

Object-Based Selection Within an Analog VLSI Visual Attention System

Tonia G. Morris, Timothy K. Horiuchi, and Stephen P. DeWeerth

Abstract—An object-based analog very large-scale integration (VLSI) model of selective attentional processing has been implemented using a standard 2.0- μm CMOS process. This chip extends previous work on modeling a saliency-map-based selection and scanning mechanism to incorporate the ability to group pixels into objects. This grouping, or segmentation, couples the circuitry of the object's pixels to act as a single, larger pixel. The grouping of pixels is dynamic, driven solely by the segmentation criterion at the input. In this demonstration circuit, image intensity has been chosen for the input saliency map and the segmentation is based on spatial low-pass filtering followed by an intensity threshold. We present experimental results from a one-dimensional implementation of the object-based analog VLSI selective-attention system.

Index Terms—Focal-plane processing, neuromorphic analog VLSI, object segmentation, subthreshold circuits, visual attention, winner-take-all.

I. INTRODUCTION

A PRIMARY obstacle to solving visual processing problems in real time is the vast amount of information in a given scene. To fully process all parts of an image in parallel, a large amount of processing circuitry and wiring is needed. In both engineering and biological systems, such computational resources are rarely available and are costly in terms of power, space, and reliability. Most tasks performed by visual processing systems do not require information from all parts of the visual field, however, and thus much of the information processing problem can be handled by subdividing the image data in both space and time. Biological vision systems serve as excellent examples of this type of processing strategy. The varying density of photoreceptors on the retina is one simple example of how some biological systems strategically focus their processing resources while maintaining coverage of the full scene. Selective visual attention is another example where extensive processing is performed on subregions of an image [1].

We have developed a system framework for selective attention processing in analog CMOS focal-plane processing

systems. Previous implementations based upon this framework have included pixel-based processing arrays that perform a winner-take-all computation with excitatory and inhibitory feedback [2], [3]. The input to the winner-take-all computation is an array of values that represent levels of interest across the visual field. These levels of interest, or saliency values [4], are task-dependent and can be computed by combining several feature measures such as spatial derivatives, temporal derivatives, motion, and orientation selectivity. The winner-take-all computation selects a single region of interest that defines the spotlight of attention for further processing, thus enabling the visual processing system to perform complex processing on only a small region of the visual field. The excitatory feedback provides a mechanism for hysteresis, or persistence, in the selection. The inhibitory feedback induces shifts of attention even when the input levels do not change (i.e., the visual scene is static.)

From a computational perspective, it is desirable for the selected region to correspond to various object sizes within the visual field. Since the visual scene is constructed from objects, and not single points, the selective-attention processing should perform an object-based computation, as opposed to a pixel-based computation. This necessary shift to object-based processing has motivated the circuits we present in this paper. Our previously published selective-attention circuits performed all operations within a pixel-based processing paradigm [2], [3]. We have now extended the computations of the selective-attention framework such that the processing can be performed on contiguous groups of pixels. To identify these contiguous groups of pixels within the visual field, a segmentation computation is necessary for distinguishing objects from one another and from the background. The circuits we present in this paper address the necessary paradigm shift to object-based processing within selective-attention analog very large-scale integration (VLSI) systems. In addition, we present circuits for segmenting images on the focal plane.

Local pixel-based processing is commonly used in the focal-plane processing arrays that have been developed in analog VLSI [5]–[8]. Global image-based systems have also been implemented using the same collective architecture to compute a single measure for the entire visual field [9]–[11]. Bridging the gap between these two forms of processing, a dynamic wires approach has been implemented that separates regions of the visual field into multiple-pixel sections [12]. The dynamic wires implementation enables the application of global processing to smaller regions of the visual field. Some of the techniques introduced by the dynamic wires approach

Manuscript received November 5, 1996; revised April 10, 1998. This paper was recommended by Associate Editor T. Fiez.

T. G. Morris was with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA. She is now with Intel Corporation, Chandler, AZ 85226-3699 USA.

T. K. Horiuchi was with Computation and Neural Systems, California Institute of Technology, Pasadena, CA, 91125 USA. He is now with the Zanvyl Krieger Mind/Brain Institute, The Johns Hopkins University, Baltimore, MD 21218 USA.

S. P. DeWeerth is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

Publisher Item Identifier S 1057-7130(98)08503-6.

