A tri-mode event-based vision sensor with an embedded wireless transmitter

Juan A. Leñero-Bardallo¹, Wei Tang², Dongsoo Kim³, Joon Hyuk Park⁴, and Eugenio Culurciello⁵

¹Nanoelectronics Group, Department of Informatics, University of Oslo, Norway
²Klipsch School of Electrical and Computer Engineering, New Mexico State University, USA
³Aptina Imaging Corporation
⁴Department of Electrical Engineering, Yale University, USA
⁵Weldon School of Biomedical Engineering, Purdue University, USA
E-mail: juanle@ifi.uio.no

Abstract—We report a vision sensor with three different operation modes (intensity, spatial, and temporal contrast), and an integrated UWB wireless transmitter. The sensor is intended for surveillance applications in Wireless Sensor Networks (WSN), where autonomous sensor nodes are required. It has event-based output, high dynamic range (120dB), and reduced power consumption (we expect less than 2.7mA for the sensor and 10mA for the wireless transmitter). Its pixels only use 5 transistors and have a size of $10 \times 10\,\mu m^2$, with fill factor of 47%. The sensor array has $256 \times 256$ pixels and was implemented in the standard IBM 8HP SiGe BiCMOS 0.13µm process. In this paper, we describe the system blocks. Simulation and preliminary experimental results are provided.

I. INTRODUCTION

There is a high demand of low-cost and compact vision systems with high dynamic range of operation and reduced output data flow. This demand is partially due to the development of WSNs (Wireless Sensor Networks). Such networks are used for plenty of tasks like surveillance, home-care, or industrial applications [1]–[3]. WSNs require smart autonomous devices with reduced power consumption and compressed data flow. These sensors should only send meaningful information to a central node when necessary. In that sense, devices that detect either spatial or the temporal contrast, trying to emulate the human retina, provide a very reduced data flow, preserving the relevant information to perform recognition. AER (Address Event Representation) sensors are quite suitable for WSNs because their outputs are typically voltage pulses (events) that can be easily conveyed into radio signals and transmitted wirelessly. Several relevant bio-inspired vision sensors with low power consumption and reduced data flow have been published [4]–[7], but none of them was specifically conceived to be a WSN autonomous node.

In this paper, we present an improved version of the tri-operation vision sensor presented by D. Kim et al. [8]. That sensor had three different operation modes (spatial, temporal and intensity mode), reduced power consumption, and good fill factor (42%). The new imager has an integrated wireless transmitter for its usage in WSNs, higher dynamic range (40dB higher employing dual rolling shutter and a floating diffusion storage capacitor [9]), smaller pixels with only 5 transistors, and higher resolution ($256 \times 256$).

The device was implemented in the IBM 8HP SiGe BiCMOS 0.13µm process. It is the first vision sensor with three operation modes and an integrated wireless transmitter. In the next sections, we will describe the main system circuit blocks. Some simulation and preliminary experimental results will be shown.

II. SYSTEM TOPOLOGY

Fig. 1 shows the main system blocks. The array has $256 \times 256$ pixels. The row control circuitry implements dual rolling electronic shutter for the row control signals activation. Contrast computation (either spatial or temporal) is done off-pixel in the column units. Each column read-out unit contains the circuitry to calculate the spatial or the temporal contrast, and also comparators to generate signed events. The sensor can detect the sign of the contrast and has two independent outputs for each sign ($ev^+$ and $ev^-$). These outputs are connected to two wireless transmitters that send out UWB pulses. A 256-bit pattern is always transmitted at the end of each frame to recover the clock synchronism and determine the end of each frame. Finally, a 14-bit counter has been added to the design. It counts the number of events of each frame. If the number of events is quite low (below 2.5% or 5% total number of
pixels) and there is no meaningful information to transmit, the wireless transmitters can be temporary disabled. Therefore, energy is saved and bandwidth consumption reduced. When the sensor generates frames with relevant information, the transmitter can be enabled again. In the two next sections, we will describe the circuitry of each main block and its functionalities.

