

Department of Electrical & Systems Engineering

Departmental Papers (ESE)

University of Pennsylvania

Year 2000

A Foveated Silicon Retina for Two-Dimensional Tracking

Ralph Etienne-Cummings*

Paul Mueller‡

Jan Van der Spiegel†

Mao-zhu Zhang**

*Johns Hopkins University

†University of Pennsylvania, jan@seas.upenn.edu

‡Corticon, Inc.

**Corticon, Inc.

Copyright 2000 IEEE. Reprinted from *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Volume 47, Issue 6, June 2000, pages 504-517.
Publisher URL: <http://ieeexplore.ieee.org/xpl/tocresult.jsp?isNumber=18384&puNumber=82>
><http://ieeexplore.ieee.org/xpl/tocresult.jsp?isNumber=18384&puNumber=82>

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

This paper is posted at ScholarlyCommons.

http://repository.upenn.edu/ese_papers/25

A Foveated Silicon Retina for Two-Dimensional Tracking

Ralph Etienne-Cummings, *Member, IEEE*, Jan Van der Spiegel, *Senior Member, IEEE*, Paul Mueller, and Mao-zhu Zhang

Abstract—A silicon retina chip with a central foveal region for smooth-pursuit tracking and a peripheral region for saccadic target acquisition is presented. The foveal region contains a 9×9 dense array of large dynamic range photoreceptors and edge detectors. Two-dimensional direction of foveal motion is computed outside the imaging array. The peripheral region contains a sparse array of 19×17 similar, but larger, photoreceptors with in-pixel edge and temporal ON-set detection. The coordinates of moving or flashing targets are computed with two one-dimensional centroid localization circuits located on the outskirts of the peripheral region. The chip is operational for ambient intensities ranging over six orders of magnitude, targets contrast as low as 10%, foveal speed ranging from 1.5 to 10K pixels/s, and peripheral ON-set frequencies from <0.1 to 800 kHz. The chip is implemented in $2\text{-}\mu\text{m}$ Nwell CMOS process and consumes 15 mW ($V_{dd} = 4\text{ V}$) in normal indoor light ($25\ \mu\text{W}/\text{cm}^2$). It has been used as a person tracker in a smart surveillance system and a road follower in an autonomous navigation system.

Index Terms—Centroid-localization chip, foveated silicon retina, motion chip, target-tracking chip, vision chip.

I. INTRODUCTION

IN ORDER to maintain camera focus on a subject who is moving around in the environment, target tracking is required. The freedom for the subject to roam is desirable in many situations such as teleconferencing, video telephony, and distant learning classrooms. Furthermore, camera pointing or target tracking has many applications in manufacturing, surveillance, navigation, and robotics [1]–[3]. Moreover, tracking a moving target stabilizes its location on the imaging plane, which makes various image analysis, understanding and visual serving tasks much more tractable [4]–[6].

For consumer electronics and mobile robotics applications, the tracking system must be inexpensive, low power, compact, and must work for a wide range of ambient lighting, target speed and distance. The dynamic range, cost, power, and size requirements eliminate traditional methods of using infra-red or CCD-and-CPU approaches in realizing these tracking systems. On the other hand, application specific integrated circuits

(ASIC's) are ideal for implementing compact computational systems. Furthermore, biology employs time-tested algorithms for oculomotor control that can inspire ASIC implementations. Consequently, in the last few years, there has been a push to mimic biology in developing VLSI computational sensors for visual information processing [7], [8]. Biological systems have similar power, space and computational constraints to compact VLSI electronics [9].

A direct imitation of biology, however, is not always desirable [10]. Horiuchi has implemented successful tracking systems that are parsimonious imitations of their biological master, but they are electronically and physically bulky, i.e., they require complicated circuitry, with many manually-set biases that occupy multiple chips [11]. The work presented in this paper attempts to model the *function* of many biological structures/subsystems involved in two-dimensional (2-D) visual gaze control on a single VLSI chip by using a hybrid of biomimetic analog circuits and traditional digital approaches. The result is a low-power compact vision chip that, when interfaced with a pair of motors, can perform target acquisition saccades and smooth pursuit tracking movements. This system is unique in its use of a spatially variant layout and arbitration between smooth pursuit (in fovea) and saccadic acquisition (in periphery) to improve the tracking process, similar to primate visual tracking. With the exception of the motor drivers, little additional circuitry is required for successful visual tracking.

II. PREVIOUS WORK AND SIGNIFICANCE OF OUR CHIP

The significance of this chip can be appreciated in four ways. First, this chip is a complete realization of visual smooth pursuit tracking and acquisition saccadic system at the focal plane. Second, the physical and control architecture takes advantage of a 2-D spatially variant layout of receptors to improve the tracking process while reducing the complexity, thus power consumption and area usage, of the focal-plane circuits. Third, smooth pursuit circuits in the fovea uses one of the most compact direction-of-motion detection circuits in literature. Lastly, the target localization in the periphery also uses a very compact centroid computation circuit based on latched row and column address spatiotemporal events.

A. System Overview

Although monocular approaches have been reported, visual tracking systems have been primarily implemented using binocular systems [12]–[16]. These systems, which use standard CCD cameras, frame-grabbers, and workstations or parallel

Manuscript received May 1999; revised March 2000. The work of R. Etienne-Cummings was supported in part by National Science Foundation CAREER Award 9896362. This paper was recommended by Associate Editor P. Thiran.

R. Etienne-Cummings is with the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA.

J. Van der Spiegel is with the Moore School of Electrical Engineering, Center for Sensor Technologies, University of Pennsylvania, Philadelphia, PA 19104 USA.

P. Mueller and M.-Z. Zhang are with Corticon, Inc., Philadelphia, PA 19104 USA.

Publisher Item Identifier S 1057-7130(00)04978-8.

computers, are attractive because of the flexibility offered by their software implementation [15]. The main drawback of the CCD-CPU approach is their large size and high power consumption, which is worse in the binocular case. However, the advent of high-speed board computers will reduce their form factors significantly, but the power consumption will not improve. There are a few commercially available systems, but as can be expected, they are quite expensive (many thousands of dollars). We do not expect our approach to be as accurate and flexible as the CCD-CPU systems, but it will be considerably smaller and, other than the motor drivers, completely self-contained. Furthermore, it will use considerably less power, which will allow its integration into systems with small power budgets, such as mobile and flying robots.

ASIC's realization of visual target tracking has mainly followed biological models [11], [17]–[19]. The most complete implementation of the primate oculomotor control system was reported by Horiuchi and collaborators [11], [19]. Their system includes saccadic, smooth pursuit, learning for post-saccadic drift correction and attention circuits. However, by virtue of his faithfulness to biology, the system is restricted to a one-dimensional array of receptors, is distributed among many chips and requires a number of manually set voltage and current references. While our approach does not have the sophistication of Horiuchi's, we have implemented a 2-D array modeling the basic functions of the primate oculomotor system.

B. Physical, Computation, and Control Architecture

It has been argued that a foveated image sensor is a required part of any vision system if limited computational hardware is available, if data compression is required and if power economy is important [20], [21]. Considerable data compression is realized with a log-polar imaging array, as found in the primate retina. In the same way, the exponential decay in receptor density with eccentricity reduces the total number of receptors while covering the same visual field, thus requiring less energy to power the array. Furthermore, the receptor layout is ideal for target tracking because only radial motion detection is required [22]. Not many log-polar imagers have been implemented, however [22]–[24]. In addition, little, if any, focal-plane processing is performed on these log-polar imagers. The main reason for the limited number of examples is that the layout of the receptors and computation circuits is difficult and not always area efficient. Our chip also exhibits a spatially variant organization of photoreceptors. However, two distinct regions, the fovea and periphery, with constant pixel densities arranged on a rectangular grid are used instead. The rectangular grid simplifies design process and reduces the complexity of the motion and centroid circuits. The density ratio between the two regions is four; power consumption is reduced approximately four-fold compared to a large array of foveal cells. Furthermore, the rectangular layout allows efficient area usage with Manhattan rules. In our system, motion detection is computed in a divergent manner to mimic the radial property of the log-polar arrays. Our chip is unique because it combines a spatially variant layout with all focal-plane computation

for saccadic generation (in the periphery) and smooth pursuit tracking (in the fovea). The direction-of-motion in the fovea and the location of spatiotemporal edges in the periphery are immediately available for motor control uses. A simple arbitration state machine is used to decide whether to pursue or saccade.

C. Direction-of-Motion Detection in the Fovea

Focal-plane detection of the direction-of-motion has been realized by either using computationally inspired gradient-based model or biologically based local correlation [25]–[32]. It is generally accepted, however, that the local correlation method is more robust and less sensitive to noise [27]. The latter statement is incontrovertible when speed *and* direction is required, but less absolute in the direction only case, as shown in [29]. Because motion-sensor designers have been driven by a desire to integrate high-density arrays of detectors at the focal plane while maintaining a wide response dynamic range, the compactness of the design (in terms of minimum transistors/pixel and capacitance/pixel) has been a major design goal [32], [34]. In our chip, the measurement of foveal speed is deemed dispensable in order to realize a tight direction-of-motion detection circuit. Because only the direction of motion is computed, the effects of the aperture problem become less relevant [35] and only motion normal to edges are detected. With speed eliminated from the motion-detection circuits, the foveal pixel is extremely small, using only 20 transistors and 0.2 pF of capacitance (2 of the 20 transistors) for two directions per pixel (+4 transistors if four directions required). With a smaller edge-detection circuit, only 13 transistors would be needed [33]. The 13 transistors implementation is potentially the smallest direction-of-motion circuit reported [7], [34].

D. Motion Centroid Localization

Implementation of centroid localization has been either completely analog or completely digital. The analog implementations, realized in the early 1990's, used focal plane current mode circuits to find a global continuous time centroid of the pixels' intensities [18], [36]. On the other hand, the digital solutions use standard CCD cameras, A/D converters, and DSP/CPU to compute the intensity centroid [37]. These software approaches offer multiple centroid localization with complex mathematical processing. However, they suffer from the usual high power consumption and nonscalability of traditional digital visual processing systems. Our approach is novel in many aspects. We benefit from the low power, compactness, and parallel organization of focal-plane analog circuits and the speed, robustness, and standard architecture of asynchronous digital circuits. Furthermore, it uses event triggered analog address readout, which simplifies the centroid localization problem. Our chip responds to translating or flashing targets only by using the ON-set detectors in each pixel. This reduces the impact of textured backgrounds on the centroid computation. Lastly, our chip models both the primate retina and two dimensional saccade motor error maps of superior colliculus on a single chip [38]. Previous work only realized a one-dimensional model using multiple chips [19].

