Active Vision Using an Analog VLSI Model of Selective Attention

Giacomo Indiveri, Robin Mürer, and Jörg Kramer, *Member, IEEE*

***Abstract*—Detailed processing of sensory information is a computationally demanding task. This is especially true for vision, where the amount of information provided by the sensors typically exceeds the processing capacity of the system. Rather than attempting to process all the sensory data simultaneously, an effective strategy is to focus on subregions of the input space, shifting from one subregion to the other, in a serial fashion. This strategy is commonly referred to as *selective attention.* We present a neuromorphic active-vision system, that implements a saliency-based model of selective attention. Visual data is sensed and preprocessed in parallel by a transient imager chip and transmitted to a selective-attention chip. This chip sequentially selects the spatial locations of salient regions in the vision sensor's field of view. A host computer uses the output of the selective-attention chip to drive the motors on which the imager is mounted, and to orient it toward the selected regions. The system's design framework is modular and allows the integration of multiple sensors and multiple selective-attention chips. We present experimental results showing the performance of a two-chip system in response to well–controlled test stimuli and to natural stimuli.**

***Index Terms*—Address-event representation, neuromorphic, selective attention, winner-take-all.**

I. Introduction

REAL-TIME processing of detailed sensory information is a computationally demanding task for both biological and artificial systems. Rather than attempting to simultaneously process all the information provided by the input sensors, an effective strategy is to select subregions of the input, and process them, shifting from one subregion to another in a serial fashion. In biology, this strategy is commonly referred to as *selective attention* [1]–[3]. The selection of the subregions appears to be driven by a competitive mechanism that facilitates the emergence of a winner from several potential targets, allowing the system to process information relevant to current goals, while suppressing the irrelevant information that the system cannot analyze simultaneously [1]. In the mammalian visual system, selective attention plays a major role in determining where to center the fovea (the region of the retina with the highest density of receptors) with respect to the subregion of interest in the visual field [4].

Recent theories suggest that the selection mechanism can be modulated by stimulus-driven and goal-driven factors [5]. Stimulus-driven attention appears to act as a rapid, bottom-up, task-independent mechanism, while goal-driven selective-attention appears to act in a slower top-down volition-controlled manner. In this paper, we present a real-time active-vision system that uses a stimulus-driven selective-attention mechanism to sequentially analyze subregions of the input space, by choosing a location for the focus of attention and changing its imaging sensor's direction of gaze accordingly. To respond to visual stimuli and select the location for the focus of attention in real time, we used two fully custom analog very large scale integration (VLSI) neuromorphic chips: an irradiance transient sensor array [6], and a selective-attention signal processing array [7]. As the characteristics of the individual chips are described in detail elsewhere [6]–[8], in this paper we focus on the description of the overall architecture, and on system-level issues such as interchip communication, motor-control strategies, and system behavior. The active-vision system that we present demonstrates how multichip networks can be interfaced effectively, and carry out computation in parallel; using an event-based asynchronous communication infrastructure which employs an *address-event representation* (AER) [9]–[11]. Multichip systems that exchange information using the AER have already been proposed [12], [13], but the one presented in this article is one of the first to use an AER transceiver (the selective-attention chip) that both receives address events, and transmits them to further processing stages. Several VLSI systems for implementing *visual* selective-attention mechanisms have also been presented [14]–[17], [8]. These systems contain photosensing elements and signal–processing elements on the same focal plane, and apply a competitive selection process to visual stimuli sensed and processed by the focal plane processor itself. Unlike these systems, the one proposed here uses a selective-attention chip able to receive input signals from any type of AER device. Therefore, input signals need not arrive only from visual sensors, but could represent a wide variety of sensory stimuli obtained from different sources; such as silicon retinas [11], silicon cochleas [18], or additional AER signal processing chips. In this paper, we show how the ability of the selective-attention chip to receive and transmit signals, using the same representation as AER neuromorphic sensors, allows us to design multichip hierarchical selective-attention systems able to interact with the real world in real time.

II. Saliency-Based Model of Selective Attention

Several computational models of selective attention have been proposed [19], [20], [5], [2], [3]. Some of these models are based on the concept of "dynamic routing" [19], by which salient regions are selected by dynamic modification of network parameters (such as neural connection patterns), under both

Manuscript received December 1999; revised May 2001. This work was supported by a Swiss National Science Foundation SPP Grant and the U.S. Office of Naval Research.

The authors are with the Institute of Neuroinformatics, University of Zürich/ETH Zürich, CH 8057 Zürich, Switzerland.

Publisher Item Identifier S 1057-7130(01)06237-1.

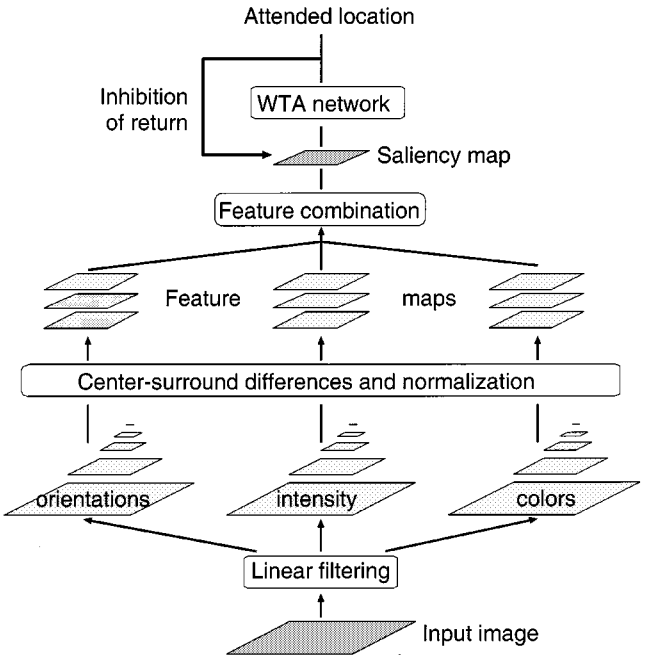
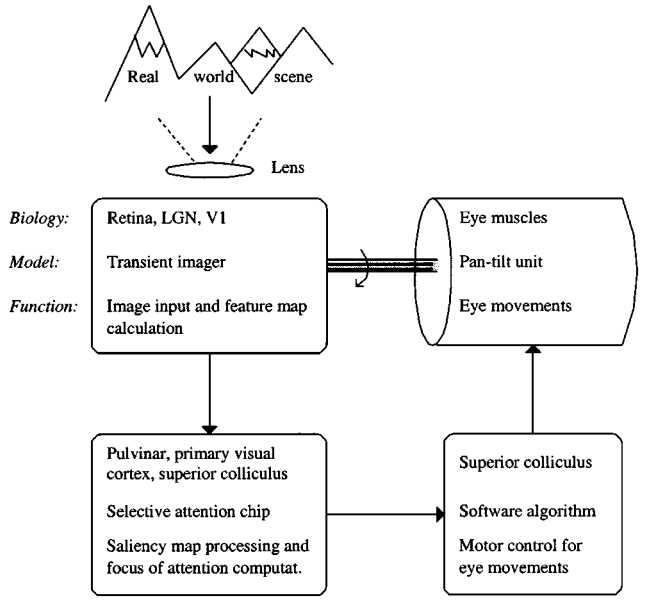


Fig. 1. Schematic diagram of a saliency–based model of selective-attention.