A. Pixel Design

Fig. 2 shows the pixel topology of the sensor. Each pixel is made up by 5 transistors and 1 floating diffusion (FD) capacitor. Pixel size is $10 \times 10 \mu m^2$ with a fill factor of 47%. All the contrast computation is done off-pixel. The main novelty of the new pixel is the partial reset analog signal, $RPD$, that allows to increase the dynamic range of operation of the sensor during two integration times: $T_i$ and $T_s$. Fig. 3 depicts how the control signal are activated to perform the partial reset operation. Fig. 4 shows a transient simulation of the photo-diode parasitic capacitance voltage of several pixels exposed to very different levels of illumination.

The principle of operation is as follows: First, the photo diode capacitance is reset completely setting $RPD = 2.5V$. Then we set $RPD = 0V$ and charge is integrated during an integration time $T_i$. After, we partially reset the integrated charge activating the analog $RPD$ signal again. We set $RPD = 1.8V$ in the simulation. Then, we continue integrating charge during a shorter time interval, $T_s$. Finally, the signals, $RFD$, $SH$, and $ROWSEL$ are activated sequentially to reset the floating diffusion, transfer the charge to the floating diffusion, and read-out the integrated charge. Dynamic range of operation is extended with this operation. Without the partial reset, some pixels high illumination would provide the same output voltage close to $0V$. After the partial reset operation, their output voltages are within the interval $V_{DD}$ and $V_{GND}$ and, therefore, dynamic range is extended. Integrated charge of pixels with low illumination is not affected significantly by the partial reset (see Fig. 4). In the simulation, $T_i/T_s = 20$. Computing the ratio of the lowest and the highest photo current, that would saturate without the partial reset operation, the increase of the dynamic range was $24dB$. With this operation, a maximum increase of $40dB$ [9] is expected when $T_i/T_s = 100$.

B. Column Units and Operation Modes

This is the main operation module of the sensor. There are 256 column units (see Fig. 1). The column unit block diagram is depicted in Fig. 5. Each one is connected to the pixels of its row and the adjacent one. The signal $COLSEL$ is a token that activates the column under selection and then is transferred to the next column with shift registers. There are one WTA (Winner Takes All) and one LTA (Looser Takes All) block to calculate the spatial contrast [8]. A VGA (Variable Gain Amplifier) is employed to amplifier the spatial difference in voltage between 4 neighboring pixels or the temporal pixel voltage variations in two consecutive integration intervals. Finally, two comparators send out positive or negative events. The circuitry of the WTA and LTA blocks, and the VGA circuitry are shown in Fig 6(a) and Fig. 6(b) respectively. The three operation modes are described below:

**Intensity mode:** The output voltage is connected with a switch, activated with the signal $COLSEL$, to a global buffer that sends out the chip the read-out voltage. The global buffer can switched-off if the intensity mode is not activated.

**Temporal mode:** The difference between the pixel read-out voltage in two consecutive integration periods is amplified. The time-line of the control signals is displayed in Fig. 7(a). Control signals $SRST$ and $ST_1$ are involved. The read-out voltage value is stored in the floating diffusion. The temporal voltage difference between the stored voltage value and the
The latest read-out voltage is amplified with the VGA. This is proportional to the temporal contrast. Each row is selected and read-out twice in consecutive intervals.

**Spatial mode**: Spatial contrast is calculated with WTA and LTA operations amplifying the difference of the output voltage of 4 neighboring pixels as explained in [8]. The WTA and LTA blocks are used to amplify the difference between the read-out voltage of four adjacent pixels: \( P_{i,j} \), \( P_{i,j+1} \), \( P_{i+1,j} \), and \( P_{i+1,j+1} \). The control signals \( SC_1a, SC_2a, SC_1b, SC_2b \) are activated as is depicted on the time-line of Fig. 7. First, the voltage of the selected pixel and the voltage of its right neighbor are stored at the input capacitors of the WTA and LTA blocks (see Fig 6.(a)). Then, when the next row is selected, the voltages of the bottom pixels \( P_{i+1,j} \), and \( P_{i+1,j+1} \) are stored in the same way. Finally, the voltage difference of the WTA and LTA blocks is amplified with the VGA. There are two tokens signals to store alternately the values of the even and the odd rows.

