A Second Generation 3D Integrated Feature-Extracting Image Sensor

Xiangyu Zhang [1], Shoushun Chen [1] and Eugenio Culurciello [2]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University
[2]Electrical Engineering Department, Yale University

***Abstract*—This paper presents a second generation 3D integrated feature-extracting CMOS image sensor. This 64×96 pixel vision sensor was designed and fabricated using a 0.18 μm 3D FDSOI process. Each pixel implements a photodiode and computation circuits on three individual tiers, which are vertically stacked and connected through the 3D inter-tier vias. The photodiode is sited on the top tier with an optimized photo sensitivity and a high fill factor (∼97%). The in-pixel analog memory and a parallel diffuser network allow this image sensor to store the previous frame image and perform a spatially smoothing operation. Hence, the proposed image sensor can deliver an intensity image, either the previous frame or the smoothed image. Using off-chip subtraction, image features including temporal motions and spatial contours can be easily extracted. Moreover, the power consumption for this image sensor is around 0.8mW at 100 fps. The low power consumption and feature extraction capability make this sensor appealing for the sensor network applications.**

I. Introduction

Surveillance applications including security monitoring and traffic enforcement require continuous image acquisition and processing in real time. However, the raw images captured by the conventional image sensors contain massive quantities of primitive and redundant information. The transmitting and processing of the raw data significantly increases system burden both on hardware resources and power consumption. Moreover, this kind of situation becomes even worse for multi-sensor network applications. Hence, there is a great demand for a feature-extracting image sensor that can perform image preprocessing on the focal plane. However, this usually leads to a large pixel size and a low fill-factor due to the increased complexity.

Spatial contour and temporal motion are two widely used features for image processing in pattern recognition and object tracking applications. Spatial contrast allows an object to be detected and segmented from the surrounding environment. Many recent research works on contour extracting image sensors have been reported. Yoshinori *et al.* proposed a contour extracting image sensor by horizontally comparing pixel's responses with those of its neighbors and producing the spatial contrast[1]. However, its main limitation is that only vertical edges can be detected. The asynchronous binary image sensor proposed by Mossimo *et al.*[2] implements a pixel-level charge transfer mechanism to estimate spatial contrast. The sensor asynchronously dispatches pixel's event using Address-Event-Representation (AER)[3] when an effective edge is detected. Despite the low power consumption, the proposed pixel architecture is fairly complex and also additional off-chip memories are required for the image reconstruction.

Moving objects can be detected and extracted from a stationary background using motion features. Many publications have reported various algorithms for motion detection. One common frame based approach is to employ temporal contrast computation by comparing consecutive images, whose differences are related to the motions in the scene[4]. However, the buffer memory (for storing the previous frame) necessitates a large silicon area and leads to a low fill factor. In contrast to the frame based scheme, Patrick *et al.*[5] reported a binary temporal contrast vision sensor, whose pixels asynchronously respond to relative changes in intensity and reduce the redundant data to be delivered. Although overall performance is efficient for motion detection, the complex pixel circuitry limits proposed sensor to a low fill factor of ∼8%.

A recently emerging 3D integrated technology offers a promising approach to alleviate the above issues. In this technology, several silicon-on-insulator (SOI) tiers with different circuits can be vertically stacked and interconnected through the 3D inter-tier vias. This high intensity integration strategy allows a vision sensor to separate photo sensitive devices with processing circuits on different tiers to achieve a high fill factor. Recent publications have reported image sensors designed and fabricated using this process[6][7]. Although sitting photodiodes on the top tier achieves a nearly 100% fill factor, the proposed vision sensors can only capture an intensity image, which does not fully exploit the advantage of the 3D technology.

In this paper, we propose a 3D integrated feature-extracting image sensor using a 3D CMOS FDSOI process. To optimize photo sensitivity, the top tier of the vision sensor is dedicated to photo detection with a maximized fill factor (∼97%). Each pixel implements an analog memory (capacitor) and diffuser network (tunable PMOS resistors), which allow the sensor to store the previous frame image or spatially smooth a captured image. Off-chip subtraction on the intensity image with the previous frame or the smoothed image can extract the temporal motion or spatial contour feature. The major contribution of this paper resides in the implementation of a feature-extracting image sensor using a 3D integration technology.

The remainder of this paper is organized as follows. Section II introduces the sensor system and pixel architectures. Contour extraction and motion detection operations are described in Section III. Section IV reports the experimental results and chip characteristics. Section V is the conclusion of this paper.