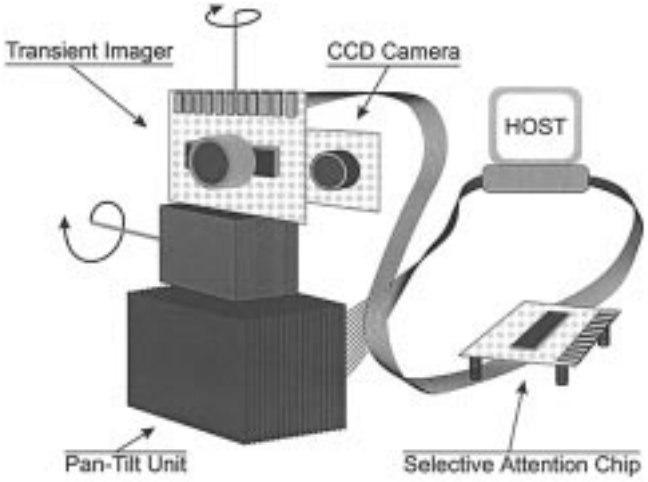
top-down and bottom-up influences. Some other models, based on similar ideas, promote the concept of "selective tuning" [20]. In these models, attention optimizes the selection procedure by selectively tuning the properties of a top-down hierarchy of winner-take-all (WTA) processes embedded within the visual processing pyramid.

The model we seek to implement in hardware is the one based on the concept of the "saliency map," originally put forth by Koch and Ullman [21]. This biologically plausible model accounts for many of the observed behaviors in neurophysiological and psychophysical experiments, and has led to several software implementations applied to machine vision and robotic tasks [22]–[25]. This model is especially appealing to us because it lends itself nicely to hardware implementation; due to its modularity and ease of expandability.

A diagram describing the main processing stages of the model is shown in Fig. 1. A set of topographic feature maps is extracted from the visual input. All feature maps are normalized and combined into a master *saliency map*, which topographically codes for local saliency over the entire visual scene. Different spatial locations then compete for largest saliency, based on how much they stand out from their surroundings. A WTA circuit selects this most salient location as the focus of attention. The WTA circuit is endowed with internal dynamics, which generate the shifts in attention based on a mechanism named *inhibition of return* (IOR) (a key feature of many selective-attention systems) [26]. As explained in Section III, the hardware system proposed in this paper only implements a subset of the model of Fig. 1. However, thanks to the modularity of the original model and to the flexibility offered by the AER communication protocol, this hardware system can scale up to arbitrarily complex selective-attention systems. We will argue in Section V that this is achievable by combining multiple instances of the presented chips, computing multiple sets of feature maps, merging



(a)



(b)

Fig. 2. (a) Block diagram of the sensory-motor selective-attention model. The figure shows the basic computational blocks used, as well as the corresponding biological analogs and their function. (b) Schematic diagram of the active-vision setup. The neuromorphic imager, mounted on a pan-tilt unit, transmits its output to the selective-attention chip. The latter sends the results of its computations to a host computer which uses this data to drive the pan-tilt unit's motors.

and normalizing feature maps into a saliency map, and implementing both the WTA and IOR mechanisms.

III. System Description

A. Overview

A block diagram of the selective-attention sensory-motor system and the correspondence between the system's computational blocks and their biological counterparts is shown in Fig. 2(a). A schematic diagram illustrating how the individual components are connected together is shown in Fig. 2(b). At the input stage, we use a neuromorphic imager that is sensitive to temporal changes in illumination (*transients*), and extracts motion or flicker as features. Since our system, in its current state, only extracts one feature map, the saliency map is identical to

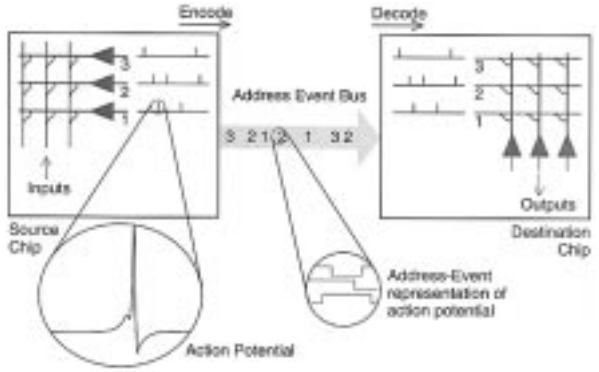


Fig. 3. Schematic diagram of an AER chip–to– chip communication example: "slow" action potentials generated by the "many" (N) neurons of the source chip are transmitted as "fast " address events, on the "few" ($\log N$) wires of the AE bus, to the synapses of the destination chip neurons.

the extracted feature map. In this case, no feature combination stage is necessary; the transient imager chip transmits its output data directly to the selective-attention chip.

Based on its inputs, the selective-attention chip computes the location of the focus of attention, and sends address events encoding this location to the host computer. In addition to managing the communication with the selective-attention chip, using the AER communication protocol, the host computer is used for data logging and, more importantly, for driving the motors of a commercial pan-tilt unit (Directed Perception, Inc., PTU 46–17.5) on which the transient imager is mounted [see Fig. 2(b)]. The pan-tilt unit is used to orient the imager chip such that the location of the focus of attention lies in its central region.

B. AER

The proposed multichip system can process signals in parallel, in real time, while exchanging data according to the asynchronous AER communication protocol. In this protocol, input and output signals are transmitted as asynchronous binary data streams which carry the analog information in their temporal structure, very much like natural spike trains do in biological systems (see Fig. 3). The time intervals between events are continuous in nature. Each event is represented by a binary word encoding the address of the sending node. In the case of single-sender/single-receiver point-to-point communication, a simple handshaking mechanism ensures that all events generated at the sender side arrive on the receiver side [9], [11]. The address of the sending element is conveyed in parallel, along with two handshaking control signals. Systems containing more than two AER chips can be constructed by implementing additional special purpose off-chip arbitration schemes [13], or using a framework such as the *silicon cortex* (SCX) [10], [12], which is a general, fully arbitered, address event communication infrastructure.

