A 1.6 mW 320x240-Pixel Vision Sensor with Programmable Dynamic Background Rejection and Motion Detection

Y. Zou, M. Gottardi, D. Perenzoni, M. Perenzoni, D. Stoppa

CMM, Fondazione Bruno Kessler, Trento, Italy.

Email:{zou, gottardi, perenzoni}@fbk.eu

Abstract—**This paper reports on a QVGA vision sensor embedding 160 column-level digital processors executing real-time tunable scene background subtraction for robust event detection. The single-ramp column-parallel ADCs are used to estimate the pixel variations and detecting anomalous behaviors against two reference images stored in on-chip. The sensor generates a 160x120 pixel bitmap associated to potential alert conditions. The chip is powered at 3.3V/1.2V for the analog/digital parts and consumes 1.6mW when operating at 15fps dispatching gray-scale image and a quarter QVGA bitmap.**

Keywords—vision sensors; low-power CMOS sensors; background subtraction; motion detection; VLSI image processing

I. INTRODUCTION

Commercial cameras are targeted to visual tasks where image quality and resolution are the most important features. However, in some applications such as surveillance and monitoring, they are not efficient since they force the processor to continuously analyze images, with a large waste of power. Embedding low-level image processing on-chip would make camera and system to be more energy-efficient. Following this approach, we present a QVGA vision sensor embedding a low-power background subtraction algorithm [1]. The sensor detects anomalous motion in the scene and generates an alert bitmap as input for high-level processing (e.g. tracking and classification) to be executed by the processor. Several implementations of on-chip motion detection have been proposed [2]-[4], which are based on frame difference technique. Although some of them can detect slow moving objects, they cannot suppress noisy zones of the scene, such as swaying vegetation or rippling water, which are not so uncommon in real scenarios. Differently from our previous fully analog implementation [5]-[6], we propose a digital approach allowing motion to be detected over a larger range and in harsh outdoor scenarios.

II. VLSI-ORIENTED ALGORITHM

The background is modeled with two thresholds [1], updated at each frame and stored into a frame buffer for subsequent operations. The embedded algorithm can be divided into two steps:

Learning step — two images are generated and updated at each frame: I_{MIN} (contains the minimum reference value for each pixel) and I_{MAX} (contains the maximum reference value for each pixel). For the generic *i*-th frame, the current value of each pixel

$P(x,y)$ is compared with its I_{MIN} and I_{MAX}; then the two reference images are updated as follows:

$$if\ I(x, y)_{MIN}\ (i\text{-}1) > P(x, y)\ then \qquad (1)$$

$$I(x, y)_{MIN}\ (i) = I(x, y)_{MIN}\ (i\text{-}1) - \Delta_{OPEN}$$

else

$$I(x, y)_{MIN}\ (i) = I(x, y)_{MIN}\ (i\text{-}1) + \Delta_{CLOSE}$$

end

$$if\ I(x, y)_{MAX}\ (i\text{-}1) < P(x, y)\ then \qquad (2)$$

$$I(x, y)_{MAX}\ (i) = I(x, y)_{MAX}\ (i\text{-}1) + \Delta_{OPEN}$$

else

$$I(x, y)_{MAX}(i) = I(x, y)_{MAX}\ (i\text{-}1) - \Delta_{CLOSE}$$

end

where Δ_{OPEN} and Δ_{CLOSE} ($\Delta_{OPEN} > \Delta_{CLOSE}$) are user-defined parameters used to update the two reference images in opening and closing conditions.

Detection step — it is used to detect if one pixel of the array is *"cold"* or *"hot"*, i.e. its behavior is normal or anomalous against its past history:

$$if\ (I(x, y)_{MIN}\ (i\text{-}1) - P(x, y)) > \Delta_{HOT}\ \ then \qquad (3)$$

$$H(x, y) = 1$$

else

$$H(x, y) = 0$$

end

$$if\ (P(x, y) - I(x, y)_{MAX}\ (i\text{-}1)) > \Delta_{HOT}\ \ \ then \qquad (4)$$

$$H(x, y) = 1$$

else

$$H(x, y) = 0$$

end

where $H(x, y)$ is the binary status (*hot-pixel*) of the pixel $P(x, y)$ and Δ_{HOT} sets the *hot-pixel* conditions. Fig. 1 shows how the
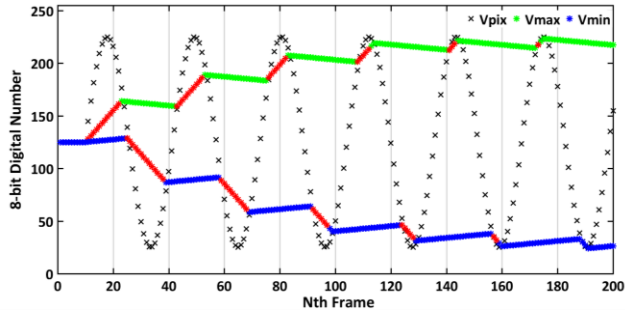
Fig. 1. Graphic representation of the pixel-level periodic background subtraction.