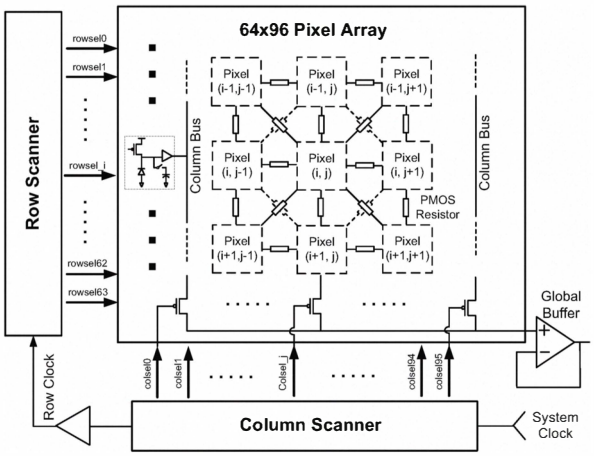

Fig. 1. System architecture of the feature-extracting image sensor. Main building blocks include 64 × 96 pixel array, PMOS resistor network, global buffer as well as row and column scanners.

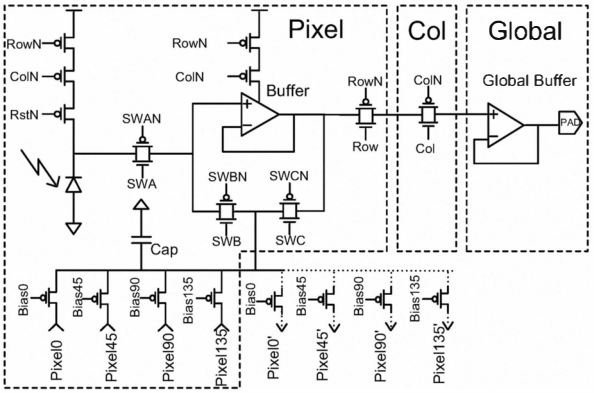

Fig. 2. Schematic of the pixel and readout path. The pixel consists of a photodiode, an analog memory (a capacitor), an in-pixel buffer, three switches (SWA, SWB and SWC) and four PMOS resistors (Bias0, Bias45, Bias90 and Bias135). Neighboring pixels are cross-connected with the tunable resistors in four orientations (0 °, 45 °, 90 °, 135 °)

II. System Overview

A. Sensor Architecture

Figure 1 shows the system architecture of the proposed 3D feature-extracting image sensor. The main building blocks include a 64 × 96 pixel array, row and column scanners, PMOS resistor network and a global analog buffer. Pixels are sequentially accessed by the row and column scanners to extract voltages to global buffer during the operation. The global analog buffer is a two-stage operational amplifier, which is in charge of driving the output pads and off-chip loads. Note that each pixel is cross-connected with eight neighbors through the PMOS resistors. These parallel connections construct a diffuser network which can perform selective orientation low pass filter (spatially smoothing) operation on the image stored in the pixel array.

B. Pixel Architecture

A pixel schematic with the global readout path is shown in Figure 2. The pixel consists of a photodiode, an analog memory (a MOS capacitor), an in-pixel buffer, three complementary switches (SWA, SWB and SWC) and four PMOS resistors (Bias0, Bias45, Bias90 and Bias135). A pure PMOS reset path allows the photodiode to be reset to Vdd. Since each pixel can be individually addressed through the row/column select transistors to be reset, the proposed image sensor benefits less row-wise mismatch. An in-pixel capacitor works as an analog memory to store the frame value. In the motion detection mode, the capacitor stores the previous frame as a reference for temporal difference computation. In the contour extraction mode, the captured raw image is copied to the capacitors where low pass filtering is performed. Instead of using a source follower, an operational amplifier is exploited as an analog buffer to drive the column bus with larger driving capability and higher speed. As to reduce the power consumption, this analog buffer is activated only when the pixel is addressed. Three complementary switches (SWA, SWB and SWC) are used to configure pixel's operation mode. Activating SWA connects the photodiode to in-pixel buffer. Turning on switch B can extract the capacitor value. Enabling both switch A and C updates the capacitor with photodiode new integrated voltage. Four PMOS resistors connect each pixel with neighbors to form a diffuser network. The resistance in each orientation is individually controlled by its gate biasing voltage.

