



1998 Special Issue

Temporal photoreception for adaptive dynamic range image sensing and encoding

Vladimir M. Brajovic*, Ryohei Miyagawa[†], Takeo Kanade*The Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*

Received and accepted 5 May 1998

Abstract

We have implemented two analog VLSI computational sensors for sensing and encoding high dynamic range images by exploiting temporal dimension of photoreception. The first sensor is a multi-integration time photoreceptor that automatically adapts to use different integration periods depending on light intensity. It exhibits a dynamic range 128 times larger than that of a single integration period photoreceptor, approximately 1:128 000. The second sensor is an intensity-to-time processing paradigm that is based on the notion that stronger stimuli elicit responses before weaker ones. The paradigm sorts pixels of sensed images by their intensities, thus achieving information-theoretic optimal encoding of images. It handles dynamic range of approximately 1:1 000 000. Both implementations can operate at standard video rate of 30 frames s⁻¹. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords: Adaptive dynamic range imaging; VLSI sensors; Image sensors; Computational sensors; Smart sensors; Robot perception; Vision

1. Introduction

When a photo detector (e.g. a photodiode or a CCD well) operates in a flux-integration mode (Weckler, 1967) it integrates photoelectrons. In the conventional image sensors, the integration process continues for a fixed integration period, followed by the measurement of accumulated charge in each pixel. If the photon flux at some pixels exceeds the limits of the sensor, the photo-charge saturates those pixels. Shortening the integration period is the usual remedy for this problem, but at the expense of not collecting sufficient photo-charge at the dark pixels. This is a well-known limited dynamic range problem of conventional image sensors, a problem that right at the acquisition level limits the entire vision system. The computer vision community reluctantly accepts this limitation and copes with the corrupt sensory data at the algorithmic level, inevitably leading to unreliable performance.

We exploit the temporal dimension of photo sensing to achieve imaging of high dynamic range scenes. Time plays a critical role in neural information processing. The neural responses are discrete in amplitude but analog in time;

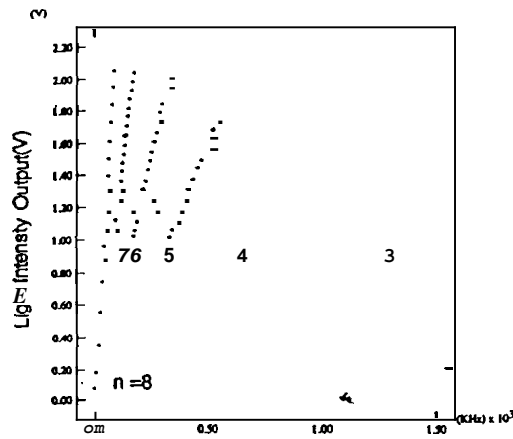
therefore, the time plays an essential role in the flow and transformation of information in biological systems (Anton et al., 1992). In addition to simply dealing with the amount of charge collected during a fixed integration interval, we deal with time intervals the detector takes to accumulate a pre-determined amount of photo-charge. Using analog VLSI, we have implemented and demonstrated two vision computational sensors (Mathur and Koch, 1991; Kanade and Bajcsy, 1993) capable of sensing and encoding large dynamic range scenes. These are good examples of how adaptive analog on-chip computation at the sensory level can extract unique, rich, and otherwise not obtainable sensory information.

The first method, a multi-integration time photoreceptor, avoids saturation by automatically choosing the "best" integration period from a set of predetermined periods. When the charge level is close to saturation, the integration stops at one of these integration periods. The receptor encodes the intensity by two signals: (1) the accumulated charge, and (2) the identifier of the selected interval. In this way, the sensor represents a wide range of light intensities.

The second method, a sorting image sensor, avoids saturation by sorting pixels by their intensities. The input image may have a large dynamic range, but the indices assigned to pixels always range from 1 to N , where N is number of pixels in the imager. The sorting in the sensor is based on the biologically inspired notion that stronger stimuli elicit responses before weaker ones. It is achieved

* Corresponding author. Tel.: (+1) 412 268 5622; fax: (+1) 412 268 5571; email: brajovic@cs.cmu.edu

[†] Present address: Toshiba Corporation, Research and Development Center, ULSI Research Laboratories, Kawasaki 210, Japan.



The photodiode B (Pd-b) integrates signal charge. The output of the latch samples and holds this charge on storage capacitor Cap1. The latch output thus controls the duration of photon integration. The latch output also samples and holds a ramp voltage in the capacitor Cap2 to memorize the identity of the period at which the Pd-b stopped integrating. The ramp voltage is incremented by a predetermined step at each pulse in New-IT; therefore, it indicates the number of the interval being used.