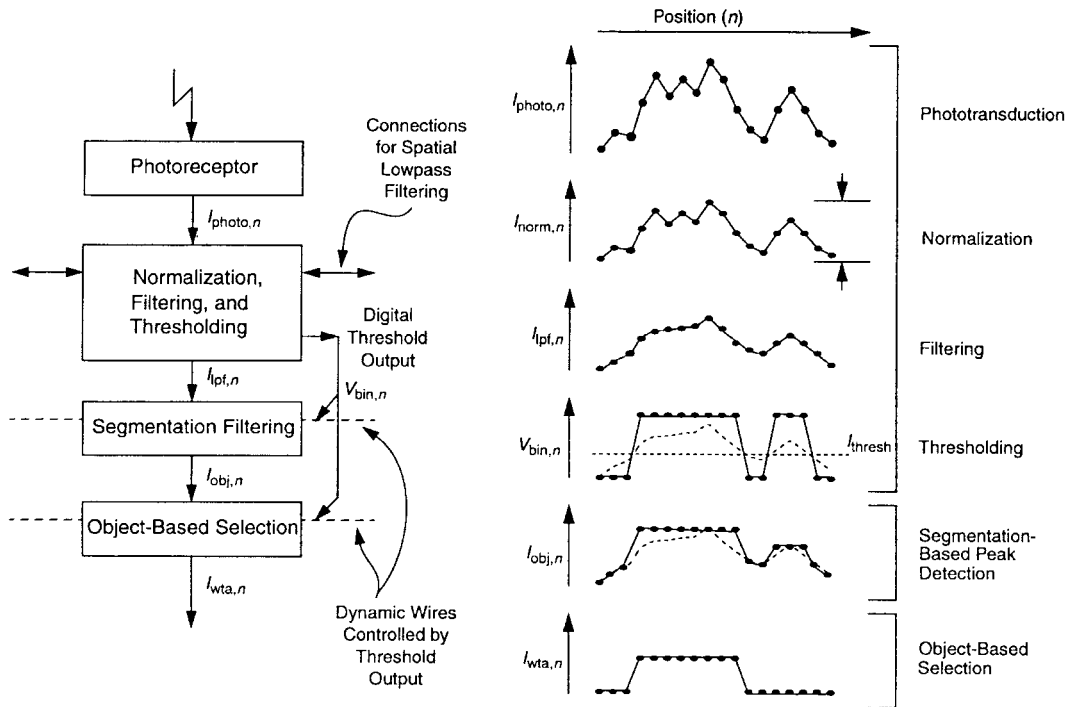


Fig. 1. Description of the object-based selection system architecture and operation. A single processing element (pixel) is shown in (a). The input to each computation is an analog current. The computations include normalization, filtering, thresholding, segmentation, and object-based selection. Communication to the nearest neighboring elements is necessary for the spatial low-pass filtering, the segmentation, and the object-based selection. The digital output from the thresholding operation is used to control the communication (dynamic wires) for the segmentation and object-based selection operations. The effect of each stage of processing on a hypothetical one-dimensional array of inputs is shown in (b).

can be applied to selective-attention processing circuits. The circuits we present utilize these dynamic wires in a unique way to implement variable-granularity selection processing circuits.

Section II describes the system that was implemented to test the object-based selective attention circuits. Section III describes the circuits that perform the segmentation and object-based selection. Section IV describes the performance of these circuits with the presentation of experimental results.

II. SYSTEM ARCHITECTURE

The design of analog systems demonstrates many advantages over digital systems in terms of power consumption and silicon area. One of the biggest disadvantages, however, is the complexity of the design and testing. The implementation we present focuses on the circuit-level issues by including only a few processing elements, which facilitates circuit characterization. The design of each processing element is such that it can be incorporated into a large two-dimensional array [13]. To concentrate on the lower-level issues, the system we discuss in this paper is a one-dimensional array of 20 object-based selective-attention processing elements. The input to the array of processing elements represents the saliency map, which is a scalar encoding of interest values across the visual field. In the implementation presented here, we use intensity levels to signify saliency. Other features used in previous work have included spatial and temporal derivatives [14].

A diagram demonstrating the organization of the processing elements and the function of each processing stage is shown in Fig. 1. These processing stages include: 1) phototransduction;

2) normalization; 3) spatial low-pass filtering; 4) thresholding; 5) segmentation-based filtering; and 6) object-based selection. The photocurrents are normalized using a linear normalization circuit and then passed through a spatial low-pass filter. The low-pass filter is used to suppress outliers in the input and emphasize larger contiguous regions. After filtering, the signal is compared against a globally set threshold, above which pixels qualify as object pixels. The thresholded output is binary and is passed to subsequent stages of processing to control the dynamic connections to neighboring processing elements. The segmentation-based filtering uses these dynamic connections to determine the peak current within each object and replicate that current value in every pixel within the object, thus defining an object-based saliency measure. In the final stage of processing, the object-based selection circuit also uses dynamic connections to neighboring processing elements. The dynamic connections couple specific nodes in the pixel-based selection circuits such that the collection of processing elements included in an object acts as a single selection processing element. In this way, all of the pixels within an object act together and compete as a single unit.

III. CIRCUIT DESCRIPTIONS

Each processing element contains analog, current-mode, subthreshold circuits that perform the object-based attentive selection processing. We partitioned the functions of the processing elements into three circuits: 1) the normalization, filtering, and thresholding circuit; 2) the segmentation-based filtering circuit; and 3) the object-based selection circuit. The

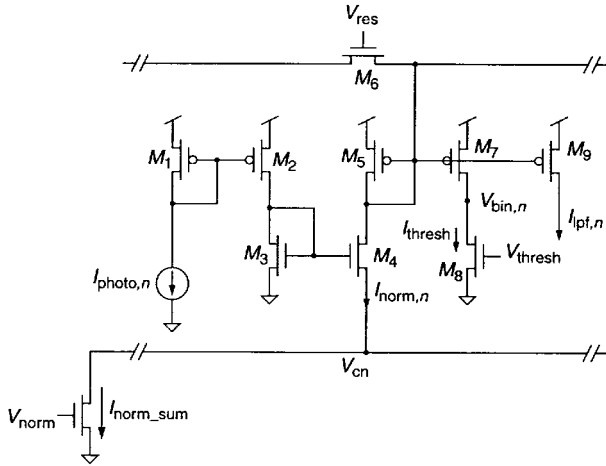


Fig. 2. Normalization, filtering, and thresholding circuit. The linear normalization computation is performed through the combination of the M_3 and M_4 transistors and a global current set by the value of V_{norm} . The filtering is implemented by a current-mode resistive network composed of transistors M_5 and M_6 . A high-gain stage is used to compare the output of the filter to a constant threshold value, which is controlled by the V_{thresh} voltage.

descriptions for each of these circuits assumes the following model of the subthreshold current–voltage relationship for the MOSFET devices operating in saturation

$$I_d = I_0 e^{(\kappa V_{gb} - V_{sb})/V_T} \tag{1}$$

I_d is the drain current, I_0 is the leakage current, V_{gb} and V_{sb} are the gate and source voltages referenced to the bulk potential, κ is the gate efficiency factor, and V_T is the thermal voltage. The transistors in these circuits were all implemented with W/L aspect ratios of one, with the channel lengths equal to 6 μm . We used larger gate lengths than the minimum for the process to avoid appreciable channel-length modulation effects. Thus, channel-length modulation is not included in the equations for analysis.

