches the data with $B = 6.0$ motion sensors, and is plotted with the data in Fig. 12. $r_{\text{sf}}(n)$ models the small-field motion sensor response when less than one pixel is stimulated, and is unity after $n = 1$. Due to the manner in which the visual stimulus expanded in relation to the two photoreceptors associated with the motion sensor of the pixel being measured, $r_{\text{sf}}(n)$ depends on the linear dimension of the stimulus, and thus is proportional to \sqrt{n} .

Due to its spatial frequency, the sinusoidal grating stimulus presented can simultaneously activate a maximum of 15 motion sensors (three columns), beyond which the response of the sensor saturates. It can be concluded from the theoretical fit that the undesired dc current response of unstimulated pixels across the entire chip sums to the equivalent of six stimulated pixels, thus reducing the strength of the peak relative to the large-field response. Also shown in Fig. 12 (in gray) is the mean response of the local motion sensor, which grows from zero to its peak at $n = 1$, and then holds constant throughout the rest of the experiment.

To get a stronger small-target response, we presented the same variable size preferred direction motion stimulus superimposed upon a full-field background moving in the null direction

TABLE I
 DETAILS OF FABRICATED SMALL-TARGET TRACKING CHIP. FILL FACTOR IS PERCENTAGE OF PIXEL AREA DEDICATED TO PHOTOTRANSDUCTION

| | |
|------------------------|---|
| Process resolution | 1.5 μm |
| Die size | 2.1 mm \times 2.1 mm (MOSIS tiny chip) |
| Array size | 5 rows \times 13 columns |
| Pixel element count | 29 transistors, 3 capacitors, and a photodiode for HR motion sensor 26 transistors for rectification, normalization, squaring, and scanout circuitry |
| Pixel dimension | 243 μm \times 106 μm (106 μm is inter-pixel spacing of HR motion detectors) |
| Fill factor | 5.7 % |
| Chip power consumption | 30 μA at 5 V = 150 μW (biased for optimal performance) |

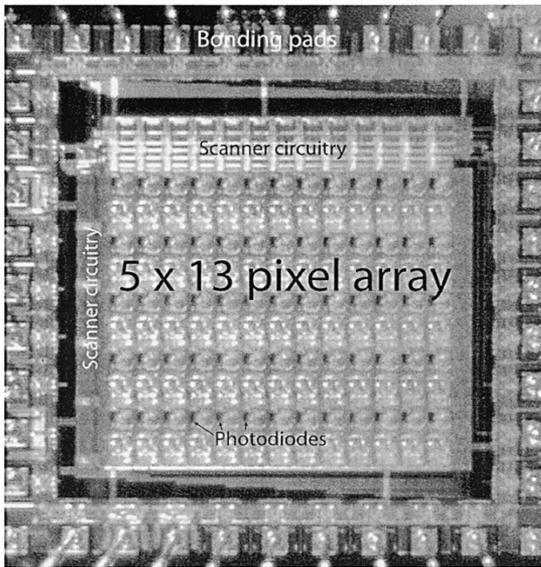


Fig. 9. Photomicrograph of experimental integrated circuit. This chip was fabricated through MOSIS in a 1.5 μm CMOS process on a 2.1 mm \times 2.1 mm die. See Table I for details.

[Fig. 11(b)]. This is meant to reduce the average response of pixels not being activated by the variable-size stimulus. This data is also shown in Fig. 12 (pluses). The null-direction background reduces the dc offset of the pixels enough that this data can be fit by (16) with $B = 4.6$. This is an effective reduction of only 1.4 stimulated motion sensors from the data with fixed background.

To show the effect of the background on the small-target response, we held constant the size of the stimulus to the pixel being measured and varied the temporal frequency of the background [Fig. 11(c)]. Fig. 13 shows the results of this experiment. When the background temporal frequency was the same as the small-field stimulus ($f_t = -3$ Hz), the small-target response was very weak. As the background temporal frequency varied, the response strength grew until the background was moving in the null direction, resulting in the strongest small-target response. Also shown in Fig. 13 (in gray) is the mean response of the small-field motion sensor, unchanged throughout the experiment.