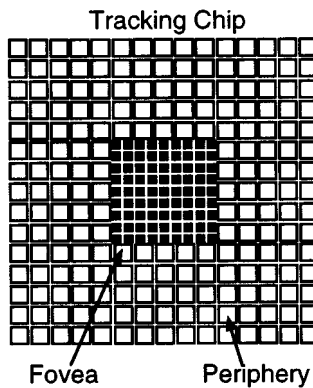


Fig. 1. Organization of the foveal and peripheral regions.

III. HARDWARE IMPLEMENTATION

A. The Visual Tracking Strategy

The tracking strategy employed in our chip follows the saccade/smooth pursuit interplay found in the primate oculomotor system [39]. As seen in Fig. 1, the imaging plane is divided into foveal and peripheral regions. The peripheral region identifies a translating or flashing (i.e., moving) target, determines the location of its centroid, and initiates a saccade to foveate it. Saccadic generation is one of the functions of the primate superior colliculus, and we model this function in a very superficial way [38]. Here, a target is defined to be one or more pixels forming the edge of the target. The centroid is computed on row and column events that carry no information about the contrast of the edges, other than the fact that they possess the spatiotemporal attributes to trigger ON-set detectors within the peripheral pixels. Hence, a positional average of moving edges is computed.

The foveal region prevents translating targets from leaving its area by matching the targets' motion with small corrective movements that force the targets toward the center of the fovea. When a stationary flashing target is foveated, the fovea does not detect translation, continues to look at the target for a prescribed time period, and then returns attention to the periphery. In primates, the smooth pursuit mechanism involves the visual cortex and is a complex system [39]. In this chip, we mimic the function of the primate smooth-pursuit mechanism, but implemented with ideas borrowed from other organisms such as the rabbit's retinal motion sensitivity and the fly's motion-detection system [26], [45]. To simplify the motion-detection circuits, velocity estimation is not performed, even though velocity matching is the basis of the primate smooth-pursuit system. Consequently, the smooth-pursuit mechanism adopted here is predicated on the detection of small displacements of the target from the center of the fovea, followed by immediate pixel-spacing sized corrections. In the limit that the pixel-spacing and correction-reaction time approach zero, the correction movements approach smooth pursuit, but realized with a discrete stop-look-measure-correct strategy. The direction-of-motion detection circuits are organized to measure only motion away from the center of the fovea. Hence, a maximum of two motion components are computed per pixel, which further reduces the size of the motion-detection circuits. To reduce the pixel spacing, all motion detection is performed outside the imaging array.

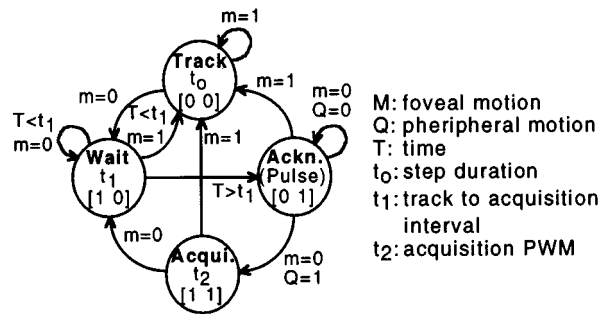


Fig. 2. Arbitration state machine for switching between foveal target tracking (smooth pursuit) and peripheral target acquisition (saccade generation).

As shown in Fig. 2, the state machine mediates the switching between saccade (target acquisition) and smooth pursuit (target tracking). The state machine favors tracking by allowing the chip to track any motion detected in the fovea. The operation scenario is described as follows.

A translating or flashing target is detected in the periphery while no motion is measured in the fovea ($Q = 1$, $M = 0$). The 2-D location of the target is computed. The motor system places the target in the fovea with a ballistic, open-loop movement. If the target is translating ($M = 1$), it is tracked. If, at any time, the target stops translating on the fovea, a timer (t_1) is started. If the target starts moving again, the timer is reset and tracking resumes. If the timer expires, then the target is assumed to have either left the fovea or is no longer of interest. The acquisition circuit is reset, by acknowledging the last target location, and a new target search begins in the periphery. At any time, if motion is detected in the fovea ($M = 1$), then the acquisition search is abandoned and tracking resumes. On the other hand, if a target is detected in the periphery before any motion is detected in the fovea ($Q = 1$, $M = 0$), then the target is acquired into the fovea and the process starts anew.

This approach allows the system to mostly execute fine smooth-pursuit tracking and to perform coarse saccadic motions only to acquire new targets or to reacquire targets that escape from the fovea. The circuits for the state machine have been implemented both on-chip and using off-chip glue logic. The most flexible approach uses a micro-controller that is part of the motor driver system. The micro-controller program sets the timer (t_1) value, converts the analog target location into a digital address, generates a pulse-width-modulated (t_2) acquisition signal (for stepper motors), controls the foveal correction step size, and communicates with the chip. All this processing is realized with a single PIC16C74A micro-controller, running at 20 MHz.

B. The Fovea

The center of the chip contains the foveal region where the photosensitive elements and edge-detection circuits are packed densely. To improve the spatial resolution of the fovea, the motion-detection circuits are separated from the fovea and placed at the bottom of the light sensitive area (see Fig. 11). The fovea computes divergent 2-D direction-of-motion, as shown in Fig. 3.

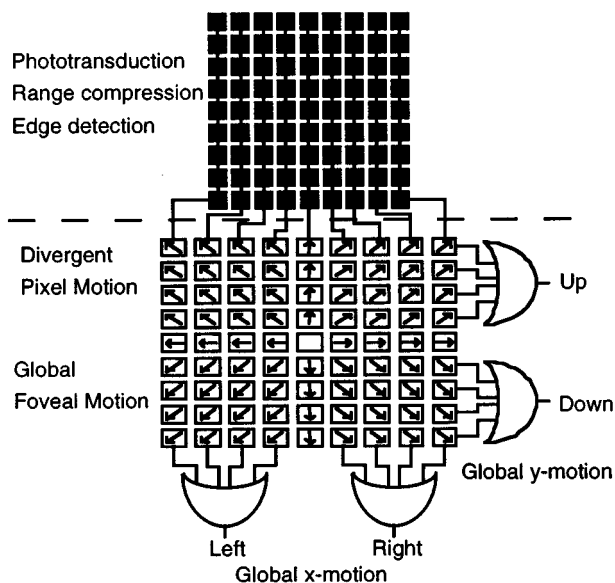


Fig. 3. Motion-detection architecture of the fovea. Only divergent motion is computed.

A global vector for the entire fovea is obtained. This vector provides the signals for fine motor control to maintain fixation on a moving target.

The details of the phototransduction and edge-detection approaches used in this chip have been described elsewhere [30], [40]. Slight modifications are made, however, to improve the compactness of the circuits. Phototransduction is realized using parasitic p-n-p transistors in an n-well CMOS process. Because the phototransistor is relatively large at $4144 \mu\text{m}^2$ (to improve the fill factor), it produces a substantial current, up to many microamperes, such that its response speed does not limit the tracking procedure severely. Compressive gain control is realized using a current mirror configuration (see Fig. 4) that provides a squared logarithmic relationship in low light, and a scaled (gain < 1) linear relationship in bright light. That is, in the former case, the diode transistor is in weak inversion while the output transistor is in strong inversion, while in the latter case they are both in strong inversion with the aspect ratio of the output transistor small than that of the diode. This local gain-control mechanism allows a scene with a wide range of light intensity to be imaged, without saturating any pixel. Pixel saturation is a problem in most integration-based imagers, such as CCD and CMOS arrays, that use global gain control mechanisms (pixel integration times) [41]. (Some CMOS imagers, however, try to solve the dynamic range problem [42].) Low-intensity details are lost in the presence of a bright source. Logarithmic range compression is found in the primate retina [43].

In our system, edge detection is used to improve the robustness of the motion-detection circuits [30]. Again, contrast-detection circuits are also found in the primate's retina. Using an approximation of a Laplacian edge detector with high gain, implemented as a difference between the image and a smoothed version, the gray-scale intensity image is converted into a binary image of edges. Edge detection normalizes contrast, accentuates spatial discontinuities, and reduces sensitivity to full-field

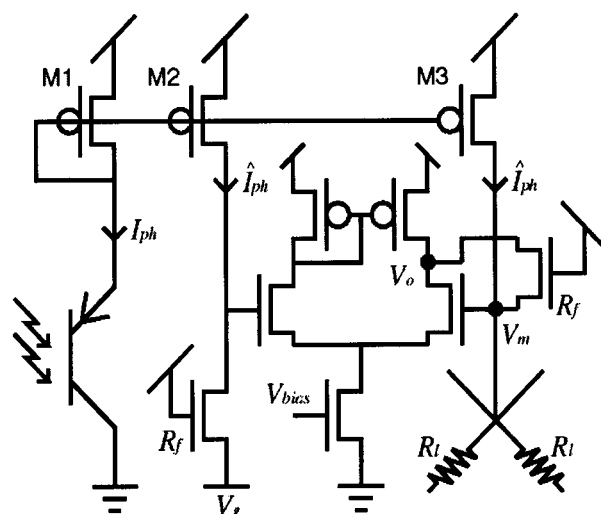


Fig. 4. Phototransduction, range compression, and edge-detection schematic.

fluctuations in ambient light (as a result of fluorescent lights, for example). The reduction in sensitivity to full-field fluctuation is obtained because the edge-detection circuits only respond contrast differences. Provided that contrast does not decrease significantly during fluctuation of the ambient intensity, the output of the edge-detection signal level remains large enough for the motion and ON-set circuits to be unaffected.