C. 3D Integration

Figure 3 illustrates the 3D integration of the proposed pixel structure. The top tier is dedicated to the photo detection and fully covered by the photodiodes with a high fill factor (∼97%). The back-side illuminated photodiode is obtained by abiding N- and P-type silicon regions, featuring a vertical PN junction. To avoid the edge effect and silicide deposition on the depletion region, the photodiode was designed in an annular shape and shielded with a ploy-silicon layer. The middle tier consists of pixel reset transistors, in-pixel buffer and the row and column scanners (not shown in this figure). Computation circuits including the analog memory and resistor network are resided on the bottom tier. Inter-connections between each tier are accomplished by the massive 3D inter-tier vias, which are highlighted with yellow bars in Figure 3

III. Feature Extraction

A. Contour Extraction

The spatial contour for a given image arises in the region where the intensities for the neighboring pixels change obviously. In our proposed vision sensor, a charge sharing mechanism is exploited to perform the spatial contour extraction. As shown in the system architecture, neighboring pixels are cross-connected through the eight tunable PMOS resistors. Due to the charge sharing effect[8], the charge stored on the capacitor for a given pixel will be redistributed among its neighboring pixels. The longer distance between two pixels induces the less charge sharing between them. Hence, the biased resistor network low-pass filters (spatially

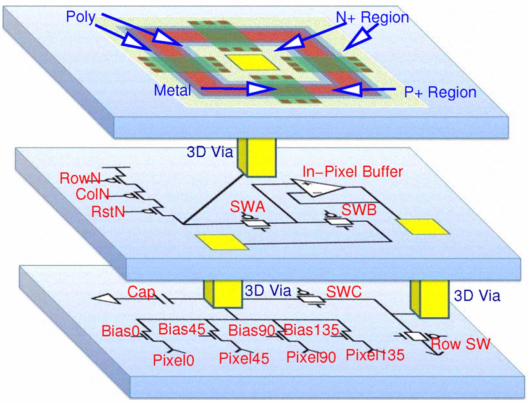

Fig. 3. 3D integration of the pixel circuit. The back-side illuminated photodiode is obtained by abiding N- and P-type silicon regions, featuring a vertical PN junction. The top tier is dedicated to the photodiodes and the processing circuits are resided on the sequential bottom tiers. Inter-connections between each tier are realized by the 3D inter-tier vias.

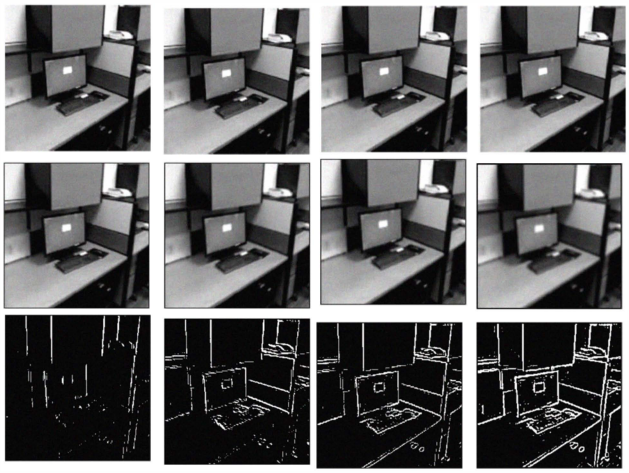

Fig. 4. Evolution of the spatial contour extraction. The sample images from top to bottom correspond to original image, smoothed image and binary spatial contrast. Column (a) is horizontally smoothed which only detects vertical edges. Both vertical and horizontal contrast can be extracted by diagonally smoothing, shown in column (b) and (c). Column (d) reports full orientation extraction when all the diffuser resistors are enabled. Note that the edges detected in column (d) are thicker due to longer smoothing time.

smoothes) the image stored on the capacitors. The comparison between the original image and its smoothed one can easily extract the spatial contour in the scene. Since both the biasing potential and pulse width of the PMOS resistor in each orientation are externally controlled, the contour extraction can be performed in a selective orientation and depth. Figure 4 shows simulation results on the evolution of the contour extraction operation. The sample images from the top to bottom correspond to the intensity image, the smoothed one and their difference (spatial contour). Example (a) shows the horizontally smoothing operation, in which the vertical edges are extracted. Both the horizontal and vertical contrast can be detected by the diagonally smoothing shown in example (b) and (c). A full orientation smoothing is shown in example (d) when all the diffuser resistors are activated. Note that the edges detected in last column are thicker due to a longer smoothing time. Because of the parallel analog processing, this architecture can perform the smoothing operation at a higher speed. Also, the charge sharing mechanism reduces the sensor's power dissipation since no DC power is consumed during the smoothing procedure.

