AN ADAPTIVE VISUAL TRACKING SENSOR WITH A HYSTERETIC WINNER-TAKE-ALL NETWORK

Giacomo Indiveri[*]*, Philip Oswald, and Jörg Kramer*

Institute of Neuroinformatics, University of Zurich and ETH Zurich
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

ABSTRACT

We present a CMOS VLSI image sensor containing a 26×26 array of differentiating adaptive photoreceptors, combined with a current-mode hysteretic winner-take-all (WTA) network, able to detect and track high-contrast targets moving within its field of view. The device, implemented using standard 0.8 μm CMOS technology, dissipates 600 μW, with a power supply voltage of 5 V. Its WTA outputs can be read out using either analog position-to-voltage circuits, or digital address encoders. Photoreceptor voltages with sustained and transient responses, can be read out serially, using an on-chip scanner, or in a random-access mode, using address decoders. Combining digital output position encoders and input address decoders it is possible to read only the location of the winning pixel, or a small window around the winning pixel (*e.g.* using a microcontroller), thus implementing a selective attention read-out mechanism.

1. INTRODUCTION

Localizing and tracking moving features in natural scenes is a computationally demanding task for machine vision systems. Conventional solutions typically require the vision sensor to transmit with high throughput all the raw image data to digital signal processors (DSPs) for further processing. If the machine vision system is part of a feedback loop on a robot or autonomous vehicle, the overall speed of computation (including the data transmission time) is critical for the stability of the control loop itself. To reduce the sensory-motor control loop latency, one can make use of custom VLSI *smart vision sensors* [1] that perform image pre-processing on the focal-plane itself and reduce the amount of data that needs to be transmitted to further processing stages. In extreme cases, the sensor carries out *all* of the vision processing necessary for the sensory-motor task on the focal plane and transmits only the result of the computation (see for example [2, 3, 4]). To further increase processing speed the smart vision sensor can be made to select the sub-regions of the image that require processing, preprocess them, and transmit only the data from those subregions. Sensors that carry out this type of operation have been recently proposed [5, 6]. In this paper we propose a similar type of device that differs from previously proposed ones in two key features: it uses adaptive photoreceptor circuits that respond to positive illuminance transients [7] to select high-contrast moving edges, independently of the absolute brightness of the scene (as opposed to simply selecting the scene's brightest region), and it uses a hysteretic winner-take-all (WTA) network [8], with positive feedback and spatial filtering properties, to lock-onto and smoothly track the selected targets, while other tracking devices use simpler types of WTA networks. The sensor described in this paper has both analog and digital outputs encoding the position of the winner, so that it can be interfaced to robots and other actuators directly or via a microcontroller. It also has on-chip scanners and address decoders to read an output of the adaptive photoreceptor array serially (*e.g.* for displaying images on monitors) or in a random-access mode (*e.g.* for reading out sub-regions of the image). The chip's input address decoders can be directly connected to the digital outputs encoding the position of the winning pixel for selectively reading the photoreceptor output of just that pixel and displaying only the part of the image that is of interest. Regions of interest can be selectively accessed by addressing small windows around the winning pixel's address.

In the next Section we describe the architecture of the system and the circuit details; then we show, in Section 3, experimental results demonstrating how the adaptive properties of the photoreceptors and the spatial filtering and hysteretic properties of the WTA circuits allow the device to reliably track stimuli in natural environments with different illumination conditions. Section 4 concludes the paper.