Finally, to demonstrate the function of the normalizer output $I_{\text{out},\beta+}$ corresponding to the I_β input, the I_β bias current was increased until significant current flowed in the corresponding output. As the stimulus size was varied with a null direction background [Fig. 11(b)], the mean value of the $I_{\text{out},\beta+}$ normalizer

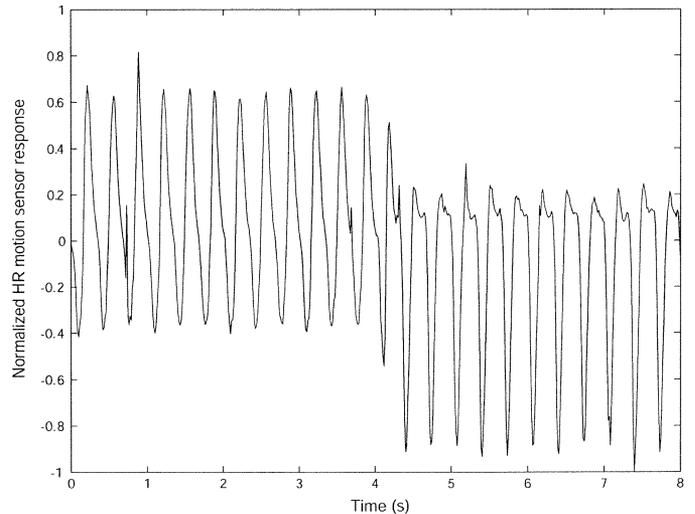


Fig. 10. Response of an individual HR motion sensor (I_{HR}) to a moving sinusoidal grating stimulus, normalized to an absolute maximum of one. The stimulus moves rightward for the first four seconds (the preferred direction, yielding a positive mean response), and leftward (the null direction, yielding a negative mean response) between 4 and 8 s.

output current was recorded. The data is shown in Fig. 14, along with the value of the local motion sensor and the small-target response, which is much reduced due to the strong I_β input. This small-target response is fit by (16) with $B = 7.5$ motion sensors. Unlike either the small-field motion response (which increases and then remains constant) or the small-target response (which increases to a peak and then falls), $I_{\text{out},\beta+}$ begins at its largest value for no preferred direction stimulation and falls in inverse proportion to the number of pixels being stimulated.

V. DISCUSSION

We have described the analog VLSI implementation of a computational model for small-target tracking based on a neuronal model from the brain of a fly, and have described how this sensor could be used for target tracking by small autonomous robots. A visual characterization of this fabricated VLSI sensor shows that, as predicted, it responds more strongly to small moving targets than to large ones, and that the strength of its response is enhanced when the target moves in opposition to the background. This output can be used to guide an autonomous robot in a tracking task with no need for significant postprocessing [15]. The power consumption, size, and weight of this monolithic system are extremely low,