Because the edge-detection circuit has a high gain, a small difference in contrast produces almost full-scale swing. This occurs at the maximum temporal response speed of the circuit. Consequently, the temporal differentiators used in motion detection in focal-plane motion-detection circuits can be found in reference [30]. Fig. 4 shows the schematic of the photoreceptors and edge-detection circuits. Equation (1a) gives the gain control relationships for the pixel under typical operating conditions, i.e., in normal room light. Typically, the aperture of the optical system is used to shift operation in the logarithmic region. Equation (1b) gives the transfer function of the edge-detection circuit, where ∇_D^2 is the discrete Laplacian operator on a square grid. An ideal discrete Laplacian is approached as A gets larger ($A = 100$ in our case) and $Vg = 0$

$$\hat{I}_{ph}(x, y, t) \approx \frac{\mu_p C_{ox}}{2} \left(\frac{nKT}{q} \right)^2 \left(\frac{W}{L} \right)_{2,3} \cdot \left[\ln \left(\frac{I_{ph}(x, y, t)}{(W/L)_1 I_{DO}} \right) \right]^2$$

or

$$\hat{I}_{ph}(x, y, t) = \frac{(W/L)_{2,3}}{(W/L)_1} I_{ph}(x, y, t) \quad (1a)$$

$$V_o(x, y, t) \approx \frac{R_f}{R_l} \left[R_f \nabla_D^2 \hat{I}_{ph}(x, y, t) - \frac{1}{A} \nabla_D^2 V_o(x, y, t) \right] + Vg. \quad (1b)$$

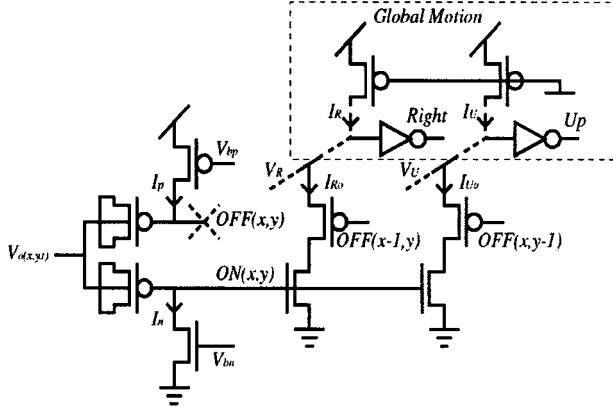


Fig. 5. The direction-of-motion detection schematic. Two directions are determined per cell.

The motion-detection scheme adopted in this chip is based on a velocity-detection circuit developed for focal-plane implementation [30]. Speed, however, is not measured. The direction-of-motion algorithm is modeled after the fly's visual motion system [26], and correlates an edge appearance event at a pixel with the disappearance events from its neighbors. The direction is gleaned from which of the neighbors the edge disappeared. The appearance of an edge at a pixel is given by the ON-set (positive temporal transient) of the edge-detector output, while the disappearance is given by the OFF-set (negative temporal transient). The correlation is implemented with a simple current NAND gate, which is implemented with pass transistors. The temporal decay times of the ON- and OFF-set detectors must account for the transit time between pixels. Because a single "vector" is obtained for the entire fovea, all the Left, Right, Up, and Down signals from all the pixels are line-OR gated into global signals, respectively, as shown in Fig. 3. Simultaneous signaling of opposite directions, which can happen as a target approaches the chip (known as looming), can be easily taken into account by motor control firmware, by computing the sign of the difference between the number of Left (Up) and Right (Down) motion pulses per unit time. Here, the firmware solution was used because it offers more flexibility. Fig. 5 shows the schematic of the motion-detection circuit. Equations (2a)–(2b)

give an approximation of the output current of the current ON gate for an edge moving with positive velocity. Both transistors, labeled ON and OFF, are assumed to be operating in the linear region. In the ideal case, the edge must disappear from a pixel before it can appear at a neighbor. Then the motion current is given by (2a)–(2b). In this case, the correct motion is detected if the ON-set detector has a short time-constant (i.e., I_n/C is large) and the ON-set detector has a large time constant (i.e., I_p/C is small). No motion is signaled for movement in the opposite direction (i.e., $-v_x$) because the second term in the numerator of (2a) will vanish at $t = 0$. A negative current in through the NAND gate is not possible. In the equation, v_x is the speed of motion, d is the displacement between pixels, I_R is the bias current for Right motion, and C , I_p , and I_n are the capacitance and discharge currents of the transient detectors. Because it is not possible to probe the transient detectors without affecting their behavior (a large series capacitance to ground will be added), we used SPICE to simulate their response for a moving thin bright bar, which is shown in Fig. 6. In the hardware system, as an edge moves from a pixel to a neighbor, the resistive grid forces the edge-detector output signal to rise at the neighbor before it can fall at the pixel. This is visible in the simulation, Fig. 6, since a neighbor reaches near maximum before the previous pixel begins to fall. Consequently, the role of the ON- and OFF-set time constants are reversed. The correct motion current is still obtained, and is now given by (3). In the equation, the effective "inter-pixel transit time" d/v_x is a fraction of the actual transit time, but is still inversely proportional to the speed of motion. An additional benefit the effective shortening of the transit time is that slower speed movements can be detected with smaller time-constants because the overlaps between OFF- and ON-sets are closer in time. Experimental results presented in the next section illustrate the relationships between neighboring pixels' edge and motion-detection outputs (see (2a), (2b), and (3), shown at the bottom of the page).

The dynamic range of the motion-detection circuit is governed by the response times of the phototransduction and temporal transient-detection circuits. For slow-moving targets, the low-frequency cutoff of the transient detector limits the detection of motion. The temporal transient detector is implemented with a current source and capacitor, as shown in Fig. 5, which

$$I_{Ro}(x_i, y_j, t) \approx \frac{V_R \beta_p \beta_n [V_R - \text{OFF}(x_{i-1}, y_j) - I_p t / C - V_{TP}] u(t) [\text{ON}(x_i, y_j) - I_n (t - d/v_x) / C - V_{TN}] u(t - d/v_x)}{\beta_p [V_R - \text{OFF}(x_{i-1}, y_j) - I_p t / C - V_{TP}] u(t) + \beta_n [\text{ON}(x_i, y_j) - I_n (t - d/v_x) / C - V_{TN}] u(t - d/v_x)} \quad (2a)$$

$$\text{Right}(t) = \text{Sign} \left[I_R - \sum_{i,j} I_{Ro}(x_i, y_i, t) \right] \quad (2b)$$

$$I_{Ro}(x_i, y_j, t) \approx \frac{V_R \beta_p \beta_n [V_R - \text{OFF}(x_{i-1}, y_j) - I_p (t - d'/v_x) / C - V_{TP}] u(t - d'/v_x) [\text{ON}(x_i, y_j) - I_n t / C - V_{TN}] u(t)}{\beta_p [V_R - \text{OFF}(x_{i-1}, y_j) - I_p (t - d'/v_x) / C - V_{TP}] u(t - d'/v_x) + \beta_n [\text{ON}(x_i, y_j) - I_n t / C - V_{TN}] u(t)} \quad (3)$$

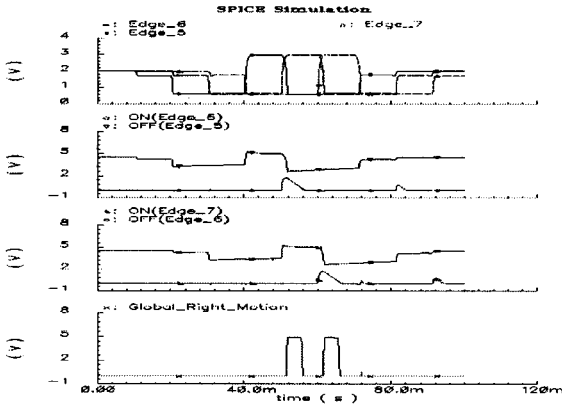


Fig. 6. SPICE simulation of the direction-of-motion detection sequence for a bright spot moving to the right. Top panel shows the edge-detection outputs of neighboring cells. Next two panels show the ON- and OFF-set detection for neighboring cells. Last panel shows the correlation of the ON- and OFF-sets to detect direction-of-motion.

gives a low-frequency cutoff of $\sim(I_n/C)$. The smallest value of the discharge current is limited by the leakage current of the source diffusion of the current source, and allows a maximum inter-pixel transit time of about 2.5 s, which corresponds to a best-case speed 0.4 pixels/s in the dark. The high end is limited by the response time of the phototransduction, which is ambient intensity dependent; the brighter the ambient, the higher the maximum speed detectable. Mainly, the time for the photoreceptor to turn off is the limiting factor. Equation (4) gives the approximate turning off response time. In the low ambient case, the load pMOS is assumed to be in weak inversion with $V_{ds} = V_{gs} \gg KT/q, \kappa \sim 2$ (a process parameter). In bright ambient, the pMOS is assumed to be in strong inversion. The worst-case response time is obtained for transitions from low light to dark. The measured dark current for the phototransistors is 18 pA and the calculated capacitance, based on process parameters, at the emitter is 0.51 pF, which predicts an off-response time of 1.5 ms, which corresponds to a maximum speed of 0.69K pixels/s. This is very close to the measured value of 0.64K pixels/s in $2.5 \mu\text{W}/\text{cm}^2$. As the intensity increases, the maximum speed also increases, but slower than linearly, because the load transistor emerges from weak inversion

$$\tau_{\text{OFF}} \approx \frac{qI_p}{2\kappa KTC_L} (\text{low light}) \rightarrow \frac{\sqrt{2I_p k_p (W/L)}}{C_L} (\text{bright light}) \quad (4)$$

C. The Periphery

Surrounding the fovea is the peripheral region, where the centroid of a moving target is computed. In the periphery, high-resolution imaging is not required. Hence, the phototransduction and spatiotemporal edge-detection circuits, required to realize the centroid computation, are placed within each cell. Centroid localization is performed outside the periphery array. The size of the phototransistors is also increased so that fewer cells can be used to cover a large area, thus reducing the power consumption of the chip. The larger phototransistors do produce larger currents; however, the increase in photocurrent is not compa-

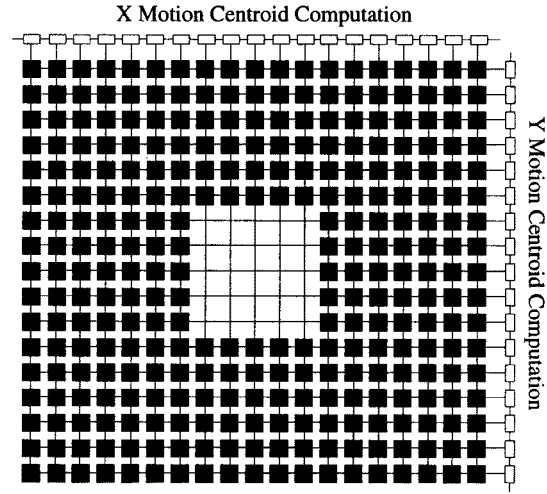


Fig. 7. Layout of the peripheral region. Centroid location is determined using row and column spatiotemporal events on the boundary of the region.

able to the bias current of the edge-detector circuit. Each cell must, however, be smaller than the fovea such that it can be successfully centered within the fovea. Fig. 7 shows the floor plan of the periphery.

