LOW-POWER TRACKING IMAGE SENSOR BASED ON BIOLOGICAL MODELS OF ATTENTION

Alexander Fish, Liby Sudakov-Boreysha, Orly Yadid-Pecht

Abstract: This paper presents implementation of a low-power tracking CMOS image sensor based on biological models of attention. The presented imager allows tracking of up to N salient targets in the field of view. Employing "smart" image sensor architecture, where all image processing is implemented on the sensor focal plane, the proposed imager allows reduction of the amount of data transmitted from the sensor array to external processing units and thus provides real time operation. The imager operation and architecture are based on the models taken from biological systems, where data sensed by many millions of receptors should be transmitted and processed in real time. The imager architecture is optimized to achieve low-power dissipation both in acquisition and tracking modes of operation. The tracking concept is presented, the system architecture is shown and the circuits description is discussed.

Keywords: Low-power image sensors, image processing, tracking imager, models of attention, CMOS sensors

ACM Classification Keywords: B.7.0 Integrated circuits: General, I.4.8 Image processing and computer vision: scene analysis: tracking

1. Introduction

Real time visual tracking of salient targets in the field of view (FOV) is a very important operation in machine vision, star tracking and navigation applications. To accomplish real time operation a large amount of information is to be processed in parallel. This parallel processing is a very complicated task that demands huge computation resources. The same problem exists in biological vision systems. Compared to the state-of-the-art artificial imaging systems, having about twenty millions sensors, the human eye has more than one hundred million receptors (rods and cones). Thus, the question is how biological vision systems succeed to transmit and to process such a large amount of information in real time? The answer is that to cope with potential overload, the brain is equipped with a variety of attentional mechanisms [1]. These mechanisms have two important functions: (a) attention can be used to select relevant information and/or to ignore the irrelevant or interfering information; (b) attention can modulate or enhance the selected information according to the state and goals of the perceiver. Most models of attention mechanisms are based on the fact that a serial selection of regions of interest and their subsequent processing can greatly facilitate the computation complexity. Numerous research efforts in physiology were triggered during the last five decades to understand the attention mechanism [2]-[10]. Generally, works related to physiological analysis of the human attention system can be divided into two main groups: those that present a spatial (spotlight) model for visual attention [2]-[4] and those following object-based attention [5]-[10]. The main difference between these models is that the object-based theory is based on the assumption that attention is referenced to a target or perceptual groups in the visual field, while the spotlight theory indicates that attention selects a place at which to enhance the efficiency of information processing.

The design of efficient real time tracking systems mostly depends on deep understanding of the model of visual attention. Thus, a discipline, named neuromorphic VLSI that imitates the processing architectures found in biological systems as closely as possible was introduced [11]. Both spotlight and object-based models have been recently implemented in analog neuromorphic VLSI design [12]-[23]. Most of them are based on the theory of selective shifts of attention which arises from a saliency map, as was first introduced by Koch and Ullman [12]. Object-based selective attention systems VLSI implementations in 1-D and lately 2-D were presented by Morris et al [13]-[16]. An additional work on an analog VLSI based attentional search/tracking was presented by Horiuchi and Niebur in 1999 [17].

Many works on neuromorphic VLSI implementations of selective attention systems have been presented by Indiveri [19]-[21] and others [22]-[23]. In 1998 Brajovic and Kanade presented a computational sensor for visual

tracking with attention. These works often use winner-take-all (WTA) [24] networks that are responsible for selection and tracking inputs with the strongest amplitude. This sequential search method is equivalent to the spotlight attention found in biological systems.

Most previously presented neuromorphic imagers utilize image processing implemented on the focal plane level and employ photodiode or phototransistor current-mode pixels. Typically, each pixel consists of a photo detector and local circuitry, performing spatio-temporal computations on the analog signal. These computations are fully parallel and distributed, since the information is processed according to the locally sensed signals and data from pixel neighbors. This concept allows reduction in the computational cost of the next processing stages placed in the interface. Unfortunately, when image quality and high spatial resolution are important, image processing should be performed in the periphery. This way a high fill factor (FF) can be achieved even in small pixels.

This paper presents implementation of a low-power tracking CMOS image sensor based on a spotlight model of attention. The presented imager allows tracking of up to N salient targets in the field of view. Employing image processing at the sensor focal plane, the proposed sensor allows parallel computations and is distributed, but on the other hand most of the image processing is performed in the array periphery, allowing image quality and high spatial resolution. The imager architecture is optimized to achieve low-power dissipation both in acquisition and tracking modes of operation. This paper is a continuation of the work presented in [25], where we proposed to employ a spotlight model of attention for the bottleneck problem reduction in high resolution "smart" CMOS image sensors and of the work presented in [26], where the basic concept for an efficient VLSI tracking sensor was presented.

Section 2 briefly describes spotlight and object-based models of attention and presents system architecture of the proposed sensor. Low-power considerations, as well imager circuits description are shown in Section 3. Section 4 discusses advantages and limitations of the proposed system. Conclusions and future work are presented in Section 5.

2. Tracking Sensor Architecture

The proposed tracking sensor operation is based on the imitation of the spotlight model of visual attention. Because this paper presents concepts taken from different research disciplines, first, a brief description of existing models of attention is presented for the readers that are not familiar with this field. Then, the proposed sensors architecture is shown.

2.1 Existing Attention Models

Much research was done in attention during the last decades and numerous models have been proposed over the years. However, there is still much confusion as to the nature and role of attention. All presented models of attention can be divided to two main groups: spatial (spotlight) or early attention and object-based, or late attention. While the object-based theory suggests that the visual world is parsed into objects or perceptual groups, the spatial (spotlight) model purports that attention is directed to unparsed regions of space. Experimental research provides some degree of support to both models of attention. While both models are useful in understanding the processing of visual information, the spotlight model suffers from more drawbacks than the object-based model. However, the spotlight model is simpler and can be more useful for tracking imager implementations, as will be shown below.