2. SYSTEM ARCHITECTURE

The device comprises a core array of 26×26 pixels arranged on a hexagonal grid, and peripheral analog and digital input/output (I/O) circuits. Each pixel contains a differentiating adaptive photoreceptor, a hysteretic WTA circuit, and

[*]Corresponding author: giacomo@ini.phys.ethz.ch

0-7803-7448-7/02/$17.00 ©2002 IEEE

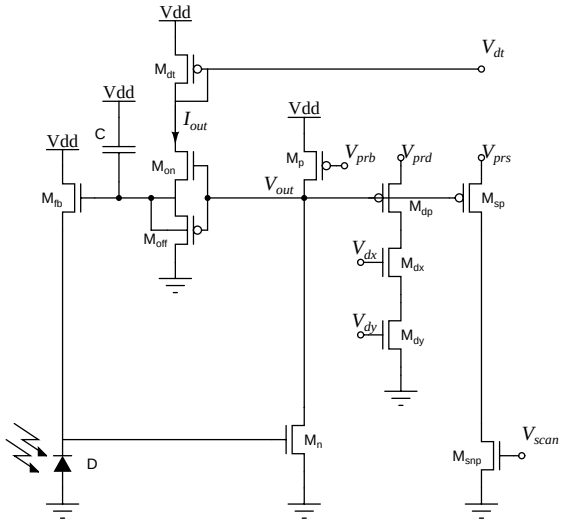

Fig. 1. Differentiating adaptive photoreceptor circuit.

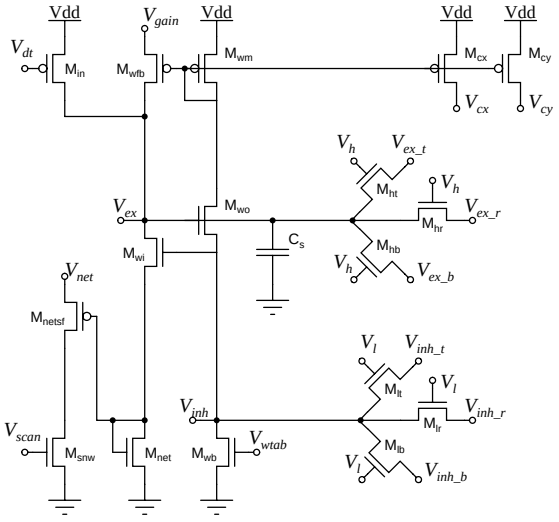

Fig. 2. Hysteretic WTA circuit with spatial coupling.

interfacing I/O circuits.

2.1. The adaptive photoreceptor

The photoreceptor circuit with its readout circuitry is shown in Fig. 1. The photoreceptor consists of a photodiode D in series with a transistor M_{fb} in source-follower configuration and a negative feedback loop from the source to the gate of M_{fb} [9]. The feedback loop consists of a high-gain inverting amplifier in common-source configuration (M_n, M_p) [9] and a thresholding and rectifying temporal differentiator stage (M_{on}, M_{off}, C) [10]. A sufficiently large positive irradiance change activates a transient current I_{out} onto capacitor C, that is converted into a voltage V_{dt} by the diode-connected transistor M_{dt}. The photoreceptor voltage V_{out} can be read out by the address decoder as V_{prd}, if the address decoder select lines V_{dx} and V_{dy} are high. The voltage V_{out} can also be read out by the on-chip scanner circuit, via V_{prs}, to display the sensor output on monitors. The photosensing sub-circuit has been analyzed in some detail and will be presented elsewhere [7]. The voltage V_{dt} is used to provide input to the locally connected WTA cell.

