

# A Compact-pixel Tri-mode Vision Sensor

Dongsoo Kim and Eugenio Culurciello  
Department of Electrical Engineering  
Yale University, CT USA  
Email: dongsoo.kim@yale.edu

**Abstract**— We present a custom image sensor capable of reporting intensity, spatial contrast, and temporal difference images at the pixel level. The smart pixel is composed of only 11 transistors, allowing tight integration of different functionalities in a  $16 \times 21 \mu\text{m}^2$  pixel area. The sensor array is  $128 \times 128$  pixels, with a fill factor of 42%, and operates at 800 fps and 13M events/s with a power consumption of 1.2mW.

## I. INTRODUCTION

Wireless Sensor Networks (WSNs) have a significant impact on advanced sensing technologies and a wide range of applications ranging from military, to scientific, to industrial, to health-care, to home. A group of wirelessly-connected sensing devices collaborate and collect raw local data, producing globally meaningful information [1], [2]. However, WSNs have severely limiting resource bottlenecks in communication bandwidth and power, especially when used with commercial-off-the-shelf image sensors. Therefore, many custom image sensors have been published in the literature to reduce the information data size and the power consumption by reporting temporal difference and spatial contrast images in the binary format [2]–[6]. The motion and contour detections can be also performed by digital signal processing on a FPGA or a  $\mu$ -processor using frame memories. However, the frame rate is limited by the analog-to-digital conversion (ADC) and digital signal processing time and the power consumption of these devices is too large for WSNs nodes. We present a custom image sensor capable of reporting intensity, spatial contour, and temporal difference images using a compact smart pixel array with low power consumption. Section II describes the architecture of the proposed image sensor. The circuit implementation and the measurement results are presented in Section III.

## II. ARCHITECTURE OF THE PROPOSED SENSOR

The most challenging function of a smart sensor that can output intensity, spatial contour (edge), and temporal difference images together is detecting a spatial contour

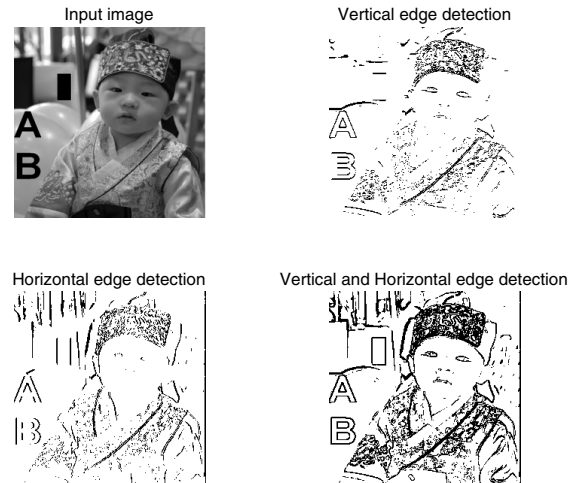


Fig. 1: Edge detection algorithm using WTA and LTA functions.

with as few transistor as possible in a pixel to keep the fill factor as high as possible. The intensity image can be easily readout by a source follower and the temporal difference information can also be obtained with an additional storage capacitor inside a pixel. However, the spatial contour image needs significant signal interactions between the neighboring pixels to get the edge information. The signal communication between the pixels increases the number of transistors and the signal lines causing an increase of the pixel size which limits the large pixel array integration. The main contribution of this work is the use of winner-takes-all (WTA) and loser-takes-all (LTA) algorithms to obtain the contour information. These algorithms do not need the multiple signal lines to communicate each other and can be implemented with a small number of the transistors.

Figure 1 shows the system level simulation using WTA and LTA functions to detect the edges. In Fig. 1, the maximum intensity signal (dark) and the minimum intensity

(bright) between the reference pixel and its neighboring pixels are calculated by the WTA and LTA functions. If the difference between the maximum and minimum values is higher than a threshold, the reference pixel can be marked as an edge. Vertical edges or horizontal edges can be detected by WTA and LTA functions between the pixels these are located on the vertical and horizontal sides, respectively. When the WTA and LTA operations are performed for the reference and its three neighboring pixels (left, bottom, and left-bottom), the vertical and horizontal edges can be computed at the same time (see Fig. 1).