2.1.1 The Spatial (Spotlight) Model

The model of spotlight visual attention mainly grew out of the application of information theory developed by Shannon. In electronic systems, similar to physiological, the amount of the incoming information is limited by the system resources. There are two main models of spotlight attention. The simplest model can be looked upon as a spatial filter, where what falls outside the attentional spotlight is assumed not to be processed. In the second model, the spotlight serves to concentrate attentional resources to a particular region in space, thus enhancing processing at that location and almost eliminating processing of the unattended regions. The main difference between these models is that in the first one the spotlight only passively blocks the irrelevant information, while in the second model it actively directs the "processing efforts" to the chosen region.

Figure 1(a) and Figure 1(b) visually clarify the difference between the spatial filtering and spotlight attention.

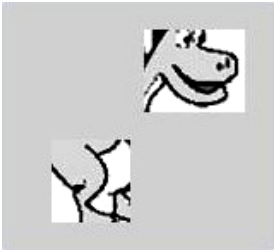

Figure 1 (a). An example of spatial filtering

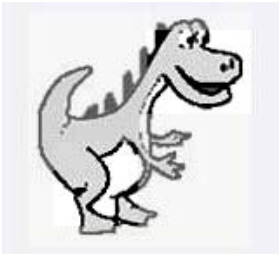

Figure 1 (b). An example of spotlight model of attention

A conventional view of the spotlight model assumes that only a single region of interest is processed at a certain time point and supposes smooth movement to other regions of interest. Later versions of the spotlight model assume that the attentional spotlight can be divided between several regions in space. In addition, the latter support the theory that the spotlight moves discretely from one region to the other.

2.1.2 Object-based Model

As reviewed above, the spotlight metaphor is useful for understanding how attention is deployed across space. However, this metaphor has serious limitations. A detailed analysis of the spotlight model drawbacks can be found in [1]. An object-based attention model suits more practical experiments in physiology and is based on the assumption that attention is referred to discrete objects in the visual field. However being more practical, in contrast to the spotlight model, where one would predict that two nearby or overlapping objects are attended as a single object, in the object-based model this divided attention between objects results in less efficient processing than attending to a single object. It should be noted that spotlight and object-based attention theories are not contradictory but rather complementary. Nevertheless, in many cases the object-based theory explains many phenomena better than the spotlight model does.

The object-based model is more complicated for implementation, since it requires objects' recognition, while the spotlight model only requires identifying the regions of interest, where the attentional resources will be concentrated for further processing.

2.2 System Architecture

The proposed sensor has two modes of operation: target acquisition and target tracking. In the acquisition mode *N* most salient targets of interest in the FOV are found. Then, *N* windows of interest with programmable size around the targets are defined. These windows define the active regions, where the subsequent processing will occur, similar to the flexible spotlight size in the biological systems. In the tracking mode, the system sequentially attends only to the previously chosen regions, while completely inhibiting the dataflow from the other regions.

The proposed concept permits choosing the attended regions in the desired order, independent on the targets saliency. In addition it allows shifting the attention from one active region to the other, independent of the distance between the targets. The proposed sensor aims to output the coordinates of all tracking targets in real time. Similar to biological systems, which are limited in their computational resources, the engineering applications are constrained with low-power dissipation. Thus, maximum efforts have been done to reduce power consumption in the proposed sensor. This power reduction is based on the general idea of "no movement – no action", meaning that minimum power should be dissipated if no change in the targets position occurred.

Figure 2 shows the architecture of the proposed tracking sensor. The sensor includes (a) A Pixel array with a two dimensional resistive network , (b) Y-Addressing and X-Addressing circuitry consisting of N digital registers for target windows definition and for image readout control, (c) an analog front end (AFE) for image readout, (d) A current Looser-take-all (LTA) circuit for target detection during the acquisition mode, (e) two analog 1-D center of mass (COM) computation circuits for X and Y COM coordinates update (each consists of 1-D analog current mode Winner-take-all (WTA) circuit), (f) Analog memory for temporal storage of loser values of all rows during the acquisition mode and digital memory for targets coordinates storage, (g) acquisition mode control and target mode control blocks.

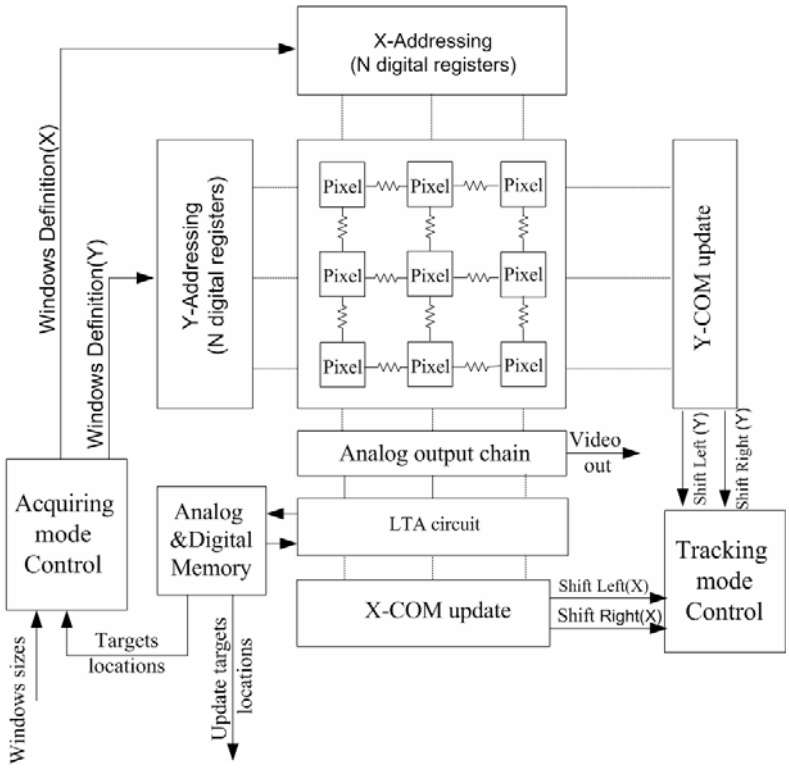

Figure 2. Architecture of the proposed tracking sensor.