2.2. The hysteretic winner-take-all circuit

The hysteretic WTA circuit is shown in Fig. 2. Together with the WTA circuits of all other pixels in the array, it forms the hysteretic WTA network. It is based on the current mode WTA circuit originally presented in [11], and has been characterized in detail in [8]. The output current I_{out} of the photoreceptor stage of Fig. 1 is mirrored by M_{in} into node V_{ex}. If the input current to the considered pixel is the strongest, the cell "wins" and transistors M_{cx} and M_{cy} source an output current proportional to the circuit's bias current, set by V_{wtab}, bringing the output voltages V_{cx} and V_{cy} high. Voltages V_{cx} of all pixels belonging to common columns are tied together, and voltages V_{cy} of all pixels belonging to a common row are tied together. A copy of the WTA bias current, attenuated exponentially by the bias voltage V_{gain} is fed back into the input node, via M_{wfb}. It is this positive feedback current that endows the circuit with hysteretic response properties [8]. Transistors M_{ht}, M_{hb}, and M_{hr} diffuse the currents coming from M_{in} and M_{wfb} to the V_{ex} nodes of the three (top, bottom, and right) neighboring cells. The bias voltage V_h is used to tune the diffusion space constant and to control the amount of lateral excitatory coupling. Conversely, transistors M_{lt}, M_{lb}, and M_{lr} implement the inhibitory coupling among neighboring cells. The bias voltage V_l is used to control the spatial extent of lateral inhibition. If V_l is set to V_{dd}, inhibition is global, and only one pixel in the whole array can win.

The current flowing through M_{net} represents the net current that the WTA cell is receiving, corresponding to sum of the input current from the photoreceptor circuit, the positive-feedback current and the diffused excitatory currents. The voltage V_{net}, logarithmically proportional to this net current, can be scanned out to image the overall network activity and view the relative effects of positive feedback current modulation (V_{gain}), and excitatory and inhibitory coupling modulations (V_h and V_l respectively).

2.3. Peripheral I/O circuits

As mentioned in Section 1, the device has several types of peripheral I/O circuits. There are analog position-to-voltage (P2V) circuits, of the type described in [12], and digital position encoding circuits for reading out the out-

put of the WTA network; there is an on-chip scanner circuit,for displaying on monitors the outputs of all photoreceptors, and/or the state of the WTA network activity (see V_{net} described above); and there are input address decoders for accessing the analog output voltage of individual photoreceptors.

WTA output: The voltages V_{cx} and V_{cy} of Fig. 2 are routed to the periphery of the architecture core, and fed into a two-input pass-transistor demultiplexer. Depending on the state of the demultiplexer, V_{cx} and V_{cy} are routed either to the analog P2V circuits, or to asynchronous digital position (address) encoders. In this way only one of the two (analog or digital) modes can be used at one time, but wiring and possible sources of cross-talk noise are minimized.

Scanner circuits: The scanner reads the output voltages V_{prs} and V_{net} of the array in the sequence used for standard electronic cameras. Each output voltage of each pixel is buffered via a source follower consisting of an input transistor ($\mathsf{M_{sp}}$ for V_{prs} and $\mathsf{M_{netsf}}$ for V_{net}) and a current source that is common to each column and signal. A vertical shift register sequentially addresses the rows with the binary voltage signal V_{scan} via switching transistors ($\mathsf{M_{snp}}$ for V_{prs} and $\mathsf{M_{snw}}$ for V_{net}). The output voltages of the column source followers are transferred to a common output line for each signal via complementary pass transistors that are sequentially opened, column by column, by a horizontal shift register. The clocks of the two shift registers are synchronized, such that the output voltages of the entire array are sequentially read out, row by row. The voltages on the common lines are buffered to be sensed off chip.

Address decoders: When properly driven, the chip's input address decoders activate the select lines V_{dx} and V_{dy} of the addressed pixel (see Fig. 1) and route the voltage V_{prd} of that pixel to a unity gain follower of an analog output pad.

3. EXPERIMENTAL RESULTS

In Fig. 3 we show experimental results obtained by enabling the analog P2V circuits and measuring their output voltages V_x and V_y encoding the x and y position of the winning pixel. The WTA network was biased in a way to have local excitation (V_h of Fig. 2 was set to 0.8V) and global inhibition (V_l was set to V_{dd}). The measurement shows the sensor's response to a target appearing in the bottom left corner of the field of view, slowly moving to the top right corner and then completing a figure-eight pattern. Before the target appeared, the sensor's output was $V_x \approx 0$V and $V_y \approx 0$V. This is because the bottom-left pixel $(1, 1)$ receives an additional input current, set by an external bias voltage, that sets a global threshold: if no visual stimulus is strong enough to overcome this threshold, the output is always "zero". As soon as the target appeared in the sensor's field of view, the WTA network switched winner, and the P2V circuits modified V_x and Vy accordingly. In this experiments the target was the light spot of a laser-pointer shone on a flat surface 30cm from the chip's focal plane. Images were focused onto the focal plane using an 8mm lens with an $f-$number of 1.2. The sensor's response does not depend on the background onto which the target is overlaid, nor does it change with absolute background illumination.

