

# Low-Power CMOS Vision Sensor for Gaussian Pyramid Extraction

Manuel Suárez, Víctor Manuel Brea, Jorge Fernández-Berni, Ricardo Carmona-Galán, *Senior Member, IEEE*, Diego Cabello, and Ángel Rodríguez-Vázquez, *Fellow, IEEE*

**Abstract**—This paper introduces a CMOS vision sensor chip in a standard  $0.18\text{ }\mu\text{m}$  CMOS technology for Gaussian pyramid extraction. The Gaussian pyramid provides computer vision algorithms with scale invariance, which permits having the same response regardless of the distance of the scene to the camera. The chip comprises  $176 \times 120$  photosensors arranged into  $88 \times 60$  processing elements (PEs). The Gaussian pyramid is generated with a double-Euler switched capacitor (SC) network. Every PE comprises four photodiodes, one 8 b single-slope analog-to-digital converter, one correlated double sampling circuit, and four state capacitors with their corresponding switches to implement the double-Euler SC network. Every PE occupies  $44 \times 44\text{ }\mu\text{m}^2$ . Measurements from the chip are presented to assess the accuracy of the generated Gaussian pyramid for visual tracking applications. Error levels are below 2% full-scale output, thus making the chip feasible for these applications. Also, energy cost is  $26.5\text{ nJ}/\text{px}$  at  $2.64\text{ Mpx/s}$ , thus outperforming conventional solutions of imager plus microprocessor unit.

**Index Terms**—CMOS vision sensors (CVISs), Gaussian filters, image pyramids, per-pixel processing, switched capacitor (SC) circuits.

## I. INTRODUCTION

THE integration of camera systems for vision applications benefits from performing scene analysis right at the sensor frontend. Such preprocessing may extract scene features and hence reduce the number of data transmitted off the sensor chip for further processing. This is quite a relevant characteristic because images contain many spare

Manuscript received March 11, 2016; revised July 29, 2016 and September 12, 2016; accepted September 13, 2016. Date of publication October 27, 2016; date of current version January 30, 2017. This paper was approved by Associate Editor Hideto Hidaka. This work was supported in part by ONR under Grant N00014-14-1-0355, in part by the Spanish Government Projects MINECO under Grant TEC2015-66878-C3-1-R and Grant TEC2015-66878-C3-3-R, in part by the Junta de Andalucía, Proyectos Excelencia-Conv. 2012 under Grant TIC 2338, and in part by FEDER under Grant EM2013/038 and Grant EM2014/012.

M. Suárez is with Atomos GmbH, Villingen-Schwenningen 78048, Germany (e-mail: manuel@atomos.com).

V. M. Brea and D. Cabello are with the Centro Singular de Investigación en Tecnologías de la Información, University of Santiago de Compostela, Rua Jenaro de la Fuente 15782, Spain (e-mail: victor.brea@usc.es).

J. Fernández-Berni and A. Rodríguez-Vázquez are with the University of Seville, Spain, and also with the Instituto de Microelectrónica de Sevilla, National Microelectronics Center, Consejo Superior de Investigaciones Científicas, Spain (e-mail: angel@imse-cnm.csic.es).

R. Carmona-Galán is with the Instituto de Microelectrónica de Sevilla (IMSE-CNM), Consejo Superior de Investigaciones Científicas- Universidad de Sevilla, Calle Américo Vespuco 41092, Spain.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSSC.2016.2610580

data, and data transmission and storage consume significant energy and area. Also, preprocessing and reduced data transmission result in increased throughput. Actually, preprocessing is smartly implemented in natural vision systems [1], [2], a fact that has motivated Lee and Hsieh [3], Fernández-Berni *et al.* [4], Carey *et al.* [5], Park *et al.* [6], and Rodríguez-Vázquez *et al.* [7] to explore architectures for CMOS imaging frontends with per-pixel processing circuitry. These systems are recently making the transition from academic proof-of-concept prototypes to industrial products [8].

Sensory-processing front-end chips with per-pixel processors typically operate as single instruction multiple data (SIMD) processors, namely, all processors run concurrently the same operation on the data captured by the pixel photosensors, thus accelerating computation. Also, mixed-signal per-pixel processors provide speed advantages with large energy efficiency [9], [10]. As a result, image sensors with embedded mixed-signal processors emerge as suitable candidates for the frontend of vision systems with optimum size, weight, and power (SWaP) figures and large throughput. Throughout this paper, we will use the term CMOS vision sensors (CVISs) to refer to image front-end devices with embedded analysis capability, and we will retain the term CMOS image sensors (CISs) for conventional image frontends conceived to deliver just images.

Major points hampering further development of the CVIS-SIMD are as follows.

- 1) Their outcome may not be compatible with computer vision software tools, thus limiting their acceptance by system engineers and integrators.
- 2) Reduced fill factor when realized in standard 2-D technologies.
- 3) Large pitch and hence smaller resolution than the CIS per the given form factor, again in standard 2-D technologies.

Nevertheless, the loss of resolution and the image quality of CVIS-SIMD are not insurmountable barriers for vision. Nature also teaches lessons in this regard, for instance, patients with retinitis pigmentosa see with a small fraction of their photoreceptors alive [11], which suggests that large pixel counts may not be a must. Indeed, resolutions as low as  $32 \times 32$  px suffice to get the gist of complex scenes [12] and have been demonstrated for indoor elderly care [13]. Also, commercial sensors with low pixel counts (QCIF:  $176 \times 144$ ) are produced for machine vision applications [14] and have

been demonstrated for adaptive laser welding [15], among other applications. Also, reduced fill factor may be overcome with controlled illumination, as it actually happens in many machine vision applications [16]. Furthermore, many computer vision algorithms cope with inaccuracies arisen during processing [17], [18], thus easing the use of mixed-signal CVIS-SIMD. As an example, the chip in [19], which runs the earliest stages for face detection using the algorithm in [17], tolerates processing errors close to 10%. As shown in Section IV-D, the chip measurements in this paper show that inaccuracies in the Gaussian pyramid are low enough as not to be a concern for visual tracking.

Regarding compatibility with computer vision tools, it can be met by aligning the conception of CVIS-SIMD to standard computer vision procedures [20], particularly by focusing on the embedding of preprocessing functions customarily used by computer vision system engineers. This is actually the case of image pyramids, such as the Gaussian pyramid [21]. Image pyramids are found at the initial stages of the processing vision chain for a large variety of computer vision applications and algorithms such as the scale invariant feature transform (SIFT) and variations thereof. Their calculation is resource intensive because it involves repetitive operations with the whole set of image data. As a consequence, the potential benefit of calculating them with the CVIS-SIMD is huge. The CVIS-SIMD may represent a first step toward embedding complete computer vision on a single die with vision capabilities into SWaP-sensitive systems, such as vision-enabled wireless sensor networks [22] and unmanned aerial vehicles [23].

From now on, we will use the term processing element (PE) for the elementary cell of CVIS-SIMD image front-end chips. This paper reports a  $0.18 \mu\text{m}$  CMOS sensory processing chip to extract the Gaussian pyramid with per-pixel processing circuitry, the ADC, and correlated double sampling (CDS). It contains  $176 \times 120$  3T APSS arranged into  $88 \times 60$  PEs, i.e., four photosensing points per PE. Gaussian filtering is realized using a diffusive, double-Euler, and switched capacitor (SC) grid. The chip operates at  $2.64 \text{ Mpx/s}$  with an energy consumption of  $26.5 \text{ nJ/px}$  ( $0.6 \mu\text{J/frame}$ ), thus outperforming conventional architectures of imager and the MPU by several orders of magnitude. Measurements show errors below 2% full-scale output (FSO) versus Gaussian pyramid computed by software [24]; these errors are tolerated by vision applications.

## II. GAUSSIAN PYRAMID EXTRACTION

### A. Basic Concepts

The scale-space enables computer vision algorithms to give the same response regardless of the distance between camera and object. A common function for scale-space generation is the Gaussian filter [25], [26]. The scale space is a function  $L(x, y, \sigma)$  resultant from the convolution of a variable-width Gaussian function with an input image  $I(x, y)$

$$L(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} * I(x, y) \quad (1)$$

where  $*$  is the convolution operator,  $\sigma$  is the width of the Gaussian function, and  $x$  and  $y$  are the spatial coordinates of the image.

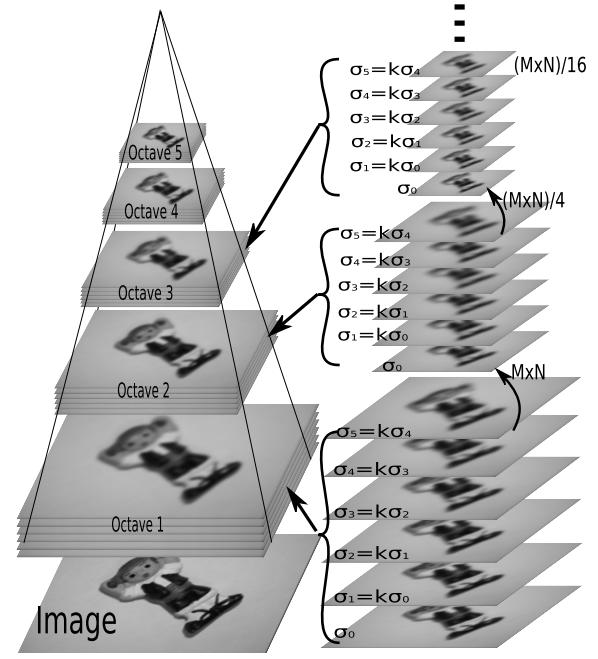


Fig. 1. Scale space through the Gaussian pyramid with octaves and scales. Each octave has one quarter the spatial resolution of the previous one, starting from the bottom. Thus, if the initial image has  $M \times N$  px, images in the second octave have  $(M \times N)/4$  and so forth.

