

A Two-Dimensional Visual Tracking Array

Stephen P. DeWeerth and Carver A. Mead

Department of Computer Science
California Institute of Technology
Pasadena, California 91125

The density and concurrency available in VLSI make it an excellent technology for implementing visual image-processing. By incorporating phototransistors and analog processing elements onto a single die, the large signal bandwidths required for real-time computations can be achieved. This paper describes a VLSI chip that computes the "center of intensity" of a two-dimensional visual field. One application for this network is the localization of a bright spot of light against a dark background. Theoretical and experimental results are presented to describe the operation of the system and its suitability as an input device for tracking servo systems.

1 Introduction

The need for highly parallel computational architectures in visual processing and sensing has been well established. VLSI has proven to be an effective implementation technology for such computation.[4] Through the use of computation on both local and global scales, high-level information can be extracted from a visual image. In particular, a class of features exists that can be encoded using a fixed number of wires, independent of the image array size. An example is the uniform velocity detector developed in our laboratory.[5]

This paper describes another feature extraction: the computation of the center of intensity of the entire visual field. This computation effectively determines the position of a bright spot in a visual image provided that the background is sufficiently dim. We have designed and fabricated an integrated circuit that employs analog processing to perform this computation. Circuitry along the periphery of a two-

dimensional array of photoreceptors is used to compute the mean or median of the position of the receptors weighted by their respective light intensities.

2 Theory

In this section, we shall present circuitry to compute a weighted mean using a resistive voltage divider. We will define the concept of a weighted median, and will show how this quantity can be computed with a similar circuit of threshold elements.

2.1 Parallel Resistor Network

A network of linear resistors configured as a voltage divider is shown in Figure 1. The current in the n th resistor is

$$i_n = g_n(V_n - V_{\text{out}}) \quad (1)$$

The sum of the currents at node V_{out} must be zero. Hence,

$$\sum_{n=0}^N g_n(V_n - V_{\text{out}}) = 0 \quad (2)$$

Solving for V_{out} gives

$$V_{\text{out}} = \frac{\sum g_n V_n}{\sum g_n} \quad (3)$$

From Equation 3, we see that the output voltage is the mean of the input voltages weighted by the corresponding conductances.

2.2 Parallel Threshold-Element Network

A threshold-element is a hypothetical infinite-gain current source with the following voltage-current relationship:

$$i = \begin{cases} -\gamma & V_1 < V_2 \\ 0 & V_1 = V_2 \\ \gamma & V_1 > V_2 \end{cases} \quad (4)$$

where γ is the magnitude of the current through the device.

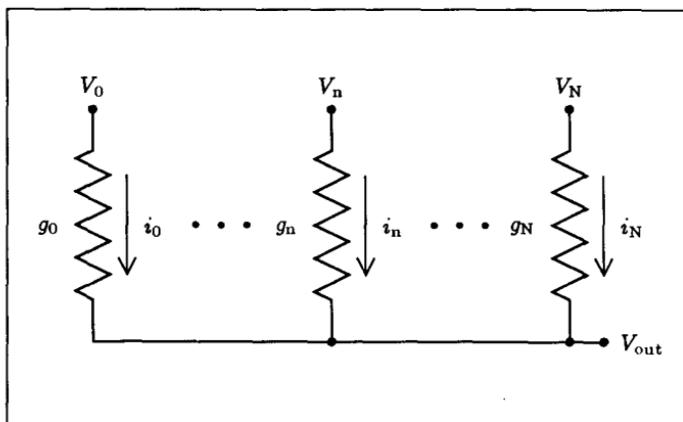


Figure 1 A parallel network of resistors. The sum of the currents through the resistors must be zero due to Kirchoff's current law at the output node. This effect and the linear voltage-current relationship of the resistors cause the output voltage to equal the mean of the input voltages weighted by the corresponding conductances.