A. Normalization, Filtering, and Thresholding

The circuit used for the normalization, filtering, and thresholding is shown in Fig. 2. The phototransistor current serves as the input to the normalization computation. The normalization is necessary to ensure that later stages of processing receive current levels within a set range of subthreshold values. Transistors M_3 and M_4 within each pixel are used to compute the normalized value [16]. A single transistor on the end of the array sets the sum of the normalized output currents I_{norm_sum} via the bias voltage V_{norm} . The relationship between the photocurrents and the normalized currents is

$$I_{norm_sum} = \sum_n I_{norm,n} = \sum_n I_{photo,n} \times e^{-V_{cn}/V_T} \tag{2}$$

V_{cn} is the voltage at the common node in the normalization circuit, which is the source node for all the parallel output transistors. The voltage on this node is set such that the normalization criteria of having a constant sum of output currents is met. The equation represents the ideal case where I_0 and κ for the transistor pairs are perfectly matched.

The output currents of the normalization feed into a current-mode resistive network [8] that performs the spatial low-pass

filtering. Transistor M_6 implements the lateral resistance of the resistive network. The resistive network can best be described according to its response to a single input current within the array, also known as its point-spread function (PSF). The PSF for this current-mode resistive network is approximated as

$$I_{lpf,n} = I_{lpf,n_0} e^{-|n-n_0|/L} \tag{3}$$

where L is the characteristic length for the spatial filter and n_0 is the location of the single input current. The characteristic length of the filter is controlled by the value of the gate voltage V_{res} on transistor M_6 . We can approximate this relationship in a closed-form solution by making an assumption that the gate efficiency factor for the p-type MOSFET's is close to one. The assumption is not very accurate, but allows us to gain intuition as to how the different voltages control the characteristic length of the filter.

$$\frac{1}{L} = \ln \left(\frac{I_{0p} e^{V_{dd} - \kappa_n V_{res}} / V_T + 1}{I_{0n}} \right) \tag{4}$$

(We have distinguished the leakage currents and gate efficiency factors for the p-type and n-type transistors by adding another level of subscripting.) As is demonstrated by the relationship in (4), an increase in the V_{res} value causes an increase in the characteristic length L of the filter. The peak value of the PSF can also be approximated by using the same set of assumptions

$$I_{lpf,n_0} = \frac{I_{norm,n_0}}{1 + \frac{2I_{0p}}{I_{0n}} e^{\kappa_n V_{res} - V_{dd}/V_T}} \tag{5}$$

As is evident from this relation, an increase in the V_{res} value causes the peak output for the PSF to decrease.

Two copies of the low-pass-filtered output current $I_{lpf,n}$ are mirrored via transistors M_7 and M_9 . Transistor M_7 is combined with M_8 to create a high-gain comparator stage for the thresholding operation. The threshold value I_{thresh} is set by the global voltage V_{thresh} . The second copy of the low-pass-filtered output current $I_{lpf,n}$ and the binary threshold output voltage $V_{bin,n}$ are sent as inputs to the next stage of processing. The combination of filtering and thresholding directly affects the spatial extent and peak saliency of the objects. An increase in the characteristic length for the resistive network increases the extent of the point-spread function of the low-pass filter, thus emphasizing objects of larger spatial extent.

There are many circuit nonidealities that can cause variations among the processing elements. Mismatch in the I_0 and κ values within the normalization transistors could cause inputs to switch their ordering in terms of relative magnitudes. The thresholding circuit can also cause slightly unpredictable behavior due to mismatches in the locally generated threshold current; one possible outcome could be the segmentation of an object into two objects. The low-pass filtering alleviates both of these problems by ensuring that single pixel values within an object do not vary significantly from one location to the next. The finite gain of the thresholding stage can impact subsequent stages of processing by failing to produce a strong binary output signal. These effects will be discussed in Sections III-B and III-C.

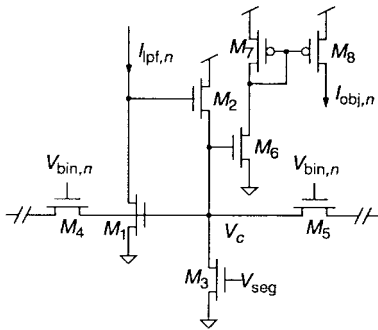


Fig. 3. Segmentation-based filtering circuit. Switches for the dynamic wires are implemented by the M_4 and M_5 transistors. The winner-take-all circuit is used to compute the maximum value among the saliency value inputs in the object. The communication (dynamic wires) of the global winner-take-all common line is controlled by the output of the thresholding circuit. The maximum value is duplicated at every pixel within the object.

B. Segmentation-Based Filtering

The segmentation-based filtering implemented in this system detects the peak input value within each object and replicates that value as the output of all the pixels. In this way, a single value (the object’s peak value) is specified as the object’s saliency value. All pixels below the threshold are isolated from their neighbors and operate as single-pixel objects. These single-pixel objects simply pass their input values as their output values. The use of a peak value as the object’s saliency value is only one of several options. Further discussion of the merits of various object saliency measures can be found in [15].