The Gaussian pyramid illustrated in Fig. 1 consists of several scale spaces arranged into octaves. Starting from the bottom of Fig. 1, images within each new octave have all one quarter the resolution of those in the previous octave. Subsampling is hence made in the transition from each octave to the next one. Regarding images contained within each octave, these images are scales obtained through Gaussian filtering with increasing widths. The width of each new scale is  $k$  times larger than that of the previous one. The range of scale widths is the same for all octaves, namely, from  $\sigma_0$  to  $2\sigma_0$ . The width  $\sigma_0$  is application dependent, and as such, it could be selected by the user. Usually, three octaves with six scales each suffice [21]. At hardware level, the issue is to provide accurate widths  $\sigma_i$  of the Gaussian function.

### B. Hardware Implementation

The Gaussian function gives the value  $I_{ij}$  of each pixel as the solution of a first-order differential equation under the driving force of the values of the four neighboring pixels along the cardinal directions, namely

$$\frac{dI_{i,j}}{dt} = D(I_{i+1,j} + I_{i-1,j} + I_{i,j+1} + I_{i,j-1} - 4I_{i,j}) \quad (2)$$

which is actually the continuous-time heat differential equation [27], with  $D$  being the diffusion coefficient, usually a constant value common to all the pixels in the image space. In the case of the Gaussian pyramid,  $D$  determines the degree of blurring through the expression  $\sigma = \sqrt{2Dt}$ , where variable  $t$  is the time. In our case, pixel values are voltages  $V_{ij}$  held at state capacitors of capacitance  $C$  and pixels are connected to the four neighbors through resistive links with resistance  $R$ .

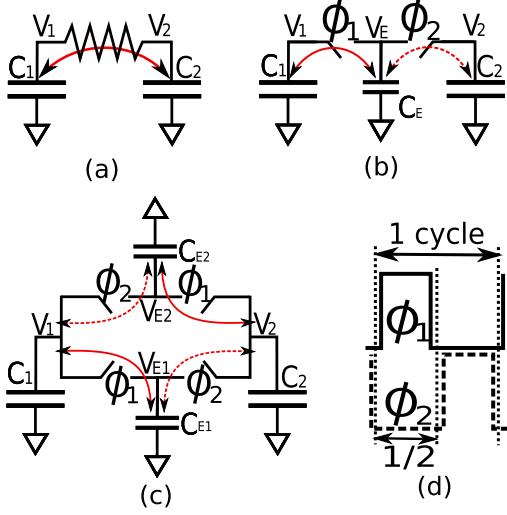


Fig. 2. Topologies for Gaussian filtering in 1-D. (a) *RC* network. (b) and (c) Simple and double-Euler SC networks, respectively. (d) Nonoverlapping control signals for SC networks.

In such a case, (2) transforms into

$$C \frac{dV_{i,j}}{dt} = \frac{V_{i+1,j} + V_{i-1,j} + V_{i,j+1} + V_{i,j-1} - 4V_{i,j}}{R} \quad (3)$$

from where  $D = 1/RC$  and the filter width  $\sigma_{RC} = \sqrt{2t/RC}$ . Resistance  $R$  can be implemented either through TMOS transistors operating in ohmic region, giving rise to *RC* networks, or through SC networks. Fig. 2 illustrates both implementation styles. The former are inherently more nonlinear than the latter. Also, *RC* networks need sampling mechanisms to stop the transient evolution of the network and thereby set the width [4]. The nonlinearity of active resistive links and the time uncertainty of sampling mechanisms degrade the accuracy of the diffusion process in *RC* networks. These problems can be overcome by emulating resistive links through SCs, giving rise to the so-called diffusive SC networks.

There are many different SC topologies to run Gaussian filters [28]. Fig. 2(b) and (c) displays simple- and double-Euler SC networks in 1-D. In both cases, an exchange capacitor  $C_E$  is sampled by two switches driven by two nonoverlapping clock signals  $\phi_1$  and  $\phi_2$  [Fig. 2(d)]. The Gaussian pyramid provided by the double-Euler configuration yields better figures of merit than those of the simple-Euler SC topology when included in the SIFT algorithm [29]. Hence, the double-Euler is the SC network implemented on the CVIS-SIMD presented in this paper.

Assuming, as in any SC circuit, that transients associated with the ON resistances of the switches are neglected, that all state capacitors have the same capacitance  $C$ , and that  $C_{E1} = C_{E2} = C_E$ , the equivalent impedance of the double-Euler SC topology is  $R = T_{clk}/nC_E$ , where  $n$  is the number of clock cycles and  $T_{clk}$  is the clock period. The resultant  $\sigma_{SC}$ , the Gaussian width of the double-Euler SC topology across the number of clock cycles, becomes

$$\sigma_{SC} = \sqrt{\frac{2nC_E}{C}}. \quad (4)$$

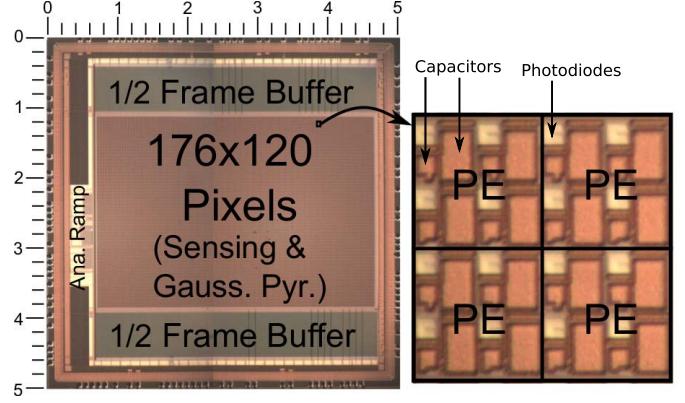


Fig. 3. Chip micrograph with dimensions (in millimeters) and a close-up of the PEs.

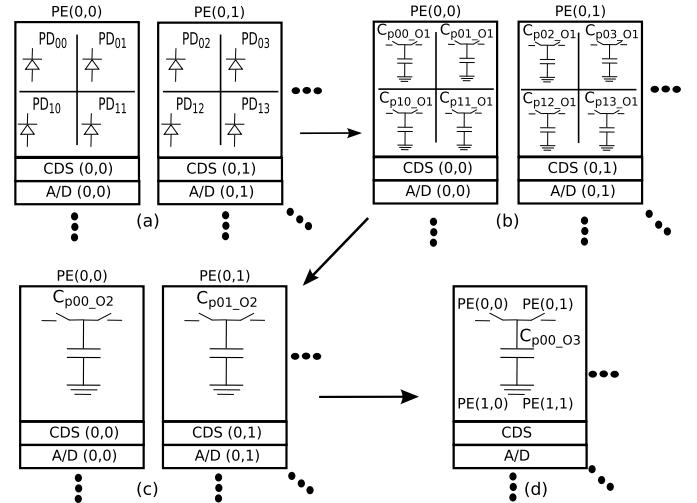


Fig. 4. PE array configuration across different functions of the chip. (a) Image acquisition, where four photodiodes share one CDS and ADC in a PE. (b) First octave, where four state capacitors share one CDS and ADC in a PE. (c) Second octave, where four state capacitors in a PE are shorted together to perform downscaling, and there is a CDS and an A/D per PE. (d) Third octave, where the state capacitors of the four PEs are combined into only one to run downscaling.

Equation (4) can be used to set the Gaussian width by design. However, deviations may be observed during fabrication, which depend on the actual device employed to implement  $C_E$  and  $C$ . It is hence convenient to extract the on-chip  $\sigma_{SC}$  value through measurements. Extracted values might be used for calibration if needed. The extraction procedure of on-chip  $\sigma_{SC}$  for our chip will be addressed in Section IV-B.

### III. CHIP DESIGN

#### A. Chip Floorplan and Processing Elements

The micrograph on the left of Fig. 3 shows the chip floorplan, consisting of a core array of PEs surrounded by a split frame buffer. The core array includes  $88 \times 60$  PEs. Each PE comprises the following:

- 1) four 3T-APS pixel-spatial resolution regarding image acquisition is hence  $176 \times 120$ ;
- 2) a comparator for in-PE A/D conversion;
- 3) four state capacitors and a CDS circuit, which is also used as part of local analog memories (LAMs) to store

