Computational Sensor for Visual Tracking with Attention

Vladimir Brajovic, *Member, IEEE*, and Takeo Kanade, *Fellow, IEEE*

***Abstract*— This paper presents a VLSI embodiment of an optical tracking computational sensor which focuses attention on a salient target in its field of view. Using both low-latency massive parallel processing and top-down sensory adaptation, the sensor suppresses interference from features irrelevant for the task at hand, and tracks a target of interest at speeds of up to 7000 pixels/s. The sensor locks onto the target to continuously provide control for the execution of a perceptually guided activity. The sensor prototype, a 24 × 24 array of cells, is built in 2-μm CMOS technology. Each cell occupies 62 μm × 62 μm of silicon, and contains a photodetector and processing electronics.**

***Index Terms*— Attention, computational sensors, low-latency vision, smart sensors, visual attention, visual tracking.**

I. Introduction

CONVENTIONAL machine vision systems are put at a disadvantage by the separation between a camera for "seeing" the world, and a computer for "figuring out" what is seen, with the result that excessive *latency* and a lack of *top-down sensory adaptation* ensues. Computational sensors [10], on the other hand, integrate sensing and processing in a very large scale integration (VLSI) chip. Such a paradigm has the potential to: 1) reduce latency through massively parallel fine-grain computation, and 2) allow for adaptation of spatiotemporal properties of sensing based on the results of on-chip processing.

Motivated by neural processes in biological retinas, a great majority of computational sensors implement *local operations* on a single light sensitive VLSI chip. Notably, there is a growing body of research in so-called neuromorphic vision chips (for examples, see [10] and [12]). Typical examples of the operations implemented in these chips include spatial filtering and motion computation.

Local operations produce preprocessed "images"; therefore, a large quantity of data must be read out and further inspected before a decision for an appropriate action is made—usually a time-consuming process. While locally computed quantities could be used for adaptation within the local neighborhood, they cannot be used for global adaptation.

Our work has investigated *global operations* in computational sensors for improved adaptability and latency in machine vision systems [4]. Unlike local operations, global operations (e.g., histogram, MAX and MIN functions) produce less data for the description of the scene. When computed on-chip, global data can be quickly routed off chip, and in many tasks will be sufficient for rapid decision making. Global quantities are also available within the chip for top-down global adaptation.

This paper presents a tracking computational sensor which uses sensory attention for low-latency adaptive and robust tracking of local intensity peaks in the sensed image. Loosely speaking, tracking may be considered to be a global operation over an image; all we are concerned about is the location of the target of interest—a global property of the image for the tracking task. The remainder of the paper describes our motivation for using sensory attention for tracking, and provides the details of the VLSI implementation and performance test results.

Manuscript received November 7, 1997; revised February 26, 1998. This work was supported in part by the Office of Naval Research (ONR) under Contract N00014-95-1-0591, and in part by the National Science Foundation (NSF) under Grant MIP-9305494.

The authors are with The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA

Publisher Item Identifier S 0018-9200(98)05537-1.

II. Tracking Versus Attention

Our goal is to design a *fast* and *reliable* computational sensor which, given a task, provides *useful* information for *coherent interaction with the environment*. Low-latency visual tracking of salient targets in the field of view (FOV) is an important task in machine vision. Several issues must be addressed: 1) the problem of selecting a target from the background, 2) the problem of maintaining the tracking engagement (i.e., target locking), and 3) the problem of shifting tracking to new targets either when the presently engaged feature leaves the FOV, or when the user or a host computer decides to select and track another target in the FOV.

The problem of visual attention in biological systems concerns similar issues: 1) select a salient location in the image, 2) transfer local data from such a location to the higher processing stage, and 3) proceed by selecting a new interesting location [11]. By selecting only relevant retinotopic information in intermediate processing levels, the visual attention protects the limited communication and processing resources from an information overload at higher levels. However, Alport suggested that the need for attention goes beyond protecting limited resources during complex object recognition: *attention is needed to ensure behavioral coherence* [1]. Namely, selective processing is necessary in order to isolate parameters (e.g., target location, velocity, size) for the appropriate action.

For behavioral coherence, the following requirements for attention are suggested [1]: 1) *low latency requirement*—attentional system must operate in fast-changing environments; 2) *locking requirement*—the appropriate attentional engagement has to be maintained while action

is executed; 3) *purposive shift requirement*—the system must be able to override the attentional engagement when faced by environmental threats or opportunities. The purposive shift requirement contradicts the locking requirement, but humans and other species adopt a combination of two partial solutions: 1) *intentional shift*—generate a shift based on a range of heuristics and experiences; and 2) *opportunistic shift*—elicit a shift in response to detecting a more "attractive" sensory cue.

The requirements for visual attention are analogous to requirements for low-latency robust tracking. In fact, if the salient features for the tracking device are also salient for the attentional system, the distinction between the tracking and attention shifts fades. It is then permissible to speak of our implementation as the implementation of a primitive sensory attention system. The term *sensory* attention (as opposed to *visual* attention) is used to emphasize the fact that the data selection in our sensor is performed over the sensory signal, rather than over the retinotopic scene representation in intermediate processing levels on which attention may operate in brains.

Our implementation of sensory attention meets both the low-latency requirement and the locking requirement while providing mechanisms for both intentional and opportunistic shifts. These features are weakly addressed in other recently reported attention-related circuits. In [14], a one-dimensional array of 19 cells electrically receives a saliency map and uses delayed transient inhibition at the selected location to model covert attentional scanning as suggested by Koch and Ullman [11]. The circuit mimics one aspect of the object recognition process in which the attention roams across the conspicuous features of an object (e.g., discontinuities, contours, etc.).