The periphery contains similar photoreceptor and edge detection as the fovea. The phototransduction load transistor is resized to accommodate the larger photocurrent. The spatiotemporal edge of the image is given by the temporal ON-set of the spatial edge. The temporal transient detectors are similar to those in the fovea. The row and column location of spatiotemporal edge is broadcast to the centroid computation circuit, where analog x and y coordinates are generated and output.

The mathematical computation of the centroid of an object determines the intensity weighted average of the position of the pixels forming the object. Equation (5) shows this formulation. The implementation of this representation can be quite involved because a product between the intensity and position is implied [36]. To eliminate this requirement, the intensity of the pixels can be normalized to a single value within the object. Normalization of the intensity using a simple threshold is not advised because the value of the threshold is dependent on the brightness of the image and the number of pixels forming the object may be altered by the thresholding process. To circumvent these problems, we take the view that the centroid of the object is defined in relation to its boundaries. This implies that spatiotemporal edge detection (second order spatial derivative of intensity and ON-set transient) can be used to highlight the boundaries of a moving object, and position labeling can be used instead of normalized intensity. Subsequently, the centroid of spatiotemporal edge events is computed. The modified centroid in (5) is then realized by projecting the edge events onto the x - and y -axis and performing two linear centroid determinations, and Fig. 8 shows this process

$$\hat{x} = \frac{\sum_{i=1}^N I_i x_i}{\sum_{i=1}^N I_i} \xrightarrow{I_i = \text{const} \tan t} \frac{\sum_{i=1}^N x_i}{N}$$

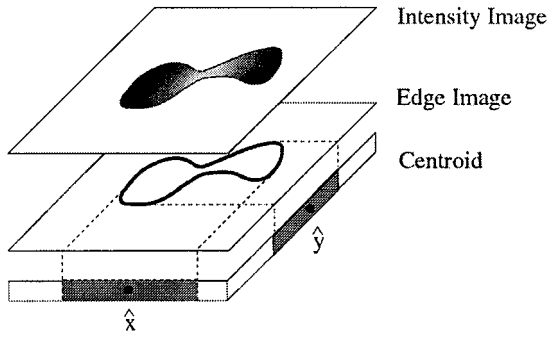


Fig. 8. Computation of centroid using the average position of spatiotemporal edges of a target.

and

$$\hat{y} = \frac{\sum_{i=1}^N I_i y_i}{\sum_{i=1}^N I_i} \xrightarrow{I_i = \cos \tan t} \frac{\sum_{i=1}^N y_i}{N}. \quad (5)$$

The centroid is computed using a resistance grid to associate the position of a column (row) with an analog voltage. In Fig. 9, the positions are given by the voltages V_i . By activating a column (row) switch when a pixel of the edge image appears in that column (row), the position voltage corresponding to the column (row) is connected to the output line through the switch impedance R_s . As more switches are activated, the position voltages are shorted through identical resistors to an output line. The average of the voltages, i.e., (5), is obtained. Clearly, because no buffers are used (to reduce power consumption) to isolate the position voltages, as more switches are activated, the position voltages will also change. This does not pose a problem because the switch resistors are designed to be larger than the position resistors (the switch currents are small compared to the grid current). Equation (6) gives the error between the ideal centroid and the switch loaded centroid in the worst case when $R_s = 0\Omega$. In the equation, N is the number of nodes, M is the number of switches set, and x_1 and x_M are the locations of the first and last set switches, respectively. This error decreases as R_s gets larger, and vanishes as $N(M \leq N)$ approaches infinity. The terms x_i represent an ascending ordered list of the activated switches; x_1 may correspond to column five, for example. This circuit is compact because it uses only a simple linear resistive grid and MOS switches. It is low power because the total grid resistance $N.R$ can be large, approximately $0.5 \text{ M}\Omega$. It can be fast when the parasitic capacitors are kept small. It provides an analog position value, but it is triggered by fast digital signals that activate the switches

$$\text{error} = \frac{V_{\max} - V_{\min}}{M(N+1)} \sum_{i=1}^M \left[x_i - \frac{x_1(N+1)}{N+1+x_1-x_m} \right]. \quad (6)$$

When an edge event is detected and the centroid is computed, the result persists until cleared by an external signal. Hence, as the object moves, a temporal average of the locations visited by the object is obtained. Typically, the circuit is allowed to accumulate spatiotemporal events over a predefined short temporal

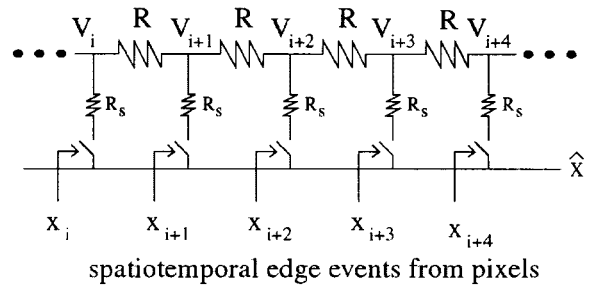


Fig. 9. Computation of the position of the edges of the target using a resistive array. The location of the centroid is represented by an analog voltage.

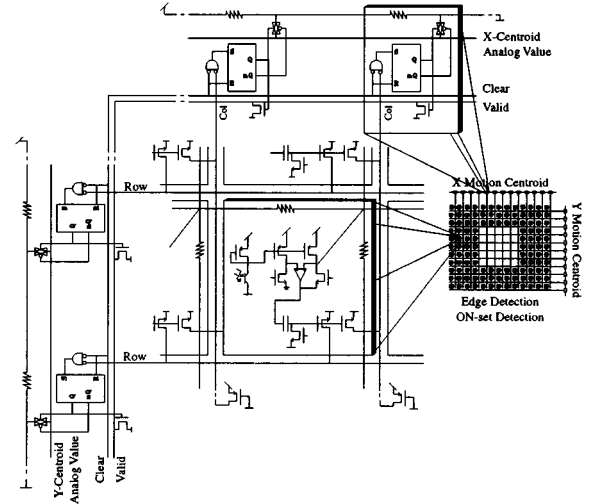


Fig. 10. Schematic of the peripheral cells and the centroid localization circuits.

window. In this way, a robust centroid is obtained which accounts for the temporally distributed response of low fill factor focal-plane image processors. That is, not all pixels belonging to the same moving object respond simultaneously. Fig. 10 shows the schematic of the centroid computation circuit.

The response dynamic range of the centroid localization circuit is also limited by the frequency cutoff of the transient detector and the turning off time of the photoreceptors. At the low end, the diode leakage current at the drain of the current source limits the minimum frequency of operation. Because space is abundant in the periphery, the capacitance of the transient detector is increased by an order of magnitude. Consequently, the minimum detectable intensity flicker frequency that will trigger the ON-set detector is smaller than 0.1 Hz . The highest frequency is ambient dependent, limited by the response time of the phototransduction circuit. The dark current and phototransistor capacitance, however, are much larger for the peripheral cell. Because the dark current is large, the strong inversion formulation of the turning off time in (4) applies. For a dim to dark transition with a measured dark current of 0.2 nA , calculated load capacitance of 4.5 pF , k_p of $15 \mu\text{A}/\text{V}^2$, and drawn load transistor aspect ratio of 9, the computed worst-case bandwidth is 51.6 kHz . This value is also close to the measured bandwidth of 63 kHz in $2.5\text{-}\mu\text{W}/\text{cm}^2$ ambient intensity. In practice, the current source is biased for a low frequency cutoff above 60 Hz . This reduces false ON-set triggering due to fluorescent lights. The false triggering is further reduced by the edge-detection circuit, which is

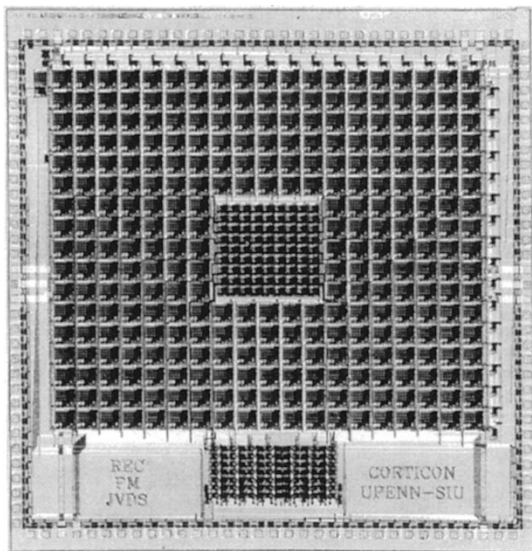


Fig. 11. Micrograph of the foveated tracking chip.