Fig. 2 shows an early part of a frame. In this example, the illumination was such that the second integration period is chosen. Following the reset to a high voltage, the two photodiodes integrate the signal charge. Pd-a decays, passes the first integration timing (IT1), and then reaches the threshold for the inverter at t_0 . The inverter trips from low to high. However, the integration gate (IG) does not change because the new state of the inverter output is still not visible at the latch output. Only after the integration timing pulse IT2 makes the latch transparent for a short time (e.g. $1 \mu s$) does the new state of the inverter affect the sample/hold gate IG. Then, the integration in Cap1 stops and the identity of the IT2 is recorded. The blanking gate **Blk** forces the integration gate (IG) to open at the end of the frame if the light intensity is too low to do so earlier.

2.1. Experimental evaluation

The photoreceptor circuit was fabricated by MOSIS with the Orbit 2m Analog process. The size of a photoreceptor is $109 \times 110 \mu m^2$. We tested the photoreceptor to investigate its behavior over a wide range of light intensities. In order to set the light intensity precisely, we varied the frequency of the LED light pulses by 50 ns each. The light intensity output L-out and the identity of the integration time output It-out are plotted against light intensity in Fig. 3 and Fig. 4. Li-out increases linearly with light intensity in the largest integration period ($n = 8$). It reaches a threshold and then switches to the second largest integration period

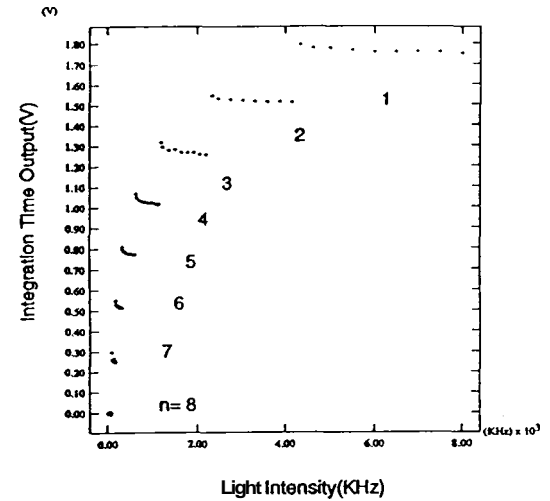


Fig. 4. Integration period identifier versus light intensity.

($n = 7$). With this transition, the Li-out decreases by half, but again increases linearly until the threshold is reached within the shorter period. Then, the receptor switches to the next integration period ($n = 6$) and so on. With each transition to the smaller integration period, It-out increases by one level; we observe in Fig. 4 that It-out has the eight distinctive output levels corresponding to the eight integration periods.

From the two signals, Li-out and It-out, we can reconstruct the actual light intensity by multiplying Li-out by $2^{(8-N)}$, where N is the value of It-out. The reproduced outputs are plotted against light intensity in Fig. 5. There are eight segments of this graph, corresponding to the different integration periods. The transition from one integration period to the next is smooth and the output as a whole shows very good linearity over a wide range. The smallest integration period is 1/128th of the largest integration time; therefore, the dynamic range is 128 times larger than that of a single integration period photoreceptor, for approximately 1:128000.

