Neuromorphic Analog VLSI Sensor for Visual Tracking: Circuits and Application Examples

Giacomo Indiveri

Abstract—This paper presents a one-dimensional visual sensor, implemented on a single VLSI chip using analog neuromorphic circuits, for selectively detecting and tracking the position of the feature with the highest spatial contrast present in the visual scene. The chip's photoreceptors adapt to stationary backgrounds and can be tuned to respond maximally to specific target velocities. The sensor drastically reduces the amount of data to be transmitted to further processing stages by encoding, in real time, the position of the target in the form of a single continuous-time analog variable. We describe the circuits implementing the sensor and show applications to three examples of tracking tasks: a stand-alone visual tracking system, an active fully analog tracking system, and a mobile platform line-following system.

I. INTRODUCTION

NEUROMORPHIC vision sensors are typically analog VLSI devices that implement hardware models of biological visual systems and can be used for machine vision tasks [1], [2]. It is only recently that these hardware models have become elaborate enough for use in a variety of engineering applications [3]. These types of devices and systems offer an attractive low-cost alternative to special-purpose digital signal processors (DSP's) for machine vision tasks. They can be used for either reducing the computational load on the digital system in which they are embedded or, ideally, for carrying out all of the necessary computation without the need for any additional hardware. They process images directly at the focal plane level. Typically, each pixel contains local circuitry that performs, in real time, different types of spatio-temporal computations on the continuous analog brightness signal. In contrast, charge coupled device (CCD) cameras or conventional complimentary metal–oxide–semiconductor (CMOS) imagers merely *measure* the brightness at the pixel level, eventually adjusting their gain to the average brightness level of the whole scene. In neuromorphic vision chips, photoreceptors, memory elements, and computational nodes share the same physical space on the silicon surface. The specific computational function of a neuromorphic sensor is determined by the structure of its architecture and by the way its pixels are interconnected. Since each pixel processes information based on locally sensed signals and data arriving from its neighbors, the type of computation being performed is fully parallel and distributed. Another important feature is the asynchronous operation of neuromorphic sensors, which is preferable to clocked operation for sensory processing, given the continuous nature of sensory signals. Clocked systems introduce temporal aliasing artifacts that can significantly compromise the time-dependent computations performed in real-time sensory processing systems.

In this paper, we present a neuromorphic sensor that consists of a one-dimensional (1-D) 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.

Tracking features of interest as they move in the environment is a computationally demanding task for machine vision systems. The control loop of active vision systems, comprising motors that steer the visual sensor, relies on the speed of the specific computation carried out. The stability of system depends on the latency of the sensory-motor control loop itself. To reduce this latency and improve the performance of the active vision system several custom VLSI sensors that pre-process the input image and extract the position of the target, have been proposed [4]–[7]. As previously proposed solutions, the tracking architecture here described reduces the computational cost of the processing stages interfaced to it by carrying out an extensive amount of computation at the focal plane itself, and transmitting only the result of this computation, rather than extensive amounts of data representing the raw input image. Despite the principle of the approach followed here being very similar in nature to the one followed by the authors cited above, the tracking architecture described here differs from previously proposed ones in two key features: 1) it selects high-contrast edges independent of the absolute brightness of the scene (as opposed to simply selecting the scene's brightest region [4], [5], [7]) and 2) it uses a hysteretic winner-take-all (WTA) network, with positive feedback and lateral coupling [8], to lock onto and smoothly track the selected targets (different from WTA networks used in other tracking devices [4]–[6]). We show, in Section III, how these features allow systems that use the architecture proposed here to reliably track natural stimuli in a wide variety of illumination conditions. Specifically, we will describe three examples of system applications that make use of the sensor proposed to track, passively and actively, the edges with the highest contrast present in the sensor's field of view.

II. THE TRACKING SENSOR

The tracking architecture proposed here is structured in a hierarchical way and can be implemented on a single-chip device. As the architecture is 1-D, we can design thin,

The author is with the Institute of Neuroinformatics, G86 8057 Zürich, Switzerland.
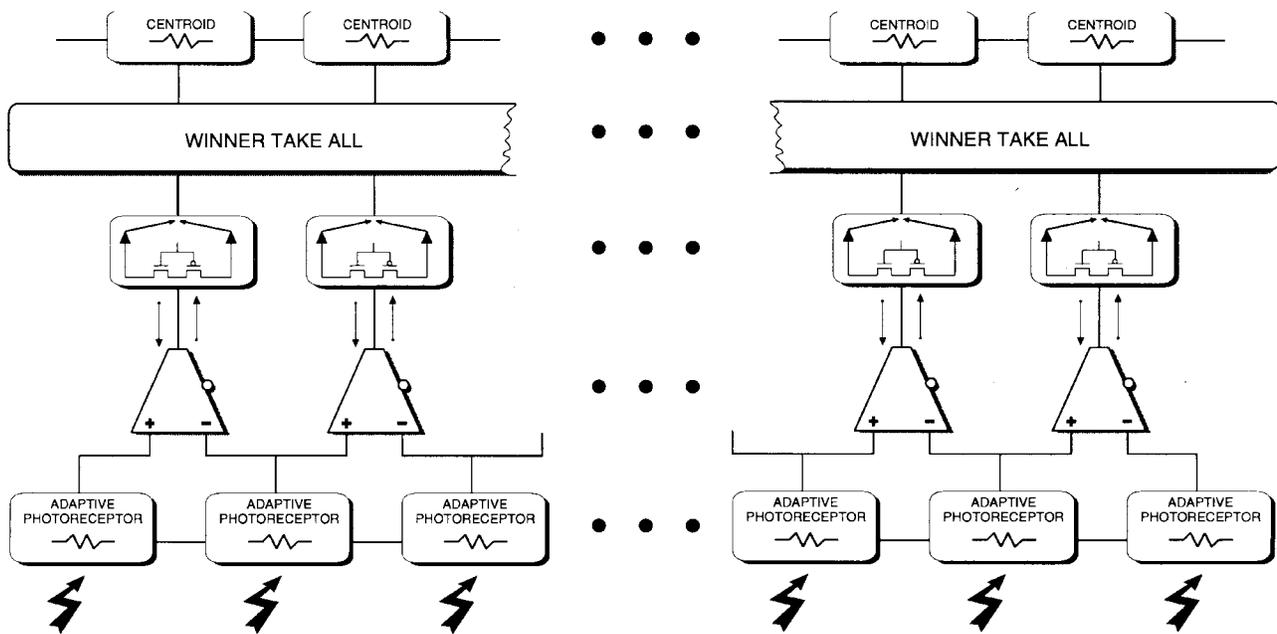
Fig. 1. Block diagram of single-chip tracking system. Spatial edges are detected at the first computational stages by adaptive photoreceptors connected to transconductance amplifiers. The edge with the strongest contrast is selected by a WTA network and its position is encoded with a single continuous analog voltage by a position-to-voltage circuit (see text for details).

long processing columns in a way to optimize the area used and increase the number of pixels on the device. Two chips of approximately 2 mm× 2 mm were fabricated using a standard 2- and 1.2-μm CMOS technology, respectively. The processing columns of each chip are 60λ wide, where λ is the scalable CMOS design rule parameter, corresponding to 1 μm for the 2-μm process and to 0.6-μm for the 1.2-μm process. As the circuits are analog and some circuit elements (such as capacitors) don't scale with λ, the layouts of the two chips are slightly different (despite the schematic diagrams are identical). The 2-μm chip has a pixel pitch of 60 μm and contains 25 processing columns, while the 1.2-μm chip has a pixel pitch of 36 μm and contains 40 processing columns.