In the acquisition mode the sensor finds the *N* most salient targets in the FOV and calculates their centroid coordinates. This is achieved in the following way: all neighboring pixels are connected by resistors (implemented by transistors), creating the resistive network. The pixel voltage becomes smaller if it is more exposed, and the local minimum of the voltage distribution can be regarded as the centroid of the target, corresponding to the exposed area. This minimum is found using an analog looser-take-all circuit (LTA). At the first stage of the acquisition mode, all pixels of the whole image are activated. The global minimum, corresponding to the brightest target is located using a one dimensional LTA circuit. To achieve this purpose, the whole image is scanned row by row (using one of the digital shift registers in the Y-addressing circuitry), finding the local minimum in each row. Then, the row local minima are input to the same LTA circuit again and the global minimum is computed. A more detailed description of this concept can be found in [27], where the two dimensional WTA computation was performed using two 1-D WTA circuits. Once the first brightest target is found, the system defines a small size programmable window, with the center located at the target centroid coordinates. The size of this window is predefined by the user before the acquisition mode starts and depends on the target size. While finding the second bright target in the FOV, all pixels of the first window, consisting of the brightest target found during the first search, are deactivated. This way, the bright pixels of the first target do not influence the result of the second search. The remains *N-1* targets are found in the same way. As a result, at the end of the acquisition mode all centroid coordinates of the *N* most salient targets in the FOV are stored in the memory and *N* small windows around these coordinates are defined. The window definition is performed using two digital shift registers. Thus, 2N shift registers are required to define N different windows. The acquisition mode control block is responsible for defining and positioning these active windows. Note, that the acquisition mode is very inefficient in terms of power dissipation because the whole sensor array is activated and the LTA operation and windows definition are power inefficient operations. On the other hand, the acquisition is a very rare operation and its time can be neglected in comparison with the tracking period.

Once the sensor has acquired *N* salient targets, the tracking mode is initiated. The predefined windows serve as a spotlight in biological systems, such that only the regions inside the windows are processed. Opposite to biological systems, these "spotlights" attend only to the regions predefined in the acquisition mode. Thus, even if new more salient objects appear during the tracking, the attention to the chosen regions is not influenced.

Because the sensor is in the tracking mode most of the time, it is very important to achieve very low-power dissipation in this mode. In the proposed system this is achieved in the following ways:

1. Only pixels of active windows and the circuitry responsible for proper centroid detection and pixels readout are active. The remaining circuits (including most pixels of the array) are disconnected from the power supply.
2. All shift registers in Y-addressing and X-addressing circuitries are optimized for low frequencies operation by leakage currents reduction.
3. During the tracking mode the sensor doesn't calculate new centroid coordinates. A simple analog circuit (COM update block in Figure 2) checks if the new centroid location differs from the centroid location of the previous frame. In the case that no difference was found, the circuit does not need to perform any action, significantly reducing system power dissipation. This principle suits the general idea of "no movement – no action". If the target changes its position, the "shift left" or "shift right" (both for x and y) signals are produced by the COM update blocks. These signals are input to the tracking mode control block and the appropriate shift register performs movement to the right direction, correcting the location of the window.
4. Each active window definition is performed using two shift registers. This windows definition method allows switching from one target of interest to another without any need in accessing the memory and loading the new target coordinates. This way the switching time between different objects does not depend on the distance between the targets and sensor power dissipation is reduced.

3. Circuits Description

In this Section we present some of the most important circuits, utilized by the sensor. This includes current mode LTA circuit, current mode WTA circuit, X-COM and Y-COM update circuits and ultra low-power shift registers. The pixel is implemented as a standard global-shutter active pixel sensor (APS) [28] with current mode readout instead of a conventional source follower amplifier inside the pixel. This current mode readout allows parallel read out and summarization of currents from all pixels in the active window at the same time. These summarized currents then are used for further processing by the X-COM and Y-COM update circuits, as described below. In addition, a conventional video data readout is available.