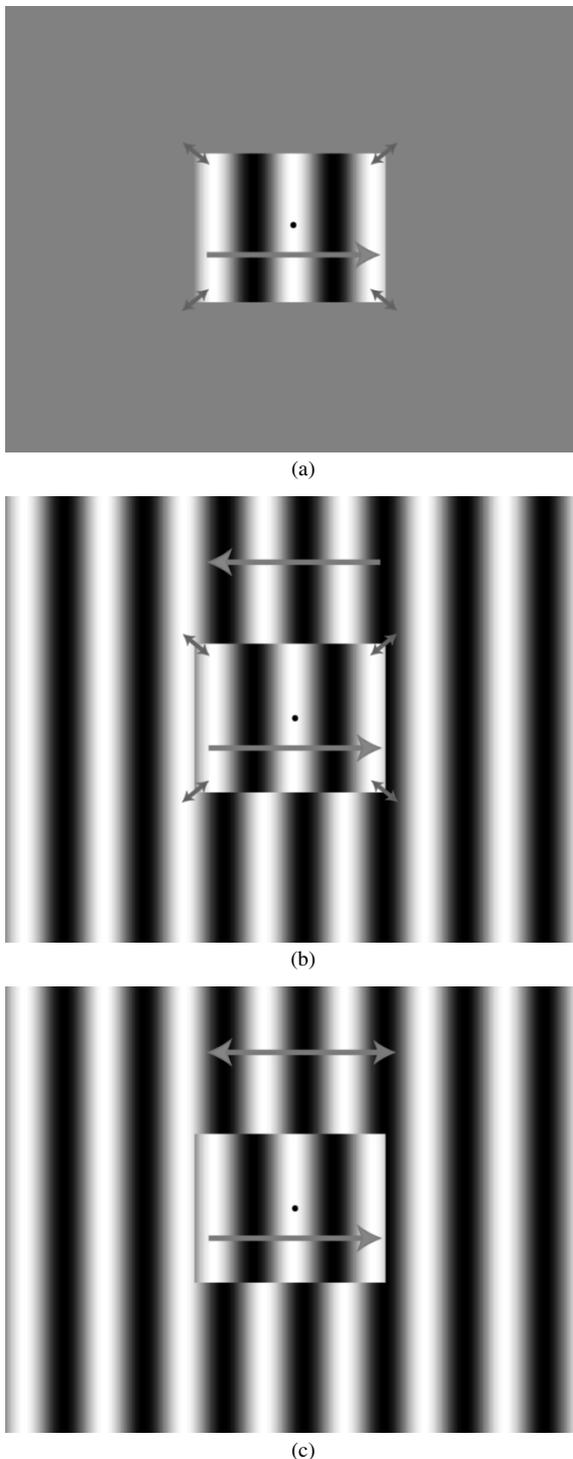


Fig. 11. Compound visual stimuli used to evaluate the small-target response. The black dot at the center indicates the the pixel being measured. (a) On a gray background covering the entire field of view of the chip, a sinusoidal grating stimulus centered on the pixel being measured moves in the preferred direction (large gray arrow) of the motion sensor. The size of this grating is varied from zero until it covers the background completely. (b) The full-field background moves in the null direction, and the same sinusoidal grating stimulus as in panel *a* moves in the preferred direction of the motion sensor. The size of this grating is again varied from zero until it covers the full field of view. (c) The size of the preferred direction sinusoidal stimulus is fixed, but the temporal frequency (and thus the speed and direction) of the wide-field background is varied.

while the equivalent computational throughput is high. This implementation has significant advantages over a conventional

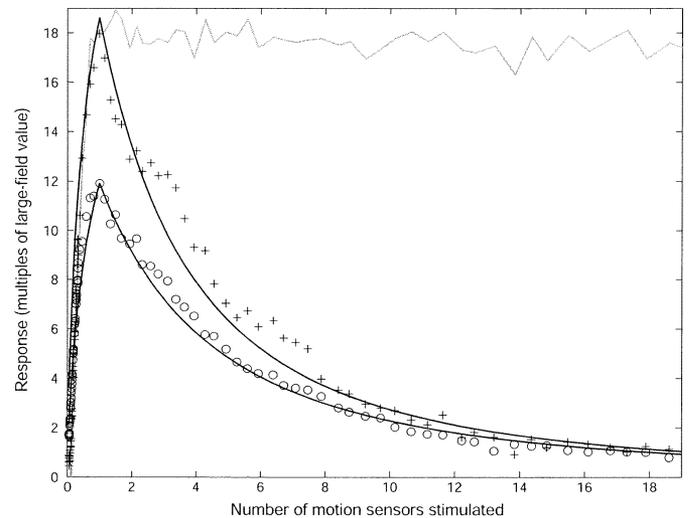


Fig. 12. Effect of visual stimulus size on single-pixel small-target response I_x . The ordinate axis has been scaled such that the small-target responses reach unity for large-field stimulation. Thus the units of the ordinate axis are “multiples of the large-field response.” Circles show the mean response of the sensor to a preferred-direction stimulus with nonmoving background. Pluses show the mean response to a preferred-direction stimulus with a null-direction moving background. The solid line through each set of data is the theoretical fit (see text). As the size of the moving visual stimulus is varied from zero to cover the entire chip, the response of the small-field output increases until it reaches a peak at one motion sensor stimulated. After peaking, it decreases as the inverse square of the number of motion sensors stimulated. Also shown in gray is the local motion sensor mean response, scaled to match the small-target response, which peaks at one motion sensor stimulated, and then holds relatively constant for the duration of the experiment.

