NEUROMORPHIC SELECTIVE ATTENTION SYSTEMS

Giacomo Indiveri

Institute of Neuroinformatics, University of Zurich and ETH Zurich
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
giacomo@ini.phys.ethz.ch

ABSTRACT

Selective attention mechanisms allow sensory systems with limited processing capacity to function in real time independent of the size of the stimulus input space. They can be particularly useful in vision systems, where the amount of information provided by the sensors typically exceeds the system's processing capacity. We present circuits for implementing models of selective attention mechanisms and show application examples, ranging from single chip focal plane processors to active multi chip vision systems, that perform active tracking and selective scanning tasks.

1. INTRODUCTION

Selective attention is a mechanisms used to sequentially select and process salient subregions of the input space, while suppressing inputs arriving from non-salient regions. By processing small amounts of sensory information in a serial fashion, rather than attempting to process all the sensory data in parallel, this mechanism overcomes the problem of flooding limited processing capacity systems with sensory inputs. It is found in many biological systems and can be a useful engineering tool for developing artificial systems that need to process in real time sensory data. The type of selective attention mechanisms that we focus on are the ones that are stimulus driven (pure bottom-up approach with no top-down influence) and based on saliency-maps [1]. One of the key computations performed by these types of selective attention mechanisms is the *winner-take-all* (WTA) operation. This highly non-linear operation is used to select the strongest input, corresponding to the most salient feature, and to suppress all other inputs. Depending on the model of selective attention mechanism implemented, the WTA locks onto the input selected and tracks it as it moves, or attends to it for a set time interval, deselects it and suppresses it, through an *inhibition of return* mechanism [1] allowing successive inputs of decreasing strength to be selected.

The systems we describe are *neuromorphic*, in the sense that they contain VLSI devices and analog circuits that use the same organizing principles found in real neural systems, such as the retina, or the cortex.

In the next Section we present a compact current-mode circuit used for implementing WTA networks; in Section 3 we describe examples of single-chip vision sensors that use WTA networks to select high contrast visual stimuli and track them in space; in Section 4 we describe a multi-chip active vision system that selects successive regions with high contrast activity in the sensor's field of view and orients the visual sensor toward them; and in Section 5 we draw the concluding remarks.

2. THE WTA CIRCUIT

The WTA circuit we use is based on the current-mode WTA circuit originally presented in [2] but contains additionally local excitatory feedback, lateral excitatory coupling, lateral inhibitory coupling, and diode-source degeneration. A detailed description of the circuit, of its additional elements and of its behavior is described in [3]. The circuit diagram of one element of the WTA network is shown in Figure 1. These elements can be arranged in one- or two-dimensional arrays. The response time of the whole network is independent of its size, as its elements are fully analog and operate collectively in parallel.

The output current of each WTA cell is null, except for the one winning the competition. In the winning cell, I_{out} is a copy of the bias current I_b. A current proportional to I_b is also sourced back into the winning cell's input node through the current-mirror formed by $M3$ and $M5$. If I_b is a subthreshold current, the proportionality factor of the local excitatory feedback current is modulated exponentially by the voltage difference $(V_{dd} - V_{gain})$. There is a hysteretic effect, because after a cell has been selected as the winner, its input current has to decrease by an additional amount equal to the local excitatory feedback current, to lose its winning status. Transistor $M6$ increases the winner selectivity gain [3] and provides a copy, with $M7$, of the cell's net input current I_{all} (namely, the input current I_{in}, the current being spread to or from the left and right nearest neighbors and the local excitatory feedback current coming from the top p-type current mirror).

Lateral coupling is implemented by means of "diffu-

0-7803-7762-1/03/$17.00 ©2003 IEEE

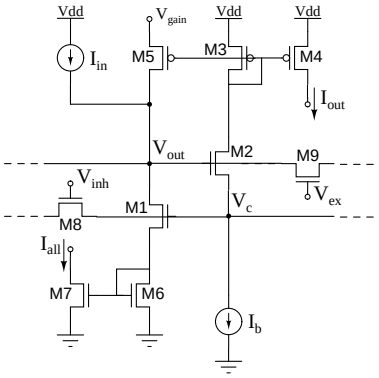

Figure 1: WTA cell, with local excitatory feedback, lateral excitatory coupling, lateral inhibitory coupling and diode-source degeneration.