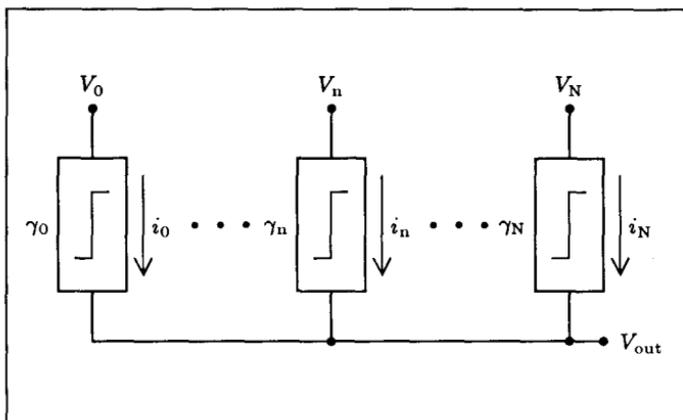


Figure 2 A parallel network of threshold elements. Kirchoff's current law mandates that the sum of currents through the devices with input voltages less than the output voltage must equal the sum of the currents through the devices with input voltages greater than the output voltage. This relationship causes the network to compute the median of the input voltages weighted by the current magnitudes for the corresponding elements.

We can construct an array of these elements (Figure 2), similar to the resistor array in Figure 1. From the definition in Equation 4, the current through the n th threshold element is

$$i_n = \begin{cases} -\gamma_n & V_n < V_{\text{out}} \\ 0 & V_n = V_{\text{out}} \\ \gamma_n & V_n > V_{\text{out}} \end{cases} \quad (5)$$

In this array, as in the resistor array, Kirchhoff's current law mandates that the sum of the currents into the V_{out} node must equal zero. This fact, combined with Equation 5, leads to the following equality:

$$\sum_{V_n < V_{\text{out}}} \gamma_n = \sum_{V_n > V_{\text{out}}} \gamma_n \quad (6)$$

If all the weights (γ_n) are equal, Equation 6 calculates the median of the input voltages. When they are not all equal, this computation defines a "weighted median."

2.3 Encoding Position with Voltages

In Sections 2.1 and 2.2, the input voltages to the parallel dividers were independent of one another. These voltages, however, can be used to encode the position of each element in the array. This encoding can be performed through the use of a linear resistive divider as a voltage divider. In this configuration, N resistors are connected in series. By applying voltages V_0 and V_N to the ends of the resistive line, a linear voltage gradient can be set up along the array, where the voltage at the n th node is given by

$$V_n = \left(\frac{V_N - V_0}{N} \right) n + V_0 \quad (7)$$

3 Implementation

In this section, we shall present the circuitry needed to implement the tracking arrays. We shall describe the transconductance amplifiers and phototransistors used in this system, and show how these elements can be used to implement both one- and two-dimensional tracking arrays.

3.1 The Photoreceptor

In this design, the vertical parasitic bipolar transistor existing in the standard CMOS fabrication process is used as a phototransistor.[1] Assuming an n -well process, this bipolar transistor is a pnp with a grounded collector.

In a phototransistor of this type, at a given wavelength of light,

$$i_{\text{photo}} \propto I \quad (8)$$

where i_{photo} is the emitter current of the phototransistor and I is the light intensity at the base-emitter junction.

3.2 The Transconductance Amplifier

The transconductance amplifier (Figure 3) is a differential stage producing an output current which is a function of the bias current (set by V_{bias}) and the differential input voltage ($V_1 - V_2$).[6] The circuit consists of a bias transistor Q_{bias} that sets the maximum output current of the amplifier, a differential pair (Q_1 and Q_2), and a current mirror (Q_3 and Q_4).¹ The bias current is divided by the differential pair according to input voltages V_1 and V_2 . The current that flows through the “minus” transistor Q_2 is subtracted directly from the output node. In the other branch, the current flowing through the “plus” transistor Q_1 also must pass through Q_3 . The current through Q_4 is the mirror of the current through Q_3 . Finally, the current through Q_4 is added to the output node. Thus, the output current is the difference between the currents in the two branches of the differential pair.

A transfer curve for the transconductance amplifier is shown in Figure 4. Saturation behavior is clearly exhibited for large input voltage differentials, indicating that the bias current is flowing through only one of the transistors in the differential pair. This amplifier often is used with very small bias currents, allowing all transistors to be operating in the subthreshold regime.[2] The transfer function when the

¹This configuration is called a current mirror because the current through the input transistor is “mirrored” by the output transistor. The gate voltage of the input transistor is set by the current applied to the device. This voltage also is applied to the output transistor gate, causing the current through the two transistors to be equivalent (ignoring drain-source voltage effects).