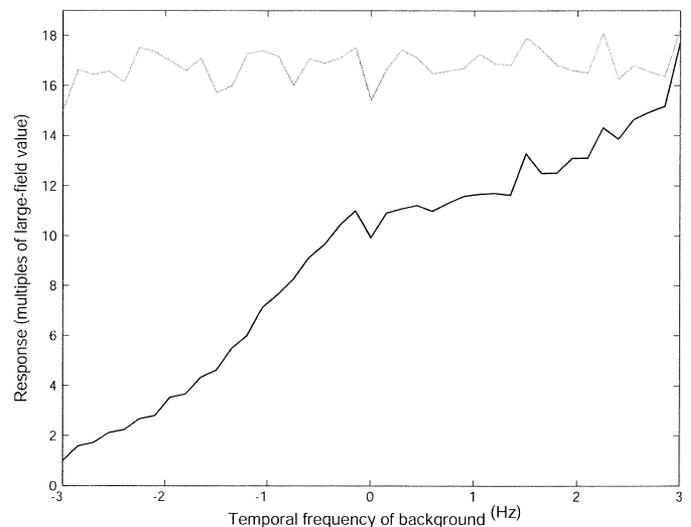


Fig. 13. Effect of the temporal frequency of a moving background on the small-target response I_x . The ordinate axis is on the same scale as Fig. 12, such that the small-target responses reach unity for large-field stimulation. The black line indicates the small-target sensor mean response. Shown in gray is the mean response of the local motion sensor (scaled to match the small-target response), which is relatively unchanged throughout the experiment. When the background temporal frequency matches that of the local motion sensor stimulus, the small-field mean output is diminished. As the background temporal frequency becomes more different, the small-field output increases.

approach when factors such as battery life and payload are critical to overall robotic system performance, as exemplified by the case of an airborne autonomous robot.

The experimental data in Fig. 12 is somewhat above the theoretical curve of (16) for low n because the theoretical model

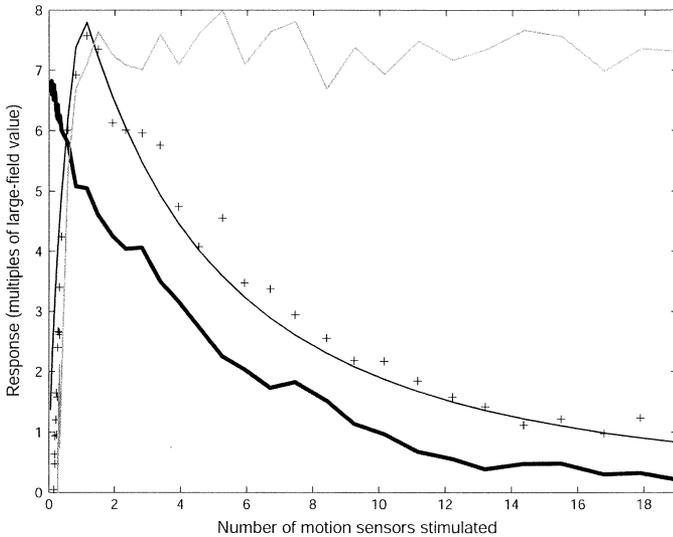


Fig. 14. Mean response of the positive normalizer circuit output $I_{out,\beta+}$ corresponding to I_β as stimulus size is varied. The ordinate axis is on the same scale as Fig. 12, such that the small-target responses reach unity for large-field stimulation. The bold black line shows the mean of $I_{out,\beta+}$ in units comparable with the small-target output. Plus signs indicate the small-target output, and the overlaid solid line the theoretical fit (see text). The local motion sensor mean output, scaled to match the small-target response, is also shown in gray. As the number of motion sensors stimulated in the preferred direction grows greater, the local motion sensor output grows and stabilizes, the small-target output peaks and then falls, but $I_{out,\beta+}$ starts at its largest value and falls as the inverse of the number of motion sensors stimulated.

does not take into account the fact that, as the rectangular motion stimulus grows, the pixels tend to become activated in row and column groups rather than singly. This leads to a discretization of the normalization response (most noticeable in the data plotted with pluses in Fig. 12) not predicted by the theory.