The system proposed here uses single-sender/single-receiver point-to-point communication. The sender chip contains a two-dimensional pixel array, with an arbiter that serially processes the requests from the different pixels in the order of their activation, latches their addresses onto the AER communication bus in the same order, and sends acknowledge pulses to the corresponding pixels [27]. As soon as a new address is ready on the bus, the handshaking cycle with the receiver chip is initiated, in the course of which the address of the sending pixel is transmitted.

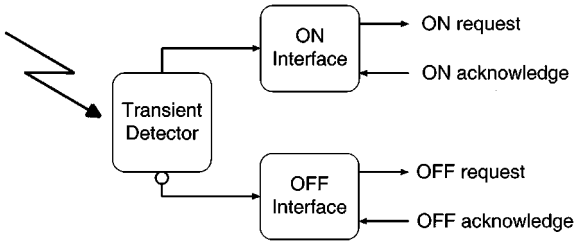


Fig. 4. Block diagram of irradiance transient detector with event-based communication interface.

The transient imager transmits its address events to the selective-attention chip using a topographic mapping. As the sender has 16×16 pixels and the receiver only 8×8 we map the addresses of 2×2 neighboring pixels on the sender to the same pixel on the receiver. This mapping was accomplished by simply discarding the least significant bit of the sender address, for each dimension.

C. Transient Imager Chip

The transient imager is a 16×16 pixel array of irradiance transient detectors that is used to generate the events that drive the system. Each pixel responds with binary pulses in real time to a local change of a brightness distribution projected through a lens onto its surface. These pulses are used as the request signals to the AER communication interface. Fig. 4 shows a block diagram of the pixel circuitry. The transient detector comprises an adaptive photoreceptor [28] with a rectifying temporal differentiator [29] in the feedback loop. Positive irradiance transients, corresponding to dark-to-bright or ON transitions, and negative irradiance transients, corresponding to bright-to-dark or OFF transitions, appear at different output terminals. The ON and OFF responses are separately amplified with tunable gains, each generating a request pulse to the on-chip arbiter if it exceeds a chosen threshold. By appropriately setting the threshold and the respective gain factors, the circuit can be made to respond only to ON transients or only to OFF transients or to both types of transients. Each acknowledge pulse from the arbiter triggers a reset pulse at the requesting terminal, whose duration determines a refractory period for the succeeding request from the same terminal. Depending on the chosen refractory period and the magnitude and duration of the irradiance transient, the pixel responds with a single spike or a burst of spikes. In the present application, a short refractory period of 140 μs was chosen to obtain bursts, and only the OFF response was used to stimulate the selective-attention chip.

The pixels are arranged on a square grid. The position of a pixel along a row is encoded with a 4-bit column address and its position along a column with a 4-bit row address. An additional address bit is used to distinguish between ON and OFF transients. The details of this circuit are described in [6].

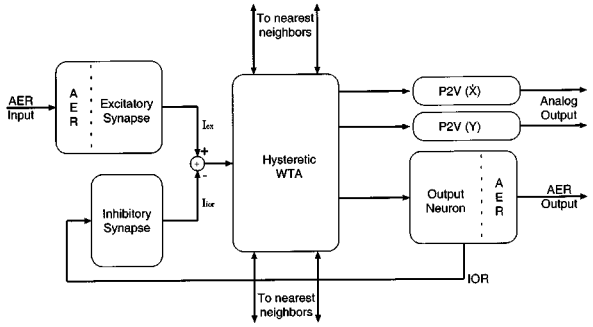


Fig. 5. Block diagram of a basic cell of the 8 × 8 selective-attention architecture.

Fig. 6. Image captured from the CCD camera mounted next to the transient imager. The outer frame shown in the image corresponds to the field of view of the transient imager, whereas the inner frame is drawn to evidence the transient imager's central region. The cross to the bottom right of the image center represents the location of the focus of attention currently computed by the selective-attention chip.

D. The Selective-Attention Chip

The selective-attention chip contains an array of 8×8 cells laid out on a square grid. Fig. 5 shows a block diagram of the cell circuitry. Each cell comprises an excitatory synapse, an inhibitory synapse, a hysteretic WTA cell [30], a local inhibitory output neuron [7], and two position-to-voltage (P2V) circuits [31]. The P2V circuits produce two analog output voltages encoding the x coordinate and the y coordinate of the winning cell. The excitatory synapse is a current-mirror integrator [27] interfaced to the input AER circuitry. It receives off-chip address events, and integrates them into an excitatory current $I_{\rm ex}$. The inhibitory synapse is a similar circuit that integrates the on-chip spikes of the same cell's output neuron into an inhibitory current $I_{\rm ior}$. The synaptic currents, $I_{\rm ex}$ and $I_{\rm ior}$, are subtracted and sourced into the input node of the WTA cell (see Fig. 5). A detailed description of these circuits, together with quantitative analysis and a description of their response properties has been recently presented in [7]. A comprehensive description of the response properties of the hysteretic WTA network as a function of circuit bias parameters and input signals can be found in [30].

In the selective-attention chip, each hysteretic WTA cell is connected to its four nearest neighbors, both with excitatory connections and inhibitory connections. The strength of the lateral inhibitory connections controls the spatial extent over which competition takes place. If lateral inhibition is maximally turned on, all WTA cells of the architecture are connected together (global inhibition) and only one winner can be selected at a time. If, on the other hand, the lateral inhibitory connections are weakened (local inhibition), the WTA cells are partially decoupled from each other and the network can select multiple (spatially distant) winners simultaneously. Similarly, the lateral excitatory connections control the amount of lateral facilitatory coupling between cells. If lateral excitation is enabled, the system tends to select new winners in the immediate neighborhood of the currently selected cell.

The winning cell supplies a current to the position-to-voltage row and column circuits. It also sources a dc current into a neuron connected to it. Each action potential generated by this neuron produces an address event. The amplitude of the injection current (and hence the frequency of the address events) is independent of the WTA's cell input.

