1st International Conference on Sensing Technology
November 21-23, 2005 Palmerston North, New Zealand

A CMOS Vision Sensor with Integrated Analog Processing

Nicola Massari, Massimo Gottardi, and Andrea Simoni
ITC-irst
via Sommarive 18, 38050 Povo, Trento, Italy
gottardi@itc.it

Abstract

A 128x64 pixels vision sensor implementing analog image processing is presented. The imager is based on a pixel-parallel architecture, executing a wide range of real-time image filtering over a kernel of 3x3 pixels. The filter's coefficients can be assigned with a large flexibility adopting two different techniques: time-based and switched-based technique. The imager's square pixel has a pitch of 32.6μm with a fill-factor of 24% and consists of two analog memories and 30 transistors, most of them dedicated to the pixel connectivity. The chip was fabricated in a CMOS 0.35μm technology; it performs a FPN of 0.7%, a total dynamic range of about 99dB and a power consumption of 14mW @ 3.3V.

Keywords: Focal-plane processing, analog image processing, vision chip, CMOS image sensors.

1 Introduction

Compact vision systems are of great interest in all those applications where low power fast features extraction are required even under harsh illuminating conditions as in surveillance, automotive and machine vision. Many architectures for real-time image processing have been proposed in the last decade. Some are related to very specialized architectures, turning into a very large pixel size [1-2], others integrate only one or few simple image processing functions with very little flexibility [3-9]. A different approach adopts analog or mixed-signal programmable processor array even though most of them occupy a considerable chip area [10-12]. The proposed architecture represents a good tradeoff between pixel dimensions and flexibility of image filtering. The imager uses a pixel-parallel architecture integrating a wide class of spatio-temporal image filters with high flexibility in the assignment of the coefficient's values by simply changing the sensor's timing diagram. The embedded image processing stems from two main analog functions which are the base for the most frequently used image processing algorithms: absolute value of the difference and accumulation of the partial results provided by the previous function.

2 Sensor Description

The sensor is organized in an array of 128x64 pixels, each one of them is connected to the 8 neighbours. A row-decoder scans the imager rows during readout, and a readout block comprising a column decoder together with a video amplifier. As shown in the schematic of figure 1, the core of the pixel consists of a charge amplifier accumulating negative voltage variations applied to the input V_{in} through the 9 MOS switches connected to C1.

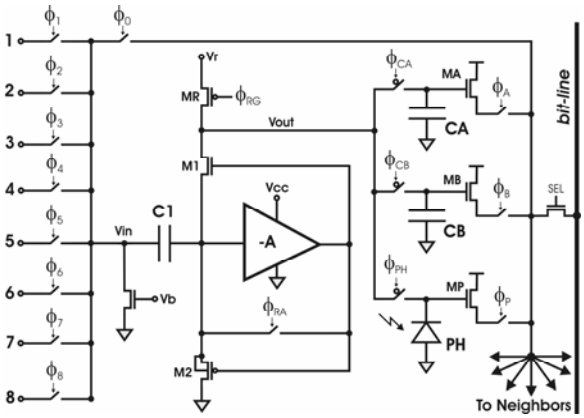

Figure1: Schematic of the pixel.

The amplifier A consists of an inverter, biased in the subthreashold region with 100nA (V_{CC}≈1.3V), which exhibits a dc gain of about 60dB. The amplifier's feedback is provided by two source followers (M1 and M2). In this way, the output voltage (V_{OUT}) is decoupled from the amplifier (A) which has to provide an output voltage range of only $2V_{GS}$ in order to guarantee the feedback.

Two analog memories (CA, CB) are integrated, storing the amplifier's output voltage and feeding it back to the input, in order to re-process the partial results. The same type of feedback is implemented for the photodiode. The voltage stored into memory CA and CB and are also available as input for the 8 neighbouring pixels, as well as the photodiode's voltage. Transistor MR serves as global reset for memories and photodiode, while the amplifier (A) can

be inhibited by means of Φ_{RA}. A bank of 9 switches (Φ_i , i= 0,1, ...,8) implements the connectivity over a 3x3 pixel kernel.

3 Operating principle

In order to better explain the sensor's functionality, a brief description of the pixel's operating mode under normal image acquisition is described.