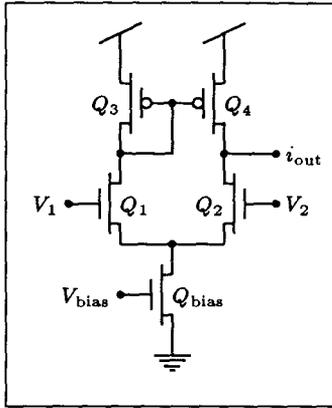


Figure 3 The transconductance amplifier. This circuit is a differential amplifier whose output current is a function of its bias current (the current through Q_{bias}) and differential input voltage. The bias current is set by V_{bias} and divided between Q_1 and Q_2 depending on V_1 and V_2 . Q_3 and Q_4 mirror the current through Q_1 giving an output current equal to the difference of the currents through Q_1 and Q_2 .

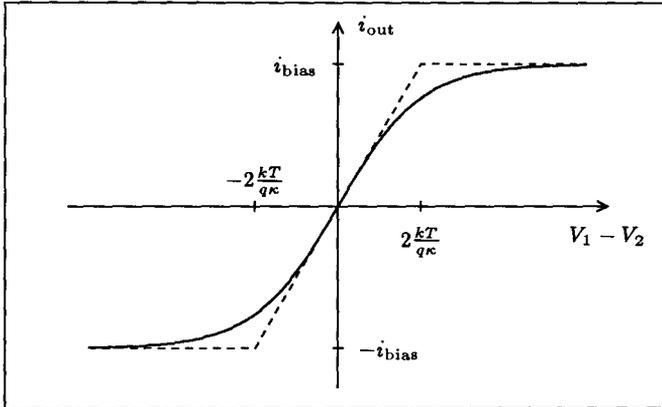


Figure 4 Transconductance amplifier transfer characteristics. The solid line is the actual transfer function for the amplifier. The dotted line is the linear approximation for the amplifier transfer curve.

bias current is subthreshold is

$$i_{\text{out}} = i_{\text{bias}} \tanh \left(\frac{V_1 - V_2}{2 \frac{kT}{q\kappa}} \right) \quad (9)$$

For the purpose of calculations in this paper, the tanh function will be approximated by the piecewise linear function described by

$$i_{\text{out}} = \begin{cases} -i_{\text{bias}} & V_1 - V_2 \leq -2 \frac{kT}{q\kappa} \\ g(V_1 - V_2) & |V_1 - V_2| < 2 \frac{kT}{q\kappa} \\ i_{\text{bias}} & V_1 - V_2 \geq 2 \frac{kT}{q\kappa} \end{cases} \quad (10)$$

where $g = i_{\text{bias}} / (2 \frac{kT}{q\kappa})$ is the transconductance of the amplifier. This approximate transfer function is superimposed in Figure 4.

3.3 A One-Dimensional Amplifier Array

Figure 5 shows a variation on the standard transconductance amplifier in which the bias transistor has been replaced with a phototransistor. Because the phototransistor produces a current (i_{photo}) that is proportional to the light intensity (I), the transfer function of this circuit becomes

$$i_{\text{out}} = i_{\text{photo}} \tanh(V_1 - V_2) \propto I \tanh(V_1 - V_2) \quad (11)$$

with voltages given in units of $2 \frac{kT}{q\kappa}$.

Figure 6 shows an array of transconductance amplifiers and phototransistors configured similarly to the arrays of resistors and threshold elements in Figures 1 and 2. The input voltages to the amplifiers are obtained from the resistive line, as described in Section 2.3.