The circuit used to implement the peak-detection segmentation-based filtering is shown in Fig. 3. In this circuit the binary threshold output voltage $V_{bin,n}$ is used to couple nodes of the pixel’s circuit to the corresponding nodes of its neighboring pixel’s circuit. If a given pixel is part of an object (i.e., its filtered input exceeds the threshold described in Section III-A), the node couples with its neighbors on either side only if each of the neighbors is also part of an object. This adaptively controlled coupling is an example of the dynamic wires approach [12]. The coupling is performed by the transistors M_4 and M_5 , each of which performs one half of the logical AND operation with its neighbor pixel to make a composite switch. When the AND condition is true, both transistors conduct and the dynamic wire is formed.

The winner-take-all subcircuit (transistors M_1 and M_2) [17] detects the peak input current among the object’s pixels via the communication along the effective V_c node within each object. Only those winner-take-all elements that share the same common node compete against one another to determine the peak value. The bias current for the winner-take-all (generated by M_3) must be included in each processing element due to the dynamic nature of the connections. The input transistor M_1 for all elements in the object shares the same gate voltage V_c , aside from any voltage drops across the switches. The feedback via the M_2 transistor sets the V_c value and the gate voltages for the M_2 transistors such that all input currents within the object are matched by the currents through their respective input transistors. The sum of the output currents going through the M_2 transistors must also equal the bias currents sourced

by the M_3 transistors. The input transistor at the peak input location within each object operates in the saturation region, while all other input transistors are pushed into the ohmic region. Thus, V_c encodes the maximum input current, which is regenerated by transistor M_6 at each pixel. For the case when the pixel is not part of an object (i.e., below threshold), the original input current is replicated at the output. The single winner-take-all element operates as a simple current mirror.

The winner-take-all computation is highly nonlinear. Thus, small mismatches in the input transistors can cause the winning location to become unpredictable when two input currents are very close in value. Due to the duplication of the peak value to all locations within the object, these mismatches are not critical. They would only cause a small difference in the assignment of the object’s saliency value. A nonbinary threshold output at any pixel’s location within an object could also cause a small change in the circuit’s intended mode of operation. If the dynamic wires begin to have a significant voltage drop across the switches, the winner-take-all begins to operate as a local winner-take-all circuit [17]. The effect is that the common node would no longer encode the same current for all pixels in the object; however, the input to the selection circuit averages these values, thus minimizing the effect.

C. Object-Based Selection

The basic winner-take-all circuit [17] used to compute the peak-saliency in the segmentation-based filtering circuit is the same compact circuit that we use in the object-based selection circuit. In contrast to the previous circuit, the dynamic wires are not used to isolate clusters of parallel winner-take-all computations. For the object-based selection, the dynamic wires are instead used to “grow” each winner-take-all element to the size of the object. The winner-take-all selection encompasses the full extent of the array, but the number of winner-take-all elements changes according to the number of objects.

The object-based selection circuit is shown in Fig. 4. The transistors M_1 and M_2 compose the winner-take-all stage for a single pixel. When the binary segmentation voltage $V_{bin,n}$ is high, the switches implemented by M_3 and M_4 close, creating a connection to neighboring pixels at the input node. The effective shorting of this node to its neighbors causes the input transistors M_1 of each pixel to be connected in parallel. Thus, the input transistors within an object operate as a single transistor with a larger aspect ratio. The same result occurs with the output transistors of the winner-take-all, M_2 . For an object N pixels wide, N input currents are summed at the common input node and passed through the N winner-take-all input transistors in parallel. Thus, the input values are averaged over the extent of each object. In this case, the average value is the peak saliency of that object since all N inputs have been set to the peak saliency. The output current going through each of the individual M_2 transistors of a winning object is equal to the bias current divided by the number of pixels in the selected object.

The mismatch among the input transistors of the winner-take-all are averaged for large objects due to the parallel combination of these input transistors. Nonbinary threshold

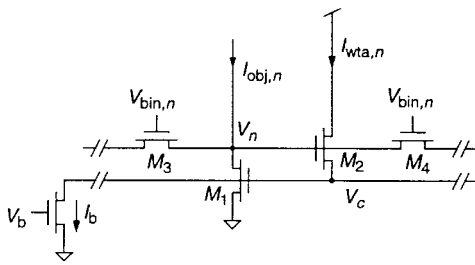


Fig. 4. Object-based selection circuit. Switches for the dynamic wires are implemented by the M_3 and M_4 transistors. The dynamic wires are used here to create a different effect from that of the segmentation-based filtering circuit. The input and output transistors are actually connected in parallel with the corresponding transistors in neighboring elements, creating one larger effective input transistor and one larger effective output transistor. Thus, the group of winner-take-all processing elements within the same object acts as a single processing element.

outputs at any location in the object could cause a variation in the output current at that location. These outputs are typically aggregated within a position-encoding circuit [2], [11], thus minimizing the effect of variation across the object. We originally had some concerns about increasing time constants for the object-based winner-take-all circuit, but we have not observed any such adverse effects during our experiments. While the capacitance on the coupled input node of the winner-take-all increases linearly with the size of the object, the total input current also increases. The bias current should be set to a reasonably high value (compared to subthreshold currents), in order to satisfy the stability criteria for the winner-take-all circuit [17].

