A Hybrid-Readout and Dynamic-Resolution Motion Detection Image Sensor for Object Tracking

Xiangyu Zhang and Shoushun Chen
Virtus IC Design Center of Excellence, School of EEE
Nanyang Technological University

***Abstract*—This paper presents a hybrid-readout and dynamic-resolution CMOS image sensor targeted for object tracking applications. The proposed vision sensor can either work in an asynchronous motion detection mode or synchronous region of interest (ROI) readout mode, with different resolutions. In the first mode, relative intensity changes are monitored by a motion detection unit formed of 2×2 pixels and further diffused by a capacitance coupling network to reduce spatial noise. After tunable thresholding, active motions are converted into binary events and asynchronously delivered to the outside using an address-event-representation (AER) protocol. In the ROI extraction mode, this image sensor sequentially accesses each pixel in the interested region to report an analog intensity image with a higher resolution. This sensor has been implemented using a standard 0.18 μm CMOS process. Each pixel contains three capacitors and around 10 transistors, occupying a silicon area of 25×25 μm^2, with a fill factor of ∼42%.**

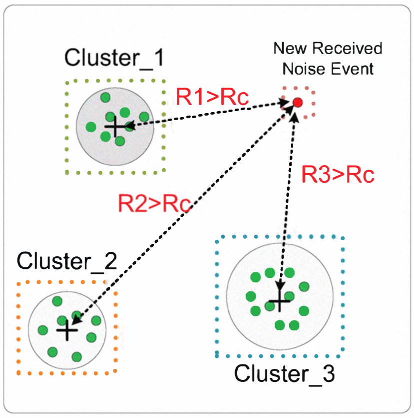

Fig. 1. Example of noise influence on the clustering algorithm. The distances between a newly received noise event and any existing cluster are larger than the distance criterion. A new temporary cluster is created for this noise event.

I. Introduction

Fast motion object detection is important for many applications including surveillance, security monitoring and traffic enforcement [1]. Conventional vision sensors without computation capability produce images with massive quantities of primitive and redundant information. In order to extract motion features, transmission and processing of these raw data significantly increase system's burdens on the hardware resources and on the power consumption. Moreover, multi-sensor network applications accentuate such issues with increasing spatial and coding resolutions. One efficient approach to reduce the redundancy and increase the efficiency is to implement motion feature-extracting algorithms on the focal plane so that image sensor can directly dispatch motion features to post-processors, leading to a low bandwidth requirement and a high working speed.

Recent publications have reported a variety of on-chip hardware-implemented motion detection image sensors. The most common approach used is the temporal difference computation algorithm [2], which compares the differences between two consecutive images to detect motions in the scene. However, this frame-based algorithm may discard fast moving objects due to the fixed integration period on capturing each frame. In contrast to the frame-based strategy, Lichtsteiner *et al.* reported an asynchronous event-based temporal contrast vision sensor [3], in which pixels continuously quantize relative intensity changes to binary events once the variations exceed predefined thresholds. The output of this sensor is a stream of address-events that directly encode scene reflectance changes. This temporal contrast sensor has been successfully implemented in several real-time moving object tracking applications [4][5]. This tracking algorithm is based on a mean-shift approach by continuously clustering each address-event. Newly received event is assigned to an existing cluster based on a distance criterion to update its weight and position information. By periodically refreshing and labeling clusters, multiple moving objects can be effectively tracked in real time. However, this algorithm requires extensive computations and memories due to a long list of clusters to be maintained and computed with each event. Moreover, the complexity of the post-processing system is quite sensitive to random noise. As shown in Fig.1, a newly received event can not be assigned to any existing cluster because its distances to them are larger than the criterion. Consequently, a new temporal separate cluster has to be created for this event. However, this individual cluster can not be immediately treated as "noise" or "active object" without computing with subsequent events to check whether it has enough nearby neighbors surrounded. These operations consume a huge amount of memory and require a lot of computations, which limits an on-chip implementation of this tracking algorithm. Therefore, a noise-suppressed motion detection image sensor is in high demand for these real-time tracking applications.