A. System Architecture

Image brightness data is processed in parallel through five main computational stages. A block diagram of the device's architecture is depicted in Fig. 1. The first stage is an array of adaptive photoreceptors [9], [10] that logarithmically map image intensity into their output voltages. The second stage is composed of an array of simple transconductance amplifiers, operated in the subthreshold regime, which receive input voltages from neighboring photoreceptors [11]. The amplitude of their output currents encode the contrast intensity of edges and the sign their polarity. At the third computational stage, the polarity of each edge is gated so that the sensor selectively responds either to ON edges (dark to bright transitions), or to OFF edges (bright to dark transitions), or to both. The fourth stage uses a hysteretic WTA network [8], which selects and locks onto the feature with strongest spatial contrast moving at the speed that best matches the photoreceptor's velocity tuning. Finally, in the last stage, there is a position-to-voltage circuit, as described in [12], that allows the system to encode
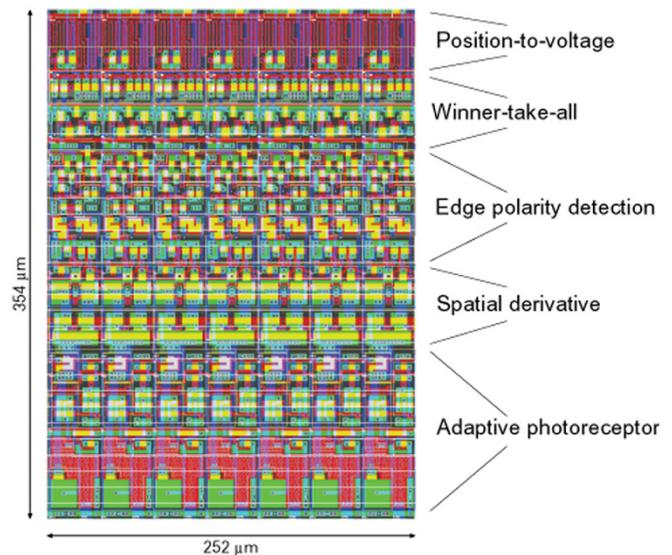
Fig. 2. Portion of layout of the 1.2-μm chip containing seven processing columns. The size of each computational stage is evidenced on the right.

the spatial position of the WTA network's output with a single analog value. The 1.2-μm chip layout of these circuits is shown in Fig. 2.

Fig. 3 summarizes the general response properties of the 2-μm chip by showing the outputs of the different computational stages described above. The top trace of Fig. 3(a) shows the responses of the array of adaptive photoreceptors to a black bar on a white background, imaged onto the chip's surface using a standard CS mount 4-mm lens with an f-number of 1.2. The two lower traces of the figure are the response of the edge polarity detector circuits, representing the spatial derivative of the input stimulus. Fig. 3(b) shows the response
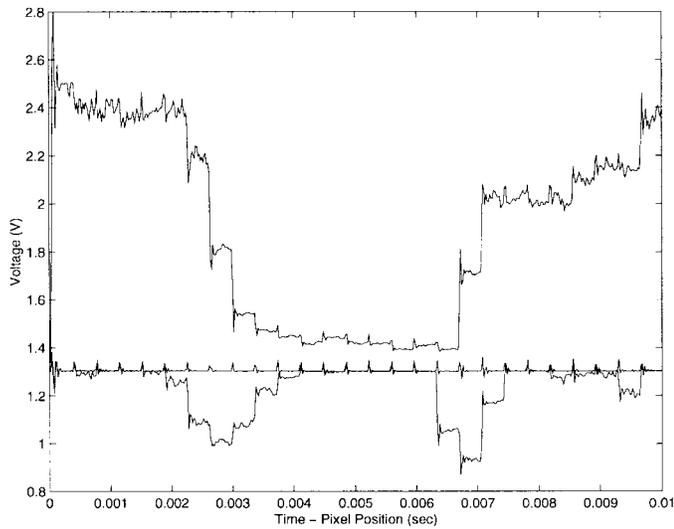
(a)
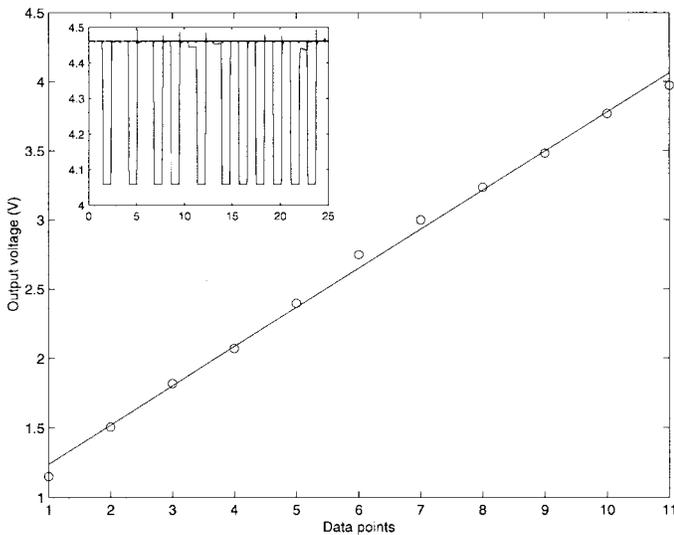
(b)

Fig. 3. (a) Response of the array of adaptive photoreceptors to a black bar on a white background (upper trace) and output traces of the edge-polarity detector circuit (lower traces). (b) Output characteristic of the position-to-voltage circuit. The figure's inset contains snapshots of many output traces of the WTA network superimposed as a stimulus was moving from left to right. The datapoints in the main figure represent the output of the circuit corresponding to the pixel position of the winner in the inset data.

(a)

(b)

Fig. 4. (a) Response of the array of photoreceptors, with a very slow adaptation rate to a dark bar on a white background moving from right to left with an on-chip speed of 31 mm/s. The dc value of the response has been subtracted. (b) Response of array of photoreceptors with a fast adaptation rate to the same bar moving at the same speed (left-pointing triangles) and at a slightly slower speed (upward-pointing triangles).

of the position-to-voltage circuit to 11 different winning pixel positions. The figure's inset displays 11 snapshots of the WTA response to the 11 corresponding spatial positions of the input stimulus.

B. Adaptive Photoreceptor Circuit

This photoreceptor circuit, originally designed by Tobi Delbrück [9] and further improved by Shih-Chii Liu [10], has been used extensively in many neuromorphic sensors. The response of the circuit is invariant to absolute light intensity (changing logarithmically with image brightness). The adaptive photoreceptor exhibits the characteristics of a temporal bandpass filter, with adjustable high- and low-frequency cutoff values. Fig. 4 shows the response of the array

of photoreceptors to a moving bar for two different adaptation settings. In Fig. 4(a), the adaptation rate was low, with adaptation time constants in the order of hundreds of milliseconds. In Fig. 4(b), the adaptation rate was very high, such that the photoreceptors adapt quickly to brightness transients. Because of its adaptation property, the photoreceptor biased in this way has a response which results in both contrast and speed dependence.