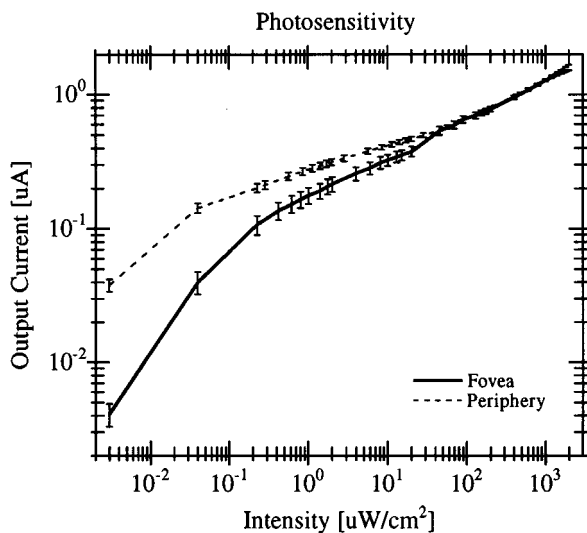


Fig. 12. Measured photosensitivity of the range compression circuit. Six decades of light intensity are compressed into six decades of output current.

mostly nonresponsive (except when the trough of the oscillation is too low and the photodetector outputs no current) to full-field fluctuations in ambient intensity. Fig. 11 shows a micrograph of the chip.

IV. RESULTS

A. The Photoreceptor

The first layer of the chip consists of the phototransistors and range compression circuits. In both the periphery and fovea, similar circuits are used. However, the larger phototransistor in the periphery is equipped with a range compression circuit that reduces the output current such that the edge-detecting circuit receives roughly the same current as the fovea for the same ambient intensity. In practice, the current from the range compression circuit is slightly larger in the periphery for all intensities. Fig. 12 shows the measured compressed photoreceptor output

current for many pixels on many chips that are fabricated in the same run. The plotted current is measured at the output of M2 or M3 in Fig. 4 and is not the direct photocurrent from the phototransistor. The variation—whose range is indicated by the error bars—among pixels is less for peripheral pixels (<10%) than fovea pixels (<20%). This is due to the larger sizes of the phototransistors and range compression circuit transistors M1–M3 in the periphery, which promotes better matching. Secondly, due to the larger currents in the periphery, the M1–M3 transistors spend little of their operation range in weak inversion, which again promotes matching. The effects of improved matching for larger currents, which is explored in reference [44], are also illustrated by the reduction in the variation in output current as ambient intensity increases. The plot shows the six orders of magnitude over which the range compression circuit provides useful output. For smaller ambient intensities, the output current is too small for the edge-detector circuit to respond. The directly measured dark currents for the fovea and peripheral phototransistors are 18 and 200 pA, respectively. These small currents, which vary among cells, coupled with other leakage currents and mismatches in the circuit produce offsets in the edge-detector outputs. In some cases, although rare, the offsets can be so large that a cell appears to be stuck high or low.

B. The Edge Detector

The second processing stage of both the foveal and peripheral cells is edge detection using the discrete approximation of a Laplacian operator, as given by 1(b). The relevant measured characteristics of this circuit are the spatial impulse response Fig. 13(a), temporal response Fig. 13(b), and contrast sensitivity Fig. 13(c). The error bars in Fig. 13(a) show the standard deviation of the output voltages. Here, the min–max range does not give a clear picture of the variation among cells because some cells, <3% for both regions, can be stuck on or off. The spatial impulse response is measured for 100% contrast white line in $25\text{-}\mu\text{W}/\text{cm}^2$ ambient light. Due to the slightly larger current in the periphery, the peak-to-peak voltage is larger and the space constant is smaller for the peripheral edge detectors.

The temporal trace is obtained for a dark to bright edge moving past the chip. The measured peak-to-peak signal as a function of calculated edge speed is also plotted in Fig. 13(b). The experiment is conducted using a flashing red LED on a white background, and the equivalent edge speed for the flashing frequency is computed by assuming that the response time corresponds to the transit time between adjacent pixels. The plot shows the slightly larger peak-to-peak voltage for the peripheral cells, which result from their slightly smaller space constant. Also visible in Fig. 13(b) is the lowering of the peak-to-peak signal value caused by an increased leakage current to ground at the source and drain of the lateral transistors (used to implement resistors). The signal shifts negatively with increasing light. The minimum value of the signal, which is constrained by the bias current of the differential pair in the edge-detector circuit, remains constant. Consequently, the magnitude of the peak-to-peak voltage decreases. Increasing ambient light intensity increases the bandwidths of both types of cells as the response time of the range compression circuit

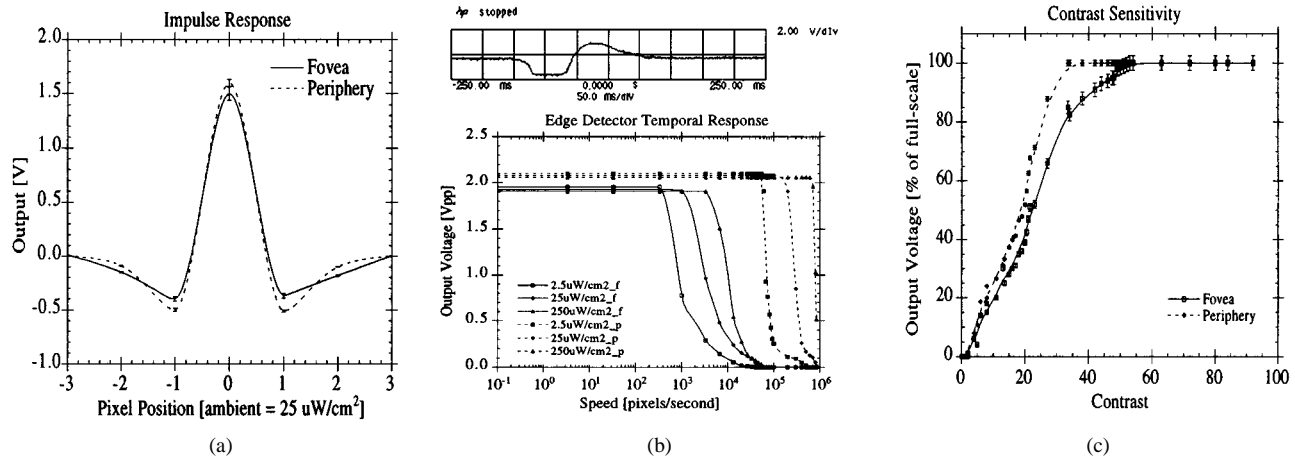


Fig. 13. Measured characteristics of the edge-detection circuits. (a) Spatial impulse response. (b) Temporal response. (Top) Oscilloscope trace of a pixel as a bright-to-dark edge moves to the right. (Bottom) Magnitude of the frequency response plotted in terms of equivalent target speed. (c) Contrast sensitivity, where contrast is defined as the magnitude of the difference/sum of intensity across an edge.

decreases. The larger bandwidth of the peripheral cells compared to the foveal cells, due to their larger phototransistors, is also visible.

Fig. 13(c) shows the contrast sensitivity of the edge-detector circuits. Here, contrast is defined as the difference/sum of brightness across an edge. The data is collected by presenting gray-scale edges to the chip and recording the output voltage of the edge-detector circuits. Each data point represents an average of many pixels in ambient intensities of 2.5-, 25-, and 250- $\mu\text{W}/\text{cm}^2$ (realized using neutral density filters). The error bars show the standard deviation of the variations observed between pixels, which are mainly introduced at the lowest ambient intensity. Consistent with the sharper spatial tuning and higher gain of the peripheral cells, their contrast sensitivity is also larger. Depending on the pull-up bias currents on the global motion and row/column event lines, the system can respond to contrast as low as 10%. Setting the minimum detectable contrast to such a low value, however, also increases the chip's sensitivity to florescent lights. Typically, the circuits are biased for minimum contrast sensitivity of 20%.

C. Foveal Motion

The motion-detection circuit in the fovea produces a pulse train on the Left, Right, Up, and Down global motion lines when a target moves away from the center in any of these directions. Clearly, a looming target will signal motion in all directions simultaneously. The directions with the fastest motion, however, will have more motion pulses per unit time because the target will traverse more pixels along the faster direction. Consequently, using temporal integration and subtraction of opposite motion allows the faster direction to dominate. In a smooth pursuit scenario, the motor action itself acts as a pulse integrator because movements in the direction with more pulses will be matched more often than the opposite direction.

During testing, we observed that the edge-detection circuit allows the edge to appear at a neighbor before it has completely disappeared from the previous pixel, as visible in Fig. 14(a). As explained above, the role of the OFF- and ON-set time constants must be reversed. The time constants for the OFF-set detectors

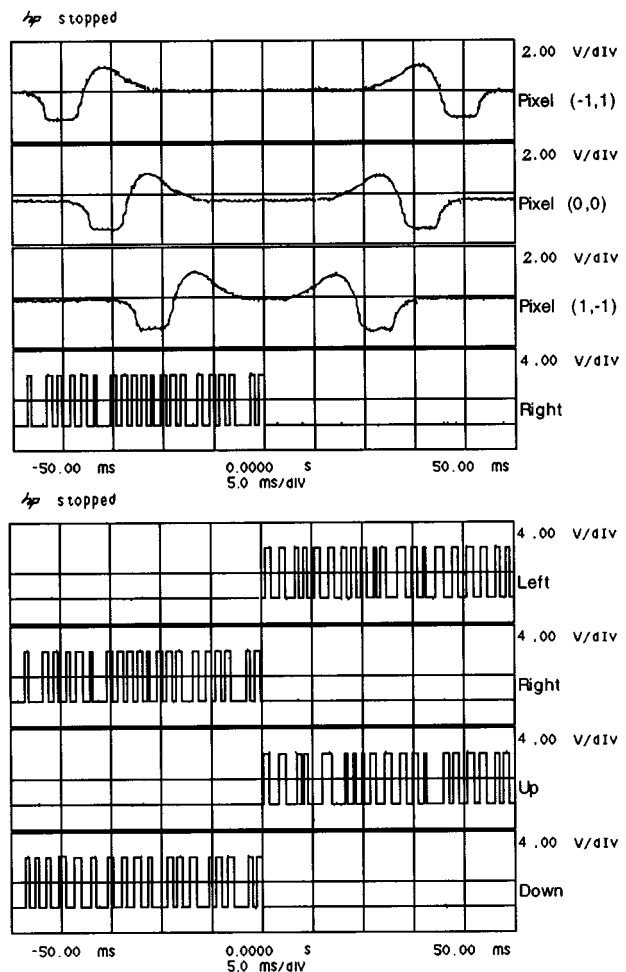


Fig. 14. (a) Oscilloscope traces of the edge-detector outputs and Right global motion signals for a broad high-contrast light-to-dark edge, oriented at 45° , moving diagonally from top-left to bottom-right and back. The top three panels show the edge-detector outputs of neighboring cells, while the bottom panel shows the Right global motion signal. (b) Oscilloscope traces of all the global motion lines. Notice that in the forward pass, Right and Down motion are detected and in the return pass, Left and Up motion is detected.