IV. EXPERIMENTAL RESULTS

The chip was implemented in a $2.0\text{-}\mu\text{m}$ CMOS process through the MOSIS silicon brokerage service. The size of the chip was 2.25 mm^2 . The system included 20 processing elements within a one-dimensional array along the width of the chip. The parasitic vertical bipolar phototransistors were extended the remaining height of the chip in order to avoid any alignment issues with the optical testing setup. By using an on-chip decoder and several additional control voltages, we were able to multiplex each intermediate current value off the chip for measurement.

We present two experiments to demonstrate system performance under different conditions. For each of the experiments, a static input image was used to determine the processing performance under well-defined conditions. The input signals were the result of imaging two LED's onto the photodetector array. Output measurements were taken of the low-pass filtering and normalization, the segmentation-based filtering, and the object-based selection. The V_{dd} voltage was set to 5.0 V for all testing. According to the MOSIS parametric test results, the threshold voltages for the NMOS and PMOS devices were 0.78 and -0.94 V, respectively, for this fabrication run.

A. Experiment One

In the first experiment we looked at the effects of low-pass filtering on the segmentation and the winner-take-all selection. We set the value of V_{res} to four different values while keeping

the threshold voltage V_{thresh} constant. The normalization was also constant with V_{norm} set to 0.943 V. Each value of V_{res} causes a different space constant in the exponentially decaying impulse response. The measured currents from the normalization and filtering circuit for all four values of V_{res} are shown in Fig. 5. The solid curves indicate the interpolation of the measured values across the array. For comparison to later stages of processing, the threshold value used in this experiment is indicated by the horizontal dashed line, and the theoretical expectation of the peak-saliency segmentation filtering is indicated by the dotted line. The first value of V_{res} is 0.00 V [Fig. 5(a)]. Thus, the spreading is turned off and the output of the filter is the same as the output of the normalization. To produce significant spreading through the resistive network, the value of V_{res} had to be set to values greater than V_{dd} . The need for such high voltages is due to the fact that the subthreshold input currents cause the source voltages of the lateral transistors to be close to V_{dd} ; the gate voltage must be much higher than the source voltage to overcome the backgate effect, as modeled by κ in (1). When V_{res} is set to 5.50 V [Fig. 5(b)], the amount of spreading in the current-mode resistive network increases, and smoothing is evident by the lower peak current values. The low-pass filtering effect is further enhanced with increased values of V_{res} , as shown in Fig. 5(c) and (d). When V_{res} is set to 5.70 V [Fig. 5(d)], the spreading is so extensive that the currents between the two peaks rise considerably.

The segmentation outputs were measured for each of the four low-pass filtered examples. The threshold value was constant; the value of V_{thresh} was set to 0.707 V, which caused I_{thresh} to be 14 nA. The measured peak-saliency currents are shown in Fig. 6. The individual data points are the measured values, and the solid curves are the theoretical expectations that were indicated in Fig. 5. The plots shown in Fig. 6(a)–(d) correspond to the V_{res} settings of 0.00, 5.50, 5.60, and 5.70 V. When no spreading occurs, with $V_{res} = 0.00$ V, the output of the segmentation-based filtering reveals two objects that have saliency values above the threshold value. The results of the thresholding operation are evident by the duplication of the peak value at each position within an object. The poor matching characteristics of the current mirrors are revealed by the difference in value from one pixel to the next. Ideally, these values would be constant over a single object. The offsets are acceptable, however, because of the averaging effect that takes place at the input of the selection circuit. When $V_{res} = 5.50$ V [Fig. 6(b)], the output of the segmentation-based filtering demonstrates little change from the previous plot in Fig. 6(a). The current levels again show the presence of two objects above threshold. A slight difference in the value of the current at position 18 demonstrates a limitation in the thresholding caused by finite gain. As the increased spreading changes the inputs to the segmentation computation, the extents of the objects increase, as shown in Fig. 6(c) for $V_{res} = 5.60$ V. When the spreading increases even further for $V_{res} = 5.70$ V, most of the current levels surpass the threshold and the two objects merge into one, as shown in Fig. 6(d).

Measurements of the winner-take-all output currents for each of these four examples reflect the ability of the selection