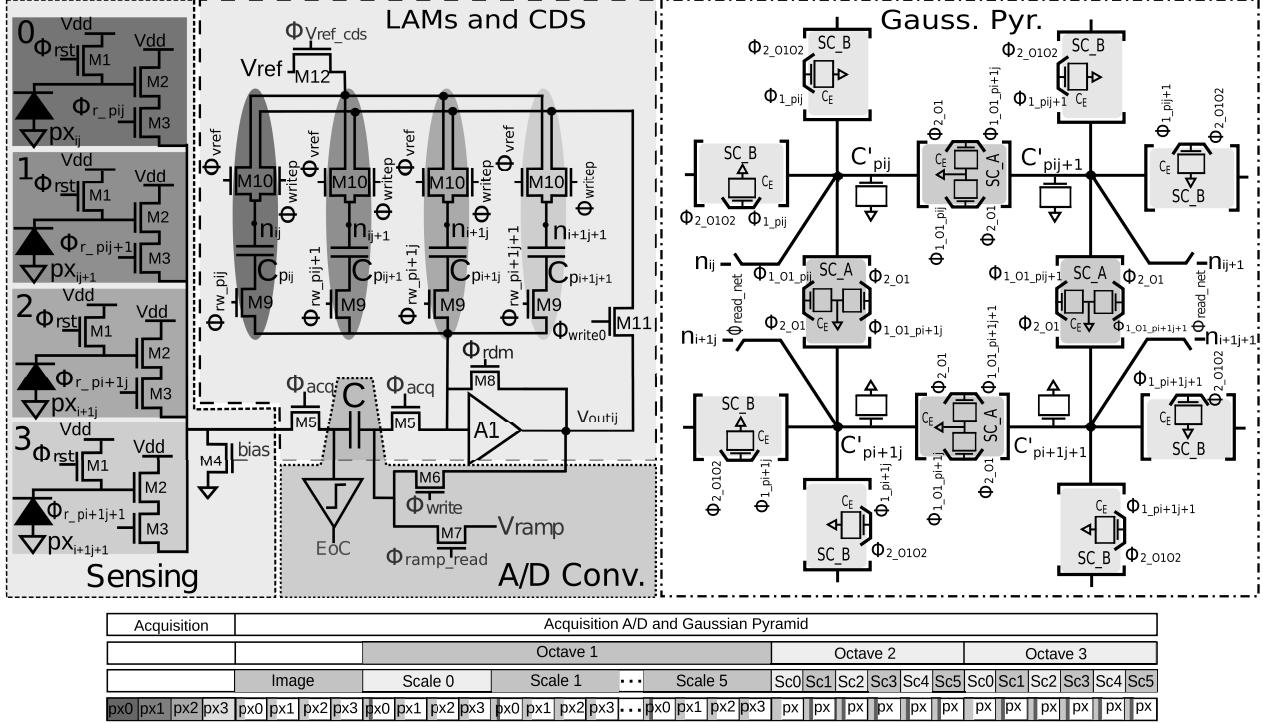


Fig. 5. PE and its associated time diagram. The PE is made up of four photosensors, four LAMs, one CDS circuit, one comparator for A/D conversion, and the local circuitry of the double-Euler SC network to build up the Gaussian pyramid.

either the acquired scene or a given scale across the Gaussian pyramid;

- 4) the double-Euler SC network made up of intra- and inter-PE switches for NEWS connectivity.

The inset on the right of Fig. 3 is a close-up of the PEs, where photodiodes and capacitors of the double-Euler diffusion network are visible.

Per-PE ADC and per-PE CDS, instead of the conventional per-column approach, increase parallelism. Also, this strategy gets favored by the retargetting of the herein proposed architecture to vertical technologies, leading to better performance metrics [30], [31]. Circuit sharing through the use of the same devices for different functions along time in part compensates for the per-PE ADC and CDS area overhead. Larger routing from the per-PE and per-CDS is alleviated by laying down the frame buffer that stores the results from the A/D conversion in two halves at the top and bottom of the PE array, which in turn diminishes power consumption.

### B. PE Array Configuration

The PE array changes its configuration according to the function realized by the chip. Fig. 4 conveys such configurations. The coordinates in the PE array are indicated within brackets. The origin of the coordinates is the PE on the top-left corner of Fig. 4. State capacitors of the double-Euler SC network in every octave ( $O_k$ ) are expressed as  $C_{pij\_Ok}$ .

The input image and the scales in the first octave are stored at state capacitors ( $C_{pij\_O1}$ ). As seen in Fig. 4(a) and (b), as there is only one ADC and CDS circuit per four pixels and four state capacitors, image acquisition and scales read out are performed for four cycles. State capacitors are shunted

across octaves. Fig. 4(c) shows the configuration during the second octave. In this case, the state capacitors of a PE are combined into only one to perform downscaling, which leads to one-to-one state capacitor per CDS and A/D circuit in the PE array. In the third octave, the state capacitors of four PEs are merged, and again, there is a one-to-one state capacitor per CDS and A/D circuit. The read out of the input image and the 18 scales resultant from three octaves and six scales each amounts to 40 A/D conversions of the PE array for the whole Gaussian pyramid.

### C. Circuit Implementation

Fig. 5 shows a circuit view of the PE with its time diagram. Table I lists the sizes of the transistors in Fig. 5. Switches are implemented with nMOS transistors with minimum dimensions. Circuit sharing is performed with amplifier  $A_1$  and capacitors  $C$  and  $C_{pij}$ . Every 3T-APS pixel has its corresponding capacitor  $C_{pij}$ . This is shown in Fig. 5 with the same gray shade. Capacitor  $C$  runs CDS and offset-compensation comparison during A/D conversion. Amplifier  $A_1$  and capacitors  $C_{pij}$  are part of LAMs and CDS circuits. The latter are also part of the state capacitors  $C_{pij\_Ok}$  in the SC network.

The gain stages in the PE are double-cascode topologies. Only one amplifier is included for CDS and image storing in the LAMs, while two are required in the comparator of the ADC. The amplifier can be configured in two modes of operation, namely,  $IA$  and  $IB$ , shown in Fig. 6(a) and (b), respectively. In both cases, the current can be cut off through enable ports. Switches driven by enable ports increase their output impedance close to the end of the operating range of the amplifier, increasing the gain too. Configuration  $IB$  consumes

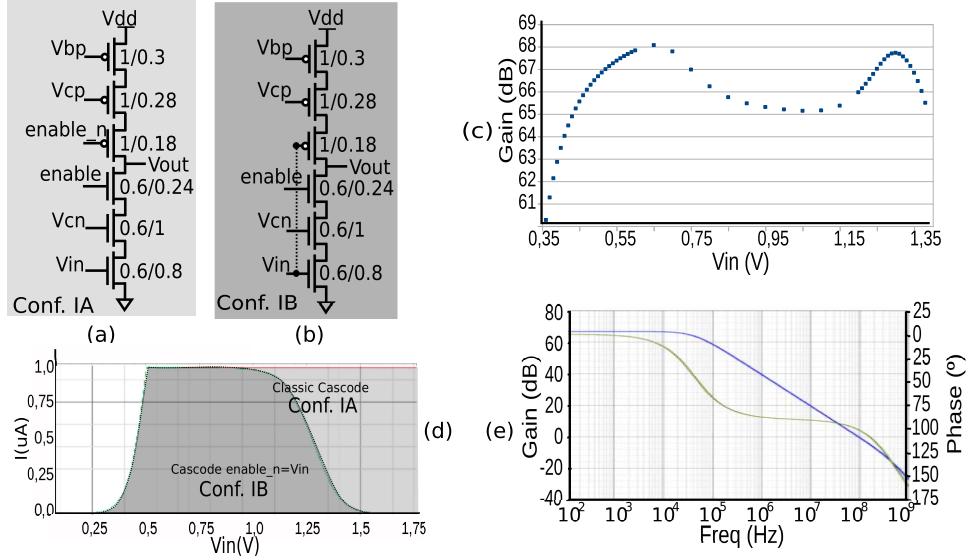


Fig. 6. Amplifier topologies used in the CDS, LAMs, and comparator circuits of Fig. 5 with some of their characteristics. (a) and (b) Cascode configurations *IA* and *IB*. (c) Gain versus input voltage. (d) Current consumption versus input voltage in configurations *IA* and *IB*. (e) Frequency response of configuration *IA* within the range of operation, [0.4, 1.3] V.

TABLE I  
PE TRANSISTOR SIZES (IN MICROMETERS)

	Width	Length		Width	Length
Photodiode	7.4	6.7	M1	0.24	1
M2	1.6	0.3	M3	0.24	0.6
M4	0.6	0.8	M5	0.24	1.4
M6	0.24	0.8	M7	0.24	1
M8	0.24	0.3	M9	0.24	0.2
M10	0.24	0.8	M11	0.24	0.2
M12	0.24	0.4			

up to 30% less power than *IA* at the cost of a narrower input range by shunting the port *enable\_n* to the input voltage  $V_{in}$  [Fig. 6(d)]. The bias current of both configurations is set to 1  $\mu$ A by  $V_{bp}$  through a wide-swing constant transconductance bias circuit trimmed with an external resistor [32], leading to a gain above 60 dB in the voltage range [0.4, 1.3] V with mismatch and process–voltage–temperature variations [Fig. 6(c) collects nominal simulations]. Bode plots are shown in Fig. 6(e).

1) *Image Acquisition*: The photodiode is an n-well over p-substrate structure in order to enhance the spectral response at longer wavelengths. The bias current of the source follower of the 3T-APS is set to 1  $\mu$ A by *M4* through a transconductance circuit with an external resistor. CDS is included to diminish reset noise and FPN from mismatch [33]. The nominal working range for the output voltage of the CDS circuit is defined by amplifier A1 in Fig. 5, namely, [0.4, 1.3] V. These are the lower and upper bounds for the voltages at the state capacitors of the double-Euler SC network.

Fig. 7 shows the CDS topology with its control signals. A similar implementation has been used in [34]. For a given

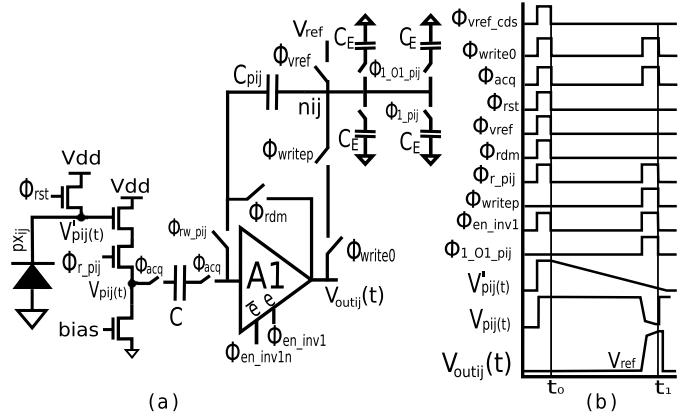


Fig. 7. Image acquisition through CDS on our chip. Signal  $\phi_{rw\_pij}$  selects the 3T-APS pixel associated with the position  $ij$  in the PE. Signal  $\phi_{rw\_pij}$  is high during the acquisition time for a pixel  $ij$ .

pixel  $ij$ , signal  $\phi_{rw\_pij}$  is high during the whole acquisition time. Reset and signal voltages for CDS are sampled at time instants  $t_0$  and  $t_1$  with signal  $\phi_{acq}$  high. The CDS output is stored in  $C_{pij}$  as well as in  $C'_{pij}$  and the four exchange capacitors  $C_E$  are connected to the node  $n_{ij}$ . Signals  $\phi_{1\_O1\_pij}$  and  $\phi_{1\_pij}$  set the initial values in the exchange capacitors used for intra-PE and inter-PE connections in the double-Euler SC network, respectively.