B. Motion Detection

Temporal difference computation is a widely used approach to detection motions in the scene. Subtracting the current frame with a previous frame can easily extract the temporal intensity changes on the focal plane, which are usually related to the motions in the scene. As to perform the computation, a previous frame has to be stored on a buffer memory. In our proposed image sensor, each pixel is implemented with a MOS capacitor as the memory to store the previous frame information. The first frame is initially stored on the capacitors by turning on the complementary switch C before the pixel finishes reading out photodiode integrated voltage. When it comes to extract the new integrated frame, pixel sequentially turns on the switch B and switch A to output the previous frame value and the current frame value. An off-chip subtraction can obtain the temporal contrast (motions) in the scene.

IV. Experimental Results

The proposed feature-extracting image sensor has been fabricated using MITLL 3D 0.18μm CMOS FDSOI process. Figure 5 shows die microphotograph on the top tier with a 64 $\times$ 96 pixel array and testing structure highlighted. The chip has a total area of 1.45 mm $\times$ 1.5 mm including the I/O pads. The top tier is fully covered by the photodiodes with optimized photo sensitivity. Excluding the vertical 3D via connected to the middle tier, the proposed pixel achieves a high fill factor of $\sim$97%. We measured photodiode's responses under different biasing and illumination conditions. Figure 6 reports the photodiode's responses under three biasing voltages (0.5v, 1v and 1.5v) with an illuminance intensity range from 2k to 200k lux. Photodiode current has a linear response to incident illuminance intensity, which can be modeled as

$$I_p = 3 \times 10^{-14} \times I_{in}$$

, where Ip is the photocurrent in ampere and Iin is the illuminance intensity in lux. Note that biasing voltage does not have a significant influence on the photosensitivity.

Figure 7 shows sample images acquired by the proposed 3D vision sensor. The sample photographs were taken on a laptop (a), a human body (b), a building (c) and vehicles (d). Due to the low resolution on this image sensor (only 64 $\times$ 96 pixels), the sample images look slightly fuzzy compared to those captured by the high resolution cameras. One can note that there are many discrete and fixed 'hot pixels' noises on the sample images, which are induced by the defects during the fabrication and package.

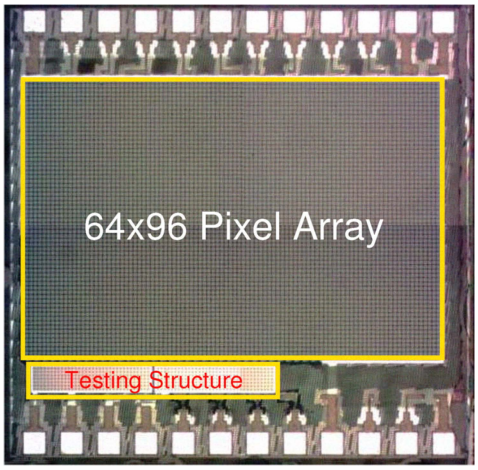

Fig. 5. Microphotograph of the vision sensor.

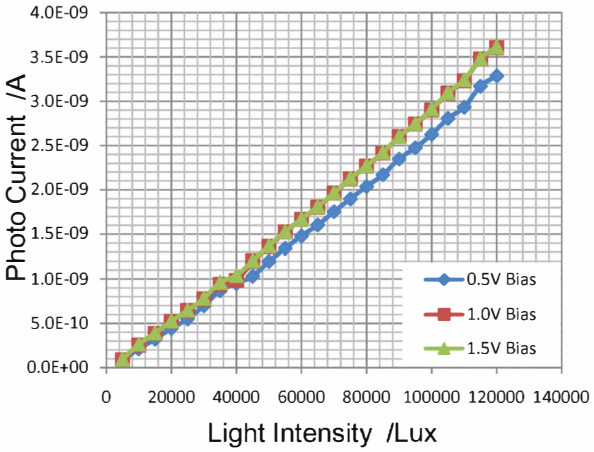

Fig. 6. Photocurrent responses versus the illuminance intensity under three biasing conditions. Due to the light source restriction, the intensity range for illuminance is limited from 5k to 120k lux.