The proposed tri-mode vision sensor (*T-Sensor*) is composed of a  $128 \times 128$  smart pixel array, a row control circuit, column readout circuits, and an event generator as shown in Fig. 2. The smart pixel contains WTA and LTA circuits to find the maximum and the minimum input values. In the intensity mode (*I-mode*), each smart pixel transfers the reset signal and the photo-integrated signal of the photodiode (PD) to the column readout circuit through the two column lines. The column readout circuit computes the difference between the signals on the two column lines and generates the intensity image through correlated double sampling (CDS). In the temporal difference mode (*T-mode*), the smart pixel outputs the previous frame signal and the current frame signal to the column readout circuit after WTA/LTA operations. The column readout circuit generates an event when the difference between the previous frame signal and the

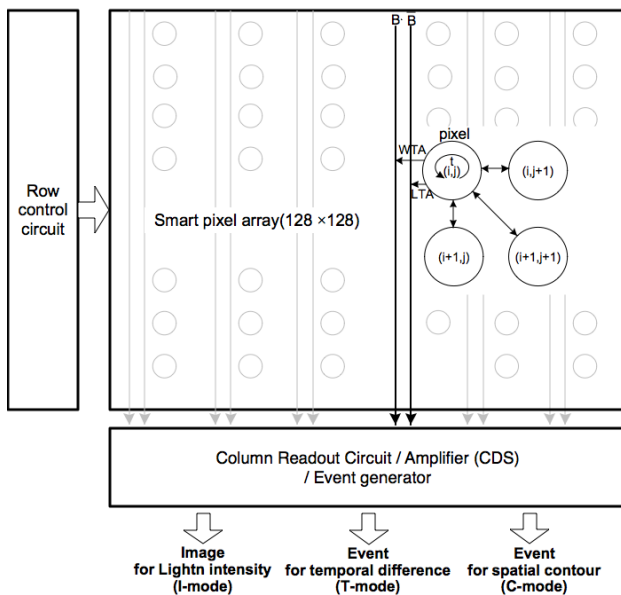


Fig. 2: Block diagram of the T-Sensor system.

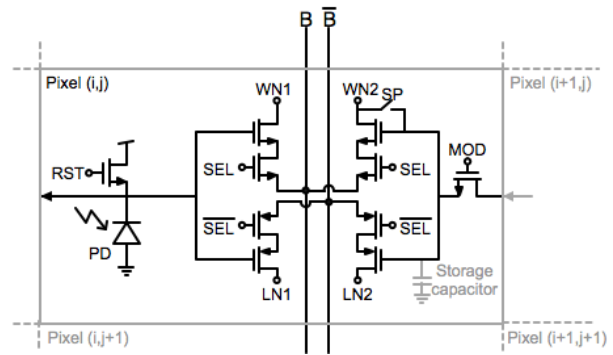


Fig. 3: Block diagram of the T-Sensor pixel.

current frame signal is greater than a certain threshold. In the spatial contrast mode (*C-mode*), the pixel $_{(i,j)}$  finds the maximum and the minimum photo-integrated signals in the 4 neighboring pixels $\{(i,j),(i+1,j),(i,j+1),(i+1,j+1)\}$  and transfers the maximum and minimum signals to the column readout circuit. If the spatial contrast in the four pixels is high, it means that a contour (edge) was found. The column readout circuit calculates the difference between the maximum and the minimum signal and generates an edge event by comparing the difference with a threshold.

### III. CIRCUIT IMPLEMENTATION AND MEASUREMENT RESULTS

The schematic diagram of the proposed compact pixel is presented in Fig. 3. Only 11-transistors/pixel (much less than 45 transistors/pixel for temporal difference and spatial contrast images in [2], and 51 transistors/pixel for spatial contrast and wide dynamic intensity images in [5]) provides intensity, temporal difference, and spatial contrast images together and enables greater pixel array integration in the same die size. Four NMOS transistors compose the WTA input circuit and four PMOS transistors form the LTA input circuit [7]. SP is turned on to transfer the signal from the PD to the parasitic storage capacitor. MOD is turned on in C-mode to connect the signal of the PD in the right side as an input of WTA/LTA circuit. Figure 4 shows the column readout circuit that makes the pixel operate for a sampling operation or WTA/LTA operations depending on the status of SP signal.

Figure 5 describes the timing diagram for I-mode, T-mode, and C-mode. In I-mode, the photo-integrated signal in the PD is sampled into the storage capacitor by turning on SP after an exposure time. Next, the pixel is reset, and the reset signal and the photo-integrated signal

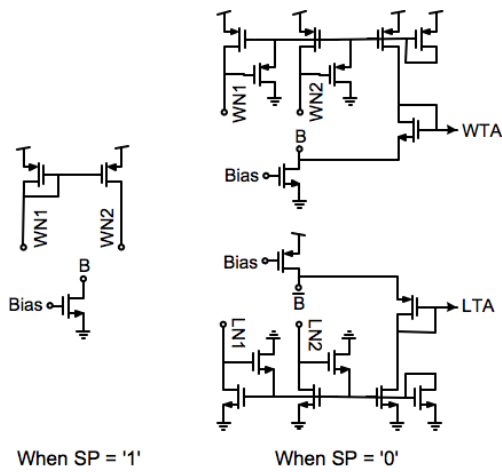


Fig. 4: Schematic diagram of the column readout circuit for WTA and LTA functions.