However, it is questionable how this implementation would meet the low latency requirement in fast-changing scenes; the attention may take a long time, if ever, before it comes across the *task-relevant* target when several equally salient (but not equally relevant) targets are in the FOV. In [8], the reduced version of this circuit is used in a one-dimensional optical tracking sensor which computes a saliency map on-chip, but has a weak mechanism for ensuring a locking requirement with complex scenes. Both designs inherently implement the opportunistic shifts, but neither provides the mechanism for the intentional shifts.

III. Implementation

Our tracking computational sensor optically receives an image, selects a local intensity peak in that input image, and continuously reports the location and magnitude of the selected peak. In the context of this paper, the selected point is called a *target*. The location of the target is global information which is reported as the output. The location of the target is also used internally to adapt the location of the attention to meet the locking requirement.

A. Target Selection and Locking

An image is optically focused onto the array of photodetectors. Generated photocurrents are fed to the winner-take-all (WTA) circuit, which is responsible for the feature

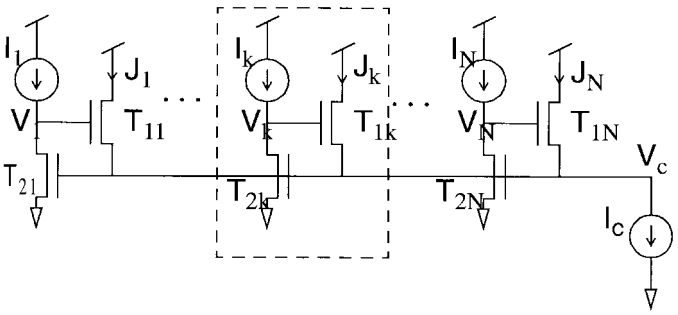


Fig. 1. Schematic diagram of the winner-take-all circuit. Boxed area indicates one cell.

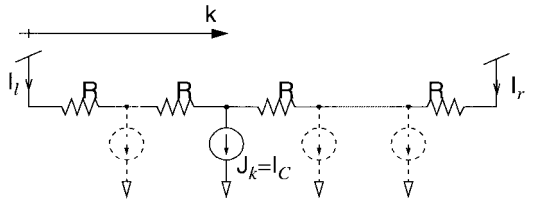


Fig. 2. Resistive network for position detection.

selection. A WTA circuit is shown in Fig. 1; it was originally proposed in [2] and [13]. Currents $I_1 \cdots I_N$ are the input photocurrents, while currents $J_1 \cdots J_N$ are the outputs of the WTA circuit. The cell receiving the largest photocurrent $I_k = \max(I_1 \cdots I_N)$ responds with nonzero output current $J_k = I_c$, while other cells respond with zero currents, i.e., $J_i = 0$ for $i \neq k$. The resolving power of our design is 1 part in 1000, i.e., a clear single winner is found if the winner is 0.1% above the next strongest input [3]. The winning photocurrent establishes and holds the common voltage V_c. For small input currents like those produced by light detection, the voltage V_c is proportional to the logarithm of the winning input current. Therefore, the intensity of the winner is accessed globally by monitoring the voltage on the common wire.

Since only the winning cell responds with nonzero current, the WTA effectively provides 1-of-N binary encoding of the winner's position. A digital on-chip decoder easily converts this code to any other binary code such as a natural binary. In addition, there are efficient analog means for winner localization [15]. One example is shown in Fig. 2. The outputs from each WTA cell are connected to nodes of a linear resistive network. The WTA ensures that only one of these currents is nonzero (i.e., I_c). The network behaves as a current divider, and the current I_c is split into I_l and I_r. By measuring currents I_l and I_r, the position k is found as

$$k = \frac{I_r}{I_c} N \qquad I_c = I_l + I_r. \tag{1}$$

The WTA cells can be physically laid out in a two-dimensional array. Still, one of the cells wins and provides nonzero output current. Using the method of projections [9], the position of this current in two dimensions is found by solving two one-dimensional problems. This concept is implemented as shown in Fig. 3. Two copies of the output current are summed into the horizontal and vertical bus, respectively. The total current in these buses represents the desired projections onto the x

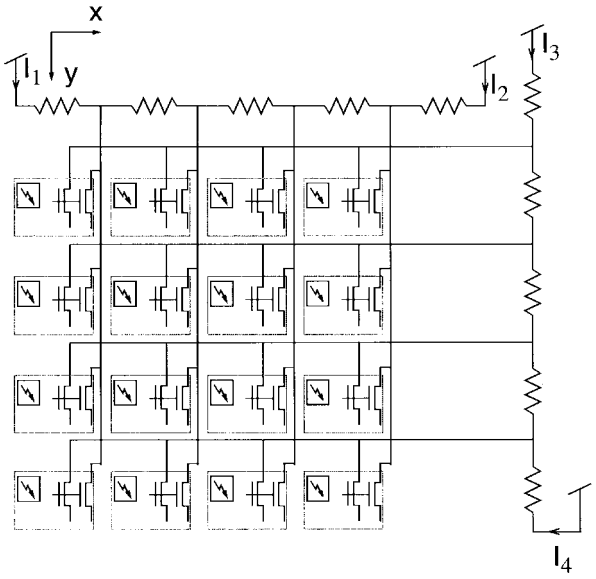


Fig. 3. Two-dimensional WTA computational sensor.

and y axes. Then, two linear resistive networks are used at the periphery of the array to locate the winner in x and y direction.

The WTA circuit locates the absolute maximum in the entire image. In practical applications, there are often several targets in the scene. The target of interest is not necessarily the strongest. We need to direct the sensor's attention toward that target. Once the target is selected, we need a mechanism that will lock and track the target while the target is of interest and/or a perceptually guided goal is being executed. Recalling the analogy with the visual attention, the selection mechanism corresponds to opportunistic attention shifts initiated by "telling" the observer where to "look," while the locking mechanism meets the locking requirement and maintains attention engagement under motion.