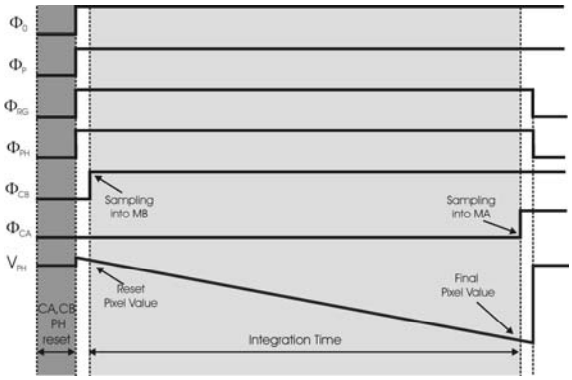

Figure 2: Timing diagram for camera mode operation.

The image acquisition process consists of a photodiode and memories reset, an integration time and a readout phase (see figure 2). After the reset, transistor MR is turned off and the reset voltage is stored on CB. The photodiode starts integrating light and its signal is accumulated onto CA, turning on the three switches: Φ_P, Φ_0, and Φ_{CA}. During the readout phase the pixel is selected (SEL), and the positive voltage difference between CB and the integrated signal (CA) is applied to the bit-line, switching from Φ_A to Φ_B. The readout procedure also implements a pixel-level CDS. Figure 4 shows an example of a real image where the integration time is set turning off Φ_{CA} before resetting the photodiode (Φ_{RG} and Φ_{PH}) in order to avoid saturation.

3.1 Spatial Filtering

3.1.1 Time-based coefficients

The novelty of the proposed imager's architecture consists of the possibility to execute the accumulation of signals, related to pixels arbitrarily chosen within the 3x3 kernel, during the integration time and not after the image acquisition, as in a conventional approach. Image smoothing is probably the most clear example describing this *on the fly* image processing performed by the sensor. In a typical mean filter, the signals of the 9 pixels of the kernel are averaged using the following algorithm:

$$V_{mean} = \frac{1}{9}\sum_{i=1}^{9} c_i V_{pi} \; ; \qquad c_i=1, \; for \; i=1,\ldots,9. \qquad (1)$$

The proposed sensor can execute equation (1) during the integration time T_0, as shown in figure 3, which takes into account the simple case of three pixels.

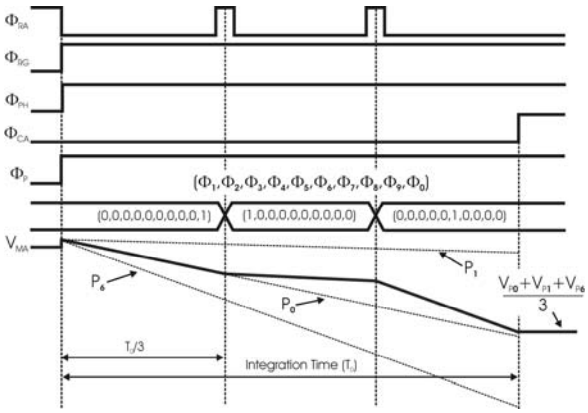

Figure 3: Timing diagram for an average filter over 3 pixels.

Let suppose the total integration time T_0 to be divided into a proper number of time base $\Delta T=(T_i-T_{i-1})$. During image acquisition, the central pixel (P_0) integrates light on CA for $\Delta T=1/3T_0$, connecting the photodiode to the amplifier's input (Φ_p and Φ_0). Then, the amplifier (A) is forced under reset (Φ_{RA}) while its input is switched to the pixel P_1 of the kernel. This prevents undesirable signal transitions which could destroy the sensor's functionality and forces the amplifier to start working always from the same bias conditions. After the amplifier's reset, the signal of the current pixel is accumulated to the previous signal, stored on CA. A third accumulation is then executed selecting, by the bank of switches, the pixel P6 and integrating the light for the same ΔT, obtaining thus $V=(V_{P0}+V_{P1}+V_{P6})/3$.