are readout through WTA and LTA circuits, respectively. In T-mode, the previous frame signal in the storage capacitor and the current frame signal in the PD are readout simultaneously through the WTA/LTA circuits. Next, the current frame signal is sampled and stored in the parasitic storage capacitor as the previous signal for the next frame by turning on SP, and the PD is reset. In C-mode, the row selection signals for two rows are turned on together and MOD is turned on. This allows the pixel circuits to be reconfigured into a 4- input WTA and LTA circuits used for detecting spatial contour in the four neighboring pixels without complicated signal connections between the neighboring pixels. The event generator including a variable gain amplifier (VGA) and a comparator is presented in Fig. 6. The VGA is implemented using a switch capacitor circuit and the gain  $\alpha$  is adjustable from 1 to 8 using a capacitor bank. The VGA can also perform CDS function for I-mode with the gain of 1. The address of the event is generated by two 7-bit counters these share the synchronized clocks with the row control circuit and the column readout circuit.

The proposed T-Sensor was fabricated with a 0.18- $\mu\text{m}$  SiGe BiCMOS 7-metals process and the microphotograph is presented in Fig. 7. The fabricated T-Sensor has  $128 \times 128$  smart pixel array and the core area is  $3.1 \times 3.1 \text{ mm}^2$ . Each smart pixel has a pitch of  $16 \mu\text{m} \times 21 \mu\text{m}$  with a fill factor of 42%. Figure 8(a) shows the test board that includes the fabricated T-Sensor, C-mount lens and USB interface components. The measured pixel sensitivity is  $2.14 \text{ V/s} \cdot (\mu\text{W}/\text{cm}^2)$  at 550-nm light wavelength and  $0.31 \text{ V/s} \cdot (\mu\text{W}/\text{cm}^2)$  at 850-nm light wavelength. Since T-Sensor was fabricated with

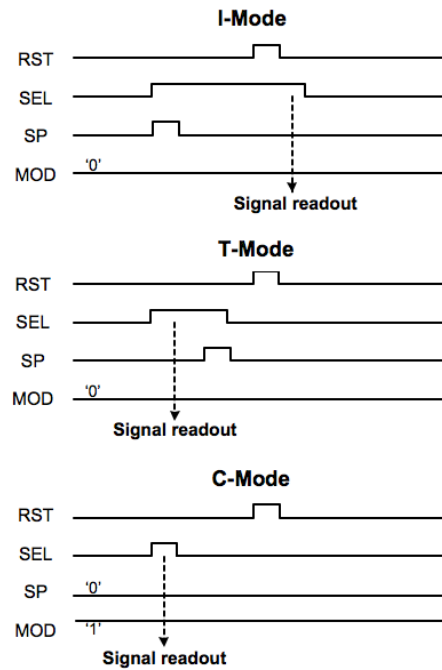


Fig. 5: Timing diagram of I-, T-, and C-modes.

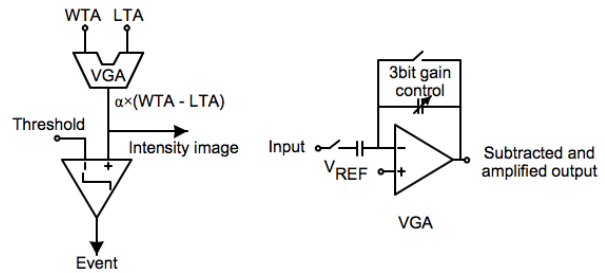


Fig. 6: Block diagram of the event generator and the variable gain amplifier.

a silicon germanium process that has better sensitivity in the longer wavelengths, the sensor can be used in dark lighting conditions with infra-red light sources. The T-Sensor operates with power supplies of 2V to 3.3V and the power consumption is 1.4 mW (I-mode) and 1.2 mW (T- and C-modes) at 3 V,  $680 \mu\text{W}$  (I-mode) and  $620 \mu\text{W}$  (T- and C-modes) at 2V, respectively. Figure 8 (b), (c), and (d) show the sample images taken from a rotating pattern on the resolution chart. The maximum frame rate is 200 frames/s (I-mode), 800 frames/s (T- and C-modes) and the maximum event rate is 13M event/s with 3-V power supply.

#### IV. CONCLUSION

The proposed smart vision sensor applied for wireless sensor networks can provide intensity, spatial contrast,

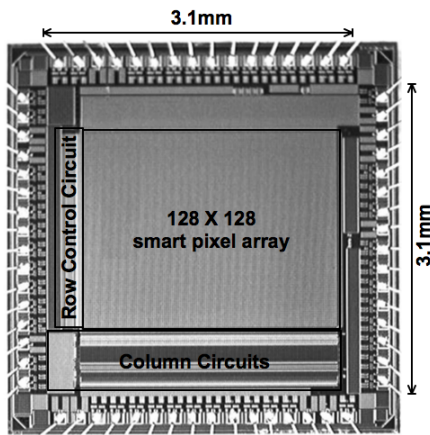


Fig. 7: Microphotograph of the fabricated T-Sensor die.

and temporal difference images with low power and high speed of 800 fps and 13M events/s. The compact size of the pixel composed of only 11 transistors enables the integration of a large pixel array in a limited sensor size and high sensitivity given its large fill factor.