Our implementation solves these issues by inhibiting a portion of the saliency map, thus restricting the activity of the WTA circuit to a programmable active region—a subset of the array. The active region is programmed by appropriate row and column addressing (see Fig. 4). There are two modes of operation: 1) select mode and 2) lock mode. In the *select mode*, the active region is user-defined by the external addressing [Fig. 4(a)]. The active region is of arbitrary size and location. The target selected by the sensor is the absolute maximum within this region. In *lock mode*, the sensor itself dynamically defines a small (e.g., 3×3 in this implementation) active region centered at the most recent location of the target [Fig. 4(b)]. The select mode directs the attention toward a feature that is useful for the task at hand. For example, a user may want to specify an initial active region, aiding the sensor in attending to the relevant local peak in the scene. Then, the lock mode is enabled for locking onto the selected feature. In the lock mode, the 3×3 cell active region is centered at the location of current attention target. If the target moves, one of the eight active neighbors in the WTA array will receive the winning intensity peak and automatically update the position of the 3×3 active region. It is now clear that the salient target is not necessarily the peak of the absolute maximum intensity in the image. The ability of the sensor to define its own active region is an example of the top-down sensory adaptation presently missing in conventional machine vision systems.

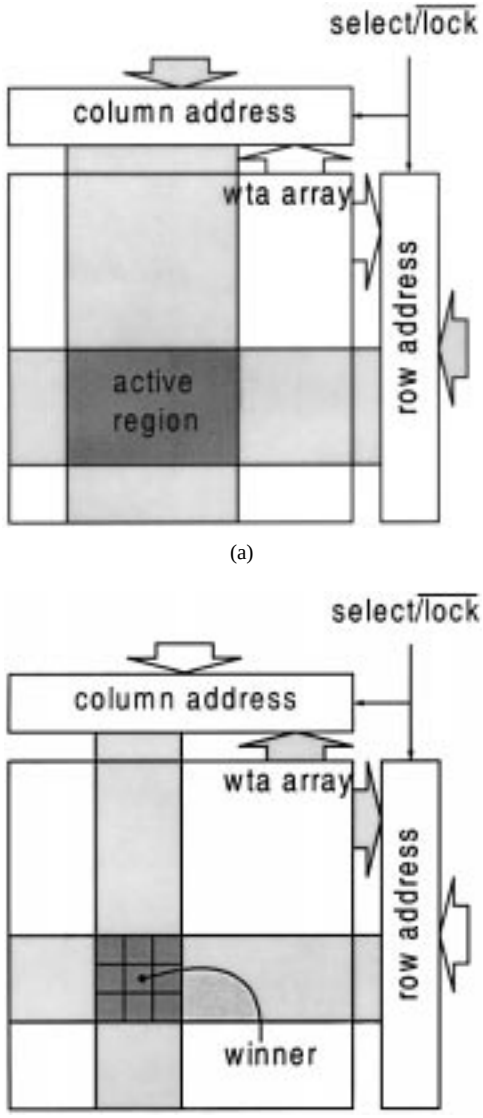


Fig. 4. Modes of operation for the sensory attention computational sensor: (a) select mode and (b) lock mode.

A circuit diagram of the WTA cell which implements the inhibition mechanism is shown in Fig. 5. The shunting path for the photocurrent is provided through the transistors T_5 and T_6. To maintain the cell as active, both $\overline{\texttt{col}}$ and $\overline{\texttt{row}}$ signals must be asserted (i.e., must be zero). This inhibition mechanism can be interpreted in two ways. One way is to say that, when the photocurrent is shunted, the cell effectively "sees" zero current and cannot win. The other way is to say that the switches T_5 and T_6 clamp the gate of T_1 to ground, thus preventing T_1 from conducting any current.

The control of the active region is achieved from the periphery of the two-dimensional WTA array. The peripheral logic across three columns is shown in Fig. 6. Similar logic is implemented for row addressing. In the select mode, the active

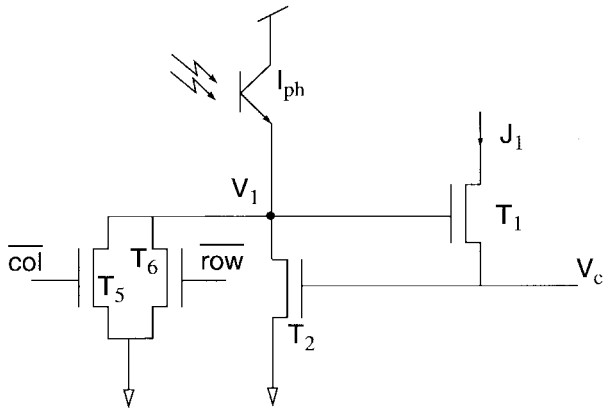


Fig. 5. WTA cell with inhibition. T_5 and T_6 implement inhibition.

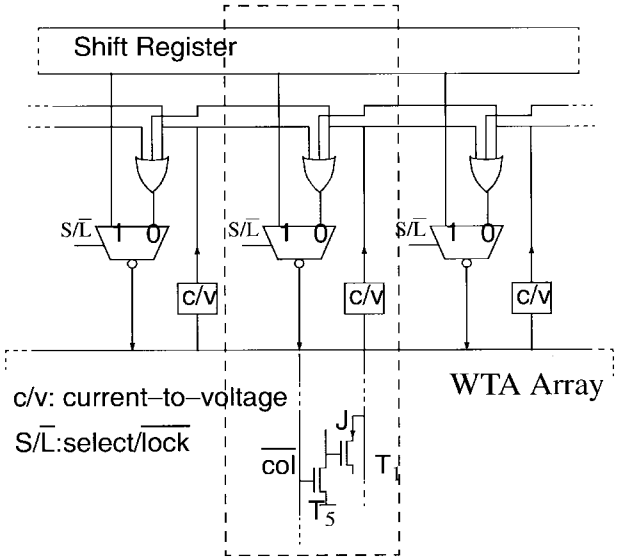


Fig. 6. Peripheral logic for central control of the active region. The boxed area indicates one column. Similar logic is used for row addressing (not shown).

column band is programmed by the content of the shift register. There are no restrictions on the width or location of the band, as any bit pattern may be entered into the shift register.