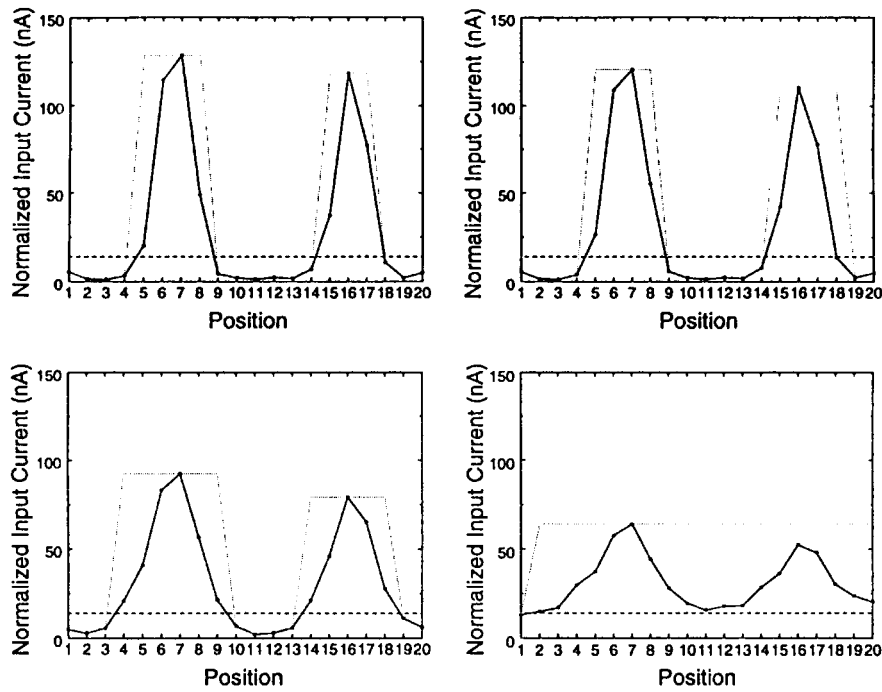


Fig. 5. Experimental results showing the performance of the filtering and normalization circuit. The exponential spreading of the filter’s impulse function is increased by increasing the value of V_{res} . The plots shown in (a)–(d) demonstrate how the filtered, normalized current distribution changes when V_{res} is set to 0.00, 5.50, 5.60, and 5.70 V. The threshold value for the segmentation is indicated by the constant horizontal line at 14 nA. The expected output of the peak-saliency computation, based on the 14-nA threshold, is also indicated by a dotted line in each graph.

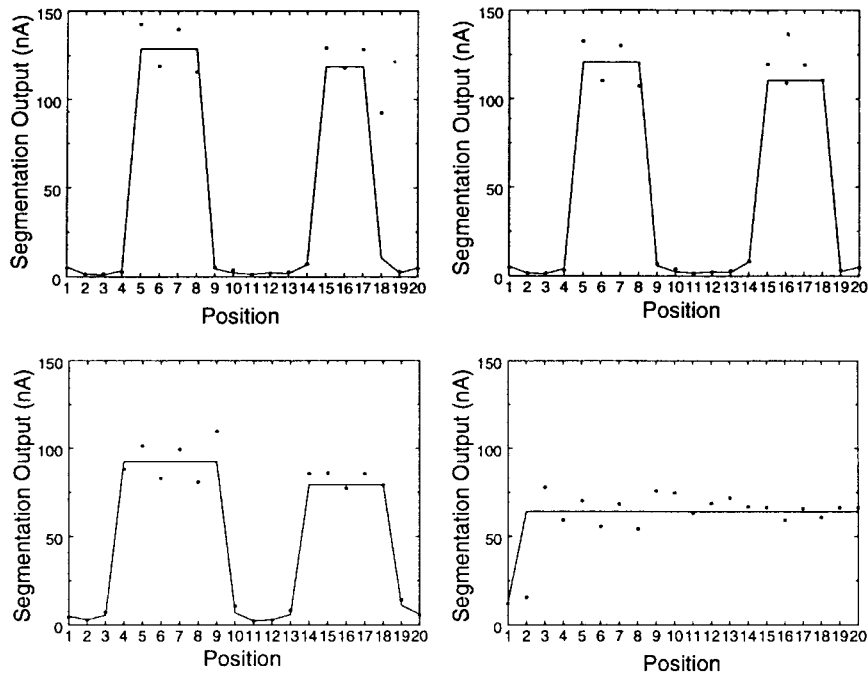


Fig. 6. Experimental results showing the performance of the segmentation-based filtering circuit. The peak-saliency values for four different filtered input distributions are shown in (a)–(d). The filtered, normalized currents are those that result when V_{res} is set to 0.00, 5.50, 5.60, and 5.70 V, as shown in Fig. 5. The measured data from the segmentation-based filtering circuit is indicated by the individual data points. The expected peak-saliency values that were indicated in Fig. 5 are repeated here as the solid curves.

circuit to change its effective processors according to the extent of the input objects. The total output current of the winner-take-all circuit was set to 114 nA, with $V_b = 0.802$ V. The output of the winner-take-all is shown for each setting of V_{res} in Fig. 7. The measured values are indicated by the individual data points, and the theoretical expectation is indicated by the

solid curve. The theoretical curve was calculated by dividing the total winner-take-all bias current by the number of pixels in the selected object. The measurements shown in Fig. 7(a) and (b) are essentially identical. The segmentation-based filtering output for the object with the highest peak value is the same for both trials, as was shown in Fig. 6(a) and (b). The level of the