#### V. ACKNOWLEDGEMENTS

This work was partially supported by NSF grants ECS-0622133 and ECCS-0901742 and the prototype

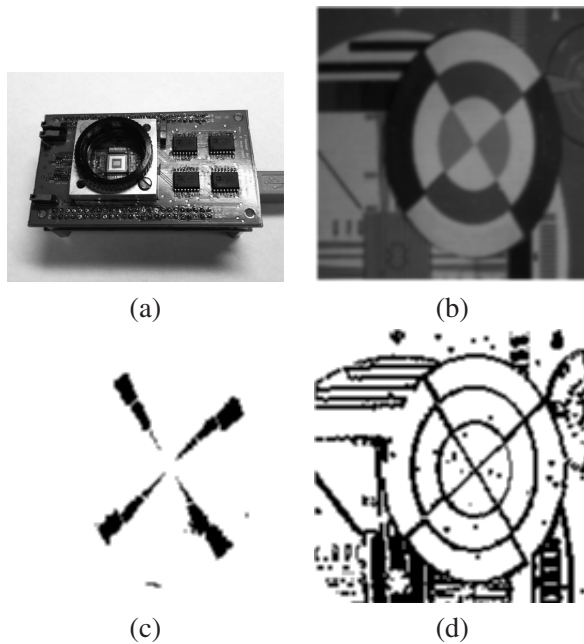


Fig. 8: Test PCB board and sample images taken from a rotating circle panel on the resolution chart, (a) shows the test board, (b) is the intensity image, (c) is the temporal difference image from the rotating circle panel, (d) is the spatial contour image.

TABLE I: Performance Summary

Process	0.18- $\mu\text{m}$ SiGe BiCMOS 7-metal
Power supply & power consumption	2V to 3.3V
N-mode	680 $\mu\text{W}$ at 2V, 1.4 mW at 3V
C- and T-mode	620 $\mu\text{W}$ at 2V, 1.2mW at 3V
Chip size	$3.1 \times 3.1 \text{ mm}^2$
Array size	$128 \times 128$
Pixel size	$16 \times 21 \mu\text{m}^2$
Fill factor	42 %
Conversion gain	$1.17 \mu\text{V}/e^-$
Full well capacity	$1.7 \times 10^6 e^-$
Sensitivity	$2.14 \text{ V/s} \cdot (\mu\text{W}/\text{cm}^2)$ @ 550 nm $0.31 \text{ V/s} \cdot (\mu\text{W}/\text{cm}^2)$ @ 850 nm
Frame(event) rate	200 fps for I-mode 800 fps (13M events/s) for C- and T-mode

fabrication was supported by MOSIS Education Program. We also thank Joon Hyuk Park for the data visualization software and help with data collection.

#### REFERENCES

- [1] C. Chong and S. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 5, August 2003.
- [2] N. Massari, M. Gottardi, and S. Jawed, "A  $100\mu\text{w}$   $64 \times 128$  pixel contrast-based asynchronous binary vision sensor for wireless sensor networks," in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, Feb. 2008, pp. 588–638.
- [3] N. Massari and M. Gottardi, "A 100 dB dynamic-range CMOS vision sensor with programmable image processing and global feature extraction," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 3, pp. 647–657, 2007.
- [4] P. Lichtsteiner, C. Posch, and T. Delbruck, "A  $128 \times 128$  120dB 30mW Asynchronous Vision Sensor that Responds to Relative Intensity Change," in *2006 International Solid State Circuits Conference (ISSCC 2006)*, 2006, pp. 508–509.
- [5] P. Ruedi, P. Heim, F. Kaess, E. Grenet, F. Heitger, P. Burgi, S. Gyger, and P. Nussbaum, "A  $128 \times 128$  pixel 120-dB dynamic-range vision-sensor chip for image contrast and orientation extraction," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 12, pp. 2325–2333, 2003.
- [6] J. Costas-Santos, T. Serrano-Gotarredona, R. Serrano-Gotarredona, and B. Linares-Barranco, "A spatial contrast retina with on-chip calibration for neuromorphic spike-based aer vision systems," *IEEE transactions on circuits and systems. I, Regular papers*, vol. 54, no. 7, 2007.
- [7] D. Kim, J. Cheon, and G. Han, "An offset cancelled winner-take-all circuit," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 92, no. 2, pp. 430–435, 2009.