B. Acquiring New Targets

In order to acquire new targets, a user or a host computer may, at any point during tracking, switch into the select mode. The strongest target within the active region specified in the peripheral shift registers will be acquired. However, one special case is of interest. With moving objects, the feature which is being tracked may reach the sensor's edge and fall out of the FOV. In order to ensure coherent transition in these situations, the following heuristic is implemented. When the tracked feature reaches one of the four edges of the array, the sensor momentarily switches from the lock mode to the select mode (despite the user-defined signal, $\texttt{user\_select}/\overline{\texttt{lock}} = 0$, dictating the lock mode). The active region specified in the shift registers is enabled and the absolute maximum is selected therein. If the newly selected feature is no longer on the edge, the sensor automatically goes back to the lock mode, shrinks

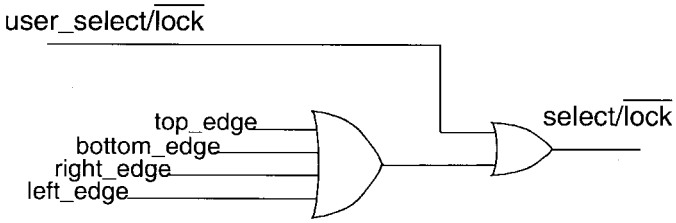


Fig. 7. Logic for automatic switching between *select* and *lock* modes.

the active region to a 3×3 size around the new target, and continues to track the new target. This heuristic is implemented by overriding the user-defined signal, $\texttt{user\_select}/\overline{\texttt{lock}}$, with the circuit shown in Fig. 7.

IV. Static Performance

The static performance has been tested on an early 1-D prototype with 20 cells fabricated in 2 μCMOS technology. A cell occupies 40×59 microns. A phototransistor occupies about 60% of the cell's area. A dark plate bisected by a bright vertical line (i.e., a stretched white wire) mounted on a calibrated translational stage served as the target scene. The optical setup was adjusted so that the target (i.e., the white wire) was smaller or comparable to the pixel size; therefore, only one photodetector or its immediate neighbors received an appreciable amount of light at a time. The illumination over the sensor FOV was approximately uniform. The target was moved horizontally throughout the entire field of view (i.e., 16 mm) in steps of 0.2 mm. The winning cell location (i.e., position) reported by the sensor was measured. The results are graphed in Fig. 8(a). Also measured was the voltage on the common wire V_c. Using computer simulation, the measured common voltage was converted to apparent input photocurrent [Fig. 8(b)].

Fig. 8 shows that the cells' winning behavior as well as the winner localization is reported as expected. Namely, a particular cell remains a winner as long as the main portions of the bright target are focused on it. In the graph, this appears as the staircase line. As the target is moved, its image leaves one cell and begins contributing photocurrent to the next one. At some point, the cell receiving the target wins and takes control of the common voltage. As the target moves toward the center of the new winning cell, the intensity of the winning input current increases. The cell continues to win as the target passes the center, but its input current diminishes. In the meantime, the next cell begins to receive an increasing amount of light and the process is repeated. Therefore, as the target passes over the winning cell, the measured common voltage increases, peaks, and then decreases. This behavior is clearly displayed in Fig. 8. The spacing of the peaks in the environmental coordinates is 0.79 mm which, for the given experimental setup, matches the pitch of the cells.

Another important observation can be made from an examination of Fig. 8. Even though a target of a constant intensity is scanned over the sensor, it does not result in equal peaks of common voltage. This is due to the device mismatch; the same target can be seen as one whose intensity apparently varies. Also observed is a periodic pattern in peak variation

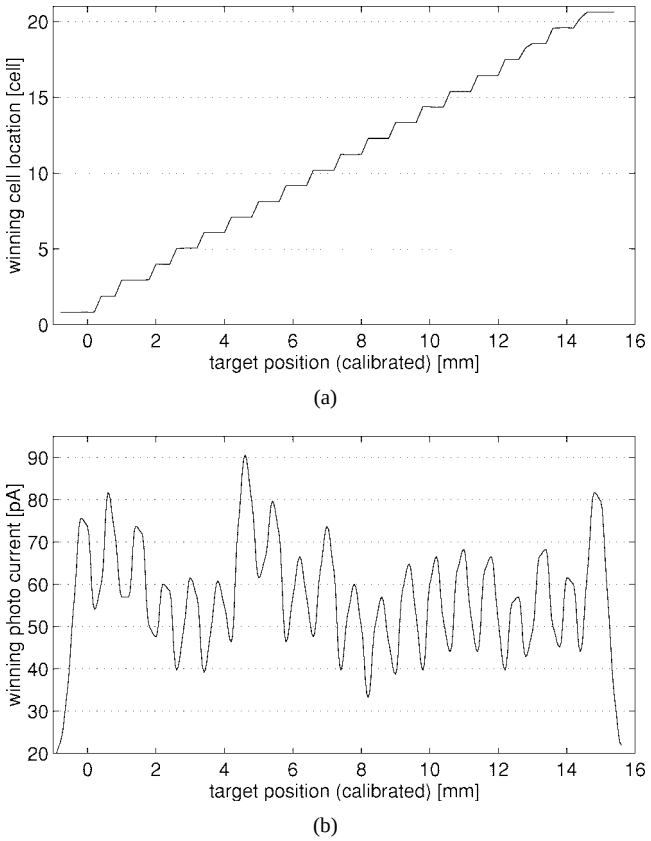


Fig. 8. Static performance: winning cell localization reported by (a) 1-D tracking sensors and (b) an apparent intensity of the winning photocurrent.

over the chip area. This is due to the striation effects. The period and relative amplitude of this variation are in good agreement with the findings in [2]. The ratio of the two extreme perceived currents (i.e., 90 pA at 4.5 mm, and 57 pA at 8.5 mm) is about 1.6. This means that if this circuit is to always correctly identify the winner, the strongest spot in the image must be about 1.6 times higher than the background. Transistor mismatch factors as high as 2 are typical for MOS devices in subthreshold operations. This imposes a serious limitation on the nature of the input image, especially in the select mode. However, the devices within a small neighborhood match better. Since the sensor activates only a small neighborhood in the lock mode, most of the time it is sufficient that the local peak is only about 20% above the immediate neighbors.