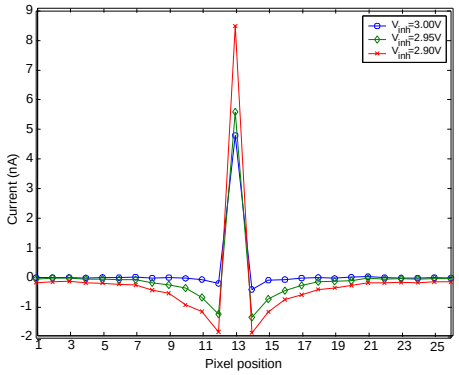

Figure 2: Center-surround response of the hysteretic WTA network to a single pixel activation. The data was taken by subtracting a response containing only lateral excitatory coupling ($V_{inh} = 5V$) to responses containing also lateral inhibitory coupling.

sor" (or "pseudo-conductance") networks implemented by transistors $M8$ and $M9$. Specifically, transistor $M9$ implements lateral excitatory coupling and transistor $M8$ lateral inhibitory coupling. Functionally, the inhibitory diffusor network can be used to spatially decouple the WTA cells, while the excitatory diffusor network can be used to smooth the input signals, combined with the local excitatory feedback current of the winning cell. Combining the effects of both excitatory and inhibitory networks, we can bias the hysteretic WTA network to exhibit center-surround response profiles of the type shown in Figure 2. These types of response profiles can be very useful for processing sensory signals and make this WTA circuit ideal for use in single- and multi-chip neuromorphic vision systems.

3. TRACKING SENSORS

Tracking salient features as they move in the environment is a computationally demanding task. Selective attention mechanisms can help in reducing the computational load of visual tracking systems [1]. The sensors we designed reduce the computational cost of the overall system, by preprocessing the most salient sub-regions of the image and transmitting only the result of the computation, rather than large amounts of data from the raw input image. These tracking sensors differ from previously proposed ones [4, 5] in two key features: they select high-contrast edges independent of the absolute brightness of the scene (as opposed to simply selecting the scene's brightest region [5]), and they use the hysteretic WTA network described in Section 2 to lock-onto and smoothly track the selected targets (different from WTA networks used in other tracking devices [4, 5]).

3.1. One dimensional sensor

This sensor consists of a one dimensional array of computational elements that detect and track in real time the position of the feature with highest spatio-temporal contrast in the visual scene [6]. Adaptive photo-receptors at its input stage report contrast variations in the visual scene. The sensor computes saliency by extracting a measure of the *spatial* derivative from the photo-receptor outputs, and by rectifying it. The WTA network, in selecting the location of the strongest input, drives a position-to-voltage (P2V) circuit at the location of the corresponding pixel. The P2V circuit output is a single analog voltage, the amplitude of which encodes the position of the winning pixel. A detailed description of the sensor and examples of tracking applications using these types of sensors are reported in [6].

3.2. Two dimensional sensor

This sensor consists of a 2D array of adaptive photo-receptor circuits that respond to positive illuminance *transients* [7] connected to WTA network elements. In this case the saliency map corresponds to a map of the *temporal* derivative of the input scene. The sensor has both analog and digital outputs encoding the position of the winner. The analog outputs are computed using two P2V circuits (one for the x component of the position, and the other for its y component). The size of the array, implemented using a standard 0.8μm CMOS technology, is of 3.2$\times$2.6mm. An example of the analog response of the sensor to a moving target is shown in Figure 3. This sensor also has address decoders to read an output of the adaptive photo-receptor array in a random-access mode (*e.g.* for reading out sub-regions of the image). The chip's digital outputs, encoding the position of the winning pixel, can be directly connected to the input