C. Spatial-Derivative Circuit

Spatial derivative is implemented using simple transconductance amplifiers operated in the subthreshold regime. The amplifiers receive input voltages from neighboring photoreceptors and provide a bidirectional output current that is proportional to the hyperbolic tangent of their differential input
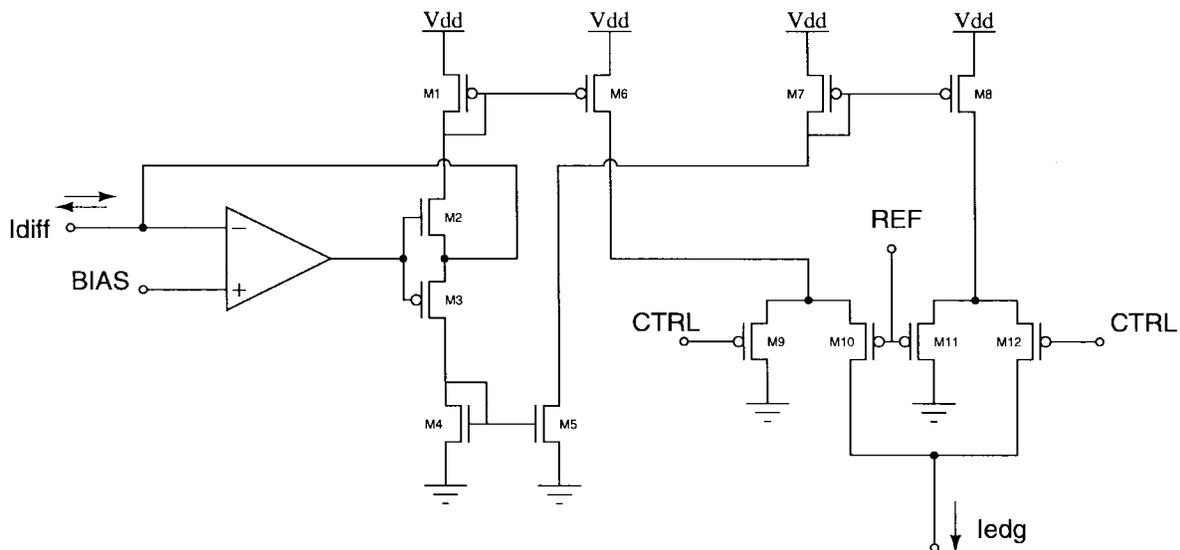
Fig. 5.   Circuit diagram of the current polarity detector. Positive I_{diff} currents are conveyed to the n-type current mirror M4, M5. Negative I_{diff} currents are conveyed to M6 through the the p-type current mirror M1, M6. Depending on the values of the control voltage signals V_{CTRL} and V_{REF}, the output current I_{edg} represents a copy of only one of the two polarities of I_{diff}, or of both polarities of I_{diff} (see text for details).

[11]. The output current saturates smoothly as the differential voltage increases (in absolute value) beyond 200–300 mV. The possibility of electronically smoothing the input image (at the adaptive-photoreceptors stage) allows the user to operate the spatial-derivative circuit always in its linear range, for a stimulus with fixed spatial frequencies. Furthermore, the presence of multiple stimuli with contrast high enough to saturate the transconductance amplifiers currents is not going to compromise the sensor's tracking performance, as the WTA network is able to lock onto the feature selected (see Section II.E).

D. Edge-Polarity Detector Circuit

The polarity of edges in the visual scene is encoded by the sign of the transconductance amplifiers' currents. Each of these currents is fed into a circuit of the type shown in Fig. 5. The amplifier in the left part of Fig. 5 together with transistors M1–M6 implement a *current conveyor* [13]. This circuit is used to separate the positive component of the input current I_{diff} from the negative one, and to decouple the spatial-derivative stage from the current-polarity selection stage. Negative input currents are conveyed to transistor M6, while positive ones are flipped through the current mirror M4, M5, and conveyed to M8. Transistors M6 and M8 source their currents to the polarity-selection circuit (transistors M9–M12) [6]. The output current of the polarity-selection circuit I_{edg} represents OFF edges (the positive component of I_{diff}), ON edges (the negative component of I_{diff}), or either type of edge (the absolute value of I_{diff}), depending on the control voltage V_{CTRL} and V_{REF} settings. The voltage V_{BIAS} on the positive node of the amplifier is a constant used to bring the circuit into its correct operating point and (in typical operating conditions) assumes values ranging from 1 to 2.5 V. The output currents I_{edg} of all edge-polarity detector circuits are sourced, in parallel, to the elements of the next processing stage: the hysteretic winner-take-all network.

E. Hysteretic WTA Network

This circuit is an extension of the basic current-mode WTA network [14]. It collectively processes all its input signals using strictly local interconnections, it operates in parallel, and it is compact, using only eight transistors per cell. Fig. 6 shows three of these cells connected together. Each cell is connected to its neighbor through a pass transistor controlled by V_{ex}. The set of four n-type transistors in the lower part of each cell implements the current-mode WTA with diode source degeneration, as described in [14]. The current I_{sum_i} of the bottom-right n-type transistor represents the sum of all of the currents converging into node i, and can be monitored to evaluate the effect of the control voltages V_b, V_{ex}, and V_{gain}. The p-type current mirror in the top part of each cell is used to provide the output of the network I_{out} and simultaneously to implement a local positive-feedback circuit [8], [15]. As the WTA network allows only one cell at a time to have a nonzero output current, only the the feedback loop of the winning cell will be active. The positive feedback introduced reinforces the choice of the winning cell by injecting into its input node a fixed amount of current, corresponding to a fraction of I_b modulated by the voltage V_{gain}. This operation introduces a hysteretic behavior which allows the WTA network to lock onto a winning pixel [8], [15], [16].

This hysteretic WTA network also contains an additional cell connected to a bias. This additional cell can be used to set a threshold for the spatio-temporal contrast of edges present in the scene; if the input from external bias is higher than all other inputs, the WTA will signal the absence of high-contrast edges in the visual scene.