are set to small values (<10 ms), while that of the ON-set detectors have large values (>100 ms), because the OFF-set occurs

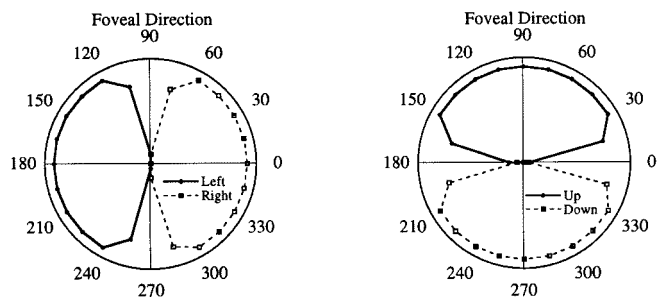
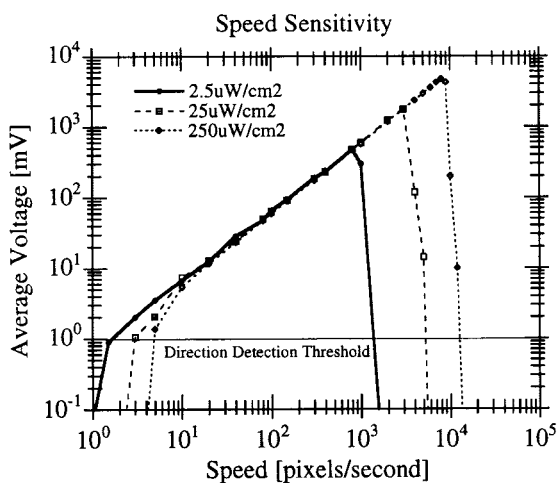
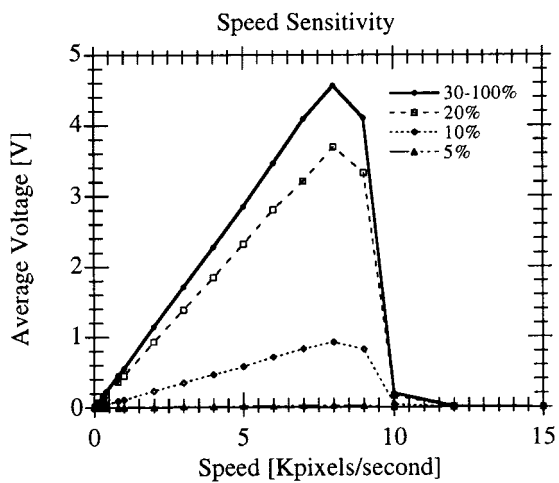


Fig. 15. Measured direction of motion, as a function of angle, for a bright spot moving over the fovea at ~ 100 pixels/s. Only divergent motion is reported.



(a)



(b)

Fig. 16. Measured speed sensitivity of the direction-of-motion detectors for (a) various ambient intensities and (b) various target contrast.

after the ON-set (see Fig. 6). Setting the times constants in this way also reduces temporal aliasing. Temporal aliasing occurs when opposite motion is reported because of false correlation between transients from a forward pass of the target with new transients during the return pass. Typically, this occurs when the time between the forward and return passes is too short. Because

the ON-set at a pixel occurs before the OFF-set from the previous pixel, a very short OFF-set transient limits the potential for false correlation greatly.

To demonstrate the motion-detection capabilities of the fovea, a broad and long, high contrast light to dark edge (left to right), oriented at 45° to the horizontal, is oscillated diagonally across the fovea from top-left to bottom-right. Fig. 14(a) shows the edge-detector outputs of three neighboring pixels: center (0,0), top-left of the center (-1, 1), and bottom-right of the center (1, -1). Because the edge has a negative spatial intensity gradient, the pixels are initially inhibited when the edge is located at their neighbors to the left. They become positive when they see the bright part of the edge, provided that at least one neighbor is seeing the dark part. As the edge passes and all neighbors see the same part of the edge, the pixels' outputs return to zero. The Right motion-detection circuit is triggered when the edge hits any pixel in the right half-plane when moving from left to the right. Because the edge is oriented at 45° , this occurs as the edge approaches the center of the fovea from the top-left. Fig. 14(a) shows the Right global motion detectors getting activated as expected. Fig. 14(b) shows all the global motion outputs. The reported motion is consistent with the edge's motion in both the forward and reverse directions. In Fig. 14, the pixel OFF-set transient detector has a 5-ms time constant, so that the individual motion pulses can be seen. During normal operation, this time constant is set to about 10–100 μ s. The ON-set transient time constant is set to approximately 100–500 ms.

To demonstrate the diverging motion-detection properties of the fovea, a high contrast bright spot, the defocused trace of an analog oscilloscope, focused to cover about 3×3 pixels, is moved at constant 100 pixels/second across the fovea at various angles. The point, whose movement is controlled by a pair of function generators, oscillates back and forth across the fovea. Fig. 15 shows the reported motion as a function of angle. The plot is obtained by integrating the voltages from the global motion directions and computing the half-wave rectified difference/sum for opposite directions during the forward and return passes separately. Each point is averaged over 10 passes. Again, the expected results are obtained with no overlap between opposite directions for most angles. As the angle of motion approaches the axis separating the half-planes, more overlap between opposite directions is observed, which in turn reduces the ratio of difference/sum.

Fig. 16 shows a plot of reported direction-of-motion as a function of target speed. Here, the OFF- and ON-set time-constants are biased for 100 μ s and about 2 s, respectively. Fig. 16(a) shows the effects of ambient intensity for a high contrast long edge. In these experiments, the edge is oriented perpendicular to the direction of motion. The average of the pulse-train voltage is computed from the rise of the first motion pulse to the fall of the last. The plot represents the average for multiple directions, cells, and chips. The variation error-bars of about 7% are not visible on the log-log plot. Each pulse is approximately 100- μ s wide and there are some overlaps, as can be seen in Fig. 14. As the speed of the target increases, so does the density of the motion pulse train. Consequently, there is a linear relationship between the density of pulses, hence the average voltage, with speed. A 1 pixel/s motion pulse train produces an

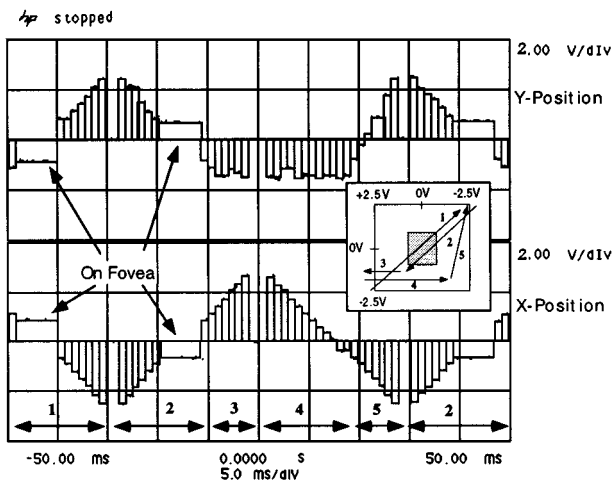


Fig. 17. Oscilloscope trace of the centroid position of a point bright target moving on the peripheral region. The inset drawing shows the actual motion of the target.

average voltage of 0.1 mV. Although this pulse train can be visually observed and a motor driver can react to it, the average voltage is below the noise floor. Hence, the detection floor of 1 mV is chosen, which corresponds to approximately 10 pulses/s (1 pixel/s with 10 consecutive cells reporting), or 1 pulse per 100 ms (10 pixels/s with a single cell reporting). The increased maximum speed as a function of ambient intensity is discussed in Section IV-B. The temporal response of the edge-detector circuit plays a major role in the high-speed cutoff for the motion cells. At low speeds, the intensity also plays a role. Increasing the ambient intensity decreases the ON- and OFF-set time constants by increasing transistor leakage. The transient detectors are shielded with metal₂, but some light still manages to hit the current source transistors in Fig. 5. Consequently, we observe that increasing light intensity also increases the minimum detectable speed of the motion cells. None the less, direction-of-motion detectors operate over three orders of target speed magnitude.

Fig. 16(b) shows the effects of contrast on the speed sensitivity. The foveal motion detectors operate at their best for contrasts of 30% or higher. For $>20\%$ contrast, the edge-detection circuit outputs near full-scale voltage swings, allowing the transient detectors to function normally. At $<20\%$, the edge-detector output starts to drop linearly, as seen in Fig. 13(c). Consequently, the transient detectors are activated with smaller voltage swings, which imply that the motion current, given in (2a), decreases quadratically to exponentially depending on the size of the input edge voltage. Hence, the motion voltages V_R , V_L , V_U and V_D in Fig. 5 are pulled down by a smaller amount and for a shorter time. The density and amplitude of the motion pulse train are decreased, resulting in smaller average voltages. The dropoff with decreasing contrast is very fast below 20%

D. Peripheral Position

The peripheral pixels report the location of the centroid of a moving or flashing target with an analog voltage. There are 19 (H) by 17 (V) cells in the periphery, excluding a 5×5 region at the center where the fovea resides. Fig. 17 shows an oscilloscope trace of the x - and y -position voltages for a high con-

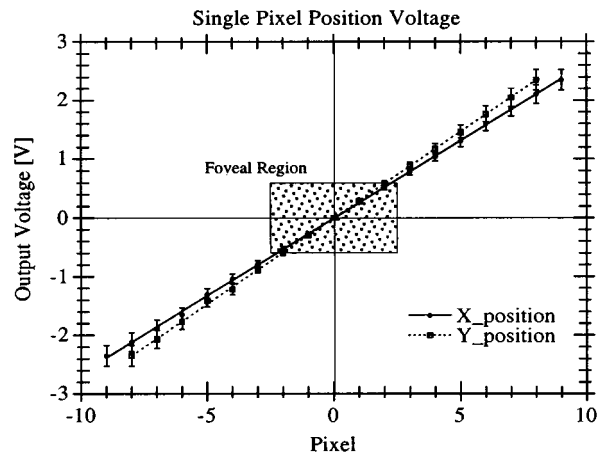


Fig. 18. The measured analog position of a bright point target. The average and variations for passes, pixels, and chips are shown.