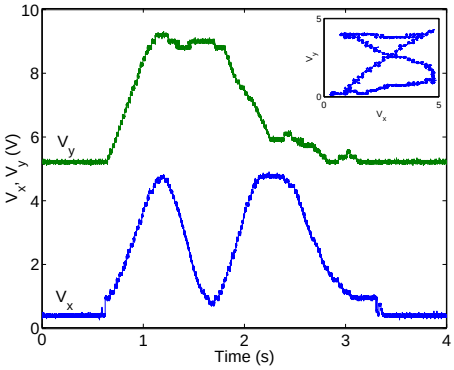

Figure 3: Output of the P2V x-circuit (lower trace) and y-circuit (top trace, offset by 4 volts) in response to a target moving from the bottom left corner to the top right one, on to the top left, to the bottom right, and back to the bottom left corner. The inset shows the x-output trace versus the y-output one.

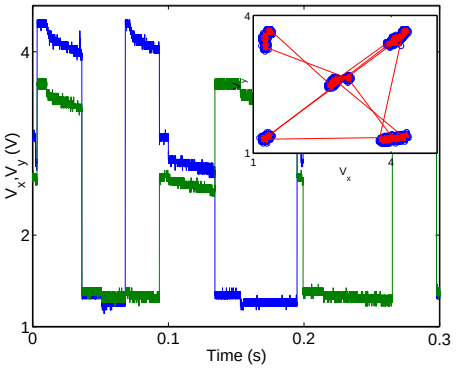

Figure 4: Output of the P2V x-circuit and y-circuit of the selective attention chip in response to a test stimulus exciting four corners of the input array at a rate of 30Hz and a central cell at a rate of 50Hz. The inset shows the x-output trace versus the y-output one.

address decoders for selectively reading the photo-receptor output of just that pixel and displaying only the salient parts of the image. Regions of interest can be selectively accessed by addressing small windows around the winning pixel's address. A detailed description of all the circuits used in this sensor is presented in [8].

4. MULTI-CHIP SELECTIVE ATTENTION MODELS

The devices described in the previous Section are example of single chip systems that implement simplified forms of *visual* selective attention. Alternative single chip systems for implementing visual selective attention mechanisms have also been proposed [4, 5, 6]. These systems contain photo-sensing elements and processing elements on the same focal plane, and typically apply the competitive selection process to *visual* stimuli. Unlike these systems, multi-chip systems decouple the sensing stage from the selective attention/competition stage. Therefore input signals need not arrive only from visual sensors, but could represent a wide variety of sensory stimuli obtained from different sources. In multi-chip selective attention systems visual sensors used for generating the saliency map could be relatively high-resolution silicon retinas and would not have the small fill factors that single-chip 2D attention systems are troubled with. Furthermore top-down modulating signals could be fused with the bottom-up generated saliency map to bias the selection process. In the following sections we present a neuromorphic selective attention chip interfaced to an adaptive visual sensor, for implementing an active vision selective attention system. Both the selective attention chip and the imaging sensor contain silicon neurons and spike-generating circuits. The communication protocol used to interface devices of this type among each other is based on the *Address-Event Representation* (AER). In this representation input and output spikes (address-events) are sent from/to VLSI devices using asynchronous non-clocked digital pulses that encode the address of the sending node. Analog information is carried in the temporal structure of the inter-pulse intervals and in their mean frequency, very much as it is believed to be conveyed in spike trains of biological neural systems.

4.1. The selective attention chip

The selective attention chip contains a 2D array of 8×8 cells layed out on a square grid. Each cell comprises an excitatory synapse, an inhibitory synapse, a WTA cell of the type described in Section 2, an output I&F neuron, and two position-to-voltage (P2V) circuit drivers. A detailed description of these circuits, together with quantitative analysis and a description of their response properties is described in [9]. The excitatory synapse of each WTA cell receives off-chip address events and integrates them into an excitatory current I_{ex}. The cell in the WTA network that wins the competition supplies a current to the P2V circuits and to its output neuron. The neuron's spikes used to transmit the pixel's address off chip are also integrated by the WTA cell's inhibitory synapse into an inhibitory current I_{ior} (to implement the inhibition of return mechanism). As the integrated inhibitory current I_{ior} increases, the cell's net input current ($I_{ex} - I_{ior}$) decreases. As soon as this net input current decreases below the value of a net input current exciting a different cell, the WTA network switches state and selects a different cell as the winner. When the old winning cell is de-selected, its corresponding output neuron stops firing and its inhibitory synapse recovers, decreasing the inhibitory current I_{ior} back to zero. Depending on the time constants and strength of the excitatory and inhibitory