We can set the voltage gradient along the amplifier inputs by changing V_0 and V_N . The output voltage (V_{out}) is a weighted combination of the amplifier input voltages, and is constrained to lie within the input voltage range. Thus, the maximum possible differential voltage at the input of any amplifier is $|V_N - V_0|$. If $|V_N - V_0| \leq 2 \frac{kT}{q\kappa}$, all the amplifiers will be operating in their "linear" regimes, and the network will calculate the mean of the input voltages of the amplifiers weighted by the corresponding light intensities:

$$V_{\text{out}} = \frac{\sum i_{\text{photo}_n} V_n}{\sum i_{\text{photo}_n}} = \frac{\sum I_n V_n}{\sum I_n} \quad (12)$$

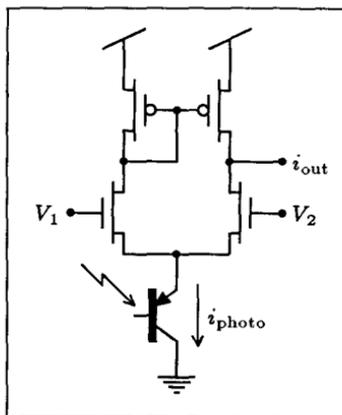


Figure 5 Transconductance amplifier modified to include phototransistor. In this circuit, the bias transistor of the amplifier is replaced with a phototransistor (a vertical parasitic bipolar transistor) resulting in an amplifier bias current that is equivalent to the photocurrent.

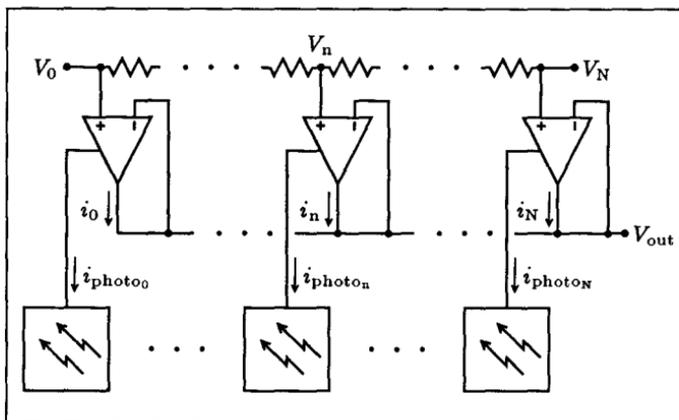


Figure 6 Transconductance amplifier array with position encoding. In this network, amplifiers with photocurrents used as their bias currents are configured as voltage followers with their outputs connected together. A resistive voltage divider encodes the position and these voltages are used as the input voltages to the amplifiers. By changing the differential input voltage ($V_N - V_0$) the network can perform the range of computations from weighted mean to weighted median.

Because the input voltages encode the positions of the receptors, this calculation is equivalent to computing the center of light intensity. Thus, if the image projected onto the chip is a bright spot of light against a dark background, the array will calculate the position of the spot, and thus, track this stimulus.

If $|V_N - V_0| \gg \frac{2kT}{q\kappa}$, the gain of each amplifier with respect to distance along the array will be very large. The transconductance amplifiers (except for those with input voltages very near the output voltage) then can be approximated as threshold elements, and the network will approximate the weighted median calculation. In this configuration, the circuit can also track a high-intensity stimulus. Because the amplifier output magnitudes are independent of the positions in the array, however, the receptors further from the bright spot will contribute less to the output than they would in the weighted mean computation. Thus, a higher differential voltage along the resistive divider leads to a more accurate bright-spot localization.

If $|V_N - V_0|$ falls between these two regimes, the calculation will be a combination of the two computation modes. In particular, the section of the array with input voltages near the output voltage (near the center of intensity) will compute a weighted mean, whereas the parts distant from the center of intensity will compute a weighted median.

3.4 A Two-Dimensional Tracking Array

The one-dimensional array of photoreceptors and amplifiers in Section 3.3 can be expanded to two dimensions. Because the two dimensions of the tracking calculations are independent, the value of the intensity at any point along each axis is taken to be the sum of the currents from the receptors in a line perpendicular to that axis. The two-dimensional array can be implemented using dual-output photoreceptors (implemented with current mirrors) and current summing along each row and column (Figure 7).