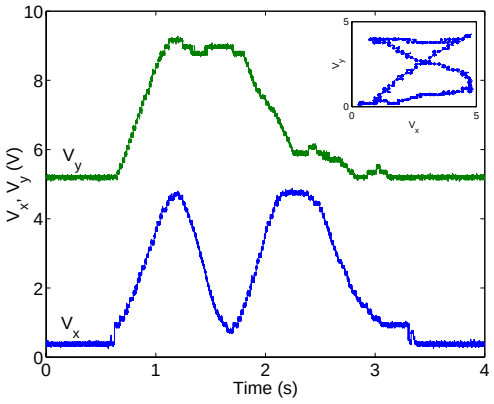

Fig. 3. Output of the analog P2V circuits in response to a target moving from the bottom left corner to the top right one, on to the top left, to the bottom right, and back to the bottom left corner.

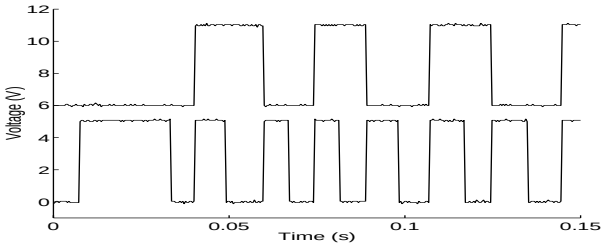

Fig. 4. Output of least significant bit (bottom trace) and second-least significant bit (top trace, displaced by 6V) of the 'X' address in response to a target moving from right to left.

By switching the state of the demultiplexer connected to the WTA outputs we disabled the analog P2V circuits and enabled the asynchronous address encoders. Figure 4 shows the the response of two address lines (the least significant and second-least significant bits of the X address) in response to the same stimulus of Fig. 3 moving from right to left. The non-uniform pulse widths are due to the asynchronous response of the circuit to the variable speed of the stimulus. In a second experiment, we placed the sensor in front of a CRT monitor, showed a white box performing a circular motion on a black background, and sampled the chip's address encoder outputs every 25ms over a period of 40s. In this period the target made 16 full revolutions. The histogram of the sampled addresses is shown in Fig. 5. As

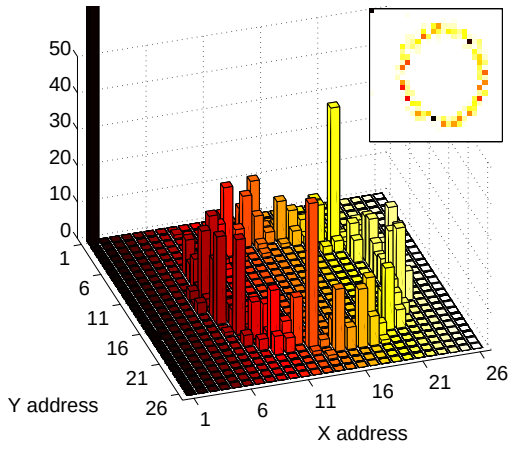

Fig. 5. Histogram of the addresses measured from the sensor's address encoders in response to a target moving on a circular trajectory.

Fabrication technology	0.8 μm CMOS 2P 2M
Resolution	26×26
Fill factor	1.2%
Pixel size	84.8μm×62.6μm
Die size	3.22 mm×2.56 mm
Power supply voltage	single 5 V
Power consumption	
digital output (scanners off)	1.1mW
analog output (scanners off)	600μW

Table 1. Characteristics of the visual tracking sensor.

the global threshold was set relatively high, address $(1, 1)$ was selected most often (193 samples, off-scale in the figure).