The option of introducing hysteresis in the WTA network might cause problems in dynamic environments for which it is necessary to update the winning pixel position continuously (e.g. in the domain of tracking applications). One solution would be to reset the WTA network manually any time it needs to be updated [15]. A more elegant solution is the one of using
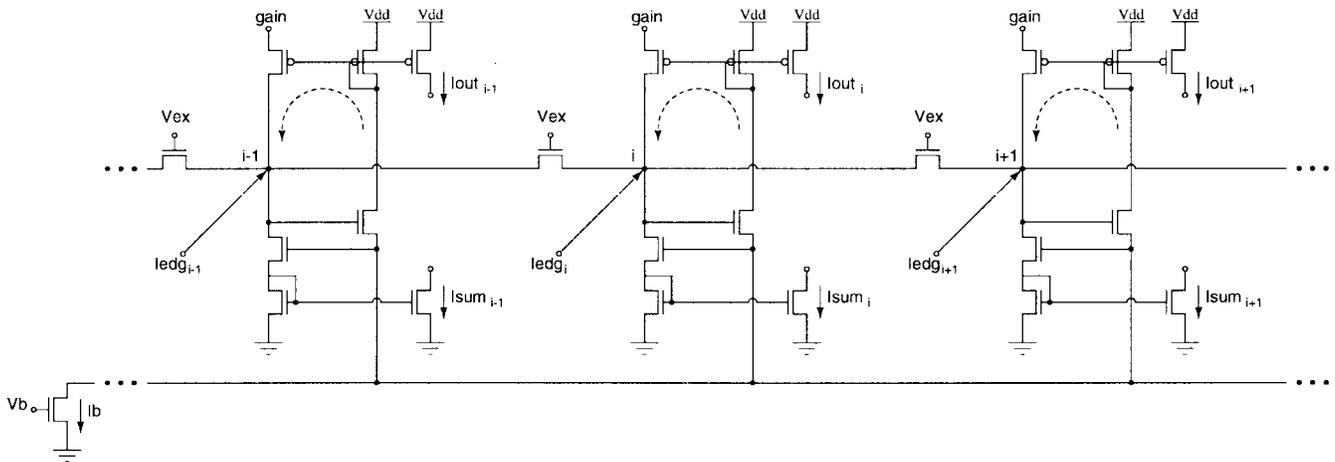
Fig. 6. Schematic diagram of the WTA circuit. Examples of three neighboring cells connected together.

lateral coupling between cells, allowing part of the hysteretic component of the winner's current to be passed to its neighbors [8], [16]. Cells adjacent to the winning pixel will, hence, be facilitated in the winner computation process, whereas cells in the periphery will be inhibited. This solution takes into account the assumption that the features being selected move continuously in space, and ensures that once the WTA network has selected a target and is engaged in visual tracking, it *locks onto* it and does not get distracted by possible distracting stimuli in the periphery. Lateral coupling between cells of the hysteretice WTA network can be accomplished by properly setting the gate voltage V_{ex} of the pass transistors in Fig. 6. These transistors allow the WTA network to spread laterally both the hysteretic current being generated at the winning cell and the input currents coming from the edge-polarity detector stage. By choosing an appropriate combination of control voltages V_b, V_{ex}, and V_{gain}, it is possible to bias the WTA network such that it produces different behavioral responses. For example, by setting V_b to a subthreshold value (e.g. $V_b = 0.7$ V) and V_{gain} to approximately 4 V, we effectively turn off the positive feedback, such that the WTA network behaves as a conventional one; the currents I_{out} of Fig. 6 are all null except for the one belonging to the winning cell and the WTA output oscillates between similar inputs. Furthermore, the array of currents I_{sum} replicates the distribution of input currents I_{edg} with a degree of spatial smoothing proportional to the value of V_{ex}. If, on the other hand, V_{gain} is set to V_{dd}, the positive feedback loop is turned fully on; the feedback current is exactly I_b (modulo device mismatch effects) and the WTA network exhibits its hysteretic properties (it selects and locks onto inputs moving continuously in space). The stability properties of this WTA network are the ones of conventional WTA circuits with positive feedback, and have been analyzed in detail in [15]. Similarly, the dynamical response properties of the network are the same ones of the current-mode WTA network described in [14] and depend on the values of I_b and of the total current entering the input nodes of the WTA cells (namely, I_{edg} summed to the hysteretic feedback current and to the currents coming from the lateral coupling transistors). It is possible to evaluate the total current entering the input
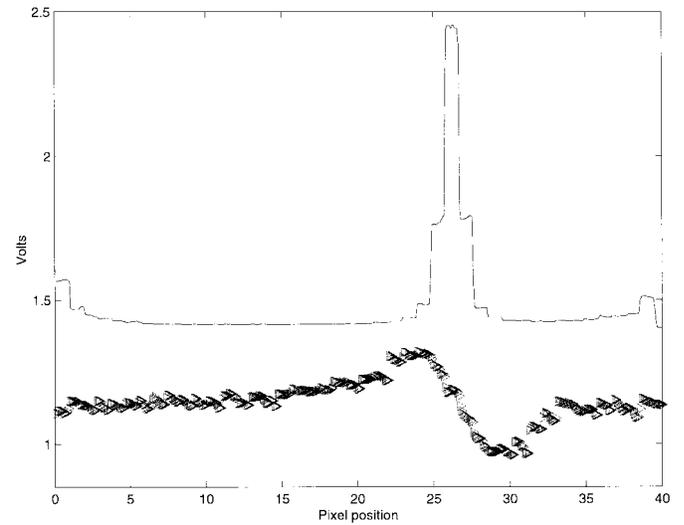
Fig. 7. Response of the WTA network to the ON-edge of a bar moving from left to right at an on-chip speed of 31 mm/s. The top trace represents the currents I_{sum} of the WTA array, while the bottom trace represents the voltage outputs of the array of adaptive photoreceptors.

nodes by measuring the currents I_{sum} at each node of the network. Fig. 7 shows an example of the WTA response to a moving bar given the following control voltage settings: $V_b = 0.75$ V, $V_{gain} = 4.65$ V, $V_{ex} = 1.85$ V. The top trace of the figure, representing the values of the currents I_{sum}, shows effect of spatial smoothing of the input currents combined with the hysteretic current coming from the positive feedback loop of the winning cell. It is clear from this figure that the active winning cell is the one corresponding to pixel 26. The bottom trace shows the response of the adaptive photoreceptors. The input stimulus was the same one used for the previous figures: a 1-cm wide black bar on a white background positioned approximately 17 cm away from the focal plane and imaged onto the chip through a 4-mm lens moving from left to right with an on-chip speed of 31 mm/s.

F. Spatial Position-Encoding Circuit

This circuit consists of a series of voltage followers, using a common global current mirror which receive inputs from a
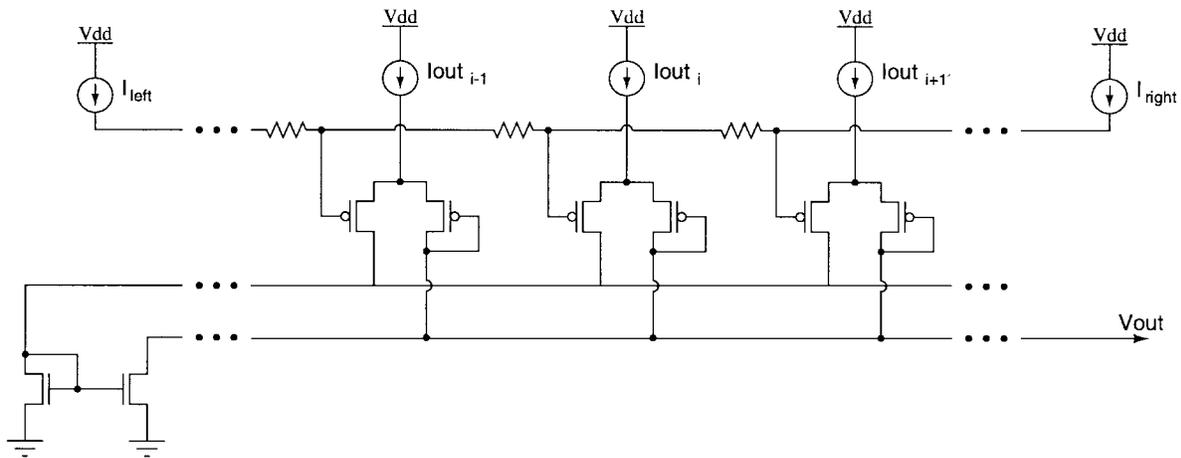
Fig. 8. Schematic diagram of position-to-voltage circuit. Example of three neighboring cells connected together.

linear resistive network [12] (see Fig. 8). The currents I_{out} being generated by the WTA network at the previous stage are used as bias currents for the followers. As only one I_{out_i} is nonnull at any given time, all followers are switched off, except for the one connected to the winning WTA cell. The output of the spatial position-encoding circuit V_{out} thus represents the position of the winning cell in the array.