This configuration has a problem, however; the layout area of a photoreceptor increases by more than a factor of two when the current mirror is added, reducing the resolution and the coverage factor of the array. For this reason, a second scheme of photoreceptor placement using single-output receptors can be used to tile the plane. In this array configuration, each photoreceptor contributes its current

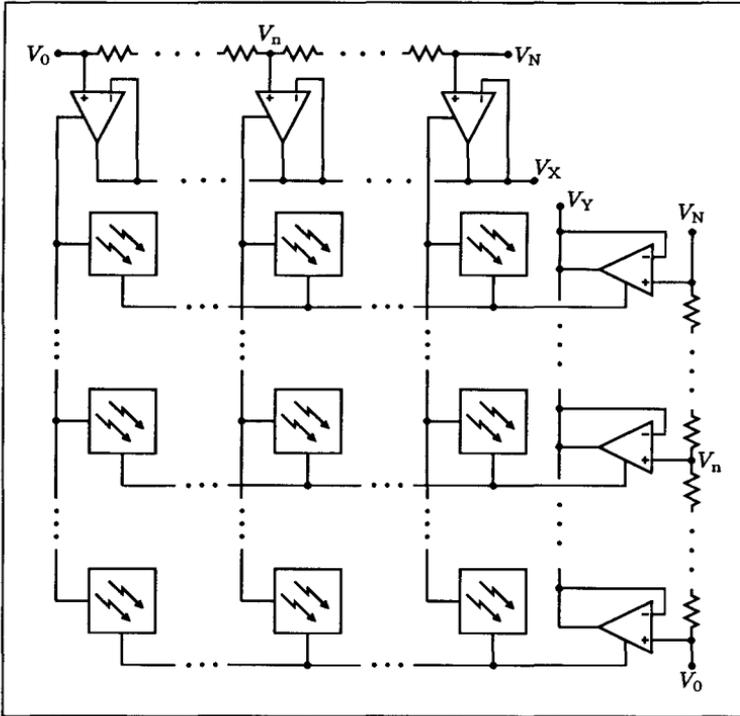


Figure 7 A two-dimensional amplifier array. The photocurrents from the receptors are replicated and summed onto wires running parallel to each axis. These current sums are used as the bias current for the amplifier arrays along the edges of the array. This edge circuitry performs the weighted mean/median computations.

to only one axis. The receptors are spatially alternated so that the currents from adjacent receptors are added to opposing axes. Using this scheme only one-half of the receptors contribute current to each axis, but the effective density of receptors is increased due to the unit size reduction made possible by removing the current mirror.

4 Experimental Set-up and Results

We chose to fabricate the alternating receptor array because of its increased receptor density and resolution. We fabricated the chip with

a 200×200 -pixel array in 2μ CMOS. We implemented the resistive dividers using polysilicon wires running the length of each of the two edges of the chip. The resistance of these lines was approximately $7 k\Omega$.

We created our experimental setup by removing the light source from a standard photographic enlarger and mounting a tracking chip in its place. We mounted a light-emitting diode (LED) on a two-dimensional precision motion table (accurate to $0.2 \mu\text{m}$) placed below the enlarger, and focused its image through the enlarger lens onto the chip. We could then use the LED as a point stimulus by moving its image across the face of the chip by moving the table. We placed a uniformly reflective background on the motion table surrounding the LED, and created a uniform background intensity by projecting light onto the background from a diffuse light source mounted above the table.

4.1 Position/Output Voltage Relationship

We investigated the output response of the chip by moving the LED parallel to one axis while holding its position constant with respect to the other axis (Figure 8). We set the input voltages (the voltage at the ends of the polysilicon wires) to 2.2 and 3.8 V. As the stimulus passed across the surface of the chip, the output voltage changed as a function of the position. At the ends of the sweep (as the image moved off the edge of the chip), the output voltage leveled off and then began to move rapidly toward the value obtained for a uniformly illuminated background.

4.2 Multiple Curves and Error

In this experiment, we took a family of curves by moving the LED along one axis of the chip at five different positions (separated by 20 pixels) on the perpendicular axis. We performed a least-squares regression on the data points in the -70 to $+70$ pixel range, and the relative error for each curve was calculated by subtracting the regressed values from the observed values and dividing by the input differential voltage. Also, the maximum deviation (defined as the largest difference between any two curves) among these curves was calculated. Both the relative error curves and the maximum deviation curve are shown in Figure 9.