The response time of the sensor to the sudden appearance of a target is 1.2μs when the digital outputs are enabled, and can be as long as 6μs when the analog outputs are enabled. Power consumption is also dependent on the output mode selected (see Table 1).

4. CONCLUSIONS

We presented a two-dimensional visual tracking sensor containing an array of differentiating adaptive photoreceptors and a hysteretic WTA network with spatial filtering capabilities. Images are sensed and processed fully in parallel. The pixel reporting the strongest positive illuminance transient (*e.g.* induced by a high-contrast moving target) is selected by the WTA network. Its position can be read out using either analog P2V circuits or digital address encoders. The sustained response of each photoreceptor and net input current to each WTA can be read out serially, using on-chip scanners, and displayed on monitors. Additionally, photoreceptor voltages can be individually sensed, using input address decoders. The WTA analog outputs can be used to drive motors and actuators, for example on small autonomous robots. The WTA digital outputs can be used to drive the input address decoders and read the photoreceptor output of only the winning pixel. This mechanism could be exploited (*e.g.* using a microcontroller) to selectively read out just the regions of the image around the position of the target, rather than reading out all the raw image data.

5. REFERENCES

[1] C. Koch and B. Mathur, "Neuromorphic vision chips," *IEEE Spectrum*, vol. 33, no. 5, pp. 38–46, May 1996.

[2] R. Etienne-Cummings, J. Van der Spiegel, and P. Mueller, "A visual smoot pursuit tracking chip," in *Advances in Neural Information Processing Systems*, D. S. Touretzky, M. C. Mozer, and Hasselmo M. E., Eds. 1996, vol. 8, MIT Press.

[3] G. Indiveri, "Neuromorphic analog VLSI sensor for visual tracking: Circuits and application examples," *IEEE Trans. on Circuits and Systems II*, vol. 46, no. 11, pp. 1337–1347, November 1999.

[4] R.R. Harrison and C. Koch, "A robust analog VLSI motion sensor based on the visual system of the fly," *Autonomous Robots*, vol. 7, pp. 211–224, 1999.

[5] V. Brajovic and T. Kanade, "Computational sensor for visual tracking with attention," *IEEE Journal of Solid State Circuits*, vol. 33, no. 8, pp. 1199–1207, Aug. 1998.

[6] C.S. Wilson, T.G. Morris, and S.P. DeWeerth, "A two-dimensional, object-based analog VLSI visual attention system," in *Proc. Conf. Advanced Research in VLSI*, S. P. DeWeerth, S.M. Wills, and A.T. Ishii, Eds., Los Alamitos, CA., 1999, vol. 20, IEEE Computer Society Press.

[7] J. Kramer, "An integrated optical transient sensor," 2001, Submitted.

[8] G. Indiveri, "A current-mode hysteretic winner-take-all network, with excitatory and inhibitory coupling," *Analog Integrated Circuits and Signal Processing*, vol. 28, no. 3, pp. 279–291, September 2001.

[9] T. Delbrück and C.A. Mead, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits," in *Vision Chips: Implementing vision algorithms with analog VLSI circuits*, C. Koch and H. Li, Eds., pp. 139–161. IEEE Computer Society Press, 1995.

[10] J. Kramer, R. Sarpeshkar, and C. Koch, "Pulse-based analog VLSI velocity sensors," *IEEE Trans. on Circuits and Systems II*, vol. 44, no. 2, pp. 86–101, Feb. 1997.

[11] 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, 1989, vol. 2, pp. 703–711, Morgan Kaufmann.

[12] S. P. DeWeerth, "Analog VLSI circuits for stimulus localization and centroid computation," *Int. J. of Comp. Vision*, vol. 8, no. 3, pp. 191–202, 1992.