In addition to transmitting the pixel's address off-chip, the output neuron is instrumental for implementing the inhibition of return (IOR) mechanism. The spikes generated by the winning cell's output neuron are integrated by the cell's inhibitory synapse. As the integrated inhibitory postsynaptic current $I_{\rm ior}$ increases, the cell's net input current $I_{\rm ex} - I_{\rm ior}$ 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 the new cell as the winner. When the old winning cell is deselected, its corresponding local output neuron stops firing, and its inhibitory synapse recovers, decreasing the inhibitory current $I_{\rm ior}$ back to zero. Depending on the time constants and strength of the excitatory and inhibitory synapses, on the input stimuli and on the frequency of the output neuron, the WTA network can exhibit different dynamic properties. The selection of the winner can switch between the largest input and the next-largest, or between the largest and more inputs of successively decreasing strength, generating focus of attention *scanpaths* [32]. Quantitative measures characterizing these scanpaths and comparisons to other selective-attention circuits/IOR mechanisms have been described in [7] and [8].

E. Motor Control Algorithm

The control algorithm that the host computer executes, is responsible for driving the motors of the pan-tilt unit in such a way as to center the location picked by the selective-attention chip within the central region of the transient imager chip. This algorithm represents a first attempt at modeling the bottom-up stimulus driven neural mechanism that generates saccadic eye movements which center the fovea with respect to the location of the focus of attention.

To evaluate, quantitatively, the response properties of the system, and test the motor control algorithm, we mounted a standard charge-coupled device (CCD) camera next to the transient imager chip and captured images on the host computer [see also Fig. 2(b)]. This allowed us to see in real-time the images projected onto the focal plane of the transient imager chip, as shown in Fig. 6. We calibrated the system so that the

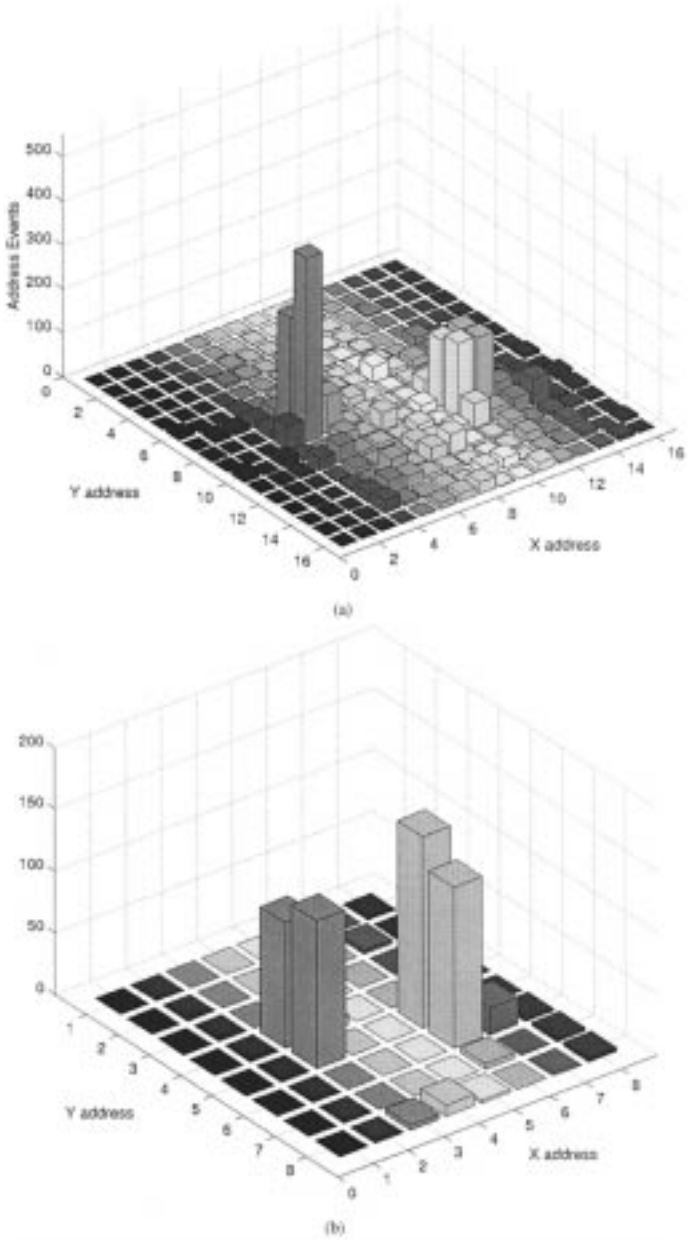


Fig. 7. (a) Histogram of events generated by the transient imager pixels in response to two diffused flashing LEDs. The LED stimulating the region around pixel (5,9) has higher contrast than the other LED. (b) Histogram of events generated by the selective-attention chip in response to the events generated by the transient imager chip.

image projected onto the transient imager array, corresponds to the central part of the image captured by the CCD camera, shown as the outer square in the center of Fig. 6. The inner square drawn in the center of Fig. 6 represents the part of the scene being projected on the central 4×4 region of the transient imager array. The location selected by the selective-attention chip is represented by a small cross, superimposed onto the CCD image.

The control algorithm produces motor commands that depend on the current position of the selected location, and its recent history; if the cross lies within the inner square of the image, no camera movements are triggered (the camera is already "foveating" the salient feature). If the cross shifts to a location outside the inner frame, the algorithm records the address of the location and increases a counter associated with that address. As soon as the counter for a particular address reaches a threshold n (i.e., when the cross revisits the same location n times), the algorithm generates a camera movement that centers the selected location within the central region of the transient imager array (the camera "saccades" to the persistent salient stimulus). In this method, camera movements are generated only if a salient location is visited more than once. The revisiting constraint ensures that the system does not saccade to all locations picked by the selective-attention chip, but orients its gaze only toward persistent salient stimuli. In the examples shown in Section IV, n was set to 5. The value of n was chosen to reproduce the characteristics of biological selective-attention systems, as reported in the neuroscience literature [3]; while the focus of attention shifts 15–20 times per second, saccadic eye movements are only made 3–5 times per second [3].

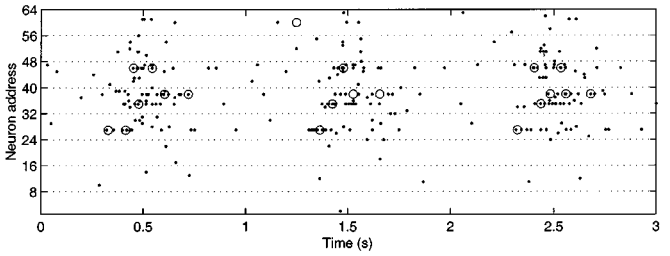


Fig. 8. Raster plot of the activity of the neurons of both transient imager chip (dots) and selective-attention chip (circles) in response to the flashing LEDs. To plot the data from both chips using an address space with the same resolution, we subsampled the addresses of the transient imager chip. The LEDs flashed approximately at 0.25, 1.25, and 2.25 s.