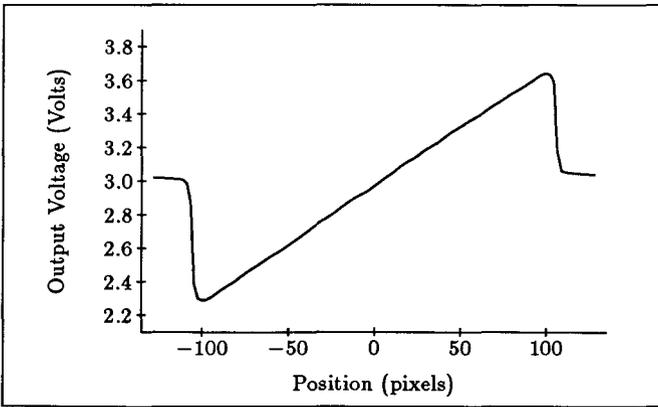


Figure 8 Single axis sweep. This figure shows the output voltage as a function of the position of the image of an LED on the chip. The curve represents a sweep of the LED through one axis while holding its position constant with respect to the other axis. As the image of the LED moves across the receptor surface, the output voltage linearly increases. As this stimulus passes off the edge of the surface, the output voltage returns to the uniform-intensity output value.

The errors appear to be quite systematic, suggesting that they were caused to a greater extent by the computational circuitry along the edges of the array than by mismatched photoreceptors. They could be a result of a nonhomogeneous resistance along the polysilicon wires of the voltage divider, or of offsets in the amplifiers due to parametric variation.

Errors due to parametric variation are very dependent upon the width of the stimulus. For a given width, the output error is the average of the errors for the individual components within the boundaries of the stimulus. Since these errors are statistically distributed around zero, the average error will decrease as the sample size becomes larger. Thus, the accuracy of the chip can be improved by increasing the stimulus width. The width of the stimulus for this experiment was approximately 6 pixels.

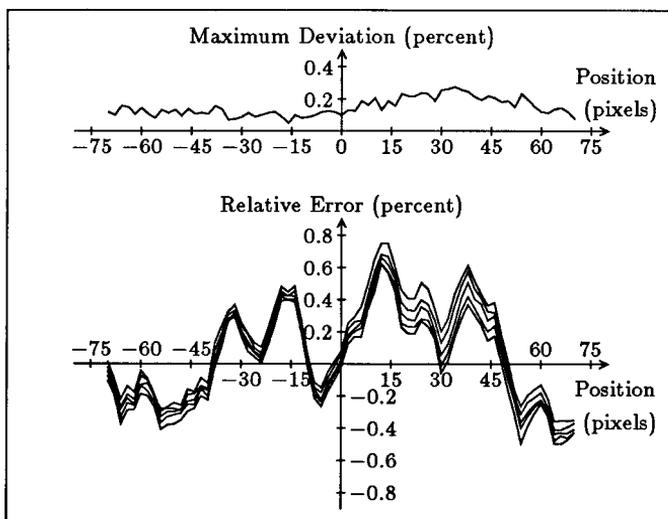


Figure 9 Tracking-array relative error. This figure displays the error for the central region of five curves taken by sweeping the LED along one axis at five different positions on the other axis. The bottom curves show the relative error from the best-fit line for the set of curves. The upper curve shows the maximum deviation between any two curves in the set.

4.3 Varying Differential Input Voltage

We tested the effects of moving from a weighted mean to a weighted median by taking a set of curves with varied differential input voltages applied across the resistive wire. These curves were all taken while holding the stimulus-to-background intensity ratio constant at approximately 1. We varied the differential input voltages from 100 mV to 1.6 V by factors of two while keeping the common mode voltage (and thus the uniform intensity output voltage) constant at 3 V. As in the previous experiments, the width of the stimulus was approximately 6 pixels.