III. SYSTEM APPLICATIONS

In this section, we describe three different application examples. These applications were developed with the intent of demonstrating the possible uses of a 1-D visual tracking device. They have not been optimized and they do not encompass all the possible application domains for such a device; yet, despite their unsophisticated nature, they have proven to perform satisfactorily in a wide range of testing conditions.

A. Stand-Alone Visual-Tracking Device

We attached a 4-mm lens to the 2-μm chip and mounted it on a board with external potentiometers, used to set its bias voltages. The board also has a 1-D LED display with its driver (see Fig. 9). The LED display is used to have visual feedback on the position of the feature selected by the chip. The power supply to the whole board is provided by a 9-V battery (attached to the back of the board) and a voltage regulator IC.

The system is able to detect and report, in real time, the position of realistic types of stimuli moving within its field of view. It performs reliably in a wide variety of illumination conditions, ranging from dim artificial room illumination to bright sunlight, thanks to the adaptive properties of the photoreceptors at the input stage. For these applications, the bias settings of the photoreceptor stage are those of fast adaptation rates, as described in Section II-B. Lateral coupling between neighboring cells was turned off at the photoreceptor stage but turned on at the WTA level (V_{ex} of Fig. 6 was set to 1.2 V). Smoothing at the WTA level was useful to reduce the offsets introduced by the spatial derivative and edge-polarity detector circuits. The hysteretic current of the WTA network (summed back into the input nodes through the positive-feedback path)
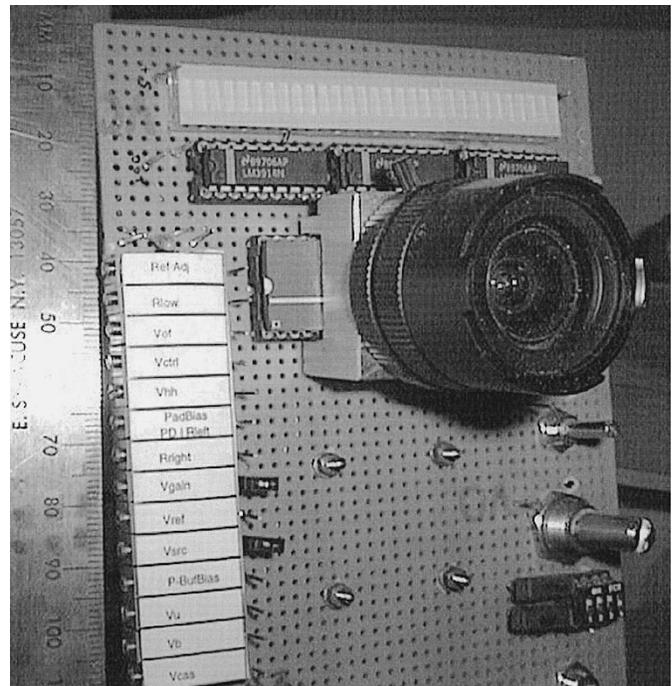
Fig. 9. Picture of the stand-alone tracker board. The neuromorphic sensor is on the chip beneath the lens. On the left part of the board there is an array of potentiometers used to bias the chip's control voltages. On the top, there is an LED display, comprising three display bar lines with their corresponding drivers. The scale in the left part of the figures is in millimeters.

was set to be a small fraction of the maximum possible feed-forward input current (controlled by the bias voltage of the spatial-derivative transconductance amplifier). All other bias parameters on the chip were not critical and were set to reasonable subthreshold voltages (i.e., 0.5 V–0.8 V for n-type transistors and 4.4 V–4.1 V for p-type transistors). The system, biased in such a way, adapts out the background of a stationary scene and selects high-contrast moving targets present in its field of view, tracking them as they move smoothly in space. Fig. 10(a) shows the output of the chip in response to a finger moving back and forth in front of the lens in a laboratory environment with cluttered background. Fig. 10(b) shows the output of the chip in response to a black pen moving at a

(a)

Stimulus velocity: 7955 pixels/sec

(b)
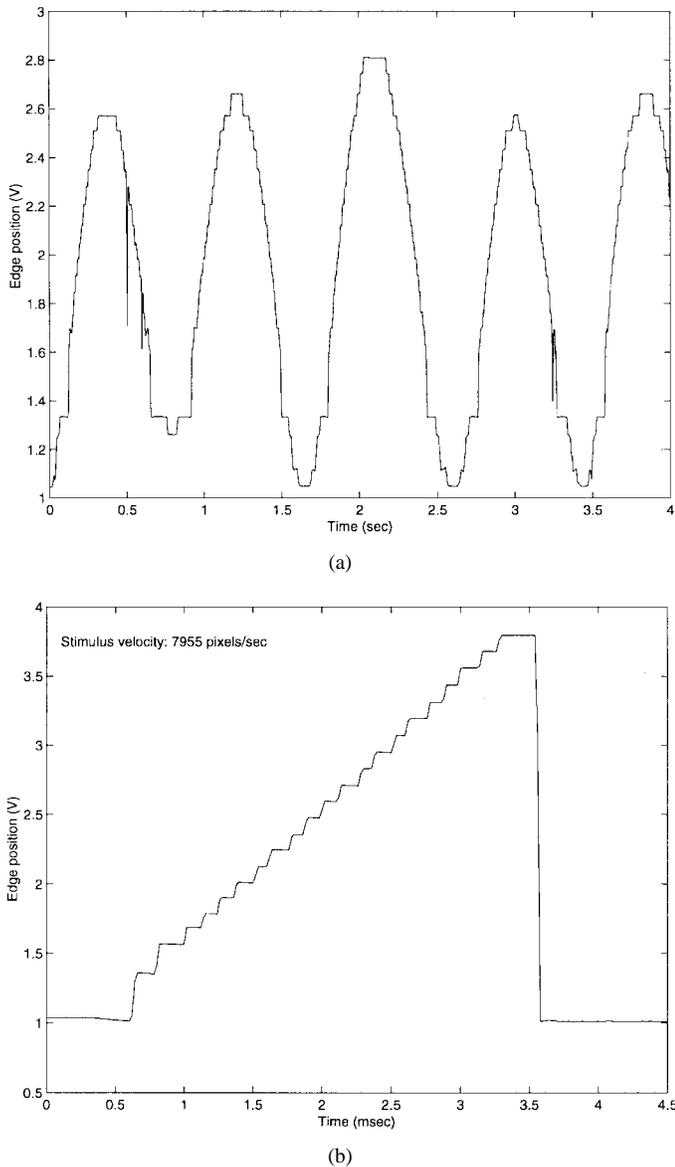
Fig. 10. (a) Output of the system in response to a finger moving back and forth in front of the chip. (b) Output of the system in response to a pen moving at approximately 8000 pixels/s on a stationary light background. Note the different time scales on the abscissae.