Another important function implemented by the motor control algorithm is that of *saccadic suppression.* During a camera movement the images projected on the focal plane of the transient imager array generate a large amount of address events. These events are not relevant for the analysis of the scene once the camera stops moving. In biology, this problem is solved by suppressing all inputs arriving from the retinas during saccadic eye movements (indeed, we are effectively blind during a saccade). In the current version of our system, the addresses generated by the transient imager chip are hardwired into the selective-attention chip [see Fig. 2(b)]. There is no way of suppressing these events at the source. During a camera movement, the selective-attention chip receives and processes all spurious events from the imager, and the addresses generated by the selective-attention chip are transmitted to the host computer. The control algorithm ignores the effect of these events, by resetting all address counters to zero after each camera movement. In this way, the recent history of all selected positions is canceled, and normal operation of the control algorithm can be resumed.

IV. Experimental Results

A. System Response in Absence of Camera Movements

Initially, we tested the system with the motors of the pan-tilt unit turned off. The input images consisted of a laboratory scene with two flashing LEDs in the foreground. The two LEDs were blinking in phase, with a frequency of 1 Hz and a duty cycle

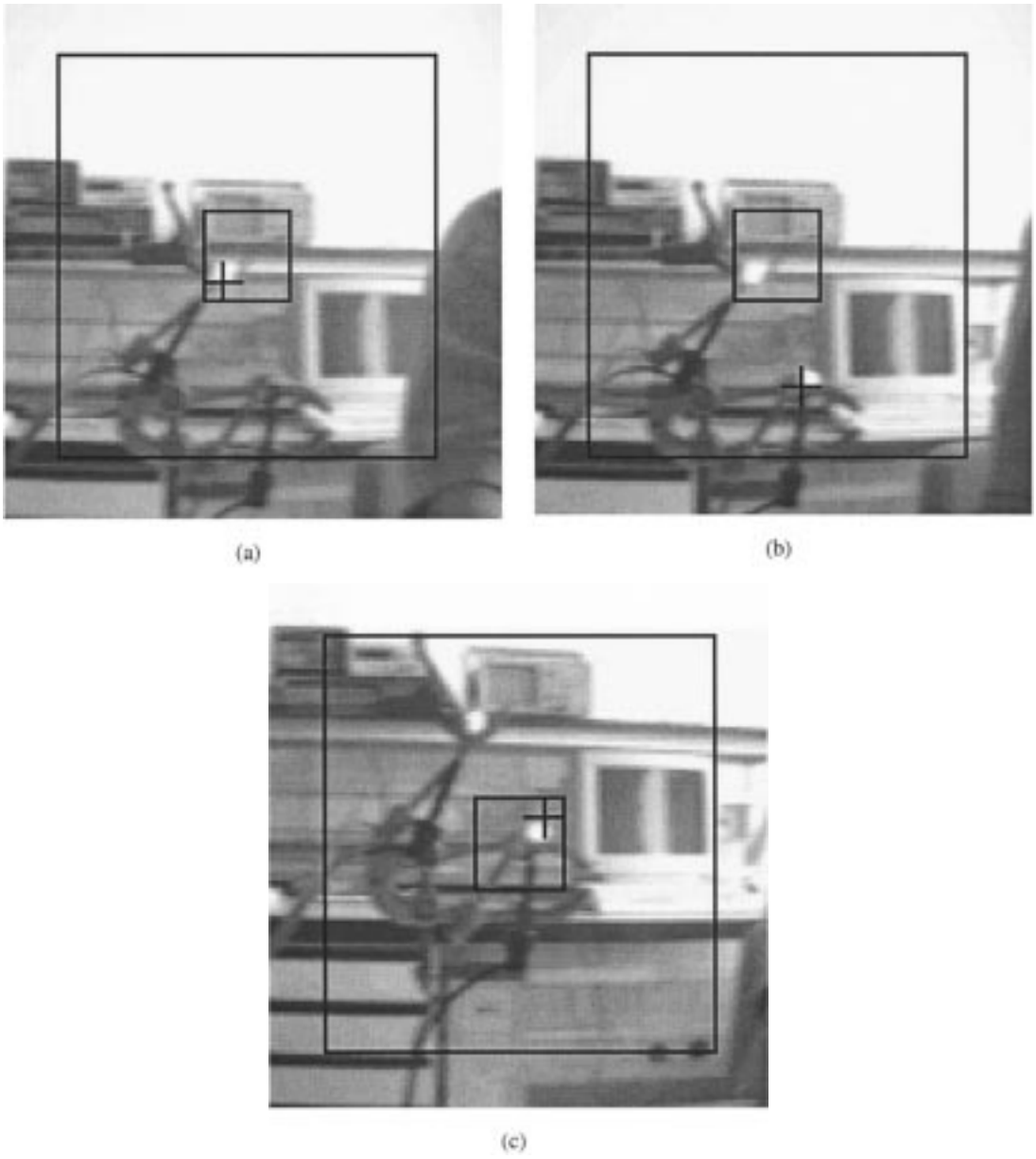


Fig. 9. Sequence of images showing the selection of a salient stimulus prior to and after a saccadic eye movement. (a) The system is attending the top LED, already centered on the central part of the imaging array. (b) The system selects the bottom LED, outside the central region of the imager. (c) The system performed a saccade toward the bottom LED, and is currently attending it.

of 50%. As the transient imager responds only to local changes in illumination, the blinking LEDs proved to be a reliable and well–controlled stimulus. The static background did not contribute to the generation of address events. We placed a diffusion glass in front of the transient imager's lens, to diffuse the projection of the two LEDs on the imager's focal plane. In this way we were able to stimulate several pixels of the imaging array with each LED. Fig. 7(a) shows the histogram of the address events generated by the transient imager array in response to the flashing LEDs, captured over a period of 2 s. The two regions with the highest occurrence of events [around pixels (5,9) and (11,11)] correspond to the locations of the LEDs. Fig. 7(b) shows the histogram of address events generated by the selective-attention chip. On average, the selective-attention chip visited pixels (3,5), (3,4) and (6,6), (6,5) most often.