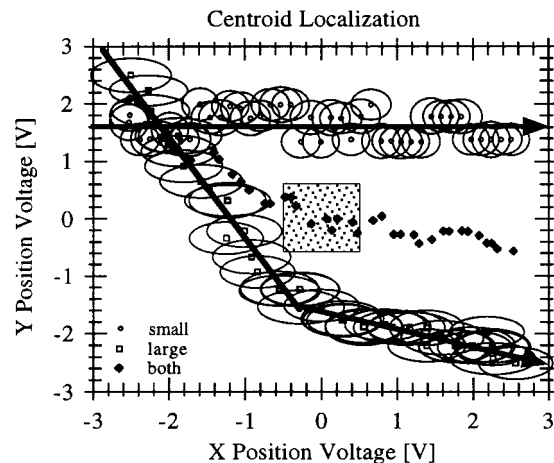


Fig. 19. Measured effects of multiple target and target sizes are shown. The target labeled "small" and "large" cover 2×2 and 5×2 peripheral pixels, respectively. When both targets move, a combined centroid is reported.

trast point-like target moving across the chip. The target, implemented with the trace of an analog oscilloscope, is moving at constant X speed of 27 foveal pixels/s and varying Y speed. The speed of the target can be extracted from the position traces by recollecting that the pitch of the peripheral pixels is twice larger than the foveal pitch. The motion of the target is also shown in the figure. In this experiment, a clear signal is sent to the periphery every $750 \mu\text{s}$, except when the target is in the fovea. The presence of foveal motion prevents the periphery from getting reset (or Acknowledged in Fig. 2). Therefore, the last position reported before entering the fovea remains until the target leaves the fovea. Because the target is point-like, each voltage level corresponds to the analog representation of its position. Fig. 18 plots the pixel position and average output voltage recorded for many passes of the target, from many pixels in a chip, and on many chips. The error-bars show the measured min-max variations, which are about 7.5% of the mean. The usual variations in analog circuits and mismatches in the branch resistance of the position resistive array can account for all these variation.

In Fig. 19, we investigate the effects of multiple targets and size on the reported centroid location. In the first experiment, we used a small target (approximately 2×2 on the array)

TABLE I
SUMMARY OF THE CHIP'S CHARACTERISTICS AND PERFORMANCE

Technology	2 μm NWELL CMOS, 2 Metal, 2 Poly	
Chip Size	6.4 x 6.8 mm ²	
Package	132 Pin DIP	
Array Sizes	Fovea: 9x9 @150 μm pitch	Peri: 19x17 @300 μm pitch
Fill Factor	Fovea: 18%	Peri: 34%
Transistors/Cell	Fovea: Receptor + Edge: 12	Peri: Receptor + Edge + ON: 12
	Fovea: Motion: 8	Peri: Centroid: 15
Photosensitivity	6 Orders of Magnitude	
Contrast	10 - 100%	
Foveal Direction Sensitivity	2.5 $\mu\text{W}/\text{cm}^2$: 1.5 - 1.5K pixels/s	
	25 $\mu\text{W}/\text{cm}^2$: 3 - 4.5K pixels/s	
	250 $\mu\text{W}/\text{cm}^2$: 5 - 10K pixels/s	
Peripheral ON-set Sensitivity	2.5 $\mu\text{W}/\text{cm}^2$: < 0.1 - 63K Hz	
	25 $\mu\text{W}/\text{cm}^2$: < 0.1 - 250K Hz	
	250 $\mu\text{W}/\text{cm}^2$: < 0.1 - 800K Hz	
Power Consumption	25 $\mu\text{W}/\text{cm}^2$: 15mW @ 4V Supply	

moving at constant Y -position and scanning back and forth in X . The heavy solid line represents the actual position of the target, while the points labeled “small” show the mean position over 20 passes. The circle shows the deviation from the mean, which is about 1 pixel in each direction. Clearly, depending where the majority of the target intensity falls immediately after the clearing of the last measurement, the edge detection of that pixel will be triggered and its position reported. In the second experiment, we used a high-contrast bar or roughly 5×2 pixels. The heavy solid line shows the actual path of the bar, while the points labeled “large” shows the average position of 20 passes. As expected, the position deviation in X is larger, about 3 pixels, than the deviation in Y , about 2 pixels. The envelope of deviation gives a profile of the target dilated by a pixel. Lastly, we presented both targets simultaneously, labeled “both.” As expected, a combination of both target locations is observed. In fact, the position skips to and from the individual targets initially, and moves toward a common centroid as more pixels are activated. The points labeled “both” shows the average positions of 20 passes. It is not useful to show the deviation as it covers the entire space between the two targets. Evidently, the merging of multiple targets into a single centroid is a limitation for this type of centroid localization system. By increasing the frequency of the clear signal, the individual centroid can be extracted. In this case, the positions of the targets immediately after the clear signal dictate which target is localized, and the output signal will randomly skip back and forth between targets.

Lastly, we investigated the temporal response characteristics of the chip. This experiment was performed using a flashing LED. The observations were identical to the data presented in Fig. 13(b). The bandwidth-limiting components are the photoreceptor and range compression circuits. Depending on ambient light intensity, the centroid localization system operated from 63–800 kHz (2.5 $\mu\text{W}/\text{cm}^2$ and 250 $\mu\text{W}/\text{cm}^2$, respectively).

Contrast sensitivity follows the same trend as the motion-detection circuit but the dropoff begins at 10% contrast for the peripheral cells. This is due to the higher contrast sensitivity and simple ON-set detection in the periphery. Correlation between two neighboring pixels is not required in the peripheral region.

E. Overall Chip Performance and Limitations

The complete chip characteristics are presented in Table I. Overall, this chip performs well in both ideal and normal environments. In addition to testing the chip with high contrast no background stimuli, we also used it for tracking people in cluttered indoor and outdoor environments. With minor readjustments of the biasing voltages, the chip works in indoor florescent lighting, as well as bright sunlight. It tracks people walking and running at close and far distances. As the target's distance from the chip increases, the limited resolution of the array becomes apparent. The fovea is not longer able to detect the target's motion because it is too small for the edge detector to respond. As can be expected, the optics is crucial in determining the useful range of the chip. Another shortcoming is the small size of the fovea, in terms of number of pixels, means that a target being smoothly pursued can escape the fovea. In that case the periphery is used to re-foveate the target. The single centroid periphery can also cause some confusion if other moving or noisy targets are present. This problem can be debilitating when a target escapes the fovea but a different moving or noisy target is localized and foveated. Lastly, the motor system is also important in the tracking system. The response characteristics and resolution of the movements must match the specifications of the chip. Because this chip only provides for position-based tracking, even during smooth pursuit, the tracking can be jerky. The chip has also been used to guide autonomous vehicles in road following tasks involving obstacle avoidance [33].

V. CONCLUSION

We have presented a 2-D foveated silicon retina that is used to track a visual target. The circuits are inspired by various parts of biological visual systems. The phototransduction, range compression, and edge-detection systems model the first two layers of the primate retina. The ON- and OFF-set detection circuits model some properties of the outer-plexiform layer of the rabbit retina. The foveal motion-detection circuits mimic the fly motion-detection system, while the centroid localization system functionally mimics the saccadic generation system in the primate superior colliculus. This oculomotor control system-on-a-chip operates over six orders of ambient intensity, contrast as low as 10%, and target speeds up to 10K pixels/s, while using only 15 mW in normal indoor lighting. The speed sensitivity dynamic range of the chip is dependent on the ambient intensity. The results show that the foveal motion-detection circuits correctly report 2-D direction of diverging motion. The peripheral circuit computes the centroid of translating or flashing edges falling on the periphery. Using a simple arbitration state machine to switch between the saccadic generation and smooth pursuit, the chip has been used to successfully track targets in cluttered environments under various lighting conditions. Lastly, the chip has also been used for autonomous navigation with obstacle avoidance.