Fig. 11. Picture of tracker chip mounted on a dc motor. The output of the chip is sent to a dual-rail power amplifier which directly drives the motor.

speed of almost 8000 pixels/s on a uniform background. As mentioned in Section II, each pixel of the 2-μm chip is 60-μm wide, and thus, the velocity of the target on the focal plane corresponds to approximately 0.5 m/s. The output of the chip is continuous in time, but discrete in space; the discrete jumps present Fig. 10 represent the shifting of the winning position from one pixel to the next.

B. Active Tracking System

We implemented a fully analog active tracking system by mounting a board with the 1.2-μm tracker chip and a 4-mm lens onto a dc motor (see Fig. 11). The bias settings of the chip were the same used in Section III-A, except for the value of the hysteretic current in the positive-feedback path of the WTA network, which was set to be greater than the feed-forward current I_{edg}. Specifically, the WTA bias voltage V_b was set

to a value slightly higher than the bias voltage of the spatial-derivative transconductance amplifier, and the source voltage of the p-type transistor of the positive-feedback current mirror (V_{gain} in Fig. 6) was set to 5 V. In this way, the WTA network locks onto the selected target and allows only the nearest-neighbor units to win if the selected stimulus moves (see also Fig. 7 is Section II-E). The position-to-voltage circuits were biased to encode the position of the winner with voltages ranging from 1 to 4 V. The analog output of the chip was rescaled and amplified (via an ST L272 power amplifier), such that the selection of features in the right part of the visual field produces positive voltages and the selection of features in the left part of the visual field produces negative voltages. The output voltage, with an amplitude directly proportional to the distance of the target's position from the center of the retina, is used to drive the dc motor. The sensory motor loop, so designed, implements a negative feedback system which attempts to zero the motion of the target on the retina: if a target appears in the periphery of the visual scene, the sensor will drive the dc motor so as to orient the sensor's

(a)

(b)
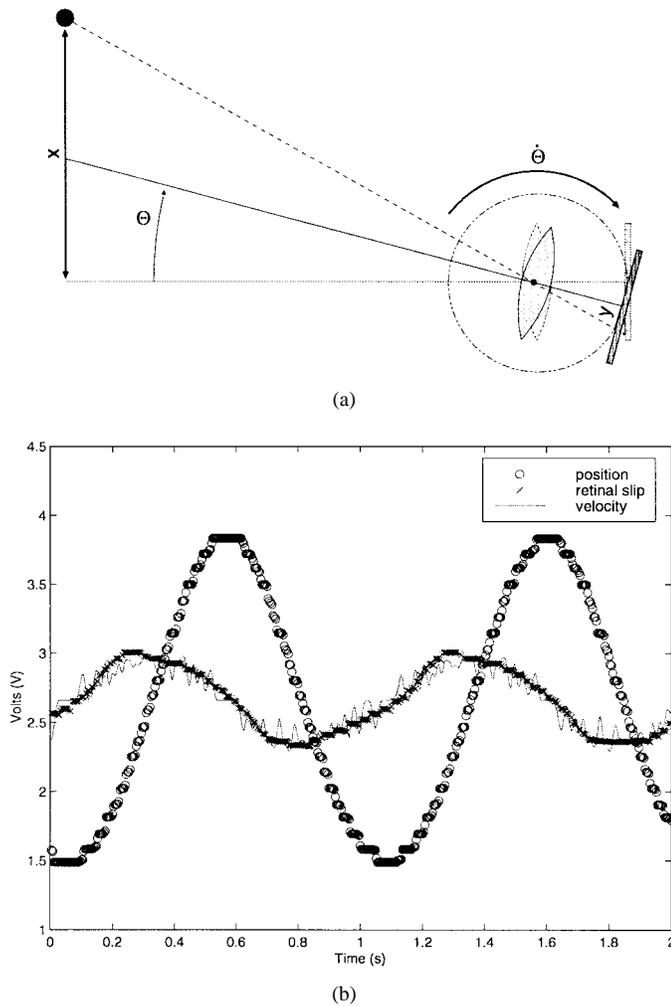
Fig. 12. (a) Setup of the active tracking system as seen from above. The angle θ represents the angular displacement produced by the dc motor, x represents the target's position in the visual space, y represents the distance of the target's projection on the retina from its center. The angular velocity $\dot{\theta}$ is proportional to y. (b) Chip data measured as the system was engaged in tracking a swinging bar. The bar's position (circles) was measured using a separate (fixed) tracking board, while its velocity (solid line) was computed off-line from the discretized position data. The crosses represent the output of the active sensor used to drive the system's dc motor.

gaze toward the target. As the projection of the target on the retina approaches the center of the pixel array, the output of the system (i.e., the motor's power supply) decreases toward zero, bringing the motor to a stop. In terms of equations we can write, to a first-order approximation

$$\begin{cases} y(t) = Fx(t) - \theta(t) \\ \dot{\theta}(t) = Ay(t) \end{cases} \quad (1)$$

where $x(t)$ represents the position of the target in the visual space, $y(t)$ represents its corresponding projection on the retina, θ the rotation angle produced by the dc motor around its axis, and F the optical magnifying factor [see Fig. 12(a)]. The term $\dot{\theta}(t)$ corresponds to the motor's angular velocity, and A to the open-loop gain of the feedback system. Solving for $\dot{y}(t)$, we obtain

$$\dot{y}(t) = F\dot{x}(t) - Ay(t). \quad (2)$$

If the system is successful in zeroing the motion of the target on the retina ($\dot{y}(t) = 0$), we should measure a retinal slip $y(t)$ directly proportional to the velocity of the target in the visual space. Fig. 12(b) shows traces obtained from the system while it was engaged in tracking a swinging target. The target stimulus was a black bar on a white background, similar to the one used to characterize the adaptive photoreceptor circuit in Section II-B. The position of the target in visual space was measured optically by the stand-alone tracker board described in Section III-A. The target's velocity was computed off-line by differentiating the discretized position signal (hence, the jitters in the figure). As shown, the measured response matches, to a first-order approximation, the theoretical prediction.

The task performed by the system here described is that of *smooth pursuit* [17]. This model does not take into account the velocity of the target, but only its position. More elaborate models of smooth pursuit tracking have been proposed [6], [7], but none using fewer components (namely a neuromorphic CMOS sensor, a dc motor, a power amplifier, and a dual power supply). The system presented here can be considered as the minimal, lowest cost and most compact solution to 1-D visual tracking of natural stimuli.