In this paper, we propose a noise-reduced motion detection image sensor with a hybrid readout strategy. There are two operating modes for this sensor: motion detection mode and analog ROI extraction mode. In the motion detection

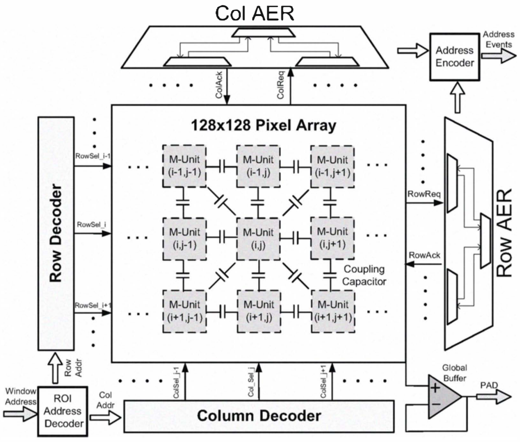

Fig. 2. System architecture of the vision sensor. The main building blocks include a 128 $\times$128 pixel array, row and column arbiters and decoders, a coupling capacitor network and a global analog buffer. In this figure, 2$\times$2 pixels are combined together as a motion detection unit.

mode, 2$\times$2 pixels are spatially combined together as a basic motion detection unit. Relative intensity changes caused by active motions are converted into electrical charges in this unit. To reduce random noise, an analog clustering circuit redistributes these motion charges among the focal plane through a coupling capacitor network. In this case, regions with multiple motions from an active moving object will create a larger bell shape response compared to that of discrete noise. After thresholding, only active motion regions can survive and asynchronously deliver its motion events to the outside data receiver. When working in the ROI extraction mode, this image sensor sequentially accesses each pixel in the interested region to synchronously report an analog intensity image with a higher resolution. Our major contributions reside in two parts: 1) an efficient on-chip analog clustering circuit to reduce discrete noise, 2) a combination of a hybrid readout strategy with a dynamic resolution structure.

The remainder of this paper is organized as follows. Section II introduces the system architecture and pixel schematic. The analog clustering algorithm is described in Section III. Section IV reports the hardware implementation for this sensor. Section V is the conclusion of this paper.

II. System Overview

A. Sensor Architecture

Fig.2 illustrates system architecture of the proposed vision sensor. The main building blocks include a 128$\times$128 pixel array, row and column arbiters and decoders, a coupling capacitor network and a global analog buffer. In this image sensor, we combine 2$\times$2 pixels as a basic motion detection cell, defined as the *"M-Unit"*. This unit is in charge of detecting relative intensity changes and converting them into the electrical charges. As shown in Fig.2, each *"M-Unit"* is cross-connected with eight neighbors in four orientations through a capacitor network, which is adopted for implementing an analog clustering algorithm to diffuse the detected motions. By comparing with tunable thresholds, final smoothed motions are converted into binary events and are asynchronously transmitted to the outside post-processer through row and column AER arbitrations. An address encoder unit labels each motion event with its firing address. Once the kinetic objects in the scene are localized by external processer, our proposed image sensor can further switch into the analog mode to synchronously scan the interested region and report an analog intensity image with a higher resolution by row and column decoders. The global analog buffer is a two-stage operating amplifier, which is used to buffer analog intensity voltage from each pixel to the loading pad.

B. Pixel Structure

Schematic of the motion detection unit excluding handshaking communication circuit is shown in Fig.3. The front end of this motion detection unit is a logarithmical photoreceptor circuit [3], which is composed of four individual photodiodes with necessary biasing and amplifying transistors. This logarithmical photoreceptor can automatically self-adapt for different incident illumination conditions. Meanwhile, the dynamic range of this logarithmical pixel can be extended to 120dB at a cost of higher fixed pattern noise (FPN). As shown in the schematic, there are four analog complementary switches (S1-S4) connecting photodiodes (P1-P4) to the peripheral circuits. These four switches are used to configure the operations on the photodiodes. Simultaneous turning on of these four switches realizes a charge binning operation among the photodiodes, resulting in a lower resolution (64$\times$64 *"M-Unit"*) and a higher photosensitivity. In ROI extraction mode, switches (S1-S4) are sequentially activated following a sequence from S1 to S4 to extract an analog intensity voltage from each photodiode to the column bus. In this case, our proposed sensor is configured with a higher resolution of 128$\times$128 pixel array.