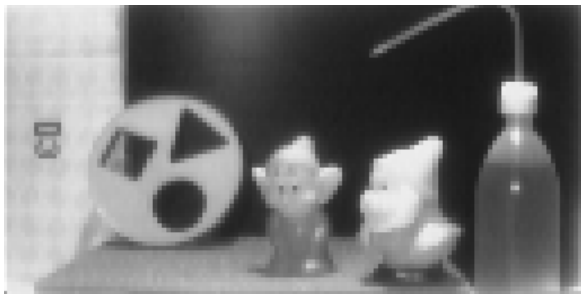

Figure 4: Example of real image.

In general, in order to achieve equation (1), integration time T_0 has to be devided into nine time slot $\Delta T=1/9T_0$, as the total number of pixels belonging to the kernel. The signal sequence need to execute a single accumulation, described above, is then executed selecting each pixel of the kernel, implementing the image mean filter. Other types of spatial filters require different values of the coefficients associated to each pixel of the kernel. In this case, the following condition has to be satisfied which guarantees the filter's normalization:

$$T_0 = \Delta T \sum_{i=1}^{N} c_i \qquad (2)$$

where *N* represents the number of pixels involved in the process. Two examples of real images, acquired with normal camera operation and with smoothing filter respectively, are shown in figure 4 and in figure 5(a) respectively.

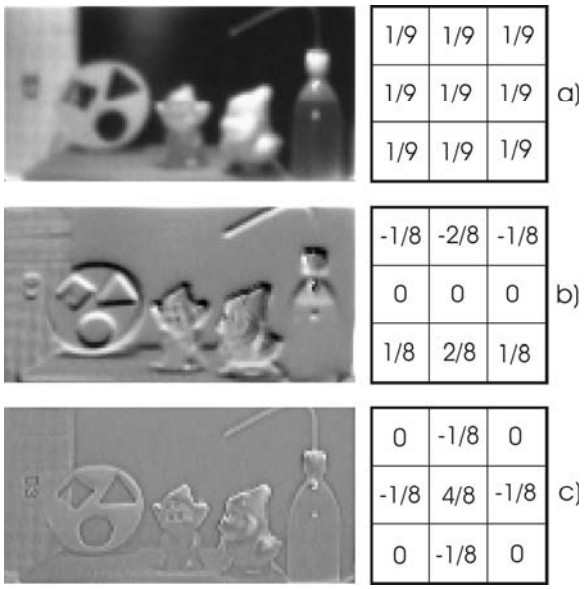

Figure 5: Examples of spatial filtering with time-based coefficient assignment: a) average filter; b) Sobel filter; c) Laplacian filter.

More in general, the type of spatial filters implementable during the acquisition time, can be expressed as follows:

$$V_{filt} = \frac{\Delta T}{T_0}\left(\sum_i c_i V_{pi} - \sum_j c_j V_{pj}\right). \qquad (3)$$

The assignment of the integer coefficient c_i and c_j is implemented associating a proper integration time (multiple of ΔT) to each pixel involved in the process. The execution of equation (3) involves the use of the two embedded memories (CA, CB), storing the two terms on the right side of (3). The subtraction of the two terms can be done during the imager's readout phase by means of the column readout circuitry. A Sobel filter along *y* direction, shown in figure 5(b), is implemented by averaging the first row of the kernel onto CA and the third row onto CB, as described in equation (1). Using the same procedure, a laplacian filter was implemented, storing the central pixel on CA and the four neighbours onto CB, as demonstrated in figure 5(c).

3.1.2 Switched-based coefficients

The second type of assignment of filter's coefficients consists of switching the amplifier's input between the reference pixel and an arbitrary pixel of the kernel, using the bank of switches shown in figure 1, after the integration time. In this way the absolute value of difference between two pixels is implemented.

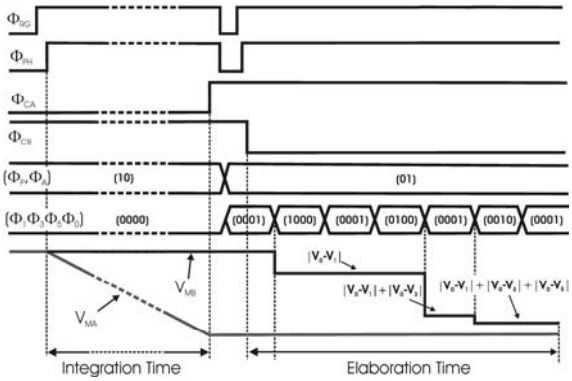

Figure 6: Timing diagram of switched-based approach referred to three pixel.