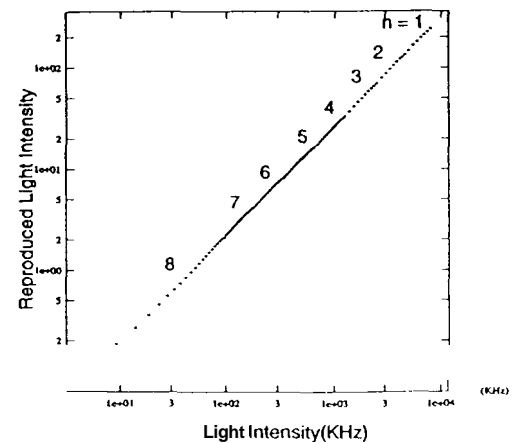


Fig. 5. Reproduced signal versus light intensity

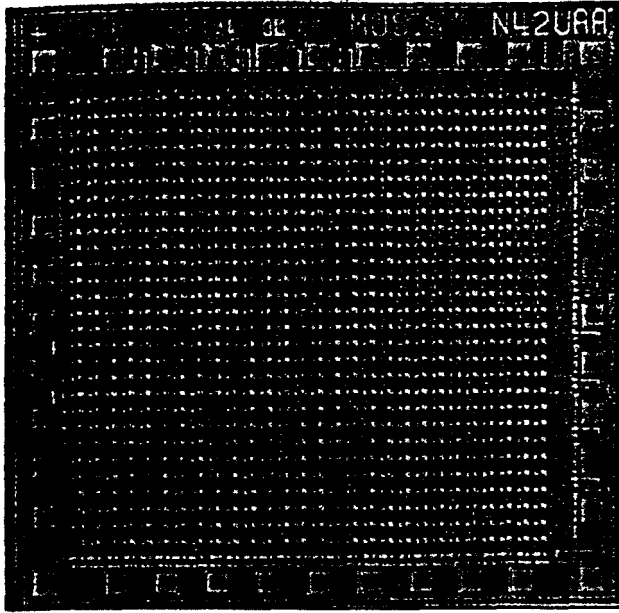


Fig. 8. Micrograph of the sorting chip

in each cell memorizes this voltage in a capacitor C , since it represents the index of the cell that is just firing its event. Once all the cells have responded, the stored indices are read out from the capacitors (a readout switch is not shown).

Fig. 7 shows the simulation of the current operation for the sorting sensor with four cells. **A** photodiode PD is charged to a high potential and left to float. Since the photocurrent discharges the diode capacitance, the voltage decreases approximately linearly to a rate proportional to the amount of light impinging on the diode (Fig. 7, top graph).

Once the diode voltage falls to the threshold of the inverter, the inverter's output changes state from low to high (Fig. 7, second graph). **A** switch **S3** is included to provide a positive feedback, force rapid latching action and produce sharp edge at the output. The edge in the inverter's output closes the switch **S6** and opens the switch **S7**.

The voltage on the wire W_{in} (Fig. 7, third graph) represents the index of a cell that is changing state. The capacitors within cells follow this voltage until they are disconnected. At each point, each capacitor C retains the index of the pixel (Fig. 7, bottom graph). The bottom graph shows that the cell with the highest intensity input receives the highest "index", the next cell one "index" lower, and so on.

The sorting sensor computes several important properties of the sensed image. First, the time when a cell k triggers is approximately inversely proportional to the input radiation it receives:

$$t_k = C \frac{(V_{DD} - V_{th})}{I_k + I_d} \quad (1)$$

where C is the diode capacitance, V_{DD} the power supply

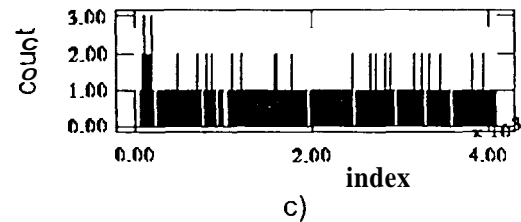
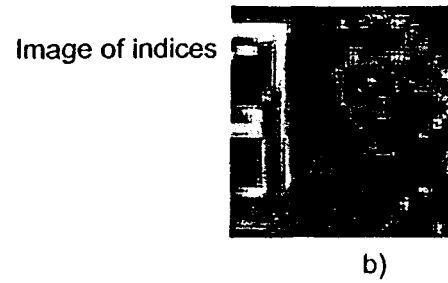
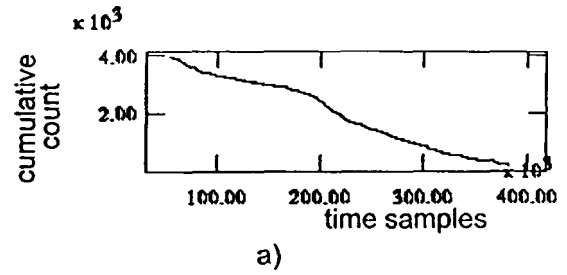


Fig. 9. An office scene imaged by the sorting sensor: (a) cumulative histogram computed by the chip (sampling period is $0.5 \mu s$); (b) image of indices; (c) histogram of indices.

voltage, V_{th} the threshold voltage of the inverter, and I_k photocurrent approximately proportional to the radiation, and I_d is the dark current.

Second, by counting pixels, the global processor knows at each given time how many cells have responded with an event. Since events are generated according to Eq. (1), the cumulative current in the wire W_{out} , or its inverse, the voltage on the wire W_{in} , are temporal representations of the cumulative histogram of the input data, with the horizontal axis being proportional to $\ln r$. The time derivative of the cumulative histogram signal is related to a histogram of the input image (Ballard and Brown, 1982). The cumulative histogram is a one scene global property that is reported by the chip with very low latency, and can be used for preliminary decision making as soon as the first responses are received (Brajovic and Kanade, 1996).

The sorting image sensor encodes high dynamic range images by producing the cumulative histogram waveform and the image of indices. During sorting, the cumulative histogram waveform maps input intensities into indices. Once collected, the cumulative histogram is used to decode the image of indices to the original high dynamic range images.