V. Dynamic Performance

A first-order small signal analysis in [13] shows that the WTA circuit is stable if $I_c > 4I_p(C_c/C)$, where C is the parasitic capacitance on the node V_1, C_c is the capacitance of the common wire, I_p is the winning input photocurrent, and I_c is the common current source. Even though the ratio C_c/C is large and scales up with number of WTA cells, the detected photocurrents I_p are small; the stable condition is easily maintained by controlling a common current I_c. In experimentation with a 24 $\times$ 24 cell two-dimensional prototype, $I_c = 10$ μA sufficed. For larger arrays, the current I_c needs to be scaled up proportionally. For example, a 100 $\times$ 100 array would require $I_c \approx 175$ μA.

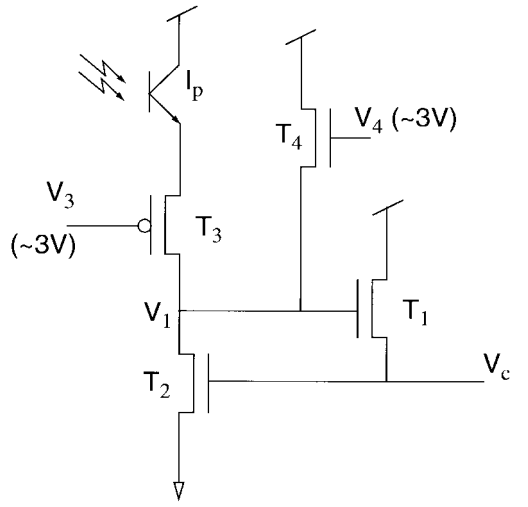


Fig. 9. WTA cell with improved dynamic performance.

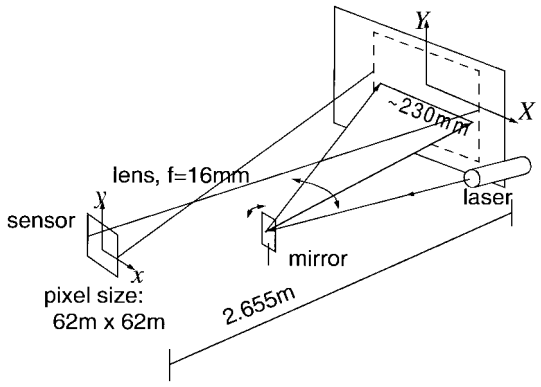


Fig. 10. Experimental setup for evaluating dynamic performance of the tracking computational sensor.

The transient response of the WTA circuit is important when tracking moving targets. A first-order small signal analysis done in [13] shows that the winning/losing dynamics of the circuit is a function of the parasitic capacitance at the input node V_1. This capacitance includes the capacitance of the photodetector and the capacitances of the gates and drains attached to this node. For a cell to win or lose, this capacitance must be charged and discharged with the photocurrent. For average room illumination, the photocurrents are very small, i.e., much less than 1 nA. Therefore, the WTA circuit in its original configuration is slow.

To improve the dynamic performance of the WTA circuit, several measures can be taken: 1) increase photocurrent, 2) decrease parasitic capacitance, and 3) reduce the voltage swing on the capacitance C. A modified WTA cell that implements all three measures is shown in Fig. 9. The phototransistor amplifies the photocurrent, T_3 isolates the capacitance of the photodetector, and T_4 acts as a pull-up and limits the voltage swing.

The dynamic performance is evaluated for a 24 $\times$ 24-cell tracking computational sensor. Each cell occupies a 62 μm $\times$ 62 μm square. The phototransistor takes about 30% of the cell's area. The experimental setup is shown in Fig. 10. A scanning mirror projects a beam of light onto a uniform