3.1 Current Mode Loser-Take-All (LTA) circuit

As previously mentioned, LTA circuit is responsible for targets detection and their COM computation during the acquisition mode. Since the COM computation is done in a serial manner and should be performed accurately, high speed and high precision LTA circuit is required. In addition, current mode LTA is required (pixels outputs are currents). Most of the previously presented LTA solutions are not suitable for the proposed sensor design. As a result, we use a LTA circuit that we have recently developed to achieve high speed and high precision [29]. Figure 3 shows cells 1 and k (out of the N interacting cells) of the LTA circuit. Cell k receives a unidirectional input current, I_{ink}, and produces an output voltage V_{outk}. This output has a low digital value if the input current I_k is identified as loser, and high, otherwise. The circuit applies two feedbacks to enhance the speed and precision: the excitatory feedback ΔI_k and inhibitory feedback ΔI_{avgk}. The basic operation of the LTA is based on input currents average computation and comparison of the input current of each cell to that average. The local excitatory feedback works to increase the compared average value in each cell, allowing that cell to be a loser. Oppositely, the inhibitory feedback works globally by reduction of the input currents average value and thus allowing inhibition of non-losing cells. A more detailed description of the circuit operation is provided below. The LTA circuit operates as follows: the drains of M_1 transistors of all N cells of the array are connected to the drains of M_4 transistors by a single common wire with voltage Vcom. The circuit starts the competition by applying Rst='1' for a short period of time. This way the excitatory feedback ΔIin and the inhibitory feedback $\Delta Iavg_k$ are cancelled. Assuming that all N cells in the array are identical and Rst='1' is applied, $\Delta Iavg_k$ = 0A and the current I'avg, through $M1_k$, is equal to the average of all input currents of the array, neglecting small deviations in the referenced input currents. I'avg is copied to $M3_k$ by the NMOS current mirror ($M1_k$ and $M3_k$) and is compared with the input current Iin_k copied by the PMOS current mirror ($M2_k$ and $M5_k$). If Iin_k=I'avg then $Vout_k$=VDD/2, assuming the same drivability factor K of $M3_k$ and $M5_k$ transistors.

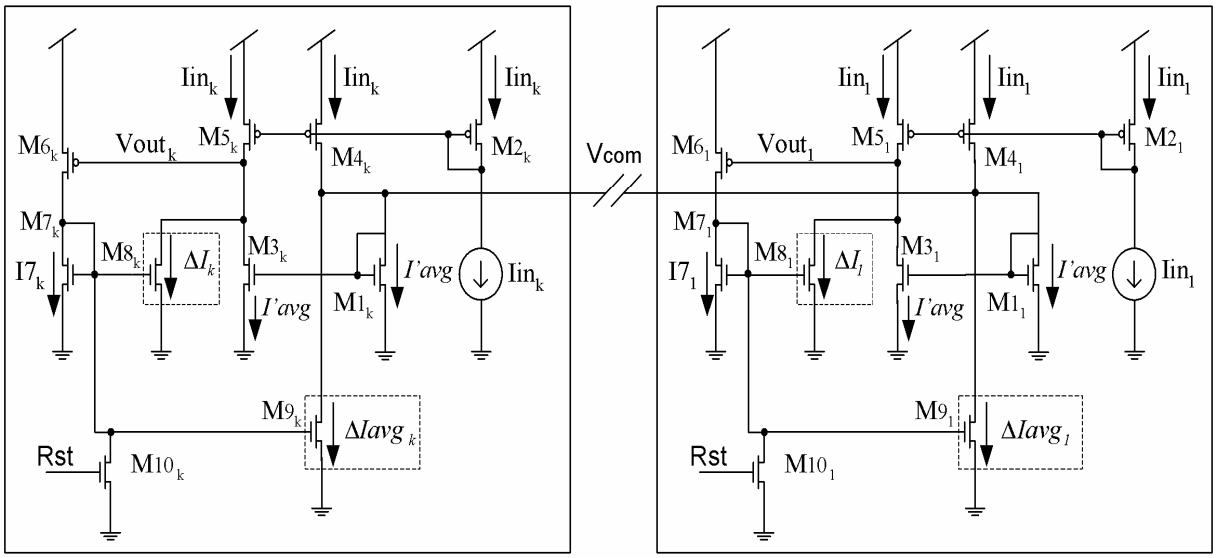

Figure 3. Cells 1 and k (out of N) of the LTA circuit.

An increase in input current Iin_k relatively to I'avg causes an increase in $Vout_k$ due to the Early effect. This way, during the reset phase, input currents of all cells are compared to the average of all input currents of the array, producing a unique output $Vout_k$ for every cell. The cell having the smallest input current value produces the smallest $Vout_k$ voltage. With the completion of the reset phase, i.e. Rst='0', the excitatory feedback ΔI_k and the inhibitory feedback $\Delta Iavg_k$ are produced. The $Vout_k$ node inputs to the gate of $M6_k$ PMOS transistor, thus the cell with the smaller Vx_k (smaller input current) produces a higher current $I7_k$ through $M6_k$ and $M7_k$. This current is copied by the NMOS current mirror ($M7_k$ and $M8_k$), creating the excitatory feedback ΔI_k. On the other hand, $I7_k$ is copied by the NMOS current mirror ($M7_k$ and $M9_k$), resulting in inhibitory feedback $\Delta Iavg_k$. ΔIk is added to the I'avg flowing through $M3_k$ and $\Delta Iavg_k$ is subtracted from the average of all input current by connection $M9_k$ transistor to the COM node, decreasing the I'avg value. This way, every cell produces a new Voutk voltage value, according to the comparison between the input current Iin_k and a sum of a current produced by the excitatory feedback ΔI_k and a new value of current I'avg, that is now given by:

$$I'avg = \frac{\sum_{k=1}^{N} Iin_k - \sum_{k=1}^{N} \Delta Iavg_k}{N} = Iavg - \frac{\sum_{k=1}^{N} \Delta Iavg_k}{N} \quad (1)$$

where Iavg is the average of all input currents of the array and N is the number of array cells. For the cell, having the smallest input current, the difference between Iin_k and a sum of ΔI_k and I'avg grows, thus decreasing $Vout_k$ value.

The computation phase is finished after one cell only is identified as a loser, producing $Vout_k$='0'. All other cells are identified as winners with $Vout_k$='1'. In this steady-state the excitatory and inhibitory feedbacks of the all winner cells and I'avg are approximately equal to zero, while $\Delta Iavg_k$ of the loser cell is approximately equal to the sum of all input currents. This way the circuit states stable preventing the selection of other potential losers unless the next reset is applied and a new computation starts. A more detailed description on the circuit operation can be found in [29].