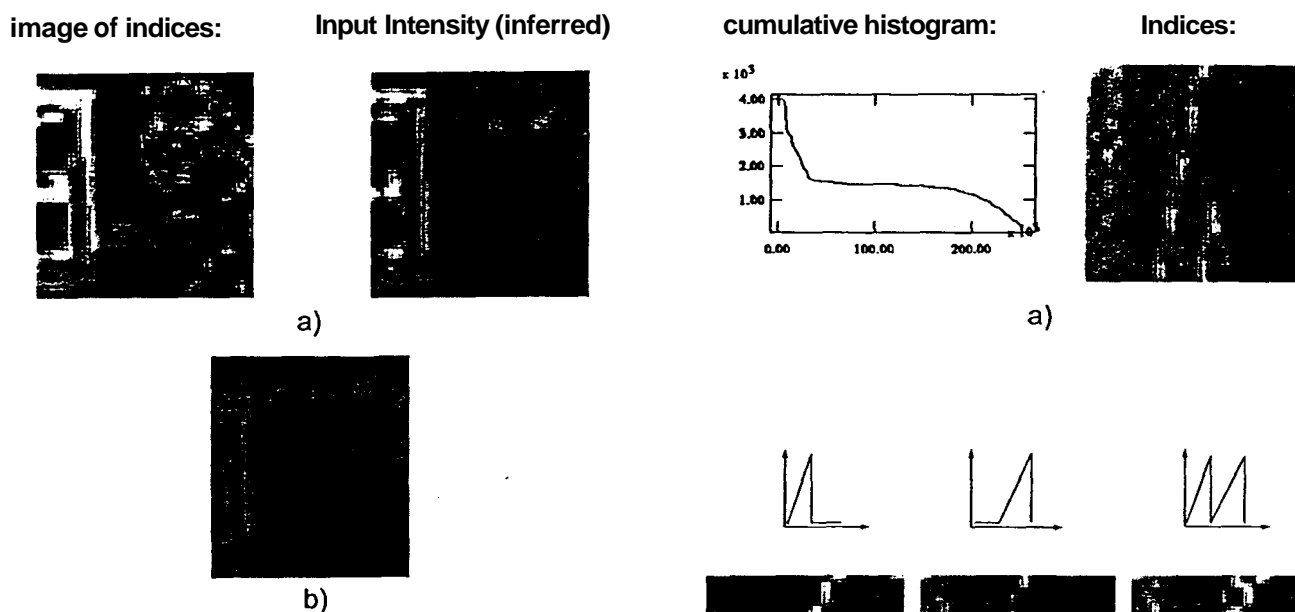


Fig. 10. Reconstructing linear images: (a) indices from the sorting sensor and inferred linear image; (b) CCD camera image.

3.1. Experimental evaluation

A 21×26 cell sorting image sensor has been built in $2 \mu\text{m}$ CMOS technology. The size of each cell is $76 \mu\text{m} \times 90 \mu\text{m}$, with 13% fill factor. The micrograph of the chip is shown in Fig. 8. The chip receives an optical image directly on its surface. The cumulative histogram waveform and the indices from the sorting sensor were digitized with 12 bit resolution.

In one experiment, the sorting sensor detected a low contrast scene from an office environment. Fig. 9(a) and Fig. 9(b) show the cumulative histogram of the scene and the image of indices both computed by the chip. Fig. 9(c) shows the histogram of the indices. Most pixels appeared to have



Fig. 11. A scene with back-illuminated objects as captured by a conventional CCD camera.

Fig. 12. Sorting sensor signal processing: (a) data from the sensors; (b) segmentation (viewing the shadowed region); (c) segmentation (viewing illuminated region); (d) segmentation and shadow removal.

different input intensities and, therefore, received different indices. Occasionally, as many as three pixels were assigned the same index. Overall, the histogram of indices is uniform, indicating that the sorting chip has performed correctly.

There are 546 pixels in the prototype described in this paper. The uniform histogram of indices (Fig. 9(c)) indicates that most of the pixels received different indices. Therefore, without special considerations as to the illumination conditions, low noise circuit design, temperature, and dark current control, our lab prototype readily provides indices with more than 9 bits of resolution. Furthermore, the range of indices remains unchanged (from 0 to 545) and the indices maintain maximum contrast and uniform histogram regardless of the range and distribution of input light intensity.

Fig. 10(a) shows the linear image reconstruction from the image of indices and the cumulative histogram. To illustrate the “natural” conditions in the scene, Fig. 10(b) includes an image taken by a commercial CCD camera. The inferred input intensities closely resemble the natural conditions in the environment. Note that since we were limited by the 8 bit dynamic range imposed by the postscript images, during the mapping, we have thrown away much of the contrast information available in the image of the indices.