algorithm works when a pixel changes regularly (e.g. swaying vegetation). In this case, the two thresholds (Vmax, Vmin) track the current signal (Vpix) at different speeds, thus modifying the safe-zone (*cold-pixel*), while outside it, the pixel is a *hot-pixel* (red). From frame to frame, the two thresholds try to suppress the pixel by reaching the max and min peaks of Vpix. After about 170 frames the hot-pixel disappears and the oscillation is effectively registered as a background.

III. VISION SENSOR ARCHITECTURE

The rolling-shutter vision sensor consists of an array of 320×240 pixels, 320 single-ramp 4MHz 8-bit column ADCs, a bank of 160 processors that implements the row-wise algorithm updating the 10b reference images (I_{MIN}, I_{MAX}), stored into a 375Kb 6T-cell SRAM. At each row readout phase, the processors generate a 160-bit hot-pixel array, which is fed into the $160 \times (3 \times 3)$ pixel kernel bank of programmable Erosion Filters before to be delivered off-chip. A QQVGA hot-pixel bitmap is generated at the end of each frame according to (1)-(4).

A. Pixel Readout and A/D Conversion

The schematic of the 3T pixel column readout and A/D conversion is shown in Fig. 2. It is implemented with a folded-cascode amplifier, which is also re-used as voltage comparator for the single ramp ADC. The readout phase starts with the pixel voltage driving the bit-line (Vbl): its value is charged on C_1 ($S=H$) and then it is amplified with a gain of 2, ($C_1/C_2=2$) ($S=L$). Fixed pattern noise is compensated by subtracting the pixel reset voltage: the reset value is stored on C_1, with inverted polarity ($Phl=H$), and added to the signal on C_2. Each of the signal and reset sampling phases can be repeated several times by pulsing S, therefore increasing the overall gain and averaging the pixel
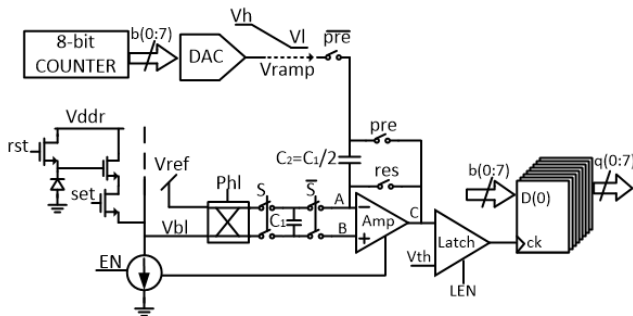

Fig. 2. Pixel readout and A/D conversion. The gain of the amplifier is programmable and is equal to 2N, where N is the number of pulses on S.
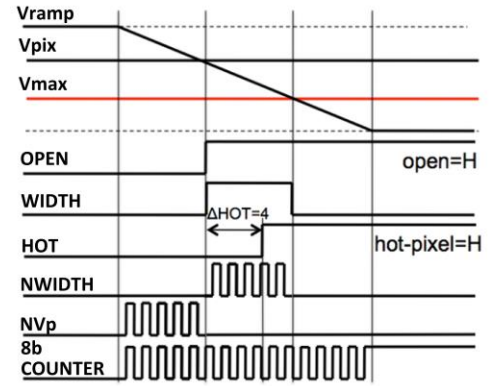

Fig. 3. Timing diagram of the pixel-level processing executed during the A/D conversion and implementing equations (2) and (4).

follower and the amplifier noise in a multiple-sampling operation. After the pixel has been read out and stored onto the feedback capacitor, the A/D conversion starts. Capacitor C_1 is disconnected from node A ($S=H$), while C_2 is connected to the DAC (*pre=L*), which provides the voltage ramp starting from V_h. This operation pulls-up the inverting node (A) of the amplifier, which is now in open-loop working as voltage comparator and forcing its output (C) to ground. The node A, connected to the global DAC through C_2, follows the decreasing voltage *Vramp* while the global counter is clocked. When the voltage on node A reaches *Vref*, the output of the amplifier switches toward Vdd and the 8-bit latch toggles, storing the value of the counter. The amplifier/comparator and ADC occupy a silicon area of 8 μm × 210 μm.