The output of this photoreceptor is connected to the first-stage delta modulator formed by transistors M6-M9 and capacitors C1-C2. The direct relationship between the input and output of this differencing circuit is given by

$$V_{out} = V_{reset} + \frac{C_1}{C_2} \times \Delta V_{in}$$

It can be seen that small voltage changes from the photoreceptor due to the intensity variations caused by the motion activities is amplified by this modulator. Transistors M7 and M8 are the global and internal reset transistors used to balance the input and output of this modulator to a reset voltage. The output voltage variation from this modulator is further buffered to the coupling capacitor network through the capacitor C3. Four coupling capacitors (Cc1-Cc4) connect each motion detection unit with its eight neighbors in four orientations. Since each coupling capacitor is shared by two *"M-Units"*, each unit only physically contains four coupling capacitors. Due to the charge coupling effect, motion detected at each "M-Unit" is spatially redistributed among the entire array. Hence, the voltage on each coupling point represents a weighted motion both from itself and neighbors. Detailed analysis of this clustering algorithm is given in the next section. Transistors

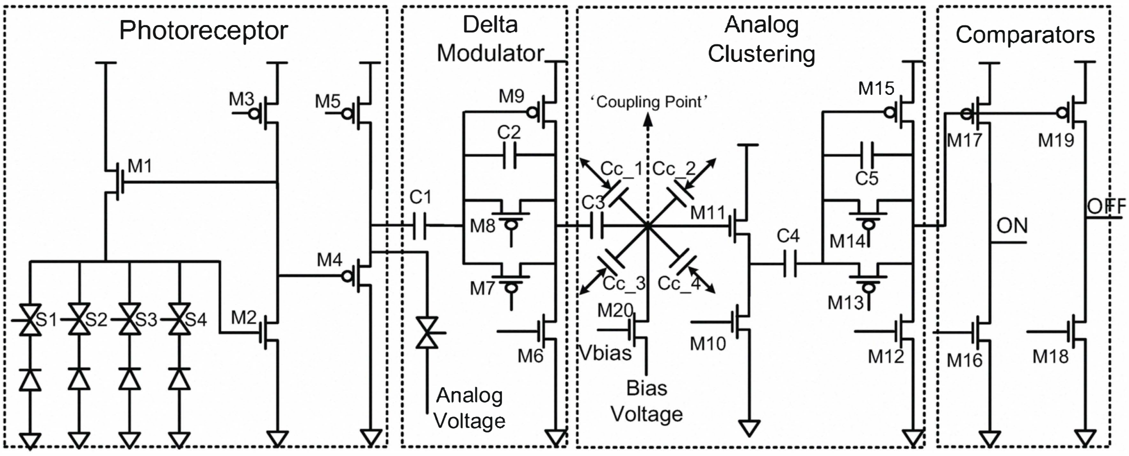

Fig. 3. Schematic of the motion detection unit. The main building blocks include photoreceptor, delta modulator, analog clustering unit and comparators. Photoreceptor is composed of logarithmic photodiodes with the peripheral biasing and readout circuits. A delta modulator is used to amplify the output voltage changes on photoreceptor. Analog clustering block diffuses the detected motion, which is further amplified by another differencing circuit. Comparators convert final motion into "ON" or "OFF" events once it exceeds threshold.

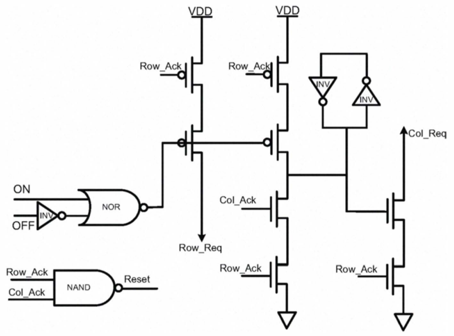

Fig. 4. Schematic of the handshaking communication circuit. Motion event is generated at the output of the NOR gate once there is an effective state change of "ON" or "OFF". A row request is immediately sent to the row arbiter once there is an active motion event. Only when this row is acknowledged can the column request be further transmitted to column arbiter. When both row and column are acknowledged, the firing unit is reset from the output of the NAND gate

M12-M15 and capacitors C4-C5 constitute the second-stage delta modulator, which amplifies the voltage variation on the coupling point. The output of this differencing circuit is connected to two compact comparators (M16-M19). Once the final motion exceeds a predefined "ON" or "OFF" threshold, active motion event is triggered and sent to the subsequent asynchronous handshaking circuit.