In another experiment, the sorting sensor detected a high

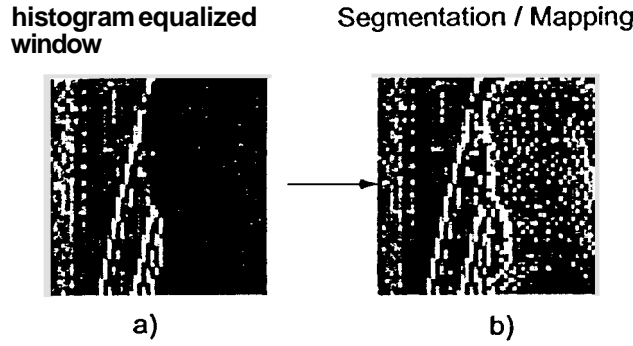


Fig. 13. Conventional CCD signal processing: (a) histogram equalization of the window; (b) segmentation and shadow removal.

dynamic range scene, a back illuminated object. Fig. 11 shows a global view of this scene as captured by a conventional CCD camera. Owing to the limited dynamic range, the CCD camera images the foreground poorly; it is mostly black. (The white box roughly marks the field-of-view for the sorting sensor.)

The sorting sensor, on the other hand, resolves detail in both dark and bright regions (Fig. 12(a)). Since all 546 indices are competing for display within the 256 levels allowed for the postscript images in this paper, one enhancement for the purpose of human viewing is to segment the image and amplify only dark pixels. Fig. 12(b) shows the result. Conversely, as shown in Fig. 12(c), the bright pixels can be spanned to the full (8 bit) output range. Finally, if these two mappings are performed simultaneously, the shadows are removed (Fig. 12(d)).

The same method can be applied to the image obtained from a standard CCD camera. If the CCD image of Fig. 13 is cropped to the white box, and such an image is histogram-equalized, we arrive at the result shown in Fig. 13(a). This image is analogous to the image of indices obtained by the sorting sensor (Fig. 12(a)). Owing to the limited dynamic range, noise and quantization, the CCD image only resolves the face with 2–3 bits. The histogram-equalized image from the CCD is mapped using the same steps as for Fig. 12(d). For obvious reasons, the result is poor in the dark region. In contrast, the sorting computational sensor optimally allocates as many levels (i.e. indices) to the dark region as there are pixels. By comparing Fig. 12(d) and Fig. 13(b), the superior utilization of the sensory signal with the sorting sensor is obvious.

Fig. 14 illustrates a rapid adaptation of the sorting sensor to accommodate changing dynamic range scenes. It shows a sequence of 93 images of indices computed by the sorting sensor. The sensor was stationary, and the only changes in the scene are due to subject movement. By observing the wall in the background, we can see the effects of adaptive dynamic range. Although the physical wall does not change the brightness, it appears dimmer in those frames in which bright levels are taken by pixels which are physically brighter (e.g. the subject's face and arm). When the subject

turns and fills the field-of-view with dark objects (e.g. hair) the wall appears brighter since it is now taking higher indices. Also, note that the maximum contrast is maintained in all the images since all images of indices have a uniform histogram.

3.2. Error analysis of temporal sorting and image encoding

Theoretically, the dynamic range of the scene detectable by the sorting sensors is unlimited. Of course, in practice the actual dynamic range of the sensor is determined by the capabilities of the photodetector, as well as by the switching speed and dark current levels.

First, we investigate the mismatch of the cells. Even when receiving equal light levels, the cells do not respond at the same time. This determines the fundamental accuracy of the imaging. Given Eq. (1), the input photocurrent can be found as

$$I_{\lambda} = \frac{Cv}{t} - I_d \quad (2)$$

where $V = V_{DD} - V_{th}$, and I_d is the dark current. The relative error can be found as

$$\left(\frac{\sigma_{I_{\lambda}}}{I}\right)^2 = \left(\frac{\sigma_{I_d}}{I}\right)^2 + \left(\frac{\sigma_C}{C}\right)^2 + \left(\frac{\sigma_v}{V}\right)^2 + \left(\frac{\sigma_t}{t}\right)^2 \quad (3)$$

where $I = I_{\lambda} + I_d$, σ_{I_d} represents the fluctuation of the dark current over the sensor area, σ_C represent the fluctuations of the photo-detector capacitance (e.g. mismatch of the photo-detectors), σ_v represents the mismatch of the threshold voltages and the diode's reset noise, and σ_t represents the fluctuation in the switching speed of the control element. After substituting Eq. (2) in the last term in Eq. (3), the relative error becomes