A limitation of the present VLSI design is the contribution of dc offset currents from each local motion detector to the normalizer network. These offset currents, inevitable to some degree due to mismatch in small MOSFETs, reduce the strength of the small-target response. Without any background stimulation, the sum of these small dc offsets from 65 motion sensors was the equivalent of six motion sensors being stimulated in the preferred direction. Even with a null-direction background, we were only able to reduce this offset figure to 4.6 motion sensors. This is due to the fact that the motion detector response in the null direction at times has positive sign (see Fig. 10), the mean magnitude of which is only slightly less than the average offset current. A solution to both of these problems would be the introduction of a current threshold in the rectifier circuit which must be overcome to contribute to the normalizer network. Such a threshold, set at a small current, would remove both the dc offset of unstimulated motion sensors and the small contribution from null-responding sensors, thus increasing the strength of the small-target response.

The fabricated die incorporates only 5×13 pixels on a MOSIS “tiny chip”. The low resolution of this sensor makes it capable of sensitivity only to relatively low image spatial frequencies without aliasing. This experimental prototype is intended only to demonstrate a proof of concept. Scaling this design to a higher resolution poses no technical challenges due to the fully parallel design architecture. The wiring complexity

of normalizer, scanner, and biasing networks does not change as the number of pixels increases. A higher pixel count would allow sensitivity to higher spatial frequencies and make operation with real-world imagery practical. Furthermore, despite an increase in the number of pixels, the magnitude of the globally-summed current outputs could be maintained at the same level by a reduction of appropriate biases, implying that chip power consumption will expand sublinearly with the number of pixels. However, the area of the small-target response pixel is rather large as compared to a motion sensor alone [21], and the fill factor rather small. Both of these problems could be addressed, at the cost of somewhat increased power consumption, by adopting a multi-chip strategy [28] which splits the computation presently done in a monolithic pixel into multiple layers of processing on separate chips.

One fundamental problem with the use of vision chips in the real world is the necessity to interpret the outputs in meaningful units. What output current indicates a “strong” small-target output? Some small output current may be measured even for no motion, so “zero” output is also nontrivial to infer. The normalizer output $I_{out,\beta}$ corresponding to I_β is present as a partial solution to this problem. $I_{out,\beta}$ is a global measure of motion activity across the entire chip. Whenever this output is larger than the small-target current, the small-target response may be considered insignificantly small since it is being overcome by I_β . This relative interpretation of outputs, rather than an absolute one, makes practical use of this sensor much easier. Due to the dc current offsets in the present implementation, use of the I_β input significantly reduced the strength of the small-target output. However, this would not be the case if the offsets were removed from the normalizer input as suggested above.

Although we have only described compensation for rotational background motion patterns, the hardware system described is also capable of compensating for expansive or contractive background motion patterns. Since such patterns would result in opposite signs of motion in the two “eyes,” this would simply require that the external current input to each normalizer be the sum of the motion sensors of opposite sign [rather than the same sign as shown in (15)] from the other eye [15].

The currents I_{P+} and I_{P-} that are used to communicate between the two “eye” chips are, at present bias settings, in the range of hundreds of nanoamperes. This is also true of the output current I_X . The small magnitude of I_{P+} and I_{P-} does not lead to a problem with slow charge and discharge of capacitances because these currents are input to a subthreshold MOSFET current mirror, which provides a very low input impedance that virtually “pins” the voltage on the wire carrying these currents. Similarly, I_X is designed to be connected to an off-chip current sense amplifier which provides a very low input impedance. Should injected noise on these off-chip currents become a problem in a practical implementation of this design, two strategies might be applied. First, the Gilbert multiplier bias current could be used to set the magnitude of output currents I_{P+} and I_{P-} to the desired level, as long as the current from each individual multiplier circuit remains subthreshold. Similarly, the bias current I_{square} in the squaring circuit can be used to set the magnitude of I_X . If this is insufficient, a current mirror with super-unity gain G might be added at the scanner

output to scale the currents up for interchip communication, and a mirror with the inverse current gain $1/G$ added at the current inputs I_{P+} and I_{P-} .