C. Roving Robot

Another application domain which is well suited for the visual tracking device is that of vehicle guidance and autonomous navigation. These types of tasks, in fact, require compact and power-efficient computing devices which should be robust to noise, tolerant to adverse conditions induced by the motion of the system (e.g., to jitter and camera calibration problems), and possibly able to adapt to the highly variable properties of the world. To test our tracking sensor within this framework, we interfaced it to a mobile robot and measured the performance of the overall system in a line-following task. The mobile robot is a *Koala* (K-Team, Lausanne). It measures 32 cm in length, 31 cm in width, and is 11-cm high. It has an on-board Motorola 68 331 processor, 12 digital I/O ports, and 6 analog inputs (with 10-bit A/D converters), 1 MByte of RAM, and 2–3 hours of autonomous operation from its battery. The tracking sensor was mounted onto a wire-wrap board together with a 4-mm lens with an f number of 1.2, and it was attached to the front of Koala with the lens tilted toward ground at an angle of approximately 60°, in a way to image onto the retinal plane the features present on the floor approximately 10-cm ahead [see Figs. 13(a) and 14(a)]. The bias settings of the chip were the same ones used in the analog active tracking system, described in Section III-B.

For this specific application example, we made use of the additional node of the WTA network with its input current set by an external potentiometer. This allowed us to set a threshold value against which we could compare the contrast of edges present in the visual scene. In the case of absence of lines to follow, the WTA network selects the external input and the sensor outputs a unique voltage different from the set of voltages generated by visual stimuli. The output voltage of the tracking chip is directly applied to one of the analog input ports of the robot and digitized. To implement the line-
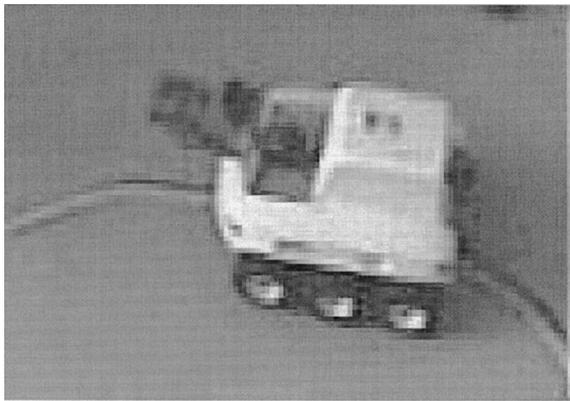
(a)
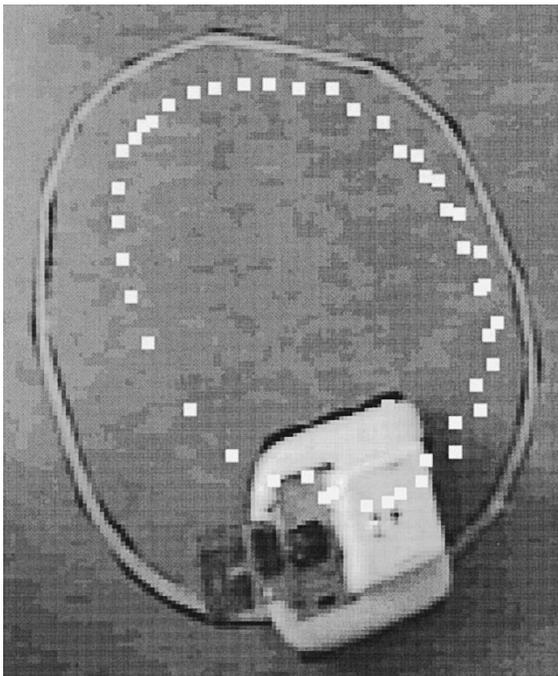
(b)

Fig. 13. (a) Koala robot with neuromorphic sensor mounted on its front. (b) Positions of Koala following a line, sampled at intervals of 0.25 s for a period of 37.5 s, in which the robot completed four loops. The features (white squares) were obtained by tracking a dark cross drawn on the white top of Koala.

(a)

(b)

Fig. 14. (a) Koala robot with neuromorphic sensor mounted on its front and a white sheet of paper with crosses attached on its top, seen from above. (b) Positions of Koala following a white line on a light-blue carpet floor, sampled at intervals of one second over a period of approximately 3 min. The features (white squares) were obtained by tracking the bars appearing on the top part of Koala (see text for explanation).

following task, Koala uses a very simple control algorithm which reads the tracking chip's output V_{out} and backs up in a random direction if no edge if found. If, on the other hand, the tracker chip detects an edge and outputs a valid voltage, the algorithm shifts and re-scales V_{out} so that the variable encoding edge position pos is zero when the target is in the center of the chip's visual field; it sets the forward component of the velocity fwd to a value weighted by a Gaussian function of pos (fwd is maximum when pos=0 and it decays as |pos| increases); it sets the rotational component of the velocity rot to a value proportional to pos; and finally, it executes motor commands sending fwd and rot directly to the robot's motors. Scaling the forward component of the velocity fwd by a Gaussian function of the line's eccentricity allows the robot to slow down in curves. If the

line goes out of the field of view of the sensor (e.g., in presence of steep curves), the algorithm forces the robot to stop and back up until it again finds a line to follow. The line-tracking algorithm makes very little use of the on-board CPU's processing power (leaving it free for other CPU-time demanding processes). The computationally expensive part of the processing (involving visual preprocessing and target selection) is done in real time by the neuromorphic sensor. Using this simple control algorithm in conjunction with these types of sensors, the robot is able to reliably track lines randomly laid out on the floor for a wide variety of conditions

(e.g., floors with different texture, cables of different colors and sizes, extreme illumination conditions, etc.) [18]. Depending on the bias settings of the edge-polarity detector circuit, the line-following robot will always make left turns at road-forks (e.g., if the circuit is selective to OFF edges and the line is darker than the background) or right-turns. The bias settings can be changed at run-time by the robot using one of its digital I/O ports.

Fig. 13 shows the robot in the process of tracking a line. The line (a high contrast bar layed onto the floor) is long approximately 323 cm and forms a closed loop of elliptic shape with major axis long, roughly 110 cm, and a minor axis long, 90 cm. The robot followed the line with an average speed of 5 loops/min (corresponding roughly to 27 cm/s). To measure quantitatively the robot's performance, we stored a sequence of images (sampled at a rate of 4 frames/s) and applied them in input to the Kanade–Lucas–Tomasi Feature Tracker [19]. The data was taken in dim natural light conditions (typical of a cloudy rainy day in Zurich, Switzerland). Fig. 13(b) shows the features tracked by the algorithm for a sequence of 150 frames (in which the robot completed four loops). The features selected by the algorithm correspond to a (moving) black cross drawn on the robot's white top. Closely grouped features indicate the re-visitation of nearby positions over time. Features are more dense in the steep parts of the curve because of the slower speed values that the robot uses, as determined by its control algorithm.