To examine the presented LTA circuit it was designed, simulated and fabricated in 0.35um, 3.3V, n-well, 4-metal, CMOS, TSMC technology process supported by MOSIS. Table 1 summarizes the main characteristics of the circuit. As can be seen, the circuit achieves both high precision and high speed.

Parameter	Typical value	Worst case value (if exists)
Range of input currents	4 – 25 [μA]	------
Voltage supply	1.8V	------
Power Dissipation	58uW per cell	75 μW per cell
Delay	5nsec	95nsec
Precision	0.1 μA	0.5 μA
Occupied area (per cell)	26μm*22μm	------

Table 1: The main characteristics of the designed circuit.

3.2 X-COM and Y-COM update circuits

As mentioned, during the tracking mode the sensor doesn't calculate the new centroid coordinates. Instead, very simple X-COM and Y-COM update circuits (see Figure 4) check if the new X or Y centroid locations (respectively) differ from the centroid locations of the previous frame. In the case that no difference was found, the sensor does not perform any action, significantly reducing system power dissipation.

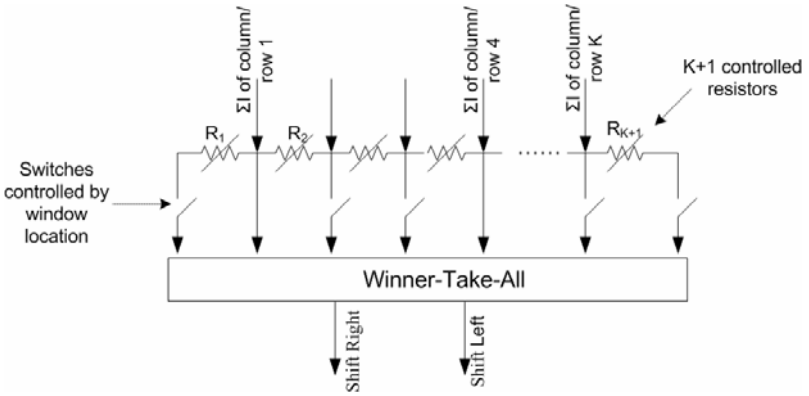

Figure 4. The X-COM and Y-COM update circuit implementation

The X-COM and Y-COM circuits have the same implementation, consisting of (K+1) controlled resistors (implemented by transistors), (K+2) digital switches, controlled by window location and (K+2) size analog current mode WTA circuit for the array, having K columns/row, respectively. Each COM update circuit receives K unidirectional input currents from the sensor array. The input current I_{ink} to the X-COM update circuit is the sum of all output pixel currents of the column K, while from the input current I_{ink} to the Y-COM update circuit is the sum of all output pixel currents of the row K. During the tracking mode (the COM update circuits are activated only in this mode), only pixels inside the windows of interest are activated. Therefore, in this mode, the input current to the COM circuit I_{ink} represents the sum of all row/column K active window output currents, for the case where the row/column K falls into the window of interest. In case, where the row/column K falls out the window of interest, the value of I_{ink} is zero. All input currents input to the resistive network, consisting of (K+1) controlled resistors (can either have a normal resistance R or very high resistance R_{high}) and then routed to the WTA circuit by switches. Both switches and the resistors are controlled by the windows of interest locations. For the active window, spread between column *i* and column (*i*+c), the resistors 1 to (*i*-1) and (*i*+2) to (K+1) have very high resistance R_{high}, while the resistors *i* to (*i*+1) are set to have a normal resistance R. This way resistive network, consisting of R value resistors is created around the active window of interest. Only two switches (*i*-1) and (*i*+1) are set to be on, while the others are set to be off. As a result, the WTA circuit receive only two non-zero input

currents, representing the X/Y COM coordinates of the target located inside the window of interest. In case these currents are equal (the WTA circuit does not succeed to find the winner), the COM coordinates are exactly in the center of the window and therefore window location update is not required. On the other hand, if (i-1) current is larger than (i+1), it means that the target has moved right relatively to its location in the previous frame and Shift Right output is activated. In case if (i-1) current is smaller than (i+1), the Shift Left output is activated.

3.3 Current Mode Winner-take-all (WTA) circuit

As mentioned, the X-COM and Y-COM update circuits utilize current mode WTA circuit. The implementation of the WTA circuit (see Figure 5) is very similar to the implementation of the LTA circuit, presented in sub-section 3.1 and detailed description of its operation can be found in [30]. Similarly to the LTA, the WTA circuit employs inhibitory and local excitatory feedbacks based on input currents average computation, enhancing precision and speed performance of the circuit. Local excitatory feedback provides a hysteretic mechanism that prevents the selection of other potential winners unless they are stronger than the selected one by a set hysteretic current. The WTA circuit can be useful for integration with circuits operating in the strong inversion region and supplying input currents of 3μA-50μA, as well as for subthreshold applications with inputs of 0nA-50nA. It achieves very high speed (32nsec for high currents of 3μA-50μA (measured) and 34nsec for subthreshold currents – (simulated)) in case when a very small difference between two input currents is applied (30nA for high currents and 1.8nA for subthreshold applications). These circuit performances are the direct result from very strong feedbacks applied in the circuit. The circuit ability to cope with wide range of input currents is very important for the COM update circuit implementation since the inputs to this circuit can have very wide range.