In the appendix, we calculated the input-output voltage relationships of the weighted mean and weighted median functions for a given stimulus width and stimulus-to-background intensity ratio. Figure 10 shows the best-fit lines (using linear regression) to the curves for this

experiment with the output voltages scaled to the number of pixels. We have also plotted the theoretical lines for the weighted mean and weighted median for this experimental set-up. As predicted, the experimental curve for a differential input voltage of 100 mV is closest in slope to the theoretical curve for the weighted mean. The slopes of the experimental curves increase as the differential voltage increases, coming closest to the theoretical weighted median curve at a differential input voltage of 1.6 V.

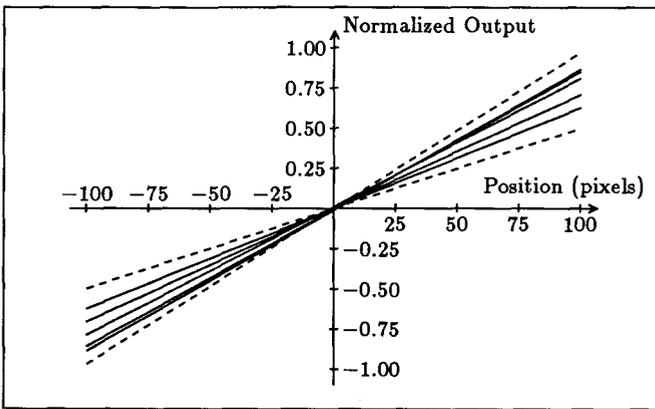


Figure 10 Varying the differential input voltage. The solid lines are the best-fit lines for a family of curves taken by varying the differential input voltage to the chip. The dotted lines are the theoretical predictions for the weighted mean and median computations. As predicted, the slopes of the experimental curves vary from the slope of the mean to the slope of the median as the amplifiers move from resistive elements to high-gain current sources.

5 Conclusions

In this paper we have presented a system that integrates both photo-sensors and processing elements on a single die to calculate the center of intensity of a visual image, or to track a bright spot on a dark background. This circuitry was implemented with a standard digital CMOS process.

For multiple sweeps along an axis, we have shown that the maximum error from a best-fit line is less than 0.8 percent, and the maximum difference among these sweeps (between any two curves) is less than 0.3 percent. In addition, the repeatability has been measured to be better than 0.05 percent, and the monotonicity (differential linearity) has been measured to be better than 0.1 percent. These facts suggest ways to increase the accuracy of the chip. If we treat the chip as a linear tracking system, we are limited by the deviation from the best-fit line. However, by creating a calibration curve for each axis, output voltages can be corrected to the accuracy of the maximum deviation among the sweeps along that axis. Finally, by creating a two-dimensional calibration map of the chip, output voltages can be corrected to within a resolution limited by the monotonicity and an accuracy equal to the repeatability. Through this scheme, it should be possible to greatly improve the accuracy of the system.

We also showed that the functionality of the system, which ranges from a weighted mean to a weighted median, can be changed by varying the differential input voltage that encodes the positions of the photoreceptors.

We are currently working on similar arrays to perform single edge detection using spatial derivatives coupled with the circuitry to compute the weighted mean/median described in this paper. We are also incorporating the tracking chip into sensory-motor feedback servo systems that will actively track objects in the visual field.

Appendix — Tracking Theory

Networks that compute weighted means and medians can be applied to calculate the position of any high-intensity stimulus on a low-intensity background. This tracking is accomplished by encoding the intensity of the stimulus and background as the weights of the computation, and the position of the elements as the values being weighted.

In order to simplify the computation of these functions, we assume that the stimulus and background intensities are constant and uniform over their areas. We define the variables α and β to be the total stimulus and background intensities, respectively, and assume that in the region of the stimulus, the measured intensity is the sum of the two. We scale the input and output voltages to a range of N

(the number of pixels) centered at zero so that the index step for the summations is 1. We define the stimulus to be of width w pixels centered at the input voltage. Therefore, the value of the background intensity at every pixel is $\frac{\beta}{N}$, the value of the stimulus intensity at each pixel in its range is $\frac{\alpha}{w}$, and the sum of all the intensities is $\alpha + \beta$.