Fig.4 shows a 2-phase asynchronous handshaking communication circuit. When there is an effective state change from "ON" or "OFF", a motion pulse is generated at the output of the NOR gate, which triggers a row-request signal to the row AER circuit for arbitration. Although many row-request signals can be received simultaneously, only one of them can be acknowledged. Active *"M-Units"* within the acknowledged row will further send column request signals to the column arbiters. After both row and column are acknowledged, active "M-Unit" with firing motion events is reset for a new motion detection cycle from the output of the NAND gate.

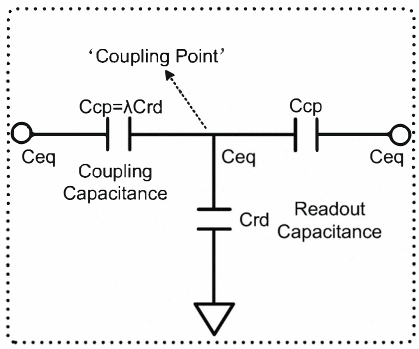

Fig. 5. One dimension capacitance coupling network on a single coupling point.

III. Analog Clustering

Each motion detection unit is cross-connected with its eight neighbors through a coupling capacitor network. The function of this capacitor network is to implement an analog clustering algorithm to reduce random noise. Fig.5 shows an extracted one-dimension network on the coupling point in one *"M-Unit"*. The readout capacitor C_{rd} refers to the inherent capacitance on each coupling point without considering the coupling capacitors, which is approximately equal to C3. For the simplicity, coupling capacitor C_{cp} can be represented as λC_{rd}, where λ is the ratio of the coupling capacitance C_{cp} to the readout capacitance C_{rd}. Since unit array is symmetrical and infinite, equivalent capacitor C_{eq} on each coupling point can be approximated as the following equation.

$$C_eq = C_{rd} + 8\frac{\lambda C_{rd} \times C_eq}{\lambda C_{rd} + C_eq}$$

The solution of this equation gives the final expression for the equivalent capacitance C_{eq}, as shown in the following.

$$C_eq = \frac{1 + 7\lambda + \sqrt{1 + 18\lambda + 49\lambda^2}}{2} C_{rd}$$

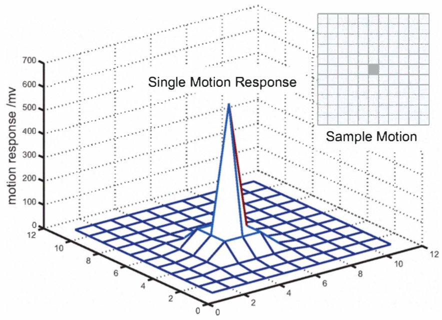

Fig. 6. Simulation response of a single motion event case. Motion image is shown in the corner, where center *"M-Unit"* detects an active motion. Voltage response exists a peak value in the center and decreases along with the distance to the firing cite

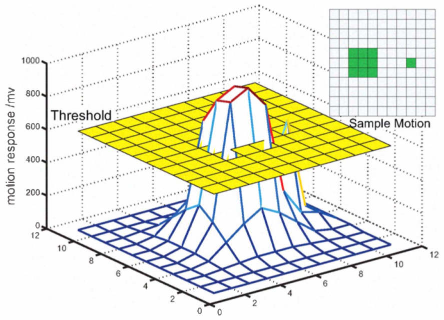

Fig. 7. Simulation response of a multiple motion events case. In the sample motion image, there are two types of motion events with different densities. It can be seen that the region with multiple motion events shows a larger response than that of the single motion event. With thresholding on the motion responses, single random noise can be removed