Because only the negative voltage variation is accumulated onto one of the two capacitors (CA or CB), the input has to be switched from the reference pixel (P_0) toward the i-th pixel and back again in order to guarantee the proper negative voltage transition. The above operation can be executed for each of the 9 pixels of the kernel accumulating the partial results onto the same storing capacitor. Figure 6 shows an example of timing diagram applied to three pixels of the kernel. At the end of the integration time, each pixel of the array samples the integrated signal into the memory CA (Φ_{CA}). After image acquisition, transistor MA, means Φ_A, provides this voltage value to all other pixels of the kernel. Selecting the involved pixels with respect to the central one (P_0), using the bank of switches, at the input of the amplifier (A), a sum of absolute value of difference among the pixels is accumulated in memory CB previously reset.

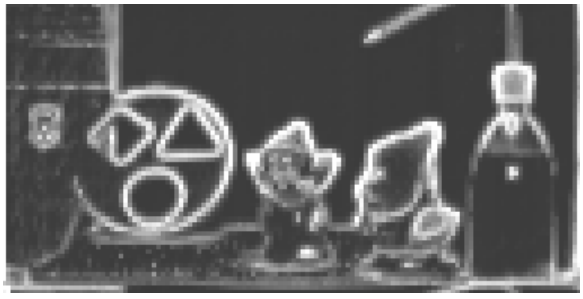

Figure 7: Full kernel edge extraction.

A gradient can be arbitrarily wheighted, switching the amplifier's input back and forth for a proper number of times between the two involved pixels. The general formula which represents the gradient extraction is:

$$V_G = \sum_i c_i \left| V_{p0} - V_{pi} \right|, \qquad (4)$$

where p_0 is the reference pixel and p_i is one of the 8 neighboring pixels of the kernel. Coefficient c_i is the

integer, representing the number of times the amplifier's input is switched between the two pixels.

Directional local gradient or full kernel gradient, as shown in figure 7, can be computed by simply changing the switching diagram of the pixel.

3.2 Time Filtering

3.2.1 Motion detection

Motion detection is implemented storing two successive images into memories CA and CB respectively [7] (see figure 8).

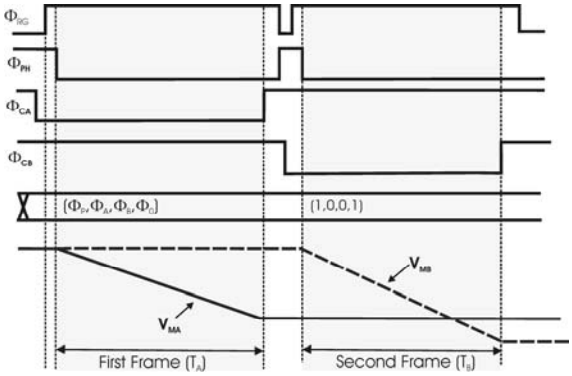

Figure 8: Timing diagram for motion detection.

The frame difference is executed during the imager readout. An example of signed motion detection is shown in figure 9 where, due to some offset between the two pixel's memories, some small shodows are present in the resulting image even though they are not clearly visible in the picture.

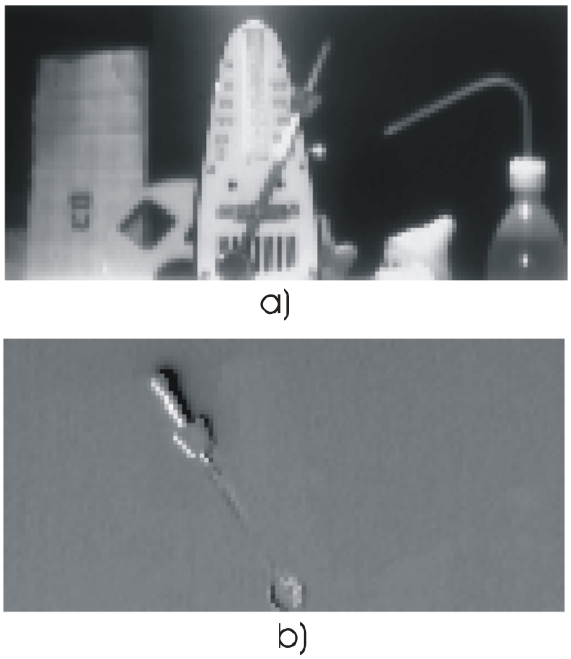

Figure 9: Motion detection: a) real image; b) motion detection.