We can calculate the tracking function for the weighted mean by separating the summations into two parts (one for stimulus and one for background) as follows:

$$\begin{aligned}
 V_{\text{out}} &= \frac{1}{\sum I_n} (\sum I_n V_n) \\
 &= \frac{1}{\alpha + \beta} \left(\sum_{V=V_{\text{in}}-\frac{w-1}{2}}^{V_{\text{in}}+\frac{w-1}{2}} \frac{\alpha}{w} V + \sum_{V=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{\beta}{N} V \right) \\
 &= \frac{1}{\alpha + \beta} \left(\frac{\alpha}{w} \sum_{V=V_{\text{in}}-\frac{w-1}{2}}^{V_{\text{in}}+\frac{w-1}{2}} V + \frac{\beta}{N} \sum_{V=-\frac{N-1}{2}}^{\frac{N-1}{2}} V \right) \\
 &= \frac{1}{\alpha + \beta} \left(\frac{\alpha}{w} (V_{\text{in}} w) + \frac{\beta}{N} (0) \right) \\
 &= \frac{\alpha}{\alpha + \beta} V_{\text{in}}
 \end{aligned}$$

We compute the tracking function for the weighted median by again dividing the summations into two parts. We must further make the assumption that when the center of the stimulus is between pixels the network will linearly interpolate. (This is acceptable because the stimulus will then fall partially on a pixel at either end.) This leads to the following two equations for the sum of the intensities on either side of the output voltage:

$$\begin{aligned}
 \sum_{V_n < V_{\text{out}}} I_n &= \frac{\beta}{N} \left[V_{\text{out}} - \left(-\frac{N-1}{2} \right) \right] + \frac{\alpha}{w} \left[V_{\text{out}} - \left(V_{\text{in}} - \frac{w-1}{2} \right) \right] \\
 \sum_{V_n > V_{\text{out}}} I_n &= \frac{\beta}{N} \left[\left(\frac{N-1}{2} \right) - V_{\text{out}} \right] + \frac{\alpha}{w} \left[\left(V_{\text{in}} + \frac{w-1}{2} \right) - V_{\text{out}} \right]
 \end{aligned}$$

Kirchhoff's current law mandates that these two sums must be equal. By combining and rearranging the equations, we get V_{out} in terms of V_{in} as follows:

$$V_{\text{out}} = \frac{N\alpha}{N\alpha + w\beta} V_{\text{in}}$$

Both of the above models are only accurate when the entire stimulus falls upon the computational surface. In addition, due to the summations used for the stimulus terms in the median computation, that computation is only valid as long as the output voltage falls within the input voltage range covered by the stimulus.

As the stimulus-to-background intensity ratio increases, the effectiveness of both computations improves (i.e., the ratio of V_{out} to V_{in} increases). As the stimulus gets narrower, the effectiveness of the median computation improves. Finally, since, by definition, the stimulus is smaller than the background, the weighted median computation is more effective at tracking than is the weighted mean.

Acknowledgements

This research was supported by grants from the System Development Foundation and the Office of Naval Research, and a hardware grant from the Hewlett-Packard Corporation. The authors are grateful to Michael Emerling, Massimo Sivilotti, and Mary Ann Maher for their careful reviews of this document. We are also indebted to Lars Nielson, Richard Lyon, and John Tanner for their many helpful discussions.

References

- [1] C. A. Mead. A sensitive electronic photoreceptor. In *1985 Chapel Hill Conference on Very Large Scale Integration*, pages 463–471, Computer Science Press, Rockville, Maryland, 1985.
- [2] C. A. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley (in preparation).
- [4] M. A. Sivilotti, M. A. Mahowald, and C. A. Mead. Real-time visual computation using analog CMOS processing arrays. In *1987 Stanford Conference on Very Large Scale Integration*, pages 295–312, MIT Press, Cambridge, MA, 1987.
- [5] J. E. Tanner. *Integrated Optical Motion Detection*. PhD thesis, California Institute of Technology, 1986. 5223:TR:86.
- [6] E. A. Vittoz. Micropower techniques. In Y. Tsvividis and P. Antognetti, editors, *Design of MOS VLSI Circuits for Telecommunications*, pages 104–144, Prentice-Hall, Englewood Cliffs, NJ, 1985.