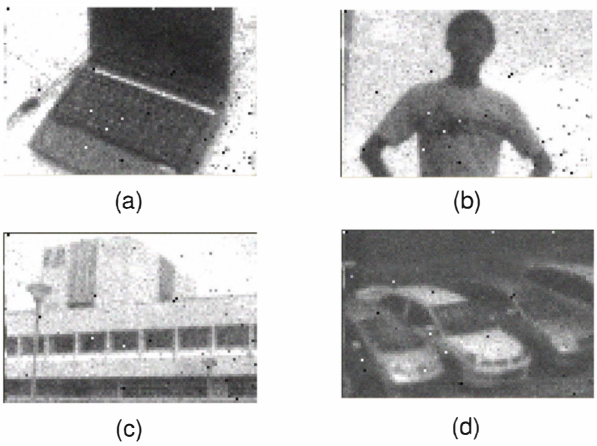

Fig. 7. Sample images taken by the 3D vision sensor. Examples (a)-(d) correspond to the sample images on a laptop, a human body, a building and vehicles. The noises in the sample images are mainly caused by the defects during fabrication and package.

Other characteristics of the 3D chip are summarized in Table I. The proposed vision sensor operates under 1.5V power supply. The power consumption for this vision sensor at 75 frame/s is around 0.8 mW. The main power dissipation comes from the row/column scanners and global buffer with total 0.6mW. The maximum frame rate of the image sensor is around 160 frame/s.

TABLE I
CHIP CHARACTERISTICS

Process Technology	MITLL 0.18 μm 3D FDSOI process
Die Size	1.45 $\times$ 1.5 mm^2
Array Size	64 $\times$ 96 Pixels
Pixel Size	14 $\times$ 14 μm^2
Pixel Complexity	22 transistors
Fill Factor	~97%
Fixed Pattern Noise	3.5%
Dark Current	~2.6fA
Maximum Frame Rate	160fps
Operating Voltage	1.5 V
Power Consumption	0.8 mW at 75 frame/s

We encountered testing difficulties on the sensor's motion detection and contour extraction functions due to the malfunctions of the in-pixel capacitors.

V. CONCLUSION

In this paper, we present a 64 $\times$ 96 pixel feature-extracting image sensor targeted for sensor network application. This vision sensor was designed and fabricated using a 3D integrated 0.18μm FDSOI CMOS process. Top tier is fully covered by photodiodes with a high fill factor of $\sim$ 97%. The in-pixel analog memory and global diffuser network allow sensor to store the previous frame and perform smoothing operation. Hence, the proposed image sensor can simultaneously deliver an intensity image, either the previous frame or the spatially smoothed image. Using off-chip subtraction, image features including temporal motions and spatial contours can easily be extracted from the scene. The power consumption for the vision sensor operating at 100 frame/s is around 0.8mW.

REFERENCES

[1] Y. Muramatsu, S. Kurosawa, M. Furumiya, H. Ohkubo and Y. Nakashiba, " A Signal-Processing CMOS Image Sensor Using a Simple Analog Operation," IEEE JSSC, 2003, Vol.38, pp.101-106
[2] M.Gottardi, M.Massari and S.A.Jawed, " A 100μw 128 $\times$ 64 Pixels Contrast-Based Asynchronous Binary Vision Sensor for Sensor Networks Applications," IEEE JSSC, 2009, Vol.44, pp.1582-1592.
[3] E. Culurciello, R. Etienne Cummings, K. A. Boahen, "A Biomorphic Digital Image Sensor," IEEE JSSC, 2003, Vol.38, pp.281-294.
[4] Shoushun Chen, W. Tang and E. Culurciello, "A 64x64 Pixels UWB Wireless Temporal-Difference Digital Image Sensor," IEEE ISCAS, 2010, pp.1404-1407,
[5] P. Lichtsteiner, C. Posch, T. Delbruck, "A 128 $\times$ 128 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," IEEE JSSC, 2008, Vol.43, pp.566-576.
[6] V. Suntharalingam, R. Berger, J. Burns, C. Chen et al., " Megapixel CMOS Image Sensor Fabricated in Three Dimensional Integrated Circuit Technology," IEEE ISSCC, 2005, pp.356-357.
[7] E. Culurciello and P. Weerakoon, " Three-Dimensional Photodetectors in 3-D Silicon-On-Insulator Technology," IEEE Electron Device Letters, 2007, Vol.28, pp.117-119.
[8] F. Zhengming and E. Culurciello, " A 3D Integrated Feature-Extracting Image Sensor," IEEE ISCAS, 2007, pp.3964