B. Column-Level Processor and SRAM

During the pixel A/D conversion, the voltage ramp is used to implement a portion of the algorithm, comparing the pixel against their thresholds (Vmax/Vmin) and checking for *opening/closing* and *hot-pixel* conditions (1)-(4). Fig. 3 shows an example of *hot-pixel* computed against the Vmax threshold. The comparison is made between Vpix (analog) and Vmax (digital) uploaded from the SRAM. Thus, an 8-bit digital comparator is used to compare Vmax against the digital code generating Vramp. Since Vpix is the first to be reached by the ramp (Vpix > Vmax), an *opening* condition is detected (OPEN=H), as in (2). In order to check if the pixel is *hot-pixel*, the difference (Vpix–Vmax) is compared with Δ_{HOT} (4). This is done with a digital counter measuring the time window WIDTH. If the counter reaches Δ_{HOT}, the pixel is a *hot-pixel* (HOT=H). Relying on OPEN and HOT signals, Vmax is increased by Δ_{OPEN}
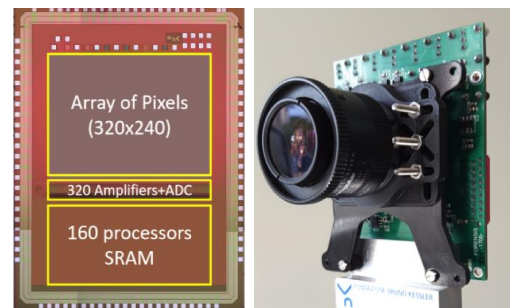

Fig. 4. Chip microphotograph together with sensor prototype.

Fig. 5. Example of the sensor operation. a) QVGA grayscale image; b) quarter QVGA hot-pixel bitmap after erosion.

and stored into the SRAM to be re-used next frame. Since the 8-bit DAC is driven with a 4 MHz clock, the voltage ramp takes 64 µs, while updating the thresholds and storing them into the SRAM takes 6µs. In order to have a precise control of the algorithm, the thresholds need to be updated with 10b resolution (0.25 LSB). The last block is the 240×160×10b, 6T-cell SRAM, storing I_{MIN} and I_{MAX}.

IV. EXPERIMENTAL RESULTS

Fig. 4 shows the microphotograph of the fully tested chip together with the sensor prototype controlled by an FPGA. A graphical user interface allows setting the sensor parameters: Δ_{OPEN}, Δ_{CLOSE}, Δ_{HOT} and the exposure time. Fig. 5 shows an example of an outdoor scenario with a moving boat. The algorithm neglects the background and clearly detects the moving boat suppressing the waves.

V. CONCLUSIONS

In this paper we presented a low-power QVGA vision sensor with programmable dynamic background subtraction. Experimental results show the capability of the sensor to robustly suppress the background (e.g. rippling water) while extracting salient moving features. The chip consumes 1.6mW while delivering QVGA gray-scale image and quarter QVGA bitmap at 15 fps. The main chip characteristics are listed in Table I.

TABLE I.

Main Chip Characteristics	Value
Technology	CMOS 0.11µm
Array Size	320 x 240 (QVGA)
Pixel pitch	8 µm
Fill Factor	67%
Supply Voltage (a, d)	3.3V , 1.2V
SNR	43.6dB
DR	53.5dB
FPN	0.9%
Frame Rate	15 fps
Power Consumption	1.6 mW
FOM	1.4 nW/pix*frame
Chip Size	3.6 mm x 4.7 mm

REFERENCES

[1] Z. Smilansky, "Miniature Autonomous Agents for Scene Interpretation," 7,489,802 B2, US Patent Application, Feb. 2009.

[2] T. Ohmaru, et al., "25.3 µW at 60fps 240x160-Pixel Vision Sensor for Motion Capturing with In-Pixel Non-Volatile Analog memory Using Crystalline Oxide Semiconductor FET," IEEE ISSCC Dig. Tech. Papers, Feb. 2015, vol. 58, pp. 118–119.

[3] J. Choi, S. Park, J. Cho, and E. Yoon, "A 3.4-w object-adaptive cmos image sensor with embedded feature extraction algorithm for motion triggered object-of-interest imaging," IEEE J. Solid-State Circuits, vol. 49, no. 1, pp. 289–300, 2014.

[4] G. Kim, M. Barangi, Z. Foo, N. Pinckney, S. Bang, D. Blaauw, and D. Sylvester, "A 467nW CMOS visual motion sensor with temporal averaging and pixel aggregation," IEEE ISSCC Dig. Tech.Papers, Feb. 2013, vol. 56, pp. 480–481.

[5] N. Cottini, M. Gottardi, N. Massari, R. Passerone, "A Bio-Inspired APS for Selective Visual Attention," IEEE Sensors Journal, vol. 13, no. 9, pp. 3341-3342, 2013.

[6] N. Cottini, M. Gottardi, N. Massari, R. Passerone, and Z. Smilansky, "A 33µW 64×64 pixel vision sensor embedding robust dynamic background subtraction for event detection and scene interpretation," IEEE J. Solid State Circuits, vol. 48, no. 3, pp. 850–863, 2013.