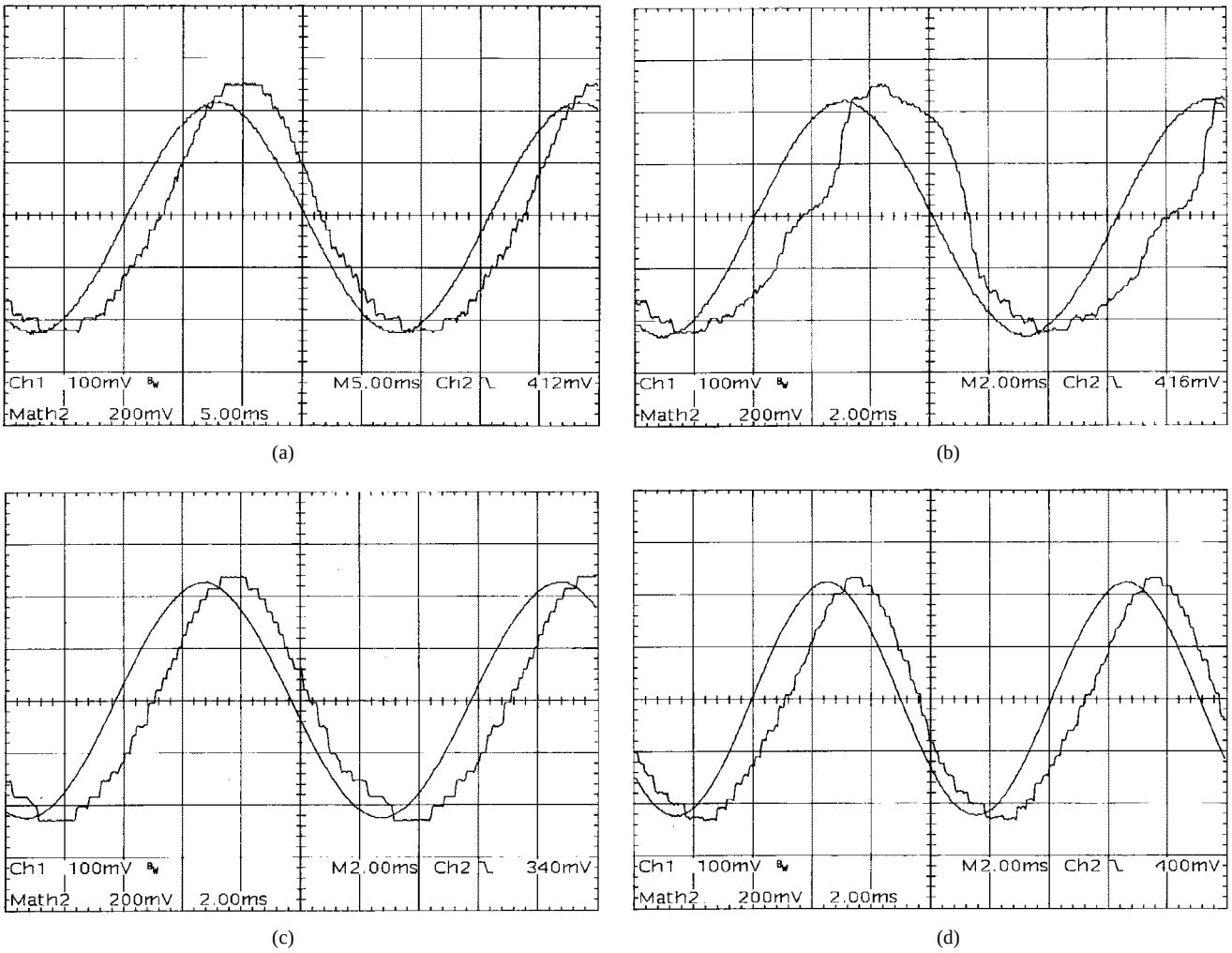


Fig. 11. Tracking performance: (a) without the current buffer and without the pull-up, $f = 33$ Hz, (b) without the current buffer and without the pull-up, $f = 83$ Hz, (c) tracking performance with the current buffer, $f = 83$ Hz, and (d) with both the current buffer and pull-up, $f = 100$ Hz.

piece of cardboard, thereby producing a dot which travels along a straight line. The sensor images the scene and tracks the moving dot. The rows of the sensor are aligned with the trajectory of the laser dot; therefore, only the x position needs to be observed. The mirror is driven from an adjustable sinusoidal oscillator. From the geometry of the setup and the scanning frequency, the maximum velocity of the target is inferred in image coordinates and expressed in cells/s.

A. Influence of the Current Buffer and the Pull-Up

The first set of tests is performed to show the contributions of the current buffer T_3 and the pull-up transistor T_4. The effects of each can be turned on or off by an appropriate biasing of V_3 and V_4, respectively. Without the buffer and the pull-up, the sensor was reliably tracking up to the scanning frequency of 33 Hz or 2304 cells/s. Fig. 11(a) shows two measured waveforms: 1) the feature's position x as reported by the tracking sensor, and 2) the sinusoid driving the mirror. If the frequency of the mirror is further increased, the reported position begins to distort. This is illustrated in Fig. 11(b) for the scanning frequency of 83 Hz. As expected, the tracking capability of the sensor starts to break down in the middle of the trajectory where the velocity of the target is the greatest.

TABLE I
SUMMARY OF THE EXPERIMENTAL FINDINGS FOR THE WINNING/LOSING DYNAMIC PERFORMANCE FOR THE WTA CIRCUIT

Basic WTA	WTA with the buffer	WTA with the buffer and pull–up	Unit
2,304	5,794	6,981	cells/s
680	1,711	2,061	ccd_pix/ccd_frame

In the next experiment, the current buffer is turned on by biasing V_3. As expected, the dynamic performance improved; the maximum tracking frequency is increased to about 83.3 Hz, or 5794 cells/s. This is shown in Fig. 11(c); the previously distorted waveform for the target's position now better resembles the sinusoid. Finally, the pull-up transistor is turned on by biasing V_4. The dynamic performance is slightly improved as shown in Fig. 11(d)—the feature tracking is improved to about 100 Hz, or 6981 cells/s. Table I summarizes the dynamic performance of the tracking sensor. For purposes of comparison illustration, the maximum target velocities are also expressed in CCD pixels per CCD frame. Conventional machine vision systems using CCD cameras cannot cope with such fast moving targets. This conversion assumes a CCD with