$$\left(\frac{\sigma_{I_{\lambda}}}{I}\right)^2 = \left(\frac{\sigma_{I_d}}{I}\right)^2 + \left(\frac{\sigma_C}{C}\right)^2 + \left(\frac{\sigma_v}{V}\right)^2 + \left(\frac{\sigma_v}{CVI}\right)^2 \quad (4)$$

$$\left(\frac{\sigma_{I_{\lambda}}}{I}\right)^2 = \frac{A^2}{I^2} + B^2 + C^2 I^2 \quad (5)$$

where A , B and C substitute constant terms in Eq. (4). This error model follows the intuition. For high levels of illumination, when the cells respond quickly, the dominant cause of error is the fluctuation in the switching speed. For low illumination levels, the dominant factor is the fluctuation in the dark current.

The constants A , B and C were experimentally determined from the prototype chip. Without the lens in front of the sensor, the sensor was illuminated by a halogen light source reflected from a white cardboard. As the cardboard was positioned several meters from the sensor, the illumination field was considered uniform over the surface of the sensor. Changing the angle between the light source and the cardboard controlled the amount of light falling on the surface of the sensor. The cumulative histogram waveform was gathered for 43 different light levels. From the cumulative histogram waveforms and Eq. (2), the mean value I , and

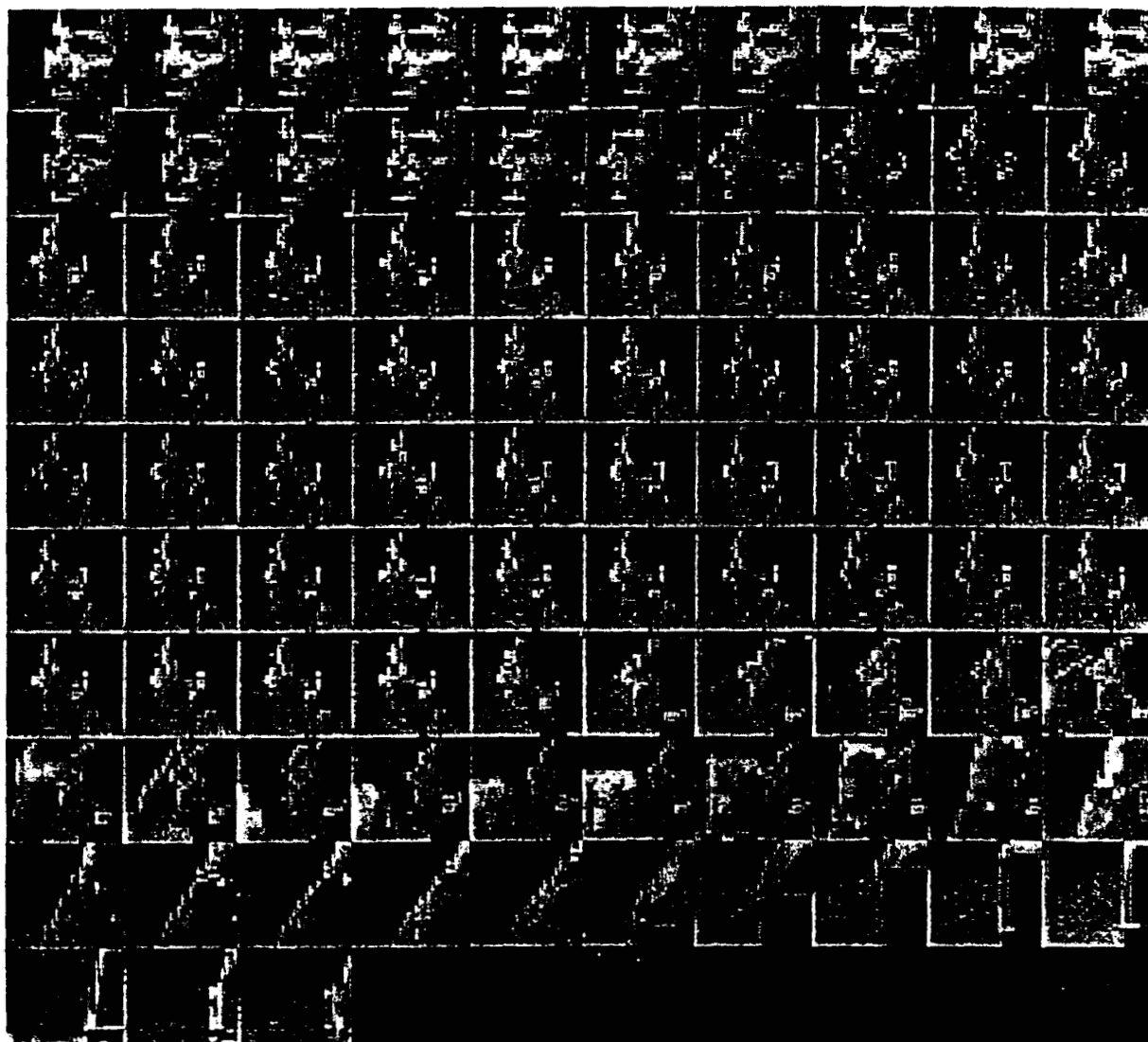


Fig. 14. Sequence of images of indices delivered by the sorting sensors.