The CDS is implemented with amplifier A1 in *IA* mode to support a wide input voltage range. Enable signal  $\phi_{en\_inv1}$  allows switching off amplifier A1 between the two samples at  $t_0$  and  $t_1$ . By assuming large enough gain A1, the CDS output voltage is given by

$$V_{outij} = V_{ref} + \frac{C}{C_{pij}}[V_{pij}(t_0) - V_{pij}(t_1)] \quad (5)$$

where  $V_{ref} = 400$  mV.

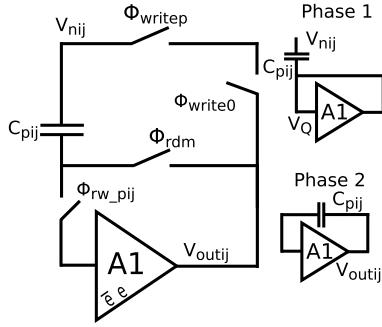


Fig. 8. LAMs working in two phases to store and read out scales across the Gaussian pyramid.

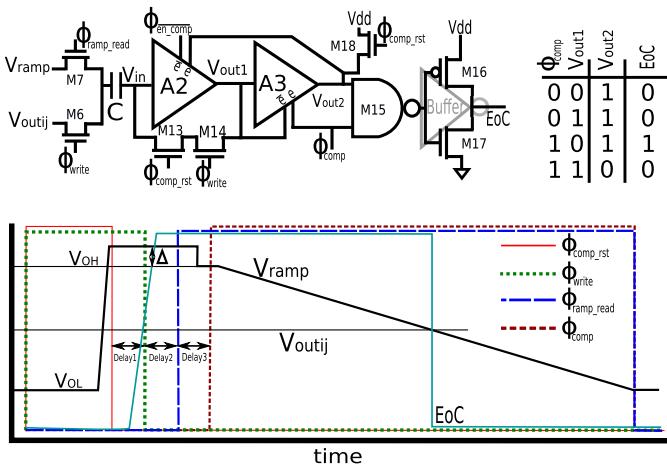


Fig. 9. Comparator of the in-PE 8 b single-slope ADC with the time diagram of its control signals.

2) *Local Analog Memories*: The LAMs store both the image after CDS and the scales across the Gaussian pyramid. The LAMs are implemented with amplifier  $A_1$ , capacitors  $C_{pij}$ , and switches  $\phi_{writep}$ ,  $\phi_{rdm}$ , and  $\phi_{write0}$  (see Fig. 5). Scales across the Gaussian pyramid are stored and read out in two phases with signal  $\phi_{rw\_pij}$  high and  $\phi_{vref\_cds}$  low. Both phases are shown in Fig. 8. During the first phase voltage,  $V_{nij} - V_Q$  is held in capacitor  $C_{pij}$  with signal  $\phi_{rdm}$  high and  $\phi_{writep}$  and  $\phi_{write0}$  low. The read out is performed during the second phase with  $\phi_{rdm}$  low and  $\phi_{write0}$  and  $\phi_{writep}$  high, leaving  $V_{outij} = V_{nij}$ , where  $V_{nij}$  is the voltage at node  $nij$ .

3) *Comparison for in-PE ADC*: Our chip embeds an 8 b single-slope in-PE ADC. Fig. 9 shows the single-input offset-compensated comparator of the in-PE ADC. Offset compensation makes the comparator less sensitive to manufacturing variability. Switches are implemented with nMOS transistors. Their sizes are collected in Table II. Label  $M15$  means the four transistors in the NAND gate of the comparator, which is implemented with complementary logic. Amplifier  $A2$  is configured as  $IA$ , while  $A3$  is in mode  $IB$  to cut power consumption, which is further decreased with the feedback loop between both gain stages. The bottom sampling technique is run with different delays between signals ( $Delay1$  -  $Delay3$  in Fig. 9).

The comparator works in two phases: reset and comparison. During reset, both the first input signal and the quiescent point

TABLE II  
COMPARATOR TRANSISTOR SIZES (IN MICROMETERS)

	W	L		W	L		W	L
M13	0.24	0.4	M14	0.24	0.8	M15	0.24	0.2
M16	2	0.2	M17	1.5	0.2	M18	0.24	0.2

of the first amplifier in the comparator are sampled. This is done with signals  $\phi_{comp\_rst}$  and  $\phi_{write}$  high. The reset phase ends by setting  $\phi_{comp\_rst}$  and  $\phi_{write}$  low, leaving  $V_Q - V_{outij}$  across  $C$ .  $V_{outij}$  can be either the input image with CDS or a given scale of the Gaussian pyramid. This voltage is compared with the voltage ramp  $V_{ramp}$  during the comparison phase, which starts with  $\phi_{comp}$  and  $\phi_{ramp\_read}$  high, giving (6) at the output of the second gain stage. The static power consumption can be cut during reset with  $\phi_{comp}$  low and  $\phi_{en\_comp}$  high. The comparator takes a falling ramp as input in the comparison phase with a downfall  $\Delta$  of signal  $V_{ramp}$  at  $V_{OH} = 1.3$  V to ensure a correct initial state for values of  $V_{outij}$  close to  $V_{dd}$

$$V_{out2} = K^2(V_{ramp} - V_{outij}) + V_Q. \quad (6)$$

$V_{outij} - V_{ramp}$  crossing triggers the signal end of conversion (EoC) to low, enabling the writing of a digital word given by an 8 b counter into the frame buffer assigned. The EoC occurs with  $V_{out2}$  low [see Fig. 9 and (6)], which in turn cuts off current in the first gain stage through a positive feedback loop. The feedback loop also reinforces logic levels. Voltage and current waveforms in the first amplifier of the comparator ( $V_{out1}$  in Fig. 9) with and without feedback loop plotted in Fig. 10(a) confirm this statement. Fig. 10(b) and (c) illustrates power savings from the feedback loop for two input voltages, corresponding to ADC output codes 250 and 40, close to the lower and upper parts of the falling ramp. Blue and pink lines are the currents integrated along the whole ramp in the first and second amplifiers of the comparator. The comparator without feedback loop consumes 1.65 and 1.7  $\mu$ W for codes 250 and 40, respectively, and the feedback loop leads to 75 nW and 1.65  $\mu$ W, resulting in large power savings for the largest ADC output codes.

4) *Gaussian Pyramid Construction*: Our double-Euler SC network with NEWS connectivity yields the Gaussian pyramid. Intra- and inter-PE connections are shown in different gray shades in Fig. 5. Fig. 11 gives a complete view of both intra- and inter-PE connections.

Downscaling across octaves in the Gaussian pyramid leads to three types of switching blocks in the SC network, labeled  $SC_A$ ,  $SC_B$ , and  $SC_C$  in Fig. 11, all of them implemented as nMOS transistors with minimum dimensions. In addition, one of the four PEs has a slightly different structure from the other three. Such a PE is shaded and marked with  $\beta$  in Fig. 11. PEs of  $\alpha$  type comprise switching blocks  $SC_A$  and  $SC_B$ . PEs of  $\beta$  type contain switching blocks  $SC_A$  and  $SC_C$ . The scales are provided by capacitors  $C_{pij\_ok}$ .  $C_{pij\_o1}$  means any of the  $176 \times 120$  state capacitors in the first octave. Similarly,  $C_{pij\_o2}$  and  $C_{pij\_o3}$  mean any state capacitor in the second and third octaves, where the resolution is downsampled to

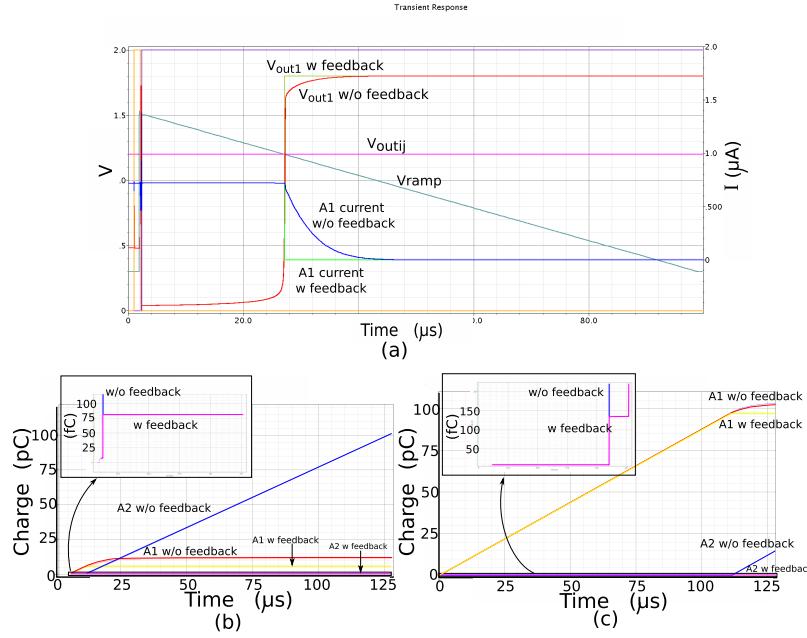


Fig. 10. Different waveforms of the in-PE comparator of Fig. 9 with and without feedback loop. (a) Currents and voltages. (b) and (c) Currents integrated for input codes 250 and 40, respectively.