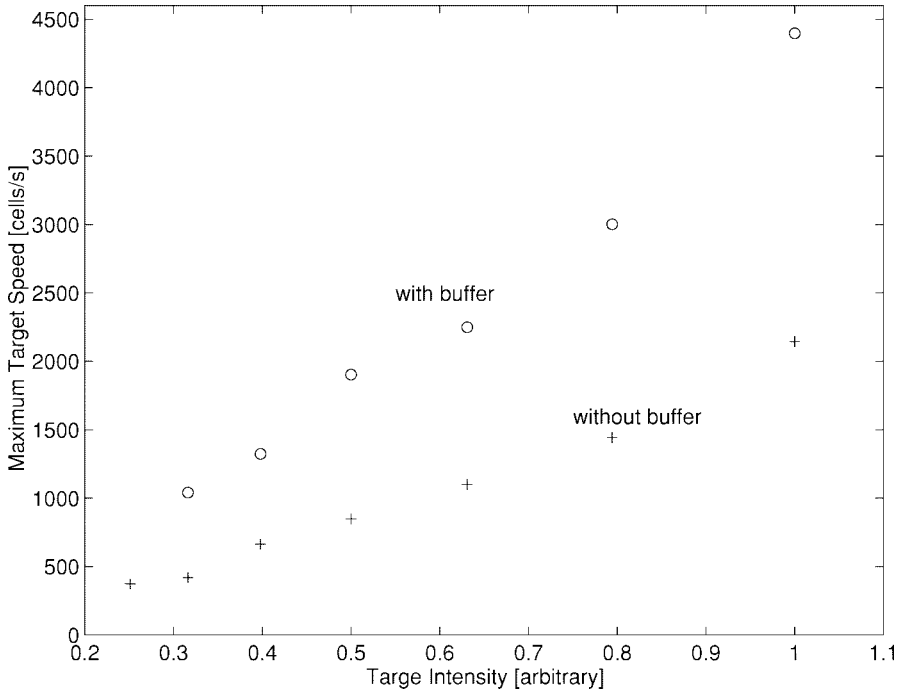


Fig. 12. Maximum angular velocity of the attention shifts as a function of the relative feature intensity.

7 μm pixel and 30 frames/s. Excluding the photocurrents, the static current consumption in these experiments for the whole chip is approximately 50 μA.

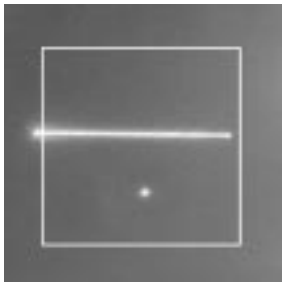
The capacitance of the photodetector is the largest parasitic; therefore, it is not surprising that the main improvement comes from isolating the photodetector's capacitance. It is not clear, however, how beneficial the use of the phototransistor is. The current gain for small photocurrents is in the range of 1–10. This benefit, however, may be offset by the fact that the base is floating, and that its capacitance cannot be pinned by the buffering transistor T_3.

B. Influence of the Target's Intensity

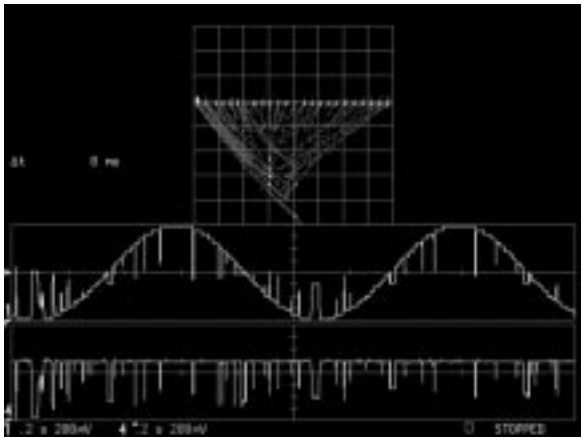
Another set of experiments is performed to evaluate how the intensity of the feature influences the dynamic performance. Using neutral density filters placed in front of the sensor's lens, the light is controllably attenuated. For each filter, the frequency of the mirror is increased until the waveform of the target's position begins to distort. This way, the maximum frequency is estimated for each intensity. Two sets of experiments are performed: 1) without the buffer and the pull-up, and 2) with the buffer and the pull-up. The results are graphed in Fig. 12.

C. Select/Lock Performance

The robust performance of the sensory attention and sensor's select/lock feature is illustrated in Figs. 13 and 14. Fig. 13(a) shows a CCD camera image of the scene viewed by the sensor in this experiment. In addition to the sinusoidally driven laser dot described in earlier experiments, the scene includes an arbitrarily roaming dot produced by a hand-held laser pointer. (The scanning dot appears as a line, since the scanning frequency exceeds the speed of the conventional CCD camera.) Fig. 13(b) shows the oscilloscope display. The top square display is the oscilloscope's X–Y display mode, while the middle and bottom displays are the temporal signals of the X and Y target positions, respectively. Fig. 13(b) shows



(a)



(b)

Fig. 13. Interference between two targets: (a) scanned laser dot and hand-held laser dot and (b) signal traces as reported by the sensor when the entire array is active.

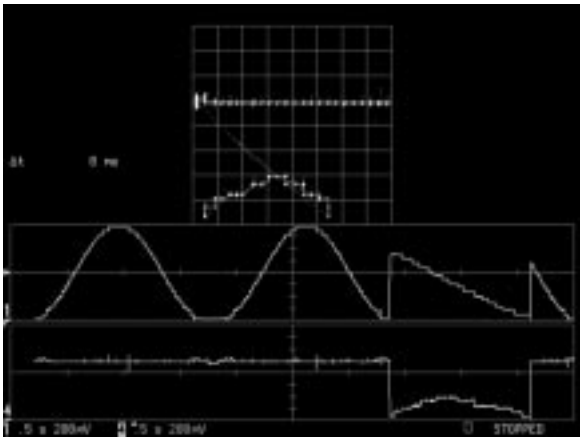


Fig. 14. Reliable tracking of two targets in the lock mode.

that, when there is no sensory selection, the WTA array exhibits unreliable behavior; the two targets interfere with each other, and the sensor erratically "jumps" between them.