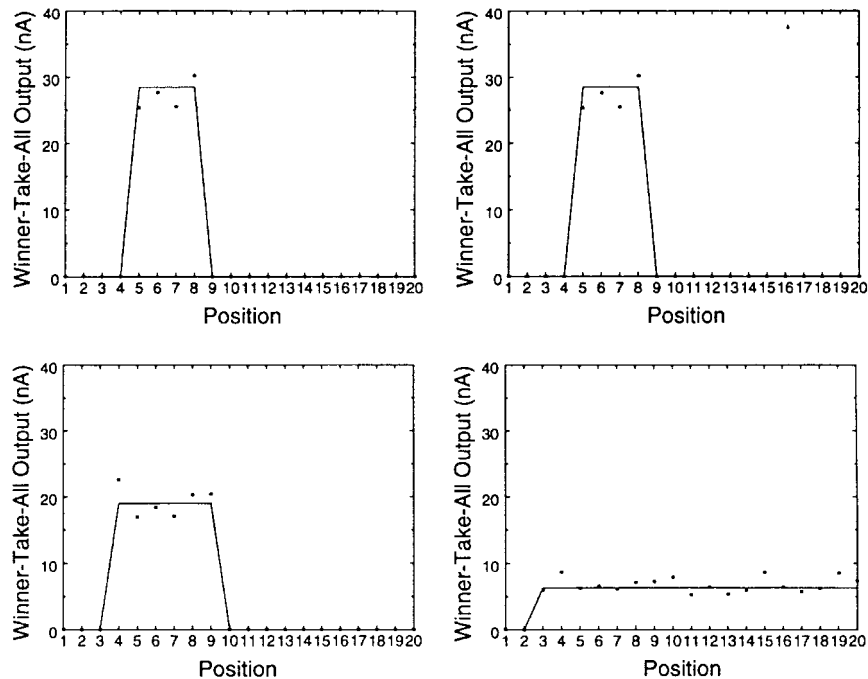


Fig. 7. Experimental results showing the performance of the object-based selection circuit. The output of the winner-take-all computation is shown for four different input distributions. The segmentation currents shown in Fig. 6 were the inputs to the selection circuit. The spatial extent of the selected object determines the spatial extent of the winner-take-all output currents. The current levels are inversely proportional to the size of the object. The measured values are indicated by the individual data points. The theoretical expectation was calculated by dividing the total winner-take-all bias current by the number of pixels included in the object. These values are indicated by the solid curve.

output currents in Fig. 7(a) and (b) is between 25 and 30 nA, which indicates that the total winner-take-all current is being distributed across an area of four pixels. The output for the case when $V_{res} = 5.60$ V is shown in Fig. 7(c). The average output current is between 15 and 20 nA, indicating that the total winner-take-all current is distributed among six pixels. For the case when the segmentation output results in a single object for $V_{res} = 5.70$ V, the winner-take-all output changes accordingly, as shown in Fig. 7(d). Again, the smaller nonzero output values of the winner-take-all circuit indicate the larger area of the selected object.

B. Experiment Two

In the second experiment we measured the output of the segmentation-based filtering stage when different values of the threshold voltage V_{thresh} were used. The results of this experiment are shown in Fig. 8. The low-pass filtered input to the segmentation processing was the same as that shown in Fig. 5(d). The first setting of the threshold voltage, $V_{thresh} = 0.71$ V, is the same value used in the previous experiment. Thus, the output of the peak-saliency computation shown in Fig. 8(a) is the same as that shown in Fig. 6(d). The threshold value was increased to 0.72 V [Fig. 8(b)], 0.73 V [Fig. 8(c)], and 0.74 V [Fig. 8(d)]. In each case the number of pixels above the threshold value decreased, thus causing a decrease in the spatial extent of the objects. The mismatch of the transistors in the high-gain stage used for the thresholding limited the accuracy when predicting the output of the segmentation-based filtering. When the threshold value was low, such as in the cases shown in Fig. 8(a) and (b), the differences between the

normalized input values near the threshold value were not large enough to overcome the noise introduced by the mismatch of the transistors. Thus, the output of the thresholding is low at position 12 [Fig. 8(b)] before it goes low at position 11 [Fig. 8(c)], even though the actual values have the opposite order and differ by about 1 nA.

V. CONCLUSIONS

We have presented analog VLSI circuits that implement object-based processing for selective-attention. The segmentation of objects within an image is a critical preprocessing stage for object-based selection. The initial implementation presented in this paper performs the segmentation with a filtering and thresholding computation. The combination of filtering and thresholding allows some flexibility for emphasizing particular object characteristics in the selective-attention competition. The thresholding output is used to segment the image into objects. The segmentation is first used to find the peak saliency within each object and later used to define the granularity of the selection operation. The implementation demonstrates an elegant method of modeling the dynamic size of the attentional spotlight. The circuits are scalable and do not require a large number of transistors beyond that of the initial pixel-based selective-attention system. While nonidealities in the fabrication of these circuits will cause some signals to degrade, all of the described effects can be referred back to the input as noise and do not cause the system to fail.

The segmentation- and object-based selection circuits have been tested with object-based excitatory and inhibitory feedback, which incorporates the remaining components of the

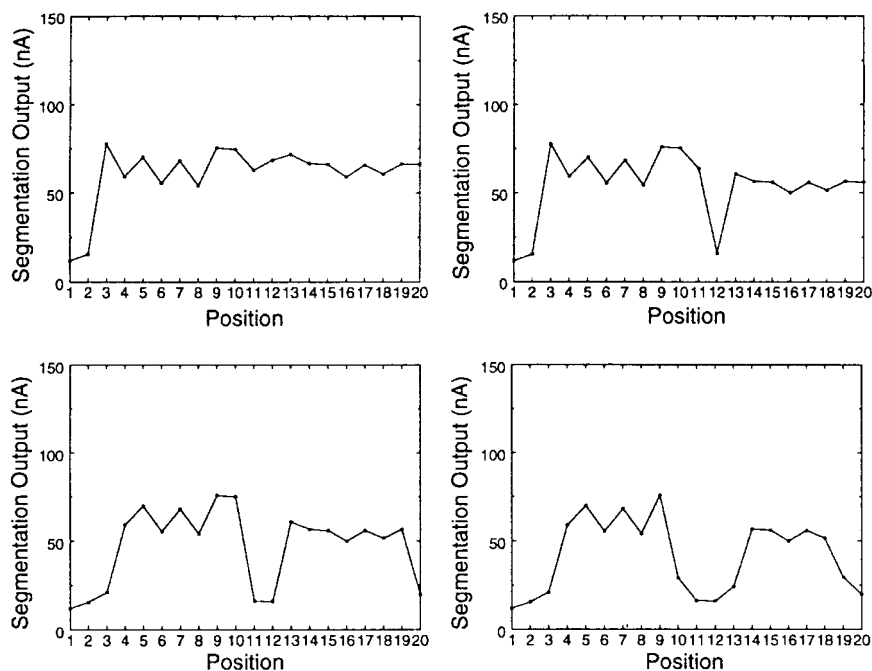


Fig. 8. Experimental results showing the performance of the segmentation-based filtering circuit when the threshold level changes. The input to the segmentation-based filtering stage is the array of inputs that was shown in Fig. 5(d). The value of $V_{i_{\text{thresh}}}$ was changed from its original value of (a) 0.71 V to (b) 0.72 V, (c) 0.73 V, and (d) 0.74 V. The extent of the segmented objects is demonstrated by the number of pixels that have similar peak-saliency output values.

selective-attention framework. The entire object-based system demonstrates the expected behavior, thus successfully implementing a selective-attention system with a dynamic spotlight size. All of these features for selective-attention processing have been integrated together in a two-dimensional chip (25×24 array of elements) that is currently undergoing extensive testing to determine the limits of operation for these analog circuits in larger two-dimensional arrays [13].