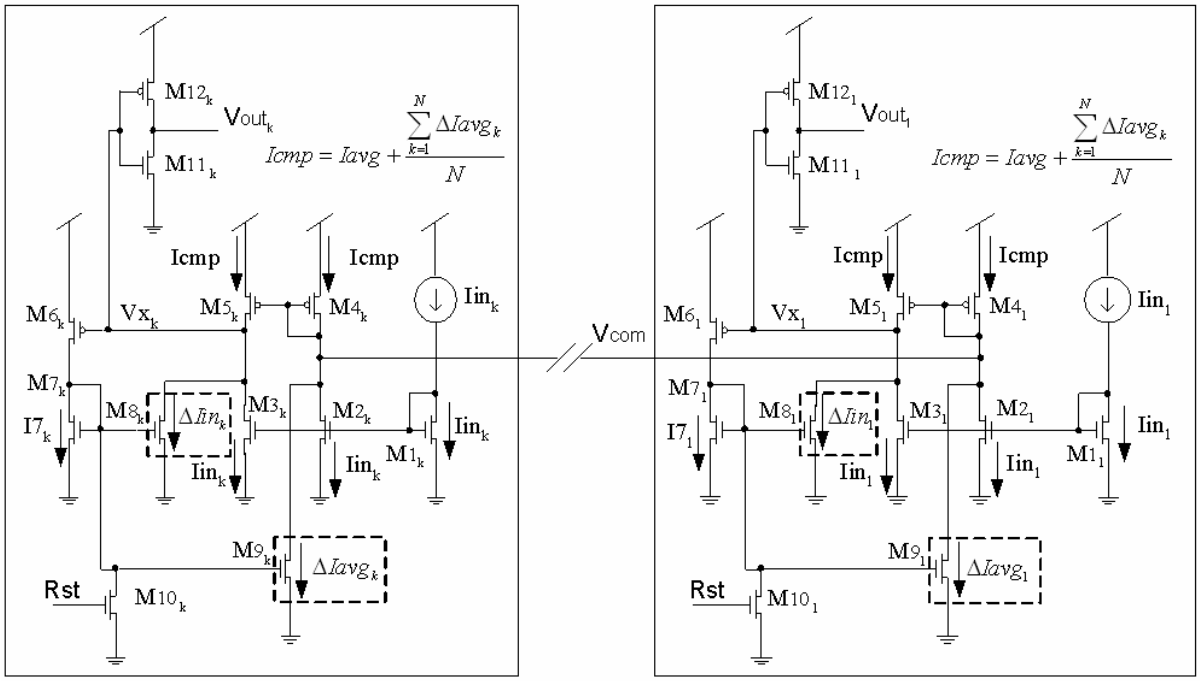

Figure 5. Cells 1 and k (out of n) of the WTA circuit

3.4 Y-addressing and X-Addressing circuitry implementation

Generally, the Y-Addressing and X-Addressing circuitry can be implemented using digital decoders. However, using shift registers for read out control reduces power dissipation and the number of global buses. In addition, windows of interest can be easily defined using shift-register. As previously mentioned, 2N shift registers are required to define N different windows. The architecture of shift register, used in the discussed tracking sensor (see Figure 6) , is based on the conventional simple shift register structure and utilizes low-power D-Flip-Flops (DFFs), described in details in [31].

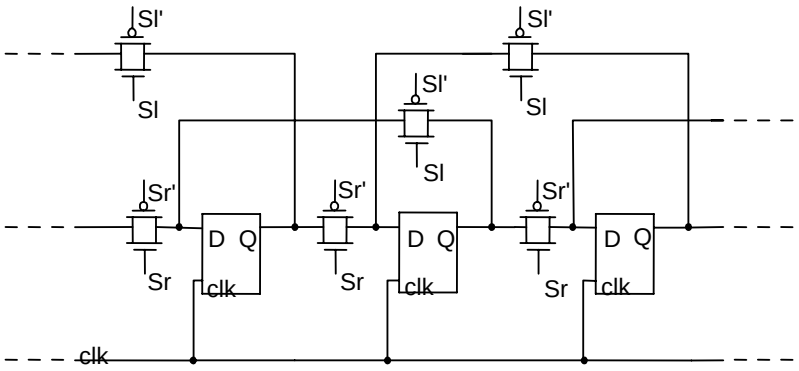

Figure 6. Architecture of the conventional shift-right and shift-left register

This register allows shifting of the vector of bits right or left – very important function in windows definition. The power dissipation of the shift register can be reduced by examining the nature of the inputs to the register. When the register is used for signal readout control, its input vector consists of a single '1' and of (N-1) digital zeros ('0'), assuming an N size register. Thus, in steady state, only one (out of N) DFF has '1' in its output. For the case, when the register is used for window definition, its input vector consists of K high digital bits and (N-K) low digital bits, assuming an N size register defines K*K size window. Usually, K<<N, resulting in the same solution for both cases. Figure 7 shows the DFF, optimized for these kinds of input vectors. This master-slave FF is constructed by cascading two different transmission gate latches: the first one is the dynamic latch and the second one is pseudo-static latch. Having only 15 transistors, this circuit is optimized for leakage current reduction for the case of '0' at the FF output. For this case all possible leakage currents in the circuit (signed by arrows in Figure 7) are reduced due to connection of two series connected "off" transistors. Note, that for the case of the shift register, having an input vector with almost all "high" bits, the presented FF can be optimized in a similar way, taking in account the high digital value in the FF output.