The VGA outputs are connected to two comparators that send out positive or negative events when a positive, \( V_{high} \), and a negative threshold, \( V_{low} \) are exceeded. The VGA gain can be adjusted with 2 control bits that set the value of the capacitor \( C_2 \).

**C. Wireless Transmitter**

Two FSK-OOK transmitters [10] are embedded in the image sensor to wirelessly send out real-time events in temporal and spatial contrast modes. One transmits the positive event output \( eiv^+ \) and the other one transmits the negative event output \( eiv^- \). They can be independently disabled or switched off. Bits "1" and "0" are up-converted to 600MHz and 2GHz, respectively with ring oscillators. A time slot is inserted between two consecutive bits. The transmitter is shut down automatically when there is no data to send. The received binary data can be recovered with a non coherent FSK-OOK demodulator. The advantage of using FSK-OOK modulation is to achieve clock and data recovery automatically without synchronization. The wireless unit can transmit data asynchronously so the event rate can change during communication. A block diagram of the transmitter is shown in Fig. 8. The maximum output rate of this wireless system is 20Mbps, which is limited by multi path and environmental noise. Each integrated transmitter occupies
A new sensor targeted for Wireless Sensor Networks has been presented. It is the first vision sensor with an integrated a wireless transmitter and three different operation modes (intensity mode, and spatial and temporal contrast extraction). It was designed for surveillance applications. Its main features are: 256 × 256 pixels resolution, 10 × 10\(\mu\)m\(^2\) pixel size, fill factor of 47%, and low power consumption. Expected power consumption is 2.7\(\mu\)A, in spatial contrast mode, and 1.2\(\mu\)A, when computes the temporal contrast. In these two operation modes, a very reduced output event flow is transmitted to the processing nodes. Dynamic range of operation it is expected to be increased 40\(\text{dB}\) using partial reset operation during two integration periods. The wireless transmitter has a maximum output rate of 20\(\text{Meps}\) and a power consumption of 10\(\text{mA}\). The chip was fabricated in the standard IBM 8HP SiGe BiCMOS 0.13\(\mu\)m process. Some simulation and preliminary experimental results are provided.

### REFERENCES


### TABLE I

**SYSTEM SPECIFICATIONS**

<table>
<thead>
<tr>
<th>Array size</th>
<th>256 × 256</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>IBM 8HP SiGe BiCMOS 0.13(\mu)m</td>
</tr>
<tr>
<td>Power supply</td>
<td>2.5(\text{V})</td>
</tr>
<tr>
<td>Chip size</td>
<td>4 × 4.2(\text{mm})^2</td>
</tr>
<tr>
<td>Pixel size</td>
<td>10 × 10(\mu)m(^2)</td>
</tr>
<tr>
<td>Fill factor</td>
<td>47%</td>
</tr>
<tr>
<td>Power consumption (I,S,T modes)</td>
<td>650(\mu)A, 2.7(\mu)A, 1.2(\mu)A</td>
</tr>
<tr>
<td>Transmitter power consumption</td>
<td>10(\text{mA})</td>
</tr>
<tr>
<td>Frame rate (I,S,T modes)</td>
<td>150/800/200(\text{fps})</td>
</tr>
<tr>
<td>Maximum event rate</td>
<td>55(\text{Meps})</td>
</tr>
<tr>
<td>Max. wireless event rate</td>
<td>20(\text{Meps})</td>
</tr>
<tr>
<td>Dynamic range</td>
<td>120(\text{dB}), extension of 40(\text{dB})</td>
</tr>
</tbody>
</table>

---

**Fig. 9.** Chip microphotograph. Dimensions are 4 × 4.2\(\text{mm}\)^2. The main circuit blocks have been highlighted.

**Fig. 10.** On the left, experimental setup. On the right, dedicated C++ interface displaying an intensity image taken with indoor illumination of 400\(\text{lux}\).