REFERENCES

- [1] M. I. Posner and S. E. Petersen, "The attention system of the human brain," *Annu. Rev. Neurosci.*, vol. 13, pp. 25–42, 1990.
- [2] T. G. Morris and S. P. DeWeerth, "Analog VLSI excitatory feedback circuits for attentional shifts and tracking," *Analog Integrated Circuits and Signal Processing*, vol. 13, pp. 79–91, 1997.
- [3] ———, "Analog VLSI circuits for covert attentional shifts," in *Proc. Fifth Int. Conf. Microelectronics for Neural Networks and Fuzzy Systems*. Lausanne, Switzerland, Feb. 1996, pp. 30–37.
- [4] C. Koch and S. Ullman, "Shifts in selective visual attention: Toward the underlying neural circuitry," *Human Neurobiol.*, vol. 4, pp. 219–227, 1985.
- [5] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
- [6] M. Mahowald, "VLSI Analogs of neuronal visual processing: A synthesis of form and function," Ph.D. dissertation, Cal. Inst. Technol., 1992.
- [7] T. Delbruck, "Silicon retina with correlation-based velocity-tuned pixels," *IEEE Trans. Neural Networks*, vol. 4, pp. 529–541, May 1993.
- [8] K. A. Boahen and A. G. Andreou, "A contrast sensitive silicon retina with reciprocal synapses," in *Advances in Neural Information Processing Systems 4*, J. E. Moody, Ed. San Mateo, CA: Morgan Kaufmann, 1991.
- [9] J. Kramer, R. Sarpeshkar, and C. Koch, "An analog VLSI velocity sensor," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 1995, pp. 413–416.
- [10] D. L. Standley, "An object position and orientation IC with embedded imager," *IEEE J. Solid-State Circuits*, vol. 26, pp. 1853–1859, Dec. 1991.
- [11] S. P. DeWeerth, "Analog VLSI circuits for stimulus localization and centroid computation," *Int. J. Comput. Vision*, vol. 8, no. 3, pp. 191–202, 1992.
- [12] S.-C. Liu and J. Harris, "Dynamic wires: An analog VLSI model for object-based processing," *Int. J. Comput. Vision*, vol. 8, no. 3, pp. 231–239, 1992.
- [13] T. G. Morris, C. S. Wilson, and S. P. DeWeerth, "An analog VLSI focal-plane processing array that performs object-based attentive selection," in *Proc. 40th Midwest Symp. Circuits and Systems*, Aug. 1997.
- [14] T. K. Horiuchi, T. G. Morris, C. Koch, and S. P. DeWeerth, "Analog VLSI circuits for attention-based visual tracking," M. C. Mozer, M. I. Jordan, and T. Petsche, Eds., *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, vol. 9, pp. 706–712, 1997.
- [15] C. S. Wilson, T. G. Morris, and S. P. DeWeerth, "Segmentation coding for object-based attentive selection systems," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 1998.
- [16] B. Gilbert, "A 16-channel array normalizer," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 956–963, 1984.
- [17] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of $O(n)$ complexity," in *Advances in Neural Information Processing Systems*, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, vol. I, pp. 703–711, 1989.



Tonia G. Morris received the B.S. degree in electrical engineering from the University of South Carolina, Columbia, in 1991, the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1993, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology in 1996.

She is a Senior Design Engineer at Intel Corporation, Chandler, AZ. Her research interests include advanced CMOS imagers and focal-plane visual processing systems.

Timothy K. Horiuchi photograph and biography not available at time of publication.



Stephen P. DeWeerth received the B.A. degree in mathematics and chemistry from Wartburg College, Waverly, IA, in 1985. He received the M.S. degree in computer science and the Ph.D. degree in computation and neural systems from the California Institute of Technology, Pasadena, in 1987 and 1991, respectively.

He is presently an Associate Professor of Electrical and Computer Engineering and Biomedical Engineering at the Georgia Institute of Technology and the Emory University Medical School, Atlanta, GA. His primary research activities are in neural modeling and neuromorphic engineering, particularly in the application of biologically inspired sensorimotor architectures and computational paradigms to the engineering of autonomous and biomedical systems. These activities are focused specifically on the use of analog VLSI circuits to explore motor pattern generation, muscular control, and oculomotor systems, and the understand the roles of sensory feedback and learning in these systems. He is also active in the application of computing and technology to engineering education with a particular emphasis on WWW-based remote laboratories.

Dr. DeWeerth served on the organizing committee of the 1996 International Symposium on Circuits and Systems and is the General Chair of the 1999 Twentieth Anniversary Conference on Advanced Research in VLSI to be held at Georgia Tech. He has received a Dupont Young Faculty Award and multiple AT&T Special-Purpose Grants. His research group is additionally funded by the National Science Foundation, the Whitaker Foundation, Hewlett Packard, and other corporations.