The fact that MOSFET's operating in subthreshold are employed in this design implies slow response times relative to modern digital circuits, and significant mismatch between nominally identical transistors. It is typical for signals processed with subthreshold MOSFET circuits to have a noise-limited dynamic range of less than 8 bits [29]. However, the response time of the present design to a motion input is less than 10 ms, fast enough to control the position of a physical system at an equivalent update rate of more than 100 Hz. The performance of the present iteration of this design is limited by transistor mismatch as described above, but this is traded for extremely low power consumption. Neuromorphic circuits including the present work are designed at the system level to work around subthreshold transistor mismatch by, for example, avoiding direct comparison of dc values of neighboring photoreceptors [30]. In processing inherently noisy visual inputs, and especially in an emulation of a biological model, high precision signals are not required [31]. Overall, the present design is well matched to a subthreshold MOSFET implementation.

REFERENCES

- [1] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
- [2] C. Koch and H. Li, *Vision Chips: Implementing Vision Algorithms with Analog VLSI Circuits*. Los Alamitos, CA: IEEE Computer Society Press, 1995.
- [3] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbruck, R. Douglas, and C. A. Mead, *Analog VLSI: Circuits and Principles*. Cambridge, MA: MIT Press, 2002.
- [4] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M. Zhang, "A foveated visual tracking chip," in *Proc. Int. Solid State Circuits Conf.*, Feb. 1997, pp. 38–39.
- [5] T. K. Horiuchi, T. G. Morris, C. Koch, and S. P. DeWeerth, "Analog VLSI circuits for attention based visual tracking," in *Advances in Neural Information Processing Systems 9*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 706–712.
- [6] V. Brajovic and T. Kanade, "Computational sensor for visual tracking with attention," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1199–1207, Aug. 1998.
- [7] G. Indiveri, "Neuromorphic analog VLSI sensor for visual tracking: Circuits and application examples," *IEEE Trans. Circuits Syst. II*, vol. 46, pp. 1337–1347, Nov. 1999.
- [8] T. Horiuchi and E. Niebur, "Conjunction search using a 1-D, analog VLSI-based attentional search/tracking chip," in *Proc. Conf. Advanced Research in VLSI*, 1999, pp. 276–290.
- [9] Z. Lu and B. E. Shi, "Visual tracking with subpixel resolution using an analog VLSI computational sensor," in *Proc. IEEE Int. Conf. Robotics and Automation*, vol. 2, San Francisco, CA, Apr. 2000, pp. 1676–1681.
- [10] G. Indiveri, P. Oswald, and J. Kramer, "An adaptive visual tracking sensor with a hysteretic winner-take-all network," in *Proc. Int. Symp. Circuit and Systems (ISCAS'02)*, vol. 2, May 2002, pp. 324–327.
- [11] V. Pant and C. M. Higgins, "A biomimetic VLSI architecture for small target tracking," in *Proc. Int. Symp. Circuit and Systems (ISCAS'04)*, vol. 3, Vancouver, BC, Canada, May 2004, pp. 5–8.
- [12] J. J. Gibson, *The Perception of the Visual World*. Boston, MA: Houghton Mifflin, 1950.
- [13] W. Reichardt, M. Egelhaaf, and A. K. Guo, "Processing of figure and background motion in the visual-system of the fly," *Biol. Cybern.*, vol. 61, pp. 327–345, 1989.
- [14] M. F. Land and T. S. Collett, "Chasing behavior of houseflies (*Fannia canicularis*): Description and analysis," *J. Comparative Physiol.*, vol. 89, pp. 331–357, 1974.
- [15] C. M. Higgins and V. Pant, "An elaborated model of fly small target tracking," *Biol. Cybern.*, to be published.
- [16] B. Hassenstein and W. Reichardt, "Systemtheoretische analyse der Zeit-, Reihenfolgen- und Vorzeichenbewertung bei der Bewegungszepherie des Rüsselkäfers *Chlorophanus*," *Zeit. Natur.*, vol. 11b, pp. 513–524, 1956.
- [17] J. P. H. Van Santen and G. Sperling, "Elaborated reichardt detectors," *J. Opt. Soc. Amer. A*, vol. 2, pp. 300–320, 1985.
- [18] R. A. Harris, D. C. O'Carroll, and S. B. Laughlin, "Adaptation and the temporal delay filter of fly motion detectors," *Vis. Res.*, vol. 39, pp. 2603–2613, 1999.