Due to the charge coupling effect, if there is a ΔV voltage change on the coupling point in one *"M-Unit"* caused by active motions, its N^{th} neighbor will have a smaller voltage response with a magnitude of ΔV_n as follows.

$$\Delta V_n \cong \left[\frac{2\lambda(C_{eq}+2\lambda C_{rd})C_{rd}}{(8\lambda+1)C_{rd}C_{eq}+\lambda(1+\lambda)C_{rd}^2}\right]^n \times \Delta V$$

It can be revealed that the farer distance from a given unit to the firing address, the smaller voltage response it will have. In this image sensor design, we select the coupling ratio λ as 0.5 by trading off the coupling performance with the silicon area consumption. Fig.6 shows the simulation response of a single motion event. In this example, the center *"M-Unit"* detects an active motion. Image sensor has a peak voltage response on the center while its neighboring units have smaller responses. In the multi-motions case as shown in Fig.7, there are two types of motions with different densities. Due to the inter-coupling effect, the region with multi-motions has a bell-shape larger voltage responses than the region with single motion event. By thresholding the final motion responses, single random noise can be effectively removed and only the region with multiple active motions from an moving object can survive. In this case, discrete spatial noise is efficiently removed by this analog clustering algorithm.

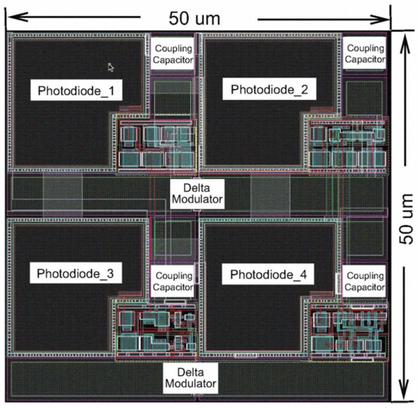

Fig. 8. Layout of the motion detection unit

IV. Hardware Implementation

The proposed vision sensor has been implemented using Global Foundry 0.18μm CMOS technology. Fig.8 shows the layout on the motion detection unit. Four photodiodes are uniformly distributed in 2×2 pixels. Biasing and readout circuits for the photoreceptor is arranged in pixel-1. Pixel-2 is dedicated to two delta modulators. Event generator and 2-phase asynchronous handshaking communication circuits are placed in pixel-3 and pixel-4. As to achieve the best matching performance, capacitors for coupling capacitor network and two delta modulators are equally divided in each pixel. The total area of this motion detection unit is around 50μm$\times$50μm. The equivalent fill factor for each pixel is about 42%.

V. Conclusion

In this paper, we present a 128$\times$128 pixel hybrid-readout and dynamic-resolution image sensor for moving object tracking application. This vision sensor can be configured as a motion detection sensor with a lower resolution, where relative intensity changes caused by motions are converted into binary events and asynchronously delivered to the outside with a short latency. When applied in a normal image acquisition mode, our image sensor can synchronously report an analog intensity image on the ROI or the entire focal plane with a higher resolution. The combination of a logarithmical photoreceptor and a coupling capacitor network feature this sensor with a high dynamic range and lower spatial noises.

References

[1] Y.-J. Wu, F.-J Lian, and T.-H, Chang, " Traffic monitoring and vehicle tracking using roadside cameras," IEEE *ICSMC*, 2006, pp.4631-4636.

[2] S.-S. Chen, W. Tang and E. Culurciello, "A 64x64 Pixels UWB Wireless Temporal-Difference Digital Image Sensor," IEEE *ISCAS*, 2010, pp.1404-1407,

[3] P. Lichtsteiner, C. Posch, T. Delbruck, "A 128 $\times$ 128 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," IEEE *JSSC*, 2008, pp.566-576.

[4] M. Litzenberger, C. Posch and etc., "Embedded vision system for real-time object tracking using an asynchronous transient vision sensor," IEEE *DSP Workshop*, 2006, pp.173-178.

[5] T. Delbruck, P. Lichtsteiner, "Fast sensory motor control based on event-based hybrid neuromorphic-procedural system," IEEE *ISCAS*, 2007, pp.845-848.