Fig. 14 shows an experiment similar to the one described in Fig. 13, but run in a different, less controlled environment. The robot was following a line of white paper-adhesive tape layed on a light-blue carpet forming an 8 figure in an area of approximately 1.3×2.5 m. The illumination conditions were of bright natural sunlight (typical of sunny summer days in Telluride, Colorado). The robot was partially covered with a sheet of paper containing bars and crosses [see Fig. 14(a)]. The Kanade–Lucas–Tomasi tracking algorithm selects different corners of the crosses as the robot changes its orientation. Fig. 14(b) shows the output of the tracking algorithm for a sequence of 200 images, sampled at intervals of approximately 1 s, in which the robot makes two full loops around the 8 figure. As in Fig. 13(b), white squares are more dense in the steeper parts of the curve because the robot slows down at those points. The robot is able to follow the line reliably in both directions, always passing the intersection of the 8 figure, for a wide selection of (maximum) speeds. At high speeds, the robot occasionally looses the line (in the steep parts of the curve), comes to a stop, backs up, and starts following the line again until it reaches the shallow parts of the curve where it speeds up again to the maximum speed.[1]

IV. CONCLUSION

We described the architecture of a neuromorphic visual sensor that selects and reports the position of the feature with highest spatio-temporal contrast present in the visual scene. We showed the response properties of the circuits implementing

its different processing stages. The device described is not merely an imaging array, but an intelligent sensor designed for tracking applications. By computing the relevant information at the focal plane and providing a single continuous-time output, the sensor selectively reduces the amount of data to transmit to further processing stages, saving both communication bandwidth and response latency, quantities that are of vital importance in real-time tracking applications.

The sensor proposed is a 1-D device; in principle, its extension to two dimensions is straightforward (the WTA network would be global, receiving inputs from all the pixels of a two-dimensional (2-D) array and providing outputs to two independent 1-D position-to-voltage circuits) but the layout of each cell would have a relatively large size. A possible alternative to the (large) circuits used to compute the absolute value of the spatial derivative of visual features and to provide input to the WTA network could be the use of *bump circuits* [20]. 2-D contrast-sensitive silicon retinas, that make use of these circuits, have recently been developed [21] and the size of their pixels (approximately 100 μm on a side) indicate that even with the addition of the WTA circuits, a 2-D tracker chip implementation would have pixels of acceptable size. Other 2-D tracking sensors have already been proposed [7], [4], but these are intensity based and don't pre-process the photo-receptor output to compute spatial derivatives, so they simply select the brightest feature in the visual scene. They are not suited for the type of applications described in this paper or, more generally, for applications in which the feature that needs to be tracked is not necessarily the brightest one present in the visual scene.

We showed three simple examples of tracking applications which make use of the sensor, in real-world scenarios. The examples, thought of as feasibility exercises, have proven to be effective and have shown the sensors' capabilities. They are examples of successful neuromorphic systems able to perform complex visual tasks using a single analog VLSI chip as a front-end preprocessor. This, and other sensors of similar nature, have proven to be efficient, compact, and low-cost solutions for real-world applications that can be considered as a viable alternative to conventional (bulky and expensive) digital-machine vision systems.

ACKNOWLEDGMENT

---

[1] An animated sequence of the robot engaged in tracking the line of Fig. 14(b) can be viewed at http://www.ini.unizh.ch/~giacomo/koala-line.html

REFERENCES

[1] M. Mahowald and C. Mead, *Analog VLSI and Neural Systems*, ch. Silicon Retina. Reading, MA: Addison-Wesley, 1989, pp. 257–278.
[2] K. Boahen and A. Andreou, "A contrast sensitive silicon retina with reciprocal synapses," in *Advances in Neural Information Processing Systems*, D. Touretzky, M. Mozer, and M. Hasselmo (Eds.). Cambridge, MA: MIT Press, 1992, vol. 4.

[3] C. Koch and B. Mathur, "Neuromorphic vision chips," *IEEE Spectrum*, vol. 33, pp. 38–46, May 1996.

[4] V. Brajovic and T. Kanade, "Computational sensor for visual tracking with attention," *IEEE J. Solid State Circuits*, vol. 33, pp. 1199–1207, Aug. 1998.

[5] 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. 79–92, May/June 1997.

[6] T. Horiuchi, T. Morris, C. Koch, and S. DeWeerth, "Analog VLSI circuits for attention-based, visual tracking," in *Advances in Neural Information Processing Systems*, M. C. Mozer, M. I. Jordan, and T. Petsche (Eds.). Cambridge, MA: MIT Press, 1997, vol. 9.

[7] 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 H. M. E. (Eds.). Cambridge, MA: MIT Press, 1996, vol. 8.

[8] G. Indiveri, "Winner-take-all networks with lateral excitation," in *Neuromorphic Systems Engineering*, T. S. Lande (Ed.), Norwell, MA: Kluwer, 1998, pp. 367–380.

[9] T. Delbrück, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits," *Tech. Rep., CNS Memo no. 30*, California Institute of Technology, Pasadena, CA, 1994.

[10] S. Liu, "Silicon retina with adaptive filtering properties," in *Advances in Neural Information Processing Systems*, M. I. Jordan, M. J. Kearns, and S. A. Solla (Eds.). Cambridge, MA: MIT Press, 1998, vol. 10.

[11] C. Mead, *Analog VLSI and Neural Systems*, Reading, MA: Addison-Wesley, 1989.

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

[13] C. Tomazou, F. J. Lidgey, and D. G. Haigh, (Eds.), *Analogue IC Design: The Current-Mode Approach*. Stevenage, U.K.: Peregrinus, 1990.

[14] J. Lazzaro, S. Ryckebusch, M. Mahowald, and C. Mead, "Winner-take-all networks of $O(n)$ complexity," in *Advances in Neural Information Processing Systems*, D. Touretzky (Ed.). San Mateo, CA: Morgan Kaufmann, 1989, vol. 2, pp. 703–711.

[15] J. A. Starzyk, and X. Fang, "CMOS current mode winner-take-all circuit with both excitatory and inhibitory feedback," *Electron. Letters*, vol. 29, pp. 908–910, May 1993.

[16] S. DeWeerth and T. Morris, "CMOS current mode winner-take-all circuit with distributed hysteresis," *Electron. Letters*, vol. 31, pp. 1051–1053, June 1995.

[17] D. Robinson, "The mechanism of human smooth pursuit eye movement," *J. Physiology*, vol. 180, pp. 569–591, 1965.

[18] G. Indiveri and P. F. M. J. Verschure, "Autonomous vehicle guidance using analog VLSI neuromorphic sensors," in *Proc. Artificial Neural Networks-ICANN'97*, vol. 1327 of *Lecture Notes in Computer Science*, M. H. W. Gerstner, A. Germond, and J.-D. Nicoud (Eds.). Berlin, Germany: Springer-Verlag, 1997, pp. 811–816.

[19] J. Shi and C. Tomasi, "Good features to track," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[20] T. Delbrück, ""Bump" circuits for computing similarity and dissimilarity of analog voltages," in *Proc. IJCNN*, June 1991, pp. I-475–479.

[21] T. Delbrück, "3 silicon retinas for simple consumer applications," in *Intelligent Vision Systems Meeting*, Santa Clara, CA, June 1999, unpublished.

Giacomo Indiveri received the Laurea degree in electrical engineering from the University of Genova, Italy, in 1992.

From 1992 to 1995, he held a doctoral Fellowship within the National Research Program on Bioelectronic Technologies. From 1994 to 1996, he was a Research Fellow at the California Institute of Technology, Pasadena, CA. Currently, he is a Reasearch Assistant at the Institute for Neuroinformatics, Zurich, Switzerland, focusing on the design and implementation of analog VLSI neuromorphic systems. He is co-organizer of the Workshop on Neuromorphic Engineering held yearly in Telluride, CO, and co-teacher of the Computation in Neuromorphic Analog VLSI Systems" class tought at ETH Zurich.