- [19] A. G. Andreou and K. Strohhorn, "Analog VLSI implementation of the Hassenstein-Reichardt-Poggio model for vision computation," in *Proc. IEEE Int. Conf. Systems, Man, Cybernetics*, Los Angeles, CA, Nov. 1990, pp. 708–710.
- [20] R. Sarpeshkar, W. Bair, and C. Koch, "An analog VLSI chip for local velocity estimation based on Reichardt's motion algorithm," in *Advances in Neural Information Processing Systems 5*, S. J. Hanson, J. Cowan, and L. Giles, Eds. San Mateo, CA: Morgan Kaufmann, 1993, pp. 781–788.
- [21] R. R. Harrison and C. Koch, "A robust analog VLSI motion sensor based on the visual system of the fly," *Autonom. Robot.*, vol. 7, pp. 211–224, 1999.
- [22] C. A. Mead and T. Delbruck, "Scanners for visualizing activity of analog VLSI circuitry," *Anal. Integr. Circuits Signal Processing*, vol. 1, pp. 93–106, 1991.
- [23] C. M. Higgins, V. Pant, and R. Deutschmann, "Analog VLSI implementation of spatio-temporal frequency based visual motion algorithms," *IEEE Trans. Circuits Syst. I*, to be published.
- [24] T. Delbruck and C. Mead, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits," Program in Computation and Neural Systems, California Institute of Technology, Pasadena, Tech. Rep. 30, 1993.
- [25] S. C. Liu, "Silicon retina with adaptive filtering properties," *Anal. Integr. Circuits Signal Processing*, vol. 18, no. 2/3, pp. 243–254, 1999.
- [26] R. Schaumann and M. E. Van Valkenburg, *Design of Analog Filters*. Oxford, U.K.: Oxford Univ. Press, 2001.
- [27] B. Gilbert, "A precise four-quadrant multiplier with subnanosecond response," *IEEE J. Solid-State Circuits*, vol. SC-3, pp. 365–373, Dec. 1968.
- [28] C. M. Higgins and C. Koch, "A modular multi-chip neuromorphic architecture for real-time visual motion processing," *Anal. Integr. Circuits Signal Processing*, vol. 24, no. 3, 2000.
- [29] R. Sarpeshkar, "Analog versus digital: Extrapolating from electronics to neurobiology," *Neural Computation*, vol. 10, pp. 1601–1638, 1998.
- [30] R. Douglas, M. Mahowald, and C. A. Mead, "Neuromorphic analogue VLSI," *Annu. Rev. Neurosci.*, vol. 18, pp. 255–281, 1995.
- [31] E. A. Vittoz, "Analog VLSI signal processing—Why, where, and how?," *J. VLSI Signal Processing*, vol. 6, pp. 27–44, 1994.
- [32] R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "A low-power wide-dynamic-range analog VLSI cochlea," *Anal. Integr. Circuits Signal Processing*, vol. 16, no. 3, pp. 245–274, 1998.
- [33] B. Gilbert, "A monolithic 16-channels analog array normalizer," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 956–963, Dec. 1984.



Charles M. Higgins (SM'04) received the Ph.D. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, in 1993.

He worked in the Radar Systems Group at the Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, MA, until 1996. He then returned to Caltech as a Postdoctoral Research Fellow in the Division of Biology studying engineering applications of visual neuroscience. In 1999, he joined the Department of Electrical and Computer Engineering, University of Arizona, Tucson, as Assistant Professor, where he holds a joint appointment in the Division of Neurobiology. His research focuses on investigation and engineering application of neurobiological computational and control architectures and data representations, particularly in the area of insect visual motion processing and associated motor control. Projects in this area range from simulation modeling to the construction of biologically inspired analog/digital custom very large-scale integrated vision chips and robotic systems.



Vivek Pant received the M.S. degree from the University of Arizona, Tucson. He is working toward the Ph.D. degree in electrical and computer engineering at the same university.

His Master's research focused on elaborating biological algorithms for target tracking and developing a small-target tracking custom very large-scale integration (VLSI) architecture. His research interests include analog/digital/mixed signal VLSI, computational modeling of neural systems, and neuromorphic VLSI design.