3.2.2 High dynamic range

A simple technique used for boosting an imager's dynamic range combines the images taken under different exposure conditions [13]. A first image is taken with a long integration time T_1. Here, the pixel's voltage is divided by a coefficient k>1 (typically k=2), preventing signal from saturation for high incident light. A second frame is acquired with a short time (T_2) and accumulated to the first one. The final result is:

$$V_0 = V_{T_2} + \frac{V_{T_1}}{k}; \qquad (5)$$

which performs a dynamic range extension of $T_1/(2 \cdot T_2)$.

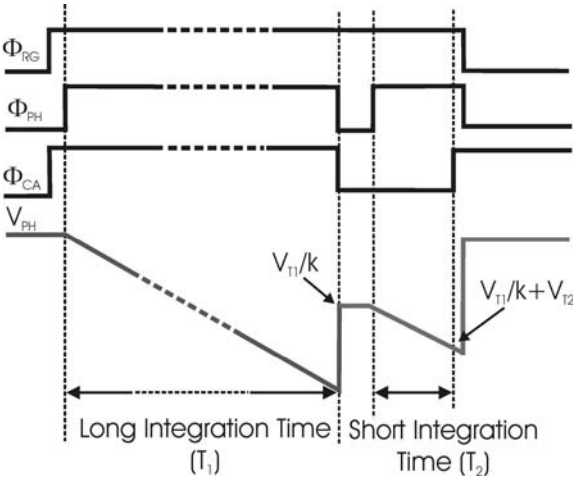

Figure 10: Timing diagram for dynamic range boosting.

This operation, described in figure 10, can be easily executed integrating the first frame for T_1 and then shorting the capacitor CA (CB), precharged to the reset voltage, with the photodiode in order to implement the dividing coefficient. After this, the photodiode integrates the second frame for T_2. Figure 11 shows an example of a real image, acquired with high dynamic range setting T_1/T_2= 500, which means a dynamic range extension of about 48dB. It has to be observed that the dividing coefficient is not precise, because photodiode and storing capacitors cannot be accurately matched.

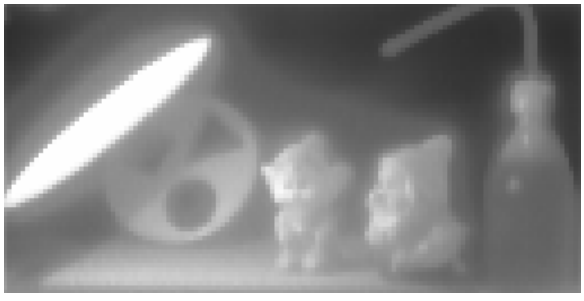

Figure 11: Image acquired in high dynamic range mode.

4 Experimental results

Table 1 summarizes the characteristics and performance of the device. The chip microphotograph is shown in figure 12. The chip has been integrated in a 0.35μm process DP TM CMOS process featuring a square pixel size of 32.6μm with a fill factor of 24 %. The maximum power consumption is 14 mW @ 3.3V. The die size is 8.7 mm^2.

Table 1: Chip Specifications

PARAMETER	VALUE
Technology	AMS CMOS 0.35μm DP TM
Array Size	128x64 pixels
Chip Size	20mm^2
Pixel Size	32.6μmx32.6μm
Fill Factor	24%
Power Consumption	14mW @ 3.3V
Saturation Level	1.8V
Dynamic Range	51dB
Dynamic Range Boosting	~48dB
Total Dynamic Range	99dB
Fixed Pattern Noise	0.7%

5 Conclusions

A 128x64 pixels image sensor with parallel-pixel image processing architecture was presented and fully tested. A large class of real-time spatio-temporal filtering can be executed taking advantage of two simple functions: absolute value of the signal difference and signal accumulation. A large flexibility in the coefficient's assignment is possible using both temporal-based and accumulation-based techniques, covering the requirements of many real-time applications.