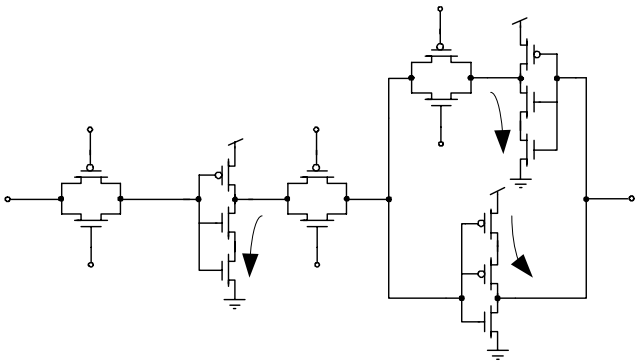

Figure 7. DFF, optimized for an input vector consisting of a large number of zeros

To check the suitability of the mentioned DFF circuit for the proposed tracking system, it has been designed and implemented in 0.18μm technology to compare the FF design with a set of representative flip-flops, commonly used for high performance design. In addition, the low leakage shift register is compared to the register, based on conventional FFs. Simulation results show up-to 10% power reduction in case of 10KHZ operation and up-to 60% reduction in case of 5 pixels size window definition at 30Hz frequency. More detailed description of the FF and SR performances can be found in [31].

4. Discussion

In this section we present the expected performance of the proposed tracking imager and briefly discuss advantages and limitations of the current architecture. Table 2 summarizes the expected characteristics of the proposed system. The sensor will be fabricated in a standard 0.18μm CMOS technology and will be operated

using 1.8V supply voltage. The pixel size is expected to be 7 x 7μm and to achieve fill factor of at least 60%. At this stage the test chip will include a relatively small array of 64 x 64. The reason for the small array is failure probability reduction and limited budget. On the other hand, this array size still allows showing the proof of concept. At the first fabrication phase the system will able to track up to 3 salient targets of interest at 30 frames per second. In future designs we are working to increase the number of the tracked targets and to improve real time operation, allowing tracking at up to 100 frames per second. As mentioned, the proposed imager employs spatial filtering version of the spotlight models of attention, where what falls outside the attentional spotlight is assumed not to be processed. The drawback of this method is that during the tracking mode the sensor filters all information outside windows of interest, including potential targets that appear in the FOV during the tracking. In our future implementations we plan to upgrade the tracking sensor with the spotlight attention model, where the spotlight serves to concentrate attentional resources to a particular region in space, thus enhancing processing at that location and almost eliminating processing of the unattended regions (but still checking theses regions). An additional limitation is that the proposed system does not utilize Correlated Double Sampling (CDS) circuit to reduce Fixed Pattern Noise (FPN). CDS implementation in such kind of tracker is not trivial and thus it will be implemented only in the next version of the system to reduce failure probability at this stage. Finally, the system is expected to achieve very low-power dissipation of less than 2mW. The next generation of this tracking system will achieve power dissipation of less than 1mW.

Parameter	Expected value
Technology	0.18μm standard CMOS technology
Array size	64 x 64
Voltage supply	1.8V
Pixel Size	7 x 7 μm
Fill Factor	> 60%
No. of tracked targets	3
Real time operation	60 frames/second
Sensor read out method	Global Shutter
Utilized Attention model	Spatial Filtering
Bad Pixel Elimination	Yes
FPN Reduction	No
Power Dissipation	< 2mW

Table 2: The expected characteristics of the tracking system.

5. Conclusions

Implementation of low-power tracking CMOS image sensor based on biological models of attention was presented. Imager architecture and principle of operation were discussed, as well designs of the most important circuits, like Winner-Take-All, Looser-Takes-All, COM update and X/Y-Addressing circuits, utilized by the tracking system were shown. A brief description of the spatial and object-based models of attention was also presented. The expected system performance was discussed, showing advantages and drawbacks of the proposed sensor. The presented imager allows tracking of up to N salient targets in the field of view. The imager architecture is optimized to achieve low-power dissipation both in acquisition and tracking modes of operation. Further research includes improvement of the current sensor architecture and its realization in an advanced CMOS technology.

Acknowledgements

We would like to thank the Israeli Ministry of Science and Technology for funding this project.

Bibliography

[1] Chun, M. M., & Wolfe, J. M. Visual Attention. In E. B. Goldstein (Ed.), Blackwell's Handbook of Perception, Vol. Ch 9, pp. 272-310). Oxford, UK: Blackwell, 2001.

[2] R. W. Remington, L. Pierce. Moving attention: Evidence for Time-invariant shifts of visual selection attention. Perception and Psychophysics, vol. 35, pp. 393-399, 1984.

[3] J. Moran, R. Desimone. Selective attention gates visual processing in the Extrastriate Cortex. Science, vol. 229, pp. 784-787, 1985.

[4] G. Sperling, E. Weichselgartner. Episodic theory of the dynamics of spatial attention. Psychological review, vol. 102, no. 3, pp. 503-532, 1995.

[5] A. Treisman. Features and targets: The fourteenth Barlett memorial lecture. Quarterly Journal of Experimental Psychology, 40A, pp. 201-237, 1988.

[6] A. Treisman. Feature binding, attention and target perception. Philosophical Transactions of the Royal Society of London, vol. 353, pp. 1295-1306, 1998.

[7] A. Treisman, G. Gelade. A feature-integration theory of attention. Cognitive Psychology, vol. 12, pp. 97-136, 1980.

[8] A. Treisman, D. Kahneman, J. Burkell. Perceptual targets and the cost of filtering. Perception & Psychophysics, vol. 33, pp. 527-532, 1983.

[9] S. Yantis. Targets, attention, and perceptual experience in “Visual Attention". R. D. Wright (Eds.), pp. 187-214. Oxford, NY: Oxford University Press, 1998.

[10] M. I. Posner. Orienting of attention. Quarterly Journal of Experimental Psychology, vol. 32,pp. 3-25, 1980.

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

[12] C. Koch and S. Ullman, “Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry,” Human Neurobiology, vol. 4, pp. 219–227, 1985.