$88 \times 60$  and  $44 \times 30$  px, respectively. Fig. 12 summarizes the states of the control signals across the Gaussian pyramid.

State capacitors  $C_{pij\_O1}$  in the first octave are the combination of MiM structures of M5–M6 metal layers  $C_{pij}$  with capacitors realized with transistors  $C'_{pij}$  in order to keep dynamic errors low, leading to  $C_{pij\_O1} = 330$  fF. Capacitors  $C'_{pij}$  are isolated from the SC network during LAMs' read out through signal  $\phi_{read\_net}$ , leaving  $C_{pij} = 200$  fF for these functions (see Fig. 5). Exchange capacitors in the first octave are set to  $C_E = 38.5$  fF and realized with transistors. According to (4), the state to exchange capacitors ratio yields  $\sigma_{SC\_O1} = 0.48\sqrt{n}$  for the scales in the first octave, with  $n$  being the number of clock cycles. Such scales are built with blocks  $SC_A$ ,  $SC_B$  and  $SC_C$ . Blocks  $SC_A$  run the two terms of the Gaussian kernel with NEWS connectivity through the switches that connect state capacitors within a given PE. The other two terms of the Gaussian kernel are executed with blocks  $SC_B$  or  $SC_C$ , correspondingly providing inter-PE connectivity of a given state capacitor with its neighbors. As an example, and as seen in Fig. 5, the state capacitor, which results from merging  $C_{pij}$  with  $C'_{pij}$  into  $C_{pij\_O1}$  within the the first octave, is connected to its eastern and southern neighbors through  $SC_A$  within the PE, while their northern and western connections comprise blocks  $SC_B$  in PEs of  $\alpha$  type and blocks  $SC_C$  in PEs of  $\beta$  type. Finally, signals  $\phi_1$  and  $\phi_2$  in the basic cell of the double-Euler SC network of Fig. 2 are implemented with signals  $\phi_{1\_O1\_pij}$  and  $\phi_{2\_O1}$  in blocks  $SC_A$ ,  $\phi_{1\_pij}$  and  $\phi_{2\_O1\_O2}$  in blocks  $SC_B$ , and  $\phi'_{1\_pij}$  and  $\phi'_{2\_O1\_O2}$  in  $SC_C$ .  $\phi_{1\_O1\_pij}$ ,  $\phi_{1\_pij}$ , and  $\phi'_{1\_pij}$  are turn on to initialize  $C_E$  and  $C'_{pij}$  capacitors during image acquisition through CDS in every PE with signal  $\phi_{read\_net}$  high, as seen in Figs. 5 and 7.

The one quarter downscaling from the first to the second octave occurs by shunting the four state capacitors  $C_{pij\_O1}$  of the first octave with the eight intra-PE exchange capacitors  $C_E$ , giving rise to larger state capacitors throughout the second

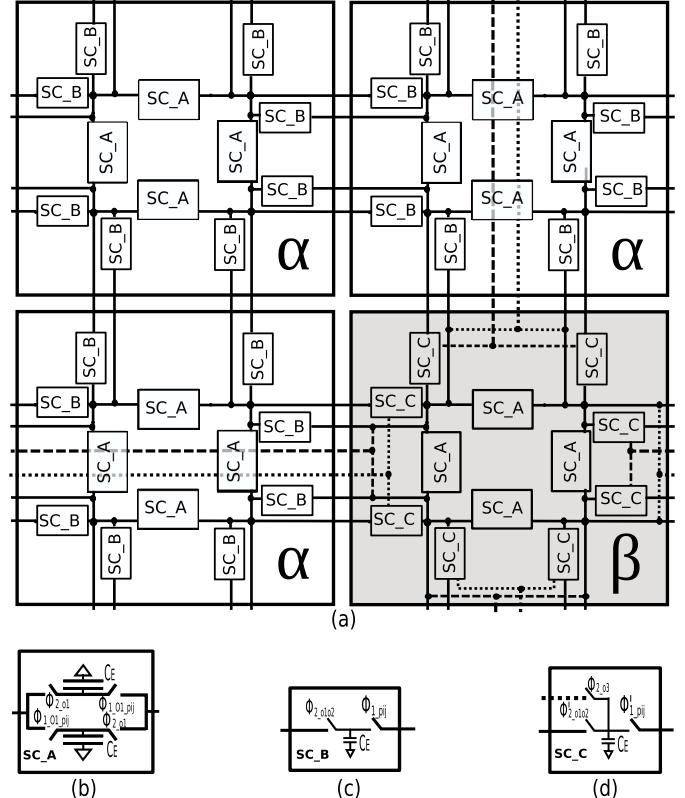


Fig. 11. (a) Double-Euler SC network for a grid of  $4 \times 4$  pixels ( $2 \times 2$  PEs) of the chip. (b–d) Internal structure of blocks  $SC_A$ ,  $SC_B$ , and  $SC_C$ . Groups of  $2 \times 2$  PEs comprise PEs of  $\alpha$  and  $\beta$  type.

octave as  $C_{pij\_O2} = 4C_{pij\_O1} + 8C_E$  for a given PE. In so doing, signals  $\phi_{1\_O1\_pij}$  and  $\phi_{2\_O1}$  in blocks  $SC_A$  are always high in the second octave. Signals  $\phi_{rw\_pij}$ ,  $\phi_{rw\_pij+1}$ ,  $\phi_{rw\_pi+1j}$ , and  $\phi_{rw\_pi+1j+1}$  are also high to shunt capacitors  $C_{pij}$  in the PE (see Fig. 5). Signals  $\phi_1$  and  $\phi_2$  in the basic

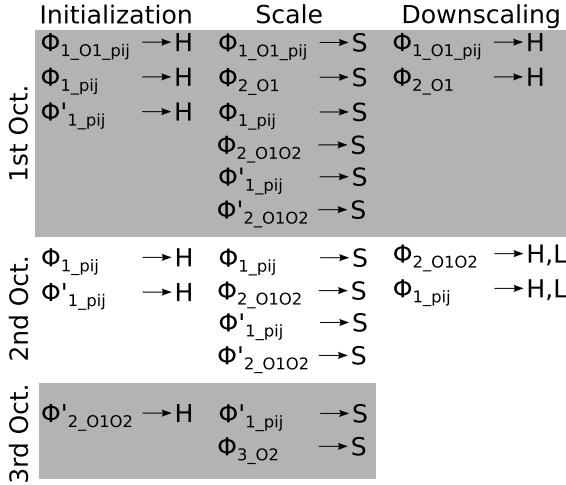


Fig. 12. State of control signals across the Gaussian pyramid. Symbols  $H$  and  $L$  refer to high and low states, respectively.  $H, L$  means that first the signal goes high and subsequently low. All the former states are found during initialization or downscaling to change between octaves. Symbol  $S$  means that the signals are switching to generate the scales of the pyramid.

cell of the double-Euler SC network of Fig. 2 are now given by the pairs  $\phi_{1\_pij}$  and  $\phi_{2\_O1O2}$ , and  $\phi'_{1\_pij}$  and  $\phi'_{2\_O1O2}$  in blocks  $SC_B$  and  $SC_C$ , respectively. Signals  $\phi_{1\_pij}$  and  $\phi'_{1\_pij}$  are used to initialize exchange capacitors for the second octave with blocks  $SC_B$  and  $SC_C$ . Also, as seen in Fig. 11, the NEWS connectivity for PEs of  $\alpha$  type is given by two  $SC_B$  blocks along each direction. Similarly, two  $SC_C$  blocks along each cardinal direction are used for PEs of  $\beta$  type. This means that now the exchange capacitors for the second octave become  $2C_E$ , which all in all leads to  $\sigma_{SC\_O2} = 0.23\sqrt{n}$ .

Finally, the one quarter downscaling from the second to the third octave is carried out in two phases. During the first step, the four state capacitors  $C_{pij\_O2}$  of four PEs are shunted together through signals  $\phi_{1\_pij}$  and  $\phi_{2\_O1O2}$  high in blocks  $SC_B$ . Subsequently, these signals turn low, disconnecting PEs of  $\beta$  type from those of  $\alpha$  type in every group of four PEs. As a consequence, the scales in the third octave are performed among PEs of  $\beta$  type through blocks  $SC_C$ , where  $\phi'_{1\_pij}$  and  $\phi_{2\_O3}$  play the role of control signals  $\phi_1$  and  $\phi_2$  in the basic cell of the double-Euler SC network of Fig. 2. Initialization of state capacitors is carried out with  $\phi'_{2\_O1O2}$  high. In this scheme, both exchange and state capacitors remain the same as in the second octave, so that  $\sigma_{SC\_O3} = \sigma_{SC\_O2}$ .

#### D. Peripheral Circuits

1) *Gaussian Pyramid Read-Out:* The Gaussian pyramid is read out through two frame buffers laid down at the top and bottom of the PE array, and labeled “1/2 frame buffer” in Fig. 3. Every register bank is assigned to the corresponding half of the PE array. The frame buffer split into two halves diminishes routing area.