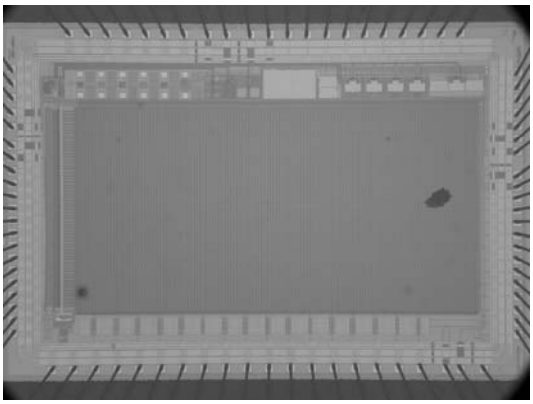

Figure 12: Chip microphotograph.

6 References

[1] M. Barbaro, P.-Y. Burgi, A. Mortara, P. Nussbaum, and F. Heitger, "A 100x100 Pixel Silicon Retina for Gradient Extraction With Steering Filter Capabilities and Temporal Output Coding", *IEEE J. of Solid-State Circuits*, vol. 37, pp. 160-172 (2002).

[2] P.-F.Ruedi, P.Heim, F.Kaess, E.Grenet, F.Heitger, P.-Y.Burgi, S.Gyger,and P.Nussbaum, "A 128 x 128 Pixel 120-dB Dynamic-Range Vision-Sensor Chip for Image Contract and Orientation Extraction", *IEEE J. of Solid-State Circuits*, vol. 38, No. 12, pp. 2325-2333 (2003).

[3] A. Dickinson, B. Ackland, E-S. Eid, D. Inglis, and E. R. Fossum, "256 x 256 CMOS active pixel image sensor with motion detection", *ISSCC Dig. Tech. Papers 1996*, pp. 178-179 (1996).

[4] H. Nomura, T. Shima, A. Kamashita, and T. Ishida, "A 256 x 256 BCAST motion detector with simultaneous video output", *ISSCC Dig. Tech. Papers 1998*, pp. 282-283 (1998).

[5] H. Miura, H. Ishiwata, Y. Iida, and Y. Matsunaga, "A 100 frames/s CMOS active pixel sensor for 3D-gesture recognition system", *ISSCC Dig. Tech. Papers 1999*, pp. 142-143 (1999).

[6] V. Gruev and R. E. Cummings, "A Pipelined Temporal Difference Imager", *IEEE J. of Solid-State Circuits*, vol. 39, No. 3, pp. 538-543, (2004).

[7] S.-Y. Ma, and L.-G. Chen, "A Single-Chip CMOS APS Camera with Direct Frame Difference Output", *IEEE J. of Solid-State Circuits*, vol. 34, pp. 1415-1418 (1999).

[8] C.-Y. Wu, and C.-F. Chiu, "A New structure of the 2-D Silicon Retina", *IEEE J. of Solid-State Circuits*, vol. 30, No. 8, pp. 890-897 (1995).

[9] D. Stoppa, A. Simoni, D. Gonzo, M. Gottardi, and G. F. Dalla Betta, "A 138dB dynamic range CMOS with new pixel architecture", *ISSCC 2002*, vol. 1, pp. 40-441 (2002).

[10] R. E. Cummings, Z. K. Kalayjian, and D. Cai "A Programmable Focal-Plane MIMD Image Processor Chip", *IEEE J. of Solid-State Circuits*, vol. 36, pp. 64-73 (2001).

[11] V. Gruev and R. E. Cummings, "Implementation of Steerable Spatiotemporal Image Filters on the Focal Plane", *IEEE Trans. Circuits and Systems*, vol. 49, No. 4, pp. 233-244 (2002).

[12] C. S. Hong, R. Hornsey, "On-Chip Binary Image Processing with CMOS Image Sensors", *Proc. of SPIE*, vol. 4669, pp. 125-136 (2002).

[13] K. Yamada, T. Nakano, and S. Yamamoto, "A Vision Sensor Having an Expanded Dynamic Range for Autonomous Vehicles", *IEEE Trans. on Vehicular Tech*, vol. 47, No. 1, pp. 332-341 (1998).