While the event histogram shows that the selective-attention chip acts *on average* like a threshold filter, picking only inputs with a high mean frequency, it does not show the more interesting aspect of the computation carried out by the chip—its dynamics. To show the dynamical aspect of the selective-attention chip's response, we plotted in Fig. 8 a *raster plot.* This plot shows the activity of the transient imager and of the selective-attention chip neurons over time, in response to the flashing LEDs. The 8×8 neurons of the selective-attention chip are labeled successively, row by row (1–64), and the events that they generated are plotted with circles. To show the events of the transient imager pixels on the same scale, we subsampled their addresses, taking into consideration only their three most significant bits (in the same way we implemented the mapping of addresses from the transient imager pixels to the selective-attention ones,

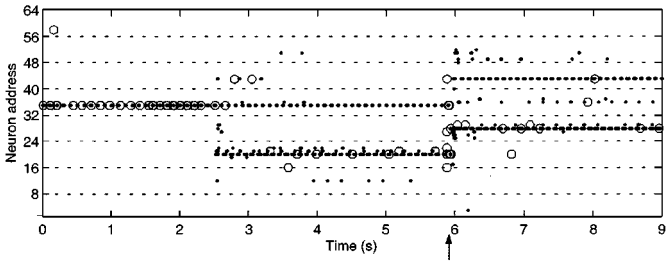


Fig. 10. Raster plot of the activity of the neurons of the transient imager chip (dots) and of the selective-attention chip (circles) in response to two flashing LEDs. The focus of attention shifts from a central region of the imaging array to a peripheral one (see circles at 2 s $\leq t <$ 6 s). Consequently, the system makes a camera movement, at the time indicated by the vertical arrow, and re-centers the attended location.

as described in Section III-B). The high density of events around time instants 0.5, 1.5, and 2.5 s is due to the flashing of the LEDs. Within a single flash, the focus of attention shifts approximately four times, moving from one region of high saliency to another. The proportion between events generated by the two chips is consistent with the data of Fig. 7. By looking at the selective-attention chip data of Fig. 8 one can extrapolate the focus of attention's scanpaths. Note how these scanpaths tend to repeat themselves over time. This characteristic will be even more evident in Section IV-C, when we analyze the response of the system to natural stimuli.

B. System Response in Presence of Camera Movements

To allow the system to make camera movements, we activated the motors of the pan-tilt unit on which the imager was mounted. The input stimulus consisted again of two flashing LEDs, but this time not in phase. Furthermore we removed the diffusion filter from the transient imager's lens, so that the two LEDs stimulated only a few pixels of the imaging array. As described in Section III-E, the selective-attention chip was driving the pan-tilt unit to orient the imager toward the attended location. Fig. 9 shows a sequence of images captured by the CCD camera mounted on the pan-tilt unit, while the system was engaged in selecting and tracking the LEDs. Initially, only the top LED was flashing, and the system selected it and oriented the central region of the imager to that location [see Fig. 9(a)]. As we turned on the bottom LED, the system changed the focus of attention location [see Fig. 9(b)] and made a camera movement centering the attended stimulus on the central region of the imager [see Fig. 9(c)].

The raster plot of Fig. 10 shows in detail the sequence of events that lead to the camera movement. The arrangement of the neuron addresses on the figure axis is the same as in Fig. 8. Initially, the selective-attention chip was attending the region of transient imager pixels that project to its 35th pixel. As the second LED flashed, the imager pixels excited also the 20th selective-attention chip pixel. After approximately 1 s, the WTA network of the selective-attention chip switched and selected the second LED as the winner. After having attended to that location for approximately 2.5 s, the system made an abrupt camera movement (saccade), and centered the attended stimulus on the imaging array.

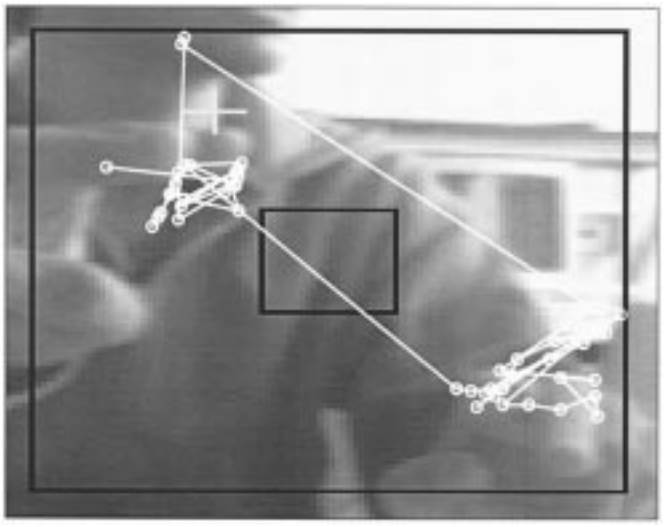


Fig. 11. Output of the P2V circuits of the selective-attention chip (see Section III-D, Fig. 5), representing the scanpath of the focus of attention, switching back and forth between the fluttering fingers of both of the experimenter's hands. The scanpath data is superimposed onto a snapshot taken from the CCD camera during the experiment.

C. System Response to Natural Stimuli

In this section, we show how the system is able to select and attend natural stimuli that were not explicitly engineered to optimally drive the imaging array. As we did in Section IV-A and -B, we initially tested the system in the absence of camera movements and subsequently tested it with the motor output activated.

Fig. 11 shows the location of the focus of attention, as measured by the P2V circuits of the selective-attention chip (see Section III-D), in response to the fluttering fingers of the experimenter, over a period of 500 ms. The x component and y component of the focus of attention are plotted against each other, and superimposed onto an image taken by the CCD camera during the experiment. Although the resolution of the selective-attention chip is 8×8 pixels, the data of Fig. 11 seems to belong to a much higher resolution architecture. This is due to the fact that the output of the P2V circuits is analog and is affected by noise [7]. These analog output signals might not be appropriate for precise quantitative measurements, but could be used to drive, via buffers or power amplifiers, motors and actuators to implement (negative feedback) sensory-motor loops [7].

Fig. 12 shows the response of the system to the same stimulus as Fig. 11, with the motors engaged. Fig. 12(a) shows the beginning of the experiment; the motors had just been activated, the imager was still in its initial position, and the selective-attention chip chose a pixel in the top left region of the transient imager array as the focus of attention. After the selective-attention chip transmitted the same pixel address to the host computer for a set number of times, specified by the motor control algorithm (see Section III-E), the control algorithm generated a camera movement and centered the focus of attention with respect to the transient imager array [see Fig. 12(b)]. If the salient stimuli were persistent (e.g., if the fingers kept on moving) and remained in the field of view of the imager, the system continuously shifted its gaze from one salient stimulus to the other.