Fig. 13(a) shows the half frame buffer. Every PE has two 8 b registers assigned in the frame buffer, allowing the read out and A/D conversion of two pixels at the same time. Such registers are named  $A$  and  $B$  in Fig. 13(b). Every frame buffer

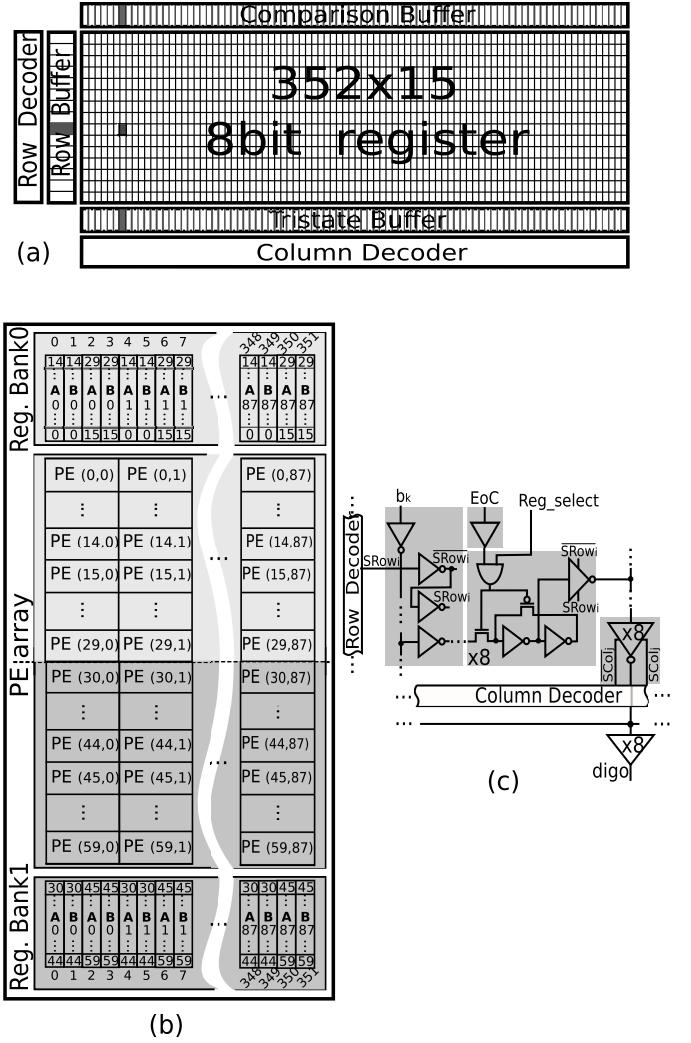


Fig. 13. (a) Half frame buffer bank for the single-slope ADC of half of the array. (b) PE register assignment. (c) Register circuitry. EoC comes from the in-PE comparator.  $b_k$  is a bit of the digital word issued by an 8 b global counter.

of the half PE array of 88 columns and 30 rows comprises 352 columns and 15 rows of registers. The 60 registers of a column of 30 PEs are placed in 4 columns of 15 rows each with the sequence  $ABAB\dots$  of Fig. 13(b). As an example of read-out procedure, for the first column of PEs of the bottom-half array—PEs across the 30th to the 59th row—the PEs from the 30th to the 44th row are A/D converted in column 0 in the register bank, while the PEs from the 45th to the 59th row are A/D converted in column 2 [both of them in reg. A in Fig. 13(b)]. At the same time, the data converted in the previous cycle are read out of the chip in columns 1, 3, ..., [reg. B in Fig. 13(b)]. Signal  $Reg\_select$  allows selecting one of the two 8 b registers, either  $A$  or  $B$ , yielding the A/D conversion. Finally, the 4 and a 9 b row and column decoders are NOR MOS decoders with pull-up transistors.

The signal EoC from the in-PE comparator enables writing of the digital word generated by a global counter into the registers, which are implemented with an nMOS transistor at the input and a pMOS transistor in their feedback loop [Fig. 13(c)]. The 8 b register of a word includes a tristate



Fig. 14. Prototype camera module to extract the on-chip Gaussian pyramid.

at the output as shown in Fig. 13. The row decoder enables these tristates in a full row and all write the stored word in a per column vertical bus. Another tristate placed at the end of each column selects the column that must be read. The column tristate writes the data in the bus that drives the digital word to a buffer. This buffer reinforces and drives the 8 b word to the output paths *digou* and *digod* (digital output up/down).

2) *Analog Ramp and Voltage Bias Generation*: The analog ramp for the 8 b single-slope ADC is produced with an 8 b current steering DAC [35]. The DAC is laid down at the left of the PE array in Fig. 3. The unity current for the DAC is set to  $2 \mu\text{A}$ . The current from the DAC is converted to voltage in an external resistor. The DAC also comprises a 5 b current steering to set up the offset of the ramp. Finally, the bias voltage generators of the gain amplifiers in the PE are implemented with wide-swing transconductance amplifiers included on the left side of the die, within the block labeled “Ana. Ramp” in Fig. 3.

#### IV. EXPERIMENTAL RESULTS

##### A. Camera Module Prototype

Fig. 14 shows a camera module prototype composed of three interconnected boards. The first of them (carrier board) hosts the sensor chip (FPGP). The second board encloses an FPGA DEO-Nano [36] to control the chip. The last one is a microPC (Raspberry Pi [37]) for visualization purposes. The optics is a C-mount type 35 mm @ f1/4 lens. The system is powered to 5 V through a plug Jack/ $\mu\text{USB}$  type.

##### B. On-Chip Gaussian Pyramid

The chip operation depends on the value of the emulated Gaussian filter width,  $\sigma_{\text{SC}}$ . This is set during design through capacitors  $C$  and  $C_E$  with (4), where  $n$  stands for the number of clock cycles. Nevertheless,  $\sigma_{\text{SC}}$  may change during physical realization. Fig. 15 displays changes measured from the chip. The black line shows the designed  $\sigma_{\text{SC}}$  as a function of the number of clock cycles  $n$ . The blue line shows the  $\sigma_{\text{SC}}$  values of the scale space extracted by iteratively comparing the outcome of the chip across the number of cycles  $n$  with an ideal scale space  $L(x, y, \sigma)$  on the image acquired by the chip through RMSE minimization. The red line is a polynomial fitted to the measured values. This experimental curve fits (4) using exchange capacitor values of  $C_E \approx 28 \text{ fF}$  and  $C_E \approx 26.5 \text{ fF}$  for the first and second octaves, instead of the designed ones, i.e.,  $C_E = 38.5 \text{ fF}$ , due to tolerances and parasitics, which do not destroy chip functionality. It should be noted that both the exchange capacitors  $C_E$  and part of

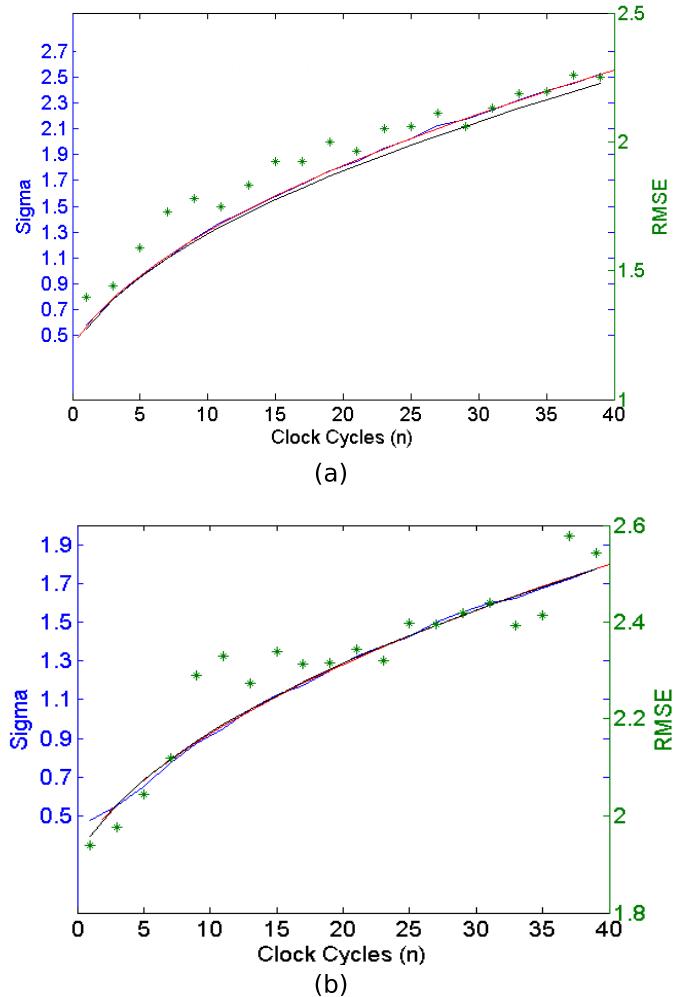


Fig. 15. On-chip  $\sigma_{\text{SC}}$  versus clock cycles  $n$  in the first and second octaves of the Gaussian pyramid.

the state capacitors  $C'_{pij}$  are implemented with transistors, while part of the state capacitors  $C_{pij}$  are MiM devices (see Fig. 5). Deviations among the experimental scales and scales designed with (4) are below 1% of the full scale, as it is illustrated by the right vertical axis in Fig. 15, where it is seen that the rms error (RMSE) saturates around 2.5 in a scale of 255 (1% of FSO). Finally, Fig. 16 further illustrates the outcome of Gaussian filters realized by the chip by showing different scales obtained within the first octave.

##### C. Implementation Comparison

The chip generates a Gaussian pyramid of three octaves with six scales each in 8 ms. Time required for A/D conversion is included in this number. Thus, the chip can provide 125 digitally encoded pyramids per second. The data conversion takes 200  $\mu\text{s}$  per conversion and the clock cycle for the double-Euler SC network is 150 ns. The relative energy consumption and throughput of our chip are 26.5 nJ/px at 2.64 Mpx/s.