When the sensory attention is enabled in the lock mode, the sensor exhibits robust and coherent behavior. In Fig. 14, we see that the sensor first tracks the scanned laser dot. As the scanned dot leaves the array at the left edge, the sensor goes to the select mode. In this experiment, the select region is defined to be the entire array; therefore, the hand held laser is the new absolute peak that is selected within the array. Since this new target is not on one of the four edges, the sensor automatically returns to the lock mode and defines the 3×3 region around the target. As the hand-held laser is being tracked, the scanning laser returns to the FOV. (To visualize this, mentally reconstruct the sinusoid in the middle display of Fig. 14.) However, the scanned laser does not interfere, as it is outside of the 3×3 active region positioned on the hand-held laser. Finally, as the hand-held laser leaves the FOV (at the bottom edge), the sensor again momentarily enables the whole array in the select mode and the scanned laser is picked up again as the next target for tracking. From the middle trace in Fig. 14, we see that the scanned laser is picked up when it is approximately in the middle of the FOV, going left.

VI. Conclusion

The proposed VLSI implementation of the tracking computational sensor exhibits several interesting features. It senses input images and produces a few global results: the position and magnitude of the target being tracked. With no latency, these global results are reported off-chip via few output pins. Furthermore, in the lock mode, the global results are used internally for programming a 3×3 active region, thus providing a low-latency top-down adaptation for securing robust performance in a rapidly changing environment. Such an adaptation, and hence reliable performance, is currently missing in conventional machine vision systems. The sensor robustly tracks targets moving up to 7000 pixels/s, while consuming only 0.25 mW of static power.

The issues addressed by the tracking sensor are analogous to issues facing the implementation of rudimentary visual attention. Therefore, the tracking computational sensor implements primitive attention. Bright spots in received images are considered salient and are potential targets for tracking. If a particular saliency is of interest, such as color or a particular intensity pattern, then optical (or electronic) preprocessing is needed. In general, the input images to the tracking chip can be considered to be optical saliency maps that encode "conspicuousness" of targets through the scene. Broad spectrum intensity images used in our experimentation are trivial saliency maps.

References

[1] A. Allport, "Visual attention,"in *Foundation of Cognitive Science*, M. Posner, Ed. Cambridge, MA: M.I.T. Press, 1989, pp. 631—682.

[2] A. G. Andreou *et al.*, "Current-mode subthreshold MOS circuits for analog VLSI neural systems," *IEEE Trans. Neural Networks*, vol. 2, pp. 205–213, Mar. 1992.

[3] V. Brajovic, "Computational sensors for global operations in vision," Ph.D. thesis, Carnegie Mellon Univ., Jan. 1996.

[4] V. Brajovic and T. Kanade, "Computational sensors for global operations," in *IUS Proc.*, pp. 621–630, 1994.

[5] ———, "A VLSI sorting image sensor: Global massively parallel intensity-to-time processing for low-latency, adaptive vision," *IEEE Trans. Robotics Automation*, 1997.

[6] ———, "Sensory attention: A computational sensor paradigm for low-latency, adaptive vision," *IEEE Trans. Robotics Automation*, Sept. 1997.

[7] ———, "A sorting image sensor: An example of massively parallel intensity-to-time processing for low-latency computational sensors," in *Proc. 1996 IEEE Int. Conf. Robotics Automation*, Minneapolis, MN, Apr. 1996.

[8] T. K. Horiuchi, T. G. Morris, C. Koch, and S. P. DeWeerth, "Analog VLSI circuits for attention-based, visual tracking," in *Advances in Neural Information Processing Systems*, Vol. 9. Cambridge, MA: M.I.T. Press, 1997.

[9] B. Horn, *Robot Vision.* Cambridge, MA: M.I.T. Press, 1986.

[10] T. Kanade and R. Bajcsy, "Computational sensors: A report from DARPA workshop," *IUS Proc.*, 1993.

[11] C. Koch and S. Ullman, "Shifts in selective visual attention: Toward the underlying neural circuitry," in *Matters of Intelligence*, L. M. Vaina, Ed. Dordrecht, The Netherlands: Reidel, 1987, pp. 115–141.

[12] C. H. Koch and H. Li, Eds., *Vision Chips: Implementing Vision Algorithms with Analog VLSI Circuits.* New York: IEEE Computer Society, 1995.

[13] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. Mead, "Winner-take-all networks of $O(n)$ complexity," in *Advances in Neural Information Processing Systems Vol. 1*, D. Tourestzky, Ed. San Mateo, CA: Morgan Kaufmann, 1988, pp. 703–711.

[14] T. G. Morris and S. P. DeWeerth, "Analog VLSI circuits for covert attentional shifts," presented at *MicroNeuro'96*, Lausanne, Switzerland.

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

Vladimir Brajovic (S'88–M'89) received the Dipl.Eng.E.E. degree from the University of Belgrade, the M.S.E.E. degree from Rutgers University, and the Ph.D. degree in robotics from Carnegie Mellon University.

He is currently a Research Scientist with the Carnegie Mellon Robotics Institute where he is the Director of the VLSI computational sensor laboratory. His research interest include computational sensors, analog and mixed-signal VLSI, machine vision, robotics, signal processing, optics, and sensors.

Dr. Brajovic received the Anton Philips Award at the 1996 IEEE International Conference on Robotics and Automation for his work on an adaptive computational image sensor.

Takeo Kanade (M'80–SM'88–F'92) received the Doctoral degree in electrical engineering from Kyoto University, Japan, in 1974.

After holding a faculty position at the Department of Information Science, Kyoto University, he joined Carnegie Mellon University in 1980, where he is currently the U. A. Helen Whitaker Professor of Computer Science and Director of the Robotics Institute. He has written more than 150 technical papers on computer vision, sensors, and robotics systems.

Dr. Kanade has been elected to the National Academy of Engineering, and is a Founding Fellow of the American Association of Artificial Intelligence. He has received several awards, including the Joseph Engelberger Award, JARA Award, and several best paper awards at international conferences. He has served on many government, industry, and university advisory or consultant committees, including the Aeronautics and Space Engineering Board (ASEB) of the National Research Council, NASA's Advanced Technology Advisory Committee (Congressionally mandate committee), and the Advisory Board of the Canadian Institute for Advanced Research.