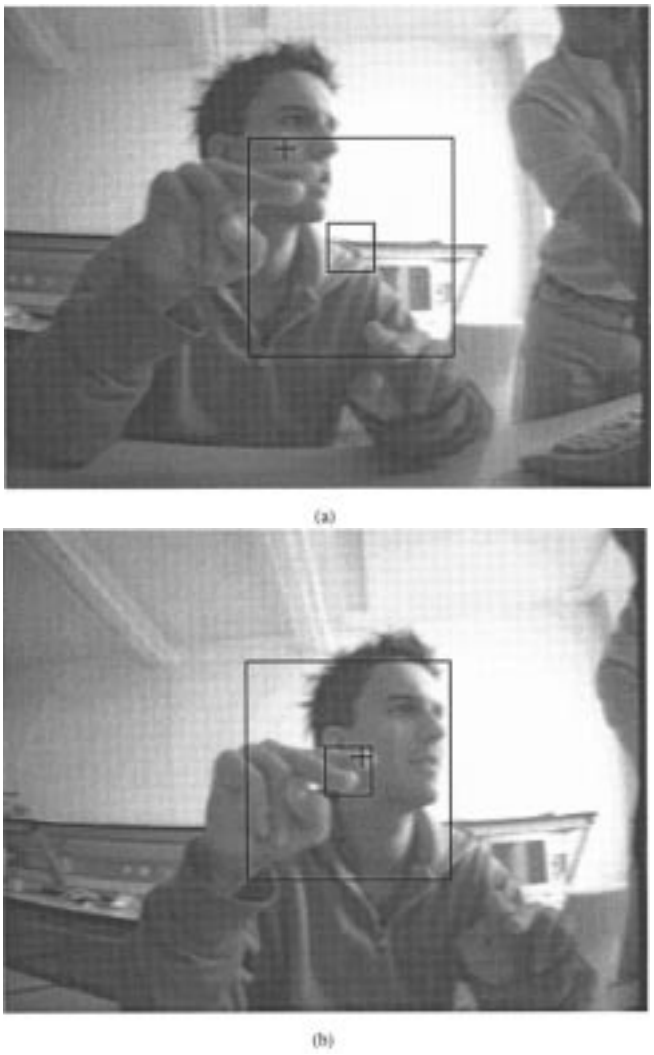


Fig. 12. Saccadic eye movements in response to moving fingers. (a) CCD camera snapshot taken before the saccadic eye movement (the focus of attention has just switched from one hand to the other). (b) CCD camera snapshot taken just after the saccadic eye movement (the focus of attention and the salient stimulus are now in the center of the imaging array).

This behavior has proven to be extremely reliable and robust. The system's response is largely invariant to illumination conditions, stimulus speed and (static) background conditions.

V. Conclusions and Outlook

We have presented a neuromorphic active-vision system comprising an imaging sensor, a selective-attention chip, and a motorized pan-tilt unit using asynchronous address event communication. The system sequentially selects the spatial locations of the most salient inputs present in the sensor's field of view, and orients the sensor toward them. It implements a simplified model of the bottom-up stimulus driven selective-attention mechanism of primates. The selective-attention chip accepts input signals in the form of address events and transmits output signals using the same representation. We demonstrated the behavior of the system using both "well-controlled" LED stimuli, and more natural types of stimuli.

Although the present system has only one visual sensor, and one selective-attention chip, its design framework allows the integration of additional AER devices for implementing selective-attention systems of arbitrary complexity. The scheme of Fig. 1 could be implemented using one single neuromorphic imager, interfaced to multiple instances of the selective-attention chip, via different connectivity mappings (see Section III-B). Several AER input sensors (silicon retinas, cochleas, etc.) could also be employed, and interfaced to additional instances of the selective-attention chip. At the first level of the hierarchy, these selective-attention chips would be biased to choose several winners simultaneously (using the local inhibitory connections of the WTA network), and produce normalized feature maps (as mentioned in Section III-D, the frequency of the spikes of the winning neuron in the selective-attention chip does not depend of the amplitude of its input signals). The normalized outputs would then converge into a top selective-attention chip, at the second level of the hierarchy. This attention chip would then determine the spatial location of the most salient stimulus, taking into account all sensory modalities used in the first level, and eventually drive the motor control components of the system. The inherent characteristics of neuromorphic sensors and of the AER ensure that the transduction of sensory signals, the competition within single-sensory modalities, and the competition across sensory modalities would all take place in parallel. The (complex) dynamics present at all levels of the hierarchy would run in continuous and in *real* time.

Equivalent selective-attention systems implemented, using conventional machine vision technology, are not able to perform these types of computations in real-time, even using powerful workstations [22]. The real-time nature of the computation carried out by neuromorphic multichip systems, and the flexibility that they offer (e.g., by selecting different sets of bias voltages for the individual chips, and by exploiting the AER to map/remap connectivity patterns) provide obvious advantages both for scientific investigation of selective-attention system properties and for engineering applications.

Acknowledgment

Fabrication of the integrated circuits was provided by MOSIS. The authors are grateful to R. Douglas and K. Martin for providing an outstanding intellectual environment at the Institute of Neuroinformatics, Zürich, Switzerland.

References

[1] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annu. Rev. Neurosci.*, vol. 18, pp. 193–222, 1995.
[2] M. C. Mozer and M. Sitton, "Computational modeling of spatial attention," in *Attention*, H. Pashler, Ed. East Sussex, U.K.: Psychology Press, 1998, pp. 341–395.
[3] E. Niebur and C. Koch, "Computational architectures for attention," in *The Attentive Brain*, R. Parasuraman, Ed. Cambridge, MA: MIT Press, 1998, pp. 163–186.
[4] J. Hoffman and B. Subramaniam., "The role of visual attention in saccadic eye movements," *Percep. Psychophys.*, vol. 57, no. 6, pp. 787–795, 1995.
[5] M. Behrmann and C. Haimson, "The cognitive neuroscience of visual attention," *Current Opinion Neurobiol.*, vol. 9, pp. 158–163, 1999.
[6] J. Kramer, "An integrated optical transient sensor," *IEEE Trans. Circuits Syst. II*, submitted for publication.