standard deviation σ , were computed in arbitrary current units (ACU). (1 ACU = 1/s, i.e. 1 ACU triggers an event according to Eq. (1) at 1 s after the beginning of the frame integration.) The error model in Eq. (5) was fitted to the data. The results are tabulated in Table 1, and plotted in Fig. 15.

Table 1
Error performance of the sorting image sensor prototype

A	2.2860×10^{-2} ACU	
B	5.0636×10^{-3}	
C	4.2862×10^{-5} ACU ⁻¹	
$\sigma/I = 1$	I_{\min}	0.0229 ACU
	I_{\max}	23 330 ACU
	Dyn. range	1:1 020 565
$3\sigma/I = 1$	I_{\min}	0.0686 ACU
	I_{\max}	7776 ACU
	Dyn. range	1:113 373

For the signal-to-noise ratio of one, the dynamic range of the sensor based on the model is over 1:1 000 000. If the 3σ rule is used for the noise limits, the dynamic range is over 1:100 000. However, the level of the dark current determines the detectable lower limit on the input photocurrent. In the experiment, we determined that the average dark current is about 0.2 ACU; therefore, the SNR = 1 and we require the lowest input photocurrent to be 0.2 ACU. Then, the dynamic range is 1:116 650 for the 1σ rule and 1:38 880 for the 3σ rule. Given the dark current, the dynamic range is limited by the error the sensor makes when detecting the high illumination levels. The dominant source for this error is the fluctuation in the turn-on time of the inverters. In our experiment, this fluctuation is about 43 ms (i.e. constant C). This is a very high switching fluctuation and is probably due to the following facts: (1) the input voltage is slowly approaching threshold level of the inverter, thus causing the long transition times at the inverter's output;

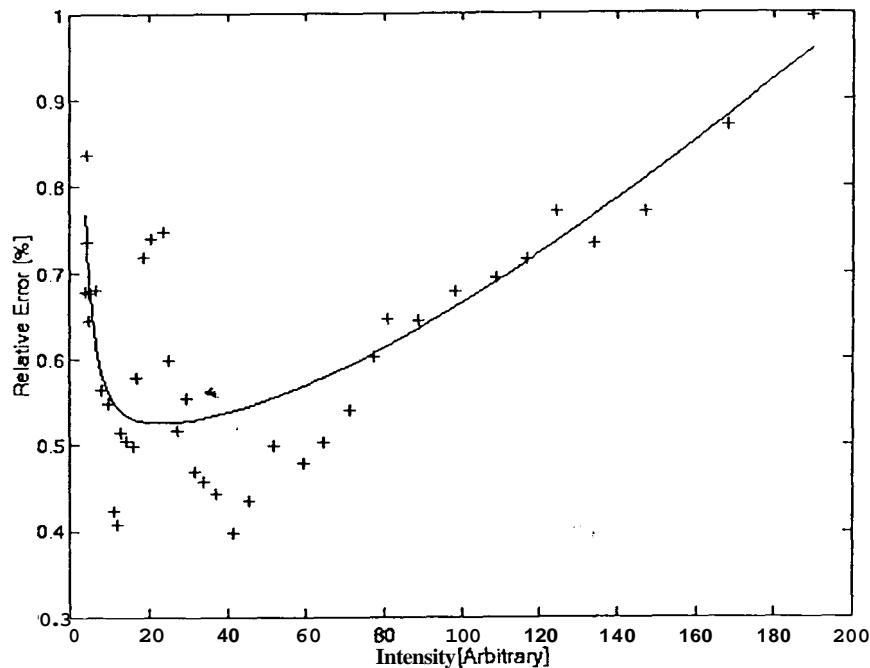


Fig. 15. Relative error σ_{II} : experimental data points and fitted error model

(2) the positive feedback transistor is active only after the “decision” to trip is made; (3) in a static CMOS inverter the p- and n-channel transistors “fight” each other for slowly changing inputs; (4) there could be some systematic limitation in our instrumentation set-up and/or the conditions under which we assume equal illumination for all pixels. In all, the switching fluctuation is approximately 10% of inverter output transition time (i.e. rise time) in the cell receiving the highest intensity in our experiment, which is reasonable. A higher gain thresholding element would probably perform better. This hypothesis will be verified with a new prototype currently being fabricated. Other sources of error, fluctuations in the dark current (i.e. constant A) and mismatch of C and V (i.e. constant B) are within reasonable limits. Relative error for the dark current is approximately 10%, whereas the lumped relative error for C and V is approximately 0.5%.