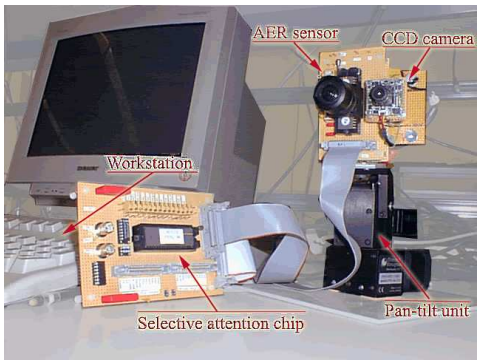

Figure 5: Multi-chip active vision system: A neuromorphic vision sensor mounted on a pan-tilt unit provides input to a selective attention chip, that controls the sensor's movements through a workstation.

synapses, on the input stimuli and on the frequency of the output neuron, the WTA network will switch the selection of the winner between the largest input and the next-largest, or between the largest and more inputs of successively decreasing strength, generating focus of attention *scanpaths*. Figure 4 shows an example of a focus of attention scanpath measured from the selective attention chip.

4.2. Active vision system

The active vision system that uses the selective attention chip is shown in Figure 5: the selective attention chip receives input from an AER imaging sensor [7], and transmits the address of the winning pixel to a workstation, that is used to drive the pan-tilt unit on which the sensor is mounted. A standard CCD camera is mounted next to the sensor, to visualize the sensor's field of view. The AER sensor responds to contrast transients and its address events report the position of moving objects. The selective attention chip selects the locations with highest contrast moving objects and cycles through them, while the workstation drives the pan-tilt unit centering the selected locations with the sensor's imaging array. An example of behavior of such a system, in response to real-world stimuli is shown in Figure 6. The system is described in detail in [10].

5. CONCLUSIONS

We presented examples of neuromorphic selective attention systems that use current-mode WTA circuits to implement the competitive selective mechanism. These systems demonstrate how VLSI-based selective attention mechanisms can be useful in engineering applications. We are currently designing more realistic models of selective attention, implemented using multi-chip AER networks of I&F neurons to investigate if they can be useful tools also for basic research in computational neuroscience.

Figure 6: Sensor view just before the movement command (left image): the focus of attention has just switched to the top left corner. Sensor view just after the camera movement (right image): the focus of attention is now centered with the salient stimulus.

6. REFERENCES

[1] L. Itti, E. Niebur, and C. Koch, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[2] 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.

[3] 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.

[4] T. G. Morris, T. K. Horiuchi, and S. P. DeWeerth, "Object-based selection within an analog VLSI visual attention system," *IEEE Trans. on Circuits and Systems II*, vol. 45, no. 12, pp. 1564–1572, 1998.

[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] 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.

[7] J. Kramer, "An ON/OFF transient imager with event-driven, asynchronous readout," in *Proc. IEEE International Symposium on Circuits and Systems*, May 2002.

[8] G. Indiveri, P. Oswald, and J. Kramer, "An adaptive visual tracking sensor with a hysteretic winner-take-all network," in *Proc. IEEE International Symposium on Circuits and Systems*, May 2002, pp. 324–327.

[9] G. Indiveri, "A neuromorphic VLSI device for implementing 2-D selective attention systems," *IEEE Trans. on Neural Networks*, vol. 12, no. 6, pp. 1455–1463, November 2001.

[10] G. Indiveri, R. Mürer, and J. Kramer, "Active vision using an analog VLSI model of selective attention," *IEEE Trans. on Circuits and Systems II*, vol. 48, no. 5, pp. 492–500, May 2001.