REFERENCES

- [1] D. Coombs, M. Herman, T.-H. Hong, and M. Nashman, "Real-time obstacle avoidance using central flow divergence, and peripheral flow," *IEEE Trans. Robot. Automat.*, vol. 14, pp. 49–59, Feb. 1998.
- [2] P. Nesi, F. Innocenti, and P. Pezzati, "RETIMAC: REEal-Time Motion Analysis Chip," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 361–375, Mar. 1998.
- [3] F. Mura and I. Shimoyama, "Visual guidance of a small mobile robot using active, biologically-inspired, eye movements," in *Proc. IEEE Int. Conf. Robotics and Automation*, Leuven, Belgium, May 16–20, 1998.
- [4] D. Irwin, "Integrating information across saccadic eye movements," *Current Direct. Psychol. Sci.*, vol. 5, no. 3, pp. 94–100, 1996.
- [5] J. Tsotsos, D. Fleet, and A. Jepson, "Toward a theory of motion understanding in man and machine," in *Motion Understanding: Robot and Human Vision*, Martin and Agarwal, Eds. Norwell, MA: Kluwer, 1998.
- [6] M. Eckert and G. Buchsbaum, "Effect of tracking strategies on the velocity structure of two-dimensional image sequences," *J. Opt. Soc. Am.*, vol. A10, no. 7, pp. 1582–1585, 1993.
- [7] C. Koch and H. Li, Eds., *Vision Chips: Implementing Vision Algorithms with Analog VLSI Circuits*. Piscataway, NJ: IEEE Computer Press, 1995.
- [8] R. Etienne-Cummings and J. Van der Spiegel, "Neuromorphic vision sensors," *Sensors and Actuators: A*, vol. SNA056, pp. 19–29, 1996.
- [9] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, pp. 1629–36, 1990.
- [10] R. Etienne-Cummings, J. Van der Spiegel, and P. Mueller, "Neuromorphic and digital hybrid systems," in *Neuromorphic Systems: Engineering Silicon from Neurobiology*, L. Smith and A. Hamilton, Eds., Singapore: World Scientific, 1997.
- [11] T. Horiuchi, "Analog VLSI-based, neuromorphic sensorimotor systems: Modeling the primate oculomotor system," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, 1997.
- [12] N. Ferrier, "The Harvard binocular head," Harvard University, Cambridge, MA, Tech. Rep. 91-9, 1991.
- [13] N. Papanikolopoulos and P. Khosla, "Adaptive robotic visual tracking: Theory and experiments," *IEEE Trans. Automat. Contr.*, vol. 38, pp. 429–445, 1993.
- [14] D. Coombs and C. Brown, "Real-time binocular smooth pursuit," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 147–164, 1993.
- [15] M. Kulaczewski and H. Siegel, "SIMD and mixed-mode implementations of a visual tracking algorithm," in *Proc. Int. Parallel Processing Symp.*, Orlando, FL, 1998.
- [16] D. Murray, K. Bradshaw, P. Mclauchlan, I. Reid, and P. Sharkey, "Driving saccade to pursuit using image motion," *Int. J. Comput. Vis.*, vol. 16, no. 3, pp. 205–228, Nov. 1995.
- [17] A. Colla, L. Trogu, and R. Zunino, "Digital neuroimplementations of visual motion-tracking systems," in *Proc. 5th IEEE Workshop Neural Networks for Signal Processing*, Cambridge, MA, 1995, pp. 571–580.
- [18] S. DeWeerth, "A VLSI-based system for tracking visual stimuli," in *Proc. IEEE Int. Conf. Robotics and Automation*, Apr. 9–11, 1991.
- [19] T. Horiuchi, T. Morris, C. Koch, and S. P. DeWeerth, "Analog VLSI circuits for attention-based visual tracking," *Adv. Neural Inform. Processing Syst.*, vol. 9, 1996.
- [20] C. Bandera, "Foveal machine vision systems," Ph.D. dissertation, SUNY Buffalo, New York, NY, 1990.
- [21] S. Xia, R. Sridar, P. Scott, and C. Bandera, "An all CMOS foveal image sensor chip," in *Proc. 11th IEEE Int. ASIC Conf.*, Rochester, NY, Sept. 1999.
- [22] G. Krieder, "A treatise on log-polar imaging using a custom computational sensor," Ph.D. dissertation, Dept. Elect. Eng., Univ. of Pennsylvania, Philadelphia, PA, 1993.
- [23] F. Ferrari, J. Nielsen, P. Questa, and G. Sandini, "Space variant sensing for personal communication and remote monitoring," in *Proc. EU-HCM Smart Workshop*, Apr. 1995.
- [24] R. Wodnicki, G. Roberts, and M. Levine, "A foveated image sensor in standard CMOS technology," in *Proc. Custom Integrated Circuits Conf.*, 1995, pp. 357–360.
- [25] B. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.
- [26] W. Reichardt, "Autocorrelation: A principle for the evaluation sensory information by the central nervous system," in *Sensory Communication*. New York, NY: Wiley, 1961.
- [27] C. Koch, "Implementing early vision algorithms in analog hardware: An overview," *SPIE Vol. 1473 Vis. Inform. Processing: From Neurons to Chips*, pp. 2–15, 1991.
- [28] R. Etienne-Cummings, S. Fernando, J. Van der Spiegel, and P. Mueller, "Real time 2-D analog motion detector VLSI circuit," in *Proc. Int. Joint Conf. Neural Networks*, vol. IV, 1992, pp. 426–431.
- [29] T. Horiuchi, W. Bair, B. Bishofberger, A. Moore, C. Koch, and J. Lazaro, "Computing motion using analog VLSI vision chips: An experimental comparison among different approaches," *Int. J. Comput. Vis.*, vol. 8, pp. 203–16, 1992.
- [30] R. Etienne-Cummings, J. Van der Spiegel, and P. Mueller, "A focal plane visual motion measurement sensor," *IEEE Trans. Circuits Syst. I*, vol. 44, pp. 55–66, Jan. 1997.
- [31] H. Jiang and C. Wu, "A 2D velocity and direction-selective sensor with BJT-based silicon retina and temporal zero-crossing detector," *IEEE J. Solid-State Circuits*, vol. 34, pp. 241–247, 1999.
- [32] R. Sarpeshkar, J. Kramer, G. Indiveri, and C. Koch, "Analog VLSI architectures for motion processing: From fundamentals to system applications," *Proc. IEEE*, vol. 84, pp. 969–987, July 1996.
- [33] R. Etienne-Cummings, V. Gruev, and M. Abdel-Ghani, "VLSI implementation of motion centroid localization for autonomous navigation," in *Advances in Neural Information Processing Systems 11*. Cambridge, MA: MIT Press, 1999, pp. 685–691.
- [34] C. Higgins, R. Deutschmann, and C. Koch, "Pulsed-based 2-D motion sensors," *IEEE Trans. Circuits Syst. II*, vol. 46, pp. 677–687, June 1999.
- [35] D. Marr, *Vision*. San Francisco, CA: Freeman, 1982.
- [36] S. DeWeerth, "Analog VLSI circuits for stimulus localization and centroid computation," *Int. J. Comput. Vis.*, vol. 8, no. 2, pp. 191–202, 1992.
- [37] P. Mansfield, "Machine vision tackles star tracking," *Laser Focus World*, vol. 30, no. 26, pp. S21–S24, 1996.
- [38] D. Sparks, C. Lee, and W. Rohrer, "Population coding of the direction, amplitude and velocity of saccadic eye movements by neurons in the superior colliculus," in *Proc. Cold Spring Harbor Quantitative Biology Symp.*, vol. LV, 1990.
- [39] D. Robinson, "Basic mechanisms of ocular motility and their clinical implications," *Oculomotor Contr. Signals*, pp. 337–374, 1975.
- [40] C. Mead, "Sensitive electronic photoreceptor," in *Proc. 1985 Chapel Hill VLSI Conf.*, 1985, pp. 463–471.
- [41] S. Mendis, S. Kemeny, and E. Fossum, "CMOS active pixel image sensor," *IEEE Trans. Electron Devices*, vol. 41, pp. 452–453, Mar. 1994.
- [42] S. Decker, D. McGrath, K. Brehmer, and C. Sodini, "A 256x256 CMOS imaging array with wide-dynamic range pixels and column-parallel digital output," *Int. J. Solid-State Circuits*, vol. 33, no. 12, pp. 2081–2091, Dec. 1998.

- [43] R. Normann and F. Werblin, "Control of retinal sensitivity," *J. Gen. Physiol.*, vol. 63, pp. 37–61, 1974.
- [44] A. Andreou, K. Boahen, P. Pouliquen, and A. Pavasovic, "Current mode sub-threshold MOS circuits for analog VLSI neural systems," *IEEE Trans. Neural Networks*, vol. 2, pp. 205–213, March 1991.
- [45] H. Barlow, *The Senses: Physiology of the Retina*. Cambridge, MA: Cambridge Univ. Press, 1982.



Ralph Etienne-Cummings (S'94–M'98) received the B.Sc. degree in physics from Lincoln University, Lincoln University, PA, and the M.S.E.E. and Ph.D. degrees in electrical engineering from the University of Pennsylvania, Philadelphia, PA, in 1988, 1991, and 1994, respectively.

Currently, he is an Assistant Professor of electrical and computer engineering at the Johns Hopkins University, Baltimore, MD. His research interest includes mixed-signal VLSI systems, computational sensors, computer vision, neuromorphic engineering,

smart structures, mobile robotics, and robotics-assisted surgery.

Dr. Etienne-Cummings is a recipient of the National Science Foundation's Career Development Award and the Office of Naval Research's Young Investigators Program Award.



Jan Van der Spiegel (M'72–SM'90) received the B.Eng. degree in electromechanical engineering and the Ph.D. degree in electrical engineering from the University of Leuven, Belgium, in 1974 and 1979, respectively.

From 1980 to 1981, he was a Post-Doctoral Fellow at the University of Pennsylvania, Philadelphia, PA, later becoming Assistant Professor, then Associate and Full Professor of electrical engineering in 1987 and 1995, respectively. He is currently Chairman of the Department and Director of the Center for Sensor

Technology. His research interests are in analog and digital integrated circuits for intelligent sensors, data acquisition, sensory data processing systems and acoustic-phonetic feature extraction for automatic speech recognition.

Dr. Van der Spiegel is the Editor for *N&S America of Sensors and Actuators* and is on the Editorial Boards of the *International Journal of High Speed Electronics* and the *Journal of the Brazilian Microelectronics Society*. He has served on several IEEE Program Committees and is currently on the Program and Executive Committees of the ISSCC. He holds the UPS Distinguished Education Term Chair, and was the recipient of the Bicentennial Chair of the Class of 1940, the Presidential Young Investigator Award, and the S.R. Warren and C & M Lindback Award for distinguished teaching. He is a member of Phi Beta Delta and Tau Beta Pi.

Paul Mueller received the M.D. degree from Bonn University, Bonn, Germany.

He was formerly at the Rockefeller University and the University of Pennsylvania, and is currently Chairman of Corticon, Inc., Philadelphia, PA. Since 1953, he has worked in molecular and systems neuroscience and has been involved in theoretical studies and hardware implementation of neural networks since the early 1960's.

Mao Zhu Zhang received the M.S. degree in electrical engineering from the University of Pennsylvania, Philadelphia, PA, in 1994

She is currently a VLSI Development Engineer with Motorola Inc., Valley Forge, PA, and has been involved in time-resolved IR spectroscopy at NIMS, Philadelphia, PA, development of the visual motion-tracking and acquisition system at Corticon, Inc., Philadelphia, PA, and subsequently the design of chips for visual and acoustical data processing for the cable television industry at General Dynamics Corporation, Valley Forge, PA.