[13] T. G. Morris, T. K. Horiuchi, and P. DeWeerth, “Target-Based Selection Within an Analog Visual Attention System”, IEEE Transactions on circuits and systems-11, Analog and Digital Signal processing, vol.45, no.12, December 1998.

[14] T. G. Moris and S. P. DeWeerth, “A Smart-Scanning Analog VLSI Visual-Attention System”, Analog Integrated Circuits and Signal Processing, vol.21, p.67-78, 1999.

[15] T. G. Morris and S. P. DeWeerth, “Analog VLSI excitatory feedback circuits for attentional shifts and tracking.” Analog Integrated Circuits and Signal Processing, 13(1/2), pp. 79–92, May/June 1997.

[16] C. S. Wilson, T. G. Morris, and P. DeWeerth, “A two-dimensional, target-based analog VLSI visual attention system", Twentieth Anniversary Conference on Advanced Research in VLSI, IEEE Computer Society Press: Los Alamitos, CA. Vol. 20. pp. 291-308. March 1999.

[17] T. Horiuchi and E. Niebur, “Conjunction search using a 1-D analog VLSI based attentional search/tracking chip,” In Wills, S. (Ed.), Proceedings of the 1999 Conference in Advanced Research in VLSI., pp. 276-90 Georgia Tech, Atlanta, 1999.

[18] V. Brajovic and T. Kanade, “Computational sensor for visual tracking with attention”, IEEE journal of Solid-State Circuits, Vol.33, No.8, August 1998.

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

[20] 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.

[21] G. Indiveri, "A 2D neuromorphic VLSI architecture for modeling selective attention", Proc. IJCNN2000, 2000.

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

[23] 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.

[24] J. Lazzaro., S. Ryckebusch, M. A. Mahowald, and C.A. Mead, “Winner-tale-all networks of O(n) complexity”, ed. D.S.Touretzky, Morgan Kaumann, San Mateo, CA, 1:703-711, 1989.

[25] A. Fish and O. Yadid-Pecht, "Bottleneck Problem Solution using Biological Models of Attention in high resolution tracking sensors", to be appear International Journal on Information Theories and Applications.

[26] A. Fish, A. Spivakovsky, A. Golberg and O. Yadid-Pecht, “VLSI Sensor for multiple targets detection and tracking”, IEEE ICECS, December 2004, CD ROM.

[27] A. Fish, D. Turchin, O. Yadid-Pecht, " An APS with 2-Dimensional winner-take-all selection employing adaptive spatial filtering and false alarm reduction", IEEE Trans. on Electron Devices, Special Issue on Image Sensors, Vol. 50, No. 1, pp. 159-165, January, 2003.

[28] O. Yadid-Pecht, R. Etienne-Cummings. CMOS imagers: from phototransduction to image processing. Kluwer Academic Publishers, 2004.

[29] A. Fish, V. Milrud and O. Yadid-Pecht, "High speed and high resolution current Loser-take-all circuit of o(N) complexity", IEEE ICECS, December 2004, CD ROM.

[30] A. Fish, V. Milirud and O. Yadid-Pecht, "High speed and high resolution current winner-take-all circuit in conjunction with adaptive thresholding", IEEE Transactions on Circuits and Systems II, vol. 52, no. 3, pp. 131-135, March 2005.

[31] A. Fish, V. Mosheyev, V. Linkovsky and O. Yadid-Pecht, "Ultra Low-Power DFF based Shift Registers Design for CMOS Image Sensors Applications", IEEE ICECS, Tel-Aviv, Israel, December 2004, CD ROM.

Authors' Information

Alexander Fish – e-mail: afish@ee.bgu.ac.il

Liby Sudakov-Boreysha –e-mail: libys@bgu.ac.il

The VLSI Systems Center, Ben-Gurion University, Beer Sheva, Israel;

Orly Yadid-Pecht – The VLSI Systems Center, Ben-Gurion University, Beer Sheva, Israel, and Dept. of Electrical and Computer Engineering, University of Calgary, Alberta, Canada; e-mail: oyp@ee.bgu.ac.il

IMAGE SENSORS IN SECURITY AND MEDICAL APPLICATIONS

Evgeny Artyomov, Alexander Fish, Orly Yadid-Pecht

Abstract: This paper briefly reviews CMOS image sensor technology and its utilization in security and medical applications. The role and future trends of image sensors in each of the applications are discussed. To provide the reader deeper understanding of the technology aspects the paper concentrates on the selected applications such as surveillance, biometrics, capsule endoscopy and artificial retina. The reasons for concentrating on these applications are due to their importance in our daily life and because they present leading-edge applications for imaging systems research and development. In addition, review of image sensors implementation in these applications allows the reader to investigate image sensor technology from the technical and from other views as well.

Keywords: CMOS image sensors, low-power, security applications, medical applications.

ACM Classification Keywords: B.8.1.i Hardware - Integrated Circuits - Types and Design Styles - VLSI

1. Introduction

Fast development of low-power miniature CMOS image sensors triggers their penetration to various fields of our daily life. Today we are commonly used to meet them in digital still and video cameras, cellular phones, web and security cameras, toys, vehicles, factory inspection systems, medical equipment and many other applications (see Figure 1). The advantages of current state-of-the-art CMOS imagers over conventional CCD sensors are the possibility in integration of all functions required for timing, exposure control, color processing, image enhancement, image compression and analog-to-digital (ADC) conversion on the same chip. In addition, CMOS imagers offer significant advantages in terms of low power, low voltage, flexibility, cost and miniaturization. These features make them very suitable especially for security and medical applications. This paper presents a review of image sensors utilization in part of the security and the medical applications.