[7] G. Indiveri, "A 2D neuromorphic VLSI architecture for modeling selective attention," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks; IJCNN2000*, S.-I. Amari, C. L. Giles, M. Gori, and V. Piuri, Eds. Los Alamitos, CA: IEEE Computer Society, 2000, vol. IV, pp. 208–213.
[8] ——, "Modeling selective attention using a neuromorphic analog VLSI device," *Neural Comput.*, vol. 12, no. 12, pp. 2857–2880, Dec. 2000.
[9] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti, and D. Gillespie, "Silicon auditory processors as computer peripherals," *IEEE Trans. Neural Networks*, vol. 4, pp. 523–528, May 1993.
[10] S. R. Deiss, R. J. Douglas, and A. M. Whatley, "A pulse-coded communications infrastructure for neuromorphic systems," in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds. Cambridge, MA: MIT Press, 1998, ch. 6, pp. 157–178.
[11] K. A. Boahen, "Multiple pathways: retinomorphic chips that see quadruple images," in *Proceedings of the 7th International Conference on Microelectronics for Neural, Fuzzy and Bio-inspired Systems; Microneuro'99*. Los Alamitos, CA: IEEE Computer Society Press, Apr. 1999, pp. 12–20.
[12] G. Indiveri, A. M. Whatley, and J. Kramer, "A reconfigurable neuromorphic VLSI multi-chip system applied to visual motion computation," in *Proc. 7th Int. Conf. Microelectronics Neural, Fuzzy and Bio-Inspired Systems; Microneuro'99*. Los Alamitos, CA: IEEE Computer Society Press, Apr. 1999, pp. 37–44.
[13] C. Higgins and C. Koch, "A modular multi-chip neuromorphic architecture for real-time visual processing," *J. Analog Integrated Circuits Signal Process.*, vol. 26, no. 3, pp. 195–211, Sept. 2000.
[14] 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: IEEE Computer Society Press, 1999, vol. 20.
[15] T. K. Horiuchi and C. Koch, "Analog VLSI-based modeling of the primate oculomotor system," *Neural Comput.*, vol. 11, pp. 243–265, 1999.
[16] T. G. Morris, T. K. Horiuchi, and S. P. DeWeerth, "Object-based selection within an analog VLSI visual attention system," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 1564–1572, Dec. 1998.
[17] V. Brajovic and T. Kanade, "Computational sensor for visual tracking with attention," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1199–1207, Aug. 1998.
[18] E. Fragnière, A. van Schaik, and E. Vittoz, "Design of an analogue VLSI model of an active cochlea," *J. Analog Integrat. Circuits Signal Process.*, vol. 13, no. 1/2, pp. 19–35, May 1997.
[19] B. A. Olshausen, C. H. Anderson, and D. C. Van Essen, "A neurobiological model of visual attention and invariant pattern recognition based on routing of information," *J. Neurosci.*, vol. 13, no. 11, pp. 4700–4719, Nov. 1993.
[20] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. J. Nuflo, "Modeling visual attention via selective tuning," *Artif. Intell.*, vol. 78, no. 1–2, pp. 507–545, 1995.
[21] C. Koch and S. Ullman, "Shifts in selective visual-attention—toward the underlying neural circuitry," *Human Neurobiol.*, vol. 4, no. 4, pp. 219–227, 1985.
[22] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 1254–1259, Nov. 1998.
[23] H. Bosch, R. Milanese, and A. Labbi, "Object segmentation by attention-induced oscillations," in *Proc. IEEE Int. Joint Conf. Neural Networks*, vol. 2, 1998, pp. 1167–1171.
[24] P. E. Trahanias, S. Velissaris, and T. Garavelos, "Visual landmark extraction and recognition for autonomous robot navigation," in *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, vol. 2, 1997, pp. 1036–1043.
[25] S. Baluja and D. A. Pomerleau, "Expectation-based selective attention for the visual monitoring and control of a robot vehicle," *Robot. Autom. Syst.*, vol. 22, pp. 329–344, 1997.
[26] B. Gibson and H. Egeth, "Inhibition of return to object-based and environment-based locations," *Percept. Psychopys.*, vol. 55, pp. 323–339, 1994.
[27] K. A. Boahen, "Communicating neuronal ensembles between neuromorphic chips," in *Neuromorphic Systems Engineering*, T. S. Lande, Ed. Norwell, MA: Kluwer Academic, 1998, pp. 229–259.
[28] 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. Los Alamitos, CA: IEEE Computer Society Press, 1995, pp. 139–161.
[29] J. Kramer, R. Sarpeshkar, and C. Koch, "Pulse-based analog VLSI velocity sensors," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 86–101, Feb. 1997.
[30] G. Indiveri, "A current-mode analog hysteretic winner-take-all network, with excitatory and inhibitory coupling," *J. Analog Integrated Circuits Signal Process.*, to be published.
[31] S. P. DeWeerth, "Analog VLSI circuits for stimulus localization and centroid computation," *Int. J. Comput. Vis.*, vol. 8, no. 3, pp. 191–202, 1992.
[32] A. L. Yarbus, *Eye Movements and Vision*. New York: Plenum, 1967.

Giacomo Indiveri graduated in electrical engineering from the University of Genova, Genoa, Italy, and completed the postgraduate fellowship of the "National Research Program on Bioelectronic Technologies" (*cum laude*) in 1992 and 1995, respectively.

From 1994 to 1996, he was a Research Fellow with at the California Institute of Technology, Pasadena, where he worked on the design of analog VLSI neuromorphic architectures for low-level visual tasks and motion detection. He is currently a Research Assistant with the Institute of Neuroinformatics, University of Zürich/ETH Zürich, Zürich, Switzerland. His current research interests include neuromorphic multi-chip systems and hardware implementations of computational models based on networks of spiking neurons. He is co-organizer of the Workshop on Neuromorphic Engineering held yearly in Telluride, CO, and co-teacher of the design and implementation of neuromorphic analog VLSI Systems, taught at ETH Zürich.

Robin Mürer was born in Scherzingen, Switzerland, in 1974. He graduated in physics at the Swiss Federal Institute of Technology Zürich (ETHZ), Zürich, Switzerland.

He is currently with the Institute of Neuroinformatics, University of Zürich/ETH Zürich.

Jörg Kramer (S'92–M'96) graduated in physics from the Swiss Federal Institute of Technology Zürich (ETHZ), Zürich, Switzerland, received the M.S. degree in applied optics from Imperial College of Science and Technology, London, U.K., and the Ph.D. degree in physics from ETHZ for a project carried out at the Paul Scherrer Institute Zürich (PSIZ), Zürich, Switzerland in 1987, 1988, and 1993, respectively.

At PSIZ, he developed a real-time video 3-D imaging system and designed optoelectronic devices for metrology applications using CMOS technology. As a post-Doctoral Fellow at California Institute of Technology, Pasadena, from 1993 to 1996, he designed analog VLSI velocity sensors and velocity-based sensing systems for image segmentation and time to contact measurements. He is currently with the Institute of Neuroinformatics, University of Zürich/ETH Zürich, where he continues to work on the design of early visual processing sensors and on their implementation on mobile platforms for orientation and navigation purposes. He also teaches courses in theory and design of analog VLSI circuits.