Table III compares these metrics versus those provided by systems where Gaussian pyramids are obtained through digital signal processing following sensor read out. Since some of these systems do not embed image sensors, energy

TABLE III  
COMPARISON OF OUR CHIP WITH CONVENTIONAL SOLUTIONS

HW Solution	Func.	Energy/frame	En./px ( $\mu$ J/px)	Mpx/s	Mpx/J	Mpx/s.mm <sup>2</sup>
This work 180 nm CMOS	Gauss. Pyr.	176 × 120 resol. 70 mW @ 8 ms 0.56 mJ/frame	0.027	2.64	37.7	0.11
Ref. [39] OV9655 + Core-i7	Gauss. Pyr.	VGA resol. 90 mW @ 30 fps + 35 W @ 136 ms 4.8 J/frame	15.5	2.26	0.064	0.007
Ref. [40] OV9655 + Core-2-Duo	Gauss. Pyr.	VGA resolution 90 mW + 35 W @ 2.1 s 73.7 J/frame	240	0.15	0.004	0.001
Ref. [41] OV6922 + Qualcomm Snapdragon S4	Gauss. Pyr.	350 × 256 resol. 30 mW + 4 W @ 98.5 ms 0.4 J/frame	4.4	0.91	0.23	—

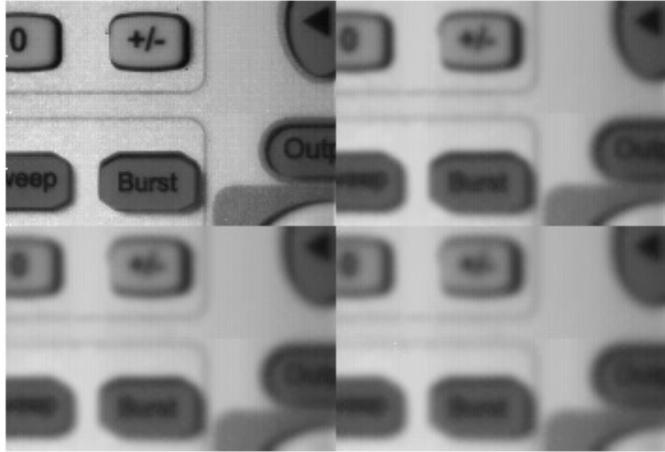


Fig. 16. Image acquisition and different snapshots of the on-chip Gaussian pyramid across the first octave. The upper left image is the input scene, the rest of the images from left to right and top to bottom correspond to  $\sigma = 1, 77$  (clock cycles  $n = 19$ ),  $\sigma = 2, 17$  ( $n = 29$ ), and  $\sigma = 2, 51$  ( $n = 39$ ).

for conventional CMOS imagers [38] scaled to the image resolution of the corresponding processor have been added for proper comparison.

Energy data in Table III do not include external memory accesses as they largely depend on the camera system. Their forecast would hence be inaccurate and similar for all the Gaussian pyramid sensory-processing subsystems, including ours. Our chip is up to four orders of magnitude better than conventional and low-power MPUs in computer performance (Mpx/J), while the throughput is similar to that of the most efficient competitor.

Table IV further illustrates the performance of the chip versus other highly efficient sensory-processing CVIS chips with per-pixel circuitry. The chip in [6] performs 2-D optic flow estimation. The PE array evaluates temporal contrast change by subtracting two frames whose gains are set by a programmable gain amplifier. The chip in [42] runs  $3 \times 3$  convolutions. The chip in [43] performs general-purpose low-level image processing. Finally, the chip in [44] performs background

subtraction. These functions are simpler than the generation of a Gaussian pyramid with three octaves @ six scales performed by the herein reported chip.

Still, the chips in [42] and [43] might compute Gaussian filters, as these are weighted convolutions. The metrics in Table IV correspond to isolated pairs of convolutions as Roberts or Prewitt edge detectors and to real-time edge detection at 25 frames/s, respectively. The evaluation of the Gaussian pyramid with these chips would certainly give different metric values and it would require additional hardware to switch between octaves. The chip in [44] performs background subtraction with two digitally programmable SC low-pass filters per pixel. The energy overhead on our chip compared with that on the chips in Table IV is partly explained by the higher complexity of the function that it runs. Differences in fill factor and pixel pitch are also due to the larger complexity of our PE. Particularly, our chip and that in [6] embed an 8 b single-slope ADC. While [6] follows a per-column ADC architecture, our chip follows a per-pixel one to achieve full parallelism and hence large speed.

#### D. Application Assessment

The accuracy of the on-chip Gaussian pyramid has been assessed by incorporating hardware errors into the interactive tool reported in [45]. This tool employs the SIFT feature detector to perform visual tracking of six 2-D textures on VGA resolution videos. Visual tracking metrics are calculated along the application of homography, defined as the matrix that captures the transformation of the 2-D textures from one frame to the next one, e.g., rotation.

Repeatability (RP) is the metric that we have calculated to assess the quality of visual tracking with the on-chip Gaussian pyramid [45]. As defined in [45] and formulated in (7), RP is the set of interest points  $S_{j-1}$  and  $S_{j-2}$  at frames  $j - 1$  and  $j - 2$  such that the geometrical distance between them after applying the corresponding homographies ( $H_{j-1}$  and  $H_{j-2}$ ) from frames  $j - 1$  and  $j - 2$  to frame  $j$  are below a certain threshold normalized to the total number of interest

TABLE IV  
COMPARISON OF OUR CHIP WITH OTHER STATE-OF-THE-ART CVIS

HW Sol.	This work	Ref. [6]	Ref. [42]	Ref. [43]	Ref. [44]
Funct.	Gauss. Pyr. w A/D (SS 8 bits)	2D Optic Flow Est. w A/D (SS 8 bits)	3×3 Conv.	General Purpose Low-level Imag.-Proc.	Back. Subt.
Tech. & Res.	0.18 $\mu\text{m}$ $176 \times 120$ px.	0.18 $\mu\text{m}$ $64 \times 64$ px.	0.35 $\mu\text{m}$ $64 \times 64$ px.	0.6 $\mu\text{m}$ $21 \times 21$ px.	0.35 $\mu\text{m}$ $64 \times 64$ px.
Fill-Fact.	10.25%	18.32%	23%	8.4%	12%
Pixel-Pitch	44 $\mu\text{m}$	28.8 $\mu\text{m}$	35 $\mu\text{m}$	98.6 $\mu\text{m}$	26 $\mu\text{m}$
En./px	26.5 nJ/px	0.89 nJ/px	0.19 nJ/px	0.52 nJ/px	0.62 nJ/px
Throughput	2.64 Mpx/s	0.49 Mpx/s	0.1 Mpx/s	0.11 Mpx/s	0.053 Mpx/s

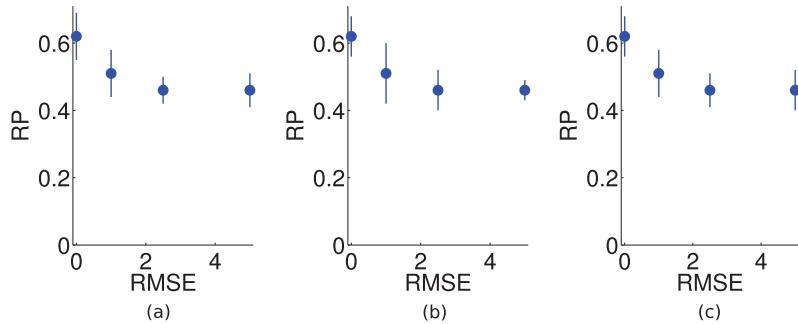


Fig. 17. RP as a function of RMSE for three image transformations. (a) Rotation. (b) Zoom. (c) Perspective distortion.

points  $S_{j-1}$  or  $S_{j-2}$ . RP gives an estimate of the percentage of interest points whose allocation in successive frames is successfully forecast with the extracted homography

$$RP = \frac{|(x_a \in S_{j-2}, x_b \in S_{j-1})| \ ||H_{j-2} \cdot x_a - H_{j-1} \cdot x_b|| < \epsilon}{|S_{j-1}|} \quad (7)$$

The RMSE values measured from the chip have been expressed as per-pixel local errors by finding the standard deviation of the normal distribution, which corresponds to the given RMSE level. The normal distribution conveys the variability from chip manufacturing. These errors have been added to every scale of the Gaussian pyramid. Fig. 17 displays RP versus RMSE for RMSEs of 0%, 1%, 2.5%, and 5%. Our on-chip RMSE levels are below 1.2% of FSO. RP is the average of the aforementioned six 2-D textures throughout all the frames of the corresponding videos with three different image transformations, namely, rotation, zoom, and perspective distortion. The error bars, calculated as the standard deviation throughout the averaged data, reports RP degradations that are tolerable for most applications. In fact, as reported in [45], the temporal distance between consecutive frames has a larger impact on RP. In this regard, the large Gaussian pyramid calculation throughput of our chip becomes an important asset as it enables reducing the baseline distance between consecutive frames.

## V. CONCLUSION

This paper presents a proof-of-concept CVIS of  $176 \times 120$  pixels for the parallel computation of the Gaussian pyramid with double-Euler SC networks. Cutting PE area through smaller state capacitors of the SC network

might be the most straightforward way to upscale our architecture while keeping performance metrics. Eventually, a given resolution could not be met with a double-Euler SC network. In that case, resorting to a simple-Euler network might be a solution if the loss of accuracy is affordable for the targeted application framework. Measurements from our chip demonstrate that sensory-processing architectures with per-pixel mixed-signal processors outperform conventional architectures consisting of an imager and an MPU in terms of both energy consumption and throughput. Our results also show that unavoidable errors of the analog circuitry do not result into unfeasible Gaussian pyramids as it has been verified by visual tracking metrics with a publicly available image data set. The main limitations posed by the type of SIMD-CVIS reported in this paper are direct consequences of the use of per-pixel circuitry and standard, planar technologies, namely: 1) enlarged pixel pitch and 2) reduced fill factors. The former might constrain the use of this type of chips to applications where the object of interest is at a short distance to the camera. The latter calls mainly for applications with controlled illumination conditions. However, these limitations can be overcome by retargetting our architecture into 3-D vertically integrated technologies, a task for which the circuits and methods reported in this paper can be reused.