The second issue we would like to consider is the error the sorting sensor makes when computing the cumulative histogram. This error is due to the mismatch of the current sources I_0 . Since there are typically thousands of cells in the sorting image sensors, the level of current I_0 is very low, pushing the corresponding transistors into the subthreshold regime. In this regime, the current sources could mismatch by 100%, i.e. one current source can be twice as large as another (Mead, 1989). Nonetheless, the monotonousness in the cumulative histogram is maintained. When the cumulative histogram is used for inverse mapping, the mapping from indices to the input intensities, the error in the cumulative histogram is not significant as it will be directly undone.

The error that could be significant when mapping from

indices to input intensities, however, is the readout error for each index. If the scene produced long horizontal segments in the cumulative histogram, such as the example in Fig. 12(a), then a small error in index can result in a large error in inferred response time for a particular cell. This problem can be handled by prohibiting the mapping process to return times within the interval of the long horizontal segments in the cumulative histogram. A few pixels may be grossly misclassified, but overall recovery of input intensities is good.

4. Conclusion

We have shown two analog VLSI computational sensors for the efficient extraction of the high dynamic range sensory data. The methods take advantage of the temporal properties of photoreception. The multiple integration time photoreceptor automatically adapts to use different integration periods depending on light intensity. It exhibits a dynamic range of about 1:128 000. The sorting sensor detects the image and sorts pixels by their intensities, thereby producing an “image of indices”. The image of indices is an information-theoretic optimal representation. Since the sorting image sensor carries out a global operation over sensed images, it adapts in each frame to deliver the “best” image as a whole.

Acknowledgements

This research has been partially funded by the ONR Grant N00014-95-1-0591 and by the NSF, Grant

MIP-9305494. The authors also acknowledge the critical and constructive comments by the reviewers.

References

- Anton, P., Granger, R. & Lynch, G. (1992). Temporal information processing in synapses, cells, and circuits. In T. McKenna, J. Davis & S. F. Zornetzer (Eds.), *Single Neuron Computation* (pp. 291–313). Boston: Academic Press.
- Ballard, D. H. & Brown, C. M. (1982). *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall.
- Boahen, K. A. & Andreou, A. G. (1992). A contrast sensitive retina with reciprocal synapses. In J. E. Moody, S. J. Hanson & R. J. P. Lippmann. (Eds.), *Advances in Neural Information Processing System 4* (pp. 764–772). San Mateo, CA: Morgan Kaufmann Publishers.
- Brajovic, V. (1996). Computational sensors for global operations in vision. Ph.D. Thesis, Carnegie Mellon University. Pittsburgh, PA.
- Brajovic, V. & Kanade, T. (1996). A sorting image sensor: an example of massively parallel intensity-to-time processing for low-latency computational sensors. *Proceedings of the 1996 IEEE International Conference on Robotics and Automation* (pp. 1638–1643), Minneapolis, MN.
- Connen, T. H., Leiserson, C. E. & Rivest R. L. (1992). *Introduction to Algorithm*. Cambridge, MA: MIT Press.
- Delbruck, T. & Mead, C. A. (1994). Analog VLSI phototransduction. California Institute of Technology. CNS Memo No. 30. May 11.
- Kanade, T. & Bajcsy, R. (1993). Computational sensors: a report from DARPA workshop. *Image Undemanding Workshop Proceedings* (pp. 335–350).
- Mathur, B. & Koch, C. (Eds.), (1991). Visual information processing: from neurons to chips. *Proc. SPIE* (p. 1473).
- Mead, C. (1989). *Analog VLSI and Neural systems*. Reading, MA: Addison-Wesley.
- Ripps, H. & Weale, R. A. (1976). Temporal Analysis and Resolution. In H. Davson (Ed), *The Eye. 2A: Visual Function in Man* (pp. 185–217). New York: Academic Press.
- Shannon, C. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL University of Illinois Press.
- Yamada, K., Tomoaki N. & Yamamoto, S. (1997). Effectiveness of video camera dynamic range expansion for lane detection. *Proc. IEEE Conference on Intelligent Transportation Systems*, Boston, MA, November.
- Weckler G. P. (1967). Operation of p–n junction photodetectors in a photon flux integrating mode. *IEEE J. Solid-State Circuits*, 2, 65–73