## ACKNOWLEDGMENT

The authors would like to thank the University of California for the Software Transfer Agreement UC Case No. 2014-453 and Dr. S. Gauglitz for his assistance with the software framework to extract the metrics for visual tracking.

## REFERENCES

- [1] B. Roska and F. Werblin, "Vertical interactions across ten parallel, stacked representations in the mammalian retina," *Nature*, vol. 410, pp. 583–587, Mar. 2001.
- [2] T. Gollisch and M. Meister, "Eye smarter than scientists believed: Neural computations in circuits of the retina," *Neuron*, vol. 65, no. 2, pp. 150–164, Jan. 2010.
- [3] C.-L. Lee and C.-C. Hsieh, "A 0.8-V 4096-pixel CMOS sense-and-stimulus imager for retinal prosthesis," *IEEE Trans. Electron Devices*, vol. 60, no. 3, pp. 1162–1168, Mar. 2013.
- [4] J. Fernández-Berní, R. Carmona-Galán, and L. Carranza-González, "FLIP-Q: A QCIF resolution focal-plane array for low-power image processing," *IEEE J. Solid-State Circuits*, vol. 46, no. 3, pp. 669–680, Mar. 2011.
- [5] S. J. Carey, A. Lopich, D. R. W. Barr, B. Wang, and P. Dudek, "A 100,000 fps vision sensor with embedded 535GOPS/W 256×256 SIMD processor array," in *Proc. Symp. VLSI Circuits (VLSIC)*, Jun. 2013, pp. C182–C183.
- [6] S. Park, J. Cho, K. Lee, and E. Yoon, "243.3pJ/pixel bio-inspired time-stamp-based 2D optic flow sensor for artificial compound eyes," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 126–127.
- [7] Á. Rodríguez-Vázquez *et al.*, "ACE16k: The third generation of mixed-signal SIMD-CNN ACE chips toward VSoCs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 5, pp. 851–863, May 2004.
- [8] *Anafocus Ltd.*, accessed on Oct. 2016. [Online]. Available: <http://www.anafocus.com>
- [9] R. Carmona-Galán *et al.*, "A hierarchical vision processing architecture oriented to 3D integration of smart camera chips," *J. Syst. Archit.*, vol. 59, no. 10, pp. 908–919, Nov. 2013.
- [10] T. Roska and A. Rodríguez-Vázquez, Eds., *Toward the Visual Microprocessor*. New York, NY, USA: Wiley, 2000.
- [11] V. Busskamp and B. Roska, "Optogenetic approaches to restoring visual function in retinitis pigmentosa," *Current Opinion Neurobiol.*, vol. 21, no. 6, pp. 942–946, 2011.
- [12] A. Torralba, "How many pixels make an image?" *Vis. Neurosci.*, vol. 26, no. 1, pp. 123–131, Jan. 2009.
- [13] M. Eldib, F. Deboeverie, W. Philips, and H. Aghajani, "Behavior analysis for elderly care using a network of low-resolution visual sensors," *J. Electron. Imag.*, vol. 25, no. 4, p. 041003, Mar. 2016.
- [14] *Toshiba-Teli*, accessed on Oct. 2016. [Online]. Available: <http://www.toshiba-teli.co.jp/en/products/industrial/sps/sps.htm>
- [15] A. Blug *et al.*, "Closed-loop control of laser power using the full penetration hole image feature in aluminum welding processes," *Phys. Procedia*, vol. 12, pp. 720–729, 2011.
- [16] E. R. Davies, *Machine Vision: Theory, Algorithms, Practicalities*. New York, NY, USA: Elsevier, 2004.
- [17] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
- [18] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Found. Trends Comput. Graph. Vis.*, vol. 3, no. 3, pp. 177–280, Jan. 2008.
- [19] J. Fernández-Berní, R. Carmona-Galán, R. del Río, and A. Rodríguez-Vázquez, "Bottom-up performance analysis of focal-plane mixed-signal hardware for Viola-Jones early vision tasks," *Int. J. Circuit Theory Appl.*, vol. 43, no. 8, pp. 1063–1079, Aug. 2015.
- [20] Á. Zarandy, Ed., *Focal-Plane Sensor-Processor Chips*. Berlin, Germany: Springer, 2011.
- [21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [22] J. Fernández-Berní, R. Carmona-Galán, and A. Rodríguez-Vázquez, *Low-Power Smart Imagers for Vision-Enabled Sensor Networks*. New York, NY, USA: Springer, 2012.
- [23] N. Michael, D. Mellinger, Q. Lindsey, and V. Kumar, "The GRASP multiple micro-UAV testbed," *IEEE Robot. Autom. Mag.*, vol. 17, no. 3, pp. 56–65, Sep. 2010.
- [24] M. Suárez, V. M. Brea, J. Fernández-Berní, R. Carmona-Galán, D. Cabello, and A. Rodríguez-Vázquez, "A 26.5 nJ/px 2.64 Mpx/s CMOS vision sensor for Gaussian pyramid extraction," in *Proc. ESSCIRC 40th Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2014, pp. 311–314.
- [25] J. J. Koenderink and A. J. van Doorn, "Representation of local geometry in the visual system," *Biol. Cybern.*, vol. 55, no. 6, pp. 367–375, 1987.
- [26] T. Lindeberg, "Scale-space for discrete signals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 3, pp. 234–254, Mar. 1990.
- [27] B. Jähne, H. Haussecker, and P. Geissler, Eds., *Handbook of Computer Vision and Applications*, vol. 2. San Diego, CA, USA: Academic, 1999, pp. 423–450.
- [28] C. B. Umminger and C. G. Sodini, "Switched capacitor networks for focal plane image processing systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, no. 4, pp. 392–400, Dec. 1992.
- [29] M. Suárez, V. M. Brea, D. Cabello, F. Pozas-Flores, R. Carmona-Galán, and A. Rodríguez-Vázquez, "Switched-capacitor networks for scale-space generation," in *Proc. 20th Eur. Conf. Circuit Theory Design (ECCTD)*, Aug. 2011, pp. 189–192.
- [30] M. Suárez *et al.*, "CMOS-3D smart imager architectures for feature detection," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 2, no. 4, pp. 723–736, Dec. 2012.
- [31] M. Suárez, V. M. Brea, F. R. Pardo, A. Rodríguez-Vázquez, and R. Carmona-Galán, "Three dimensional CMOS image processor for feature detection," U.S. Patent 8942481 B2, Jan. 27, 2015.
- [32] D. A. Johns and K. W. Martin, *Analog Integrated Circuit Design*. Hoboken, NJ, USA: Wiley, 1997.
- [33] A. El Gamal and H. Eltoukhy, "CMOS image sensors," *IEEE Circuits Devices Mag.*, vol. 21, no. 3, pp. 6–20, May/Jun. 2005.
- [34] Y. M. Chi, U. Mallik, M. A. Clapp, E. Choi, G. Cauwenberghs, and R. Etienne-Cummings, "CMOS camera with in-pixel temporal change detection and ADC," *IEEE J. Solid-State Circuits*, vol. 42, no. 10, pp. 2187–2196, Oct. 2007.
- [35] M. Suárez, V. M. Brea, C. D. Matas, R. Carmona, G. Liñán, and Á. Rodríguez-Vázquez, "In-pixel ADC for a vision architecture on CMOS-3D technology," in *Proc. IEEE Int. 3D Syst. Integr. Conf. (3DIC)*, Nov. 2010, pp. 1–7.
- [36] *Terasic*, accessed on Oct. 2016. [Online]. Available: <http://www.terasic.com.tw/>
- [37] *Raspberry*, accessed on Oct. 2016. [Online]. Available: <http://www.raspberrypi.org/>
- [38] *OmniVision*, accessed on Oct. 2016 [Online]. Available: <http://www.ovt.com/>
- [39] M. Murphy, K. Keutzer, and H. Wang, "Image feature extraction for mobile processors," in *Proc. IEEE Int. Symp. Workload Characterization*, Oct. 2009, pp. 138–147.
- [40] F.-C. Huang, S.-Y. Huang, J.-W. Ker, and Y.-C. Chen, "High-performance SIFT hardware accelerator for real-time image feature extraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 340–351, Mar. 2012.
- [41] G. Wang, B. Rister, and J. R. Cavallaro, "Workload analysis and efficient OpenCL-based implementation of SIFT algorithm on a smartphone," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 759–762.
- [42] W. Jendernalik, G. Blakiewicz, J. Jakusz, S. Szczepanski, and R. Piotrowski, "An analog sub-miliwatt CMOS image sensor with pixel-level convolution processing," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 2, pp. 279–289, Feb. 2013.
- [43] P. Dudek and P. J. Hicks, "A general-purpose processor-per-pixel analog SIMD vision chip," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 1, pp. 13–20, Jan. 2005.
- [44] N. Cottini, M. Gottardi, N. Massari, R. Passerone, and Z. Smilansky, "A 33  $\mu$ W 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, Mar. 2013.
- [45] S. Gauglitz, T. Höllerer, and M. Turk, "Evaluation of interest point detectors and feature descriptors for visual tracking," *Int. J. Comput. Vis.*, vol. 94, pp. 335–360, Sep. 2011.



**Manuel Suárez** received the B.Sc. degree in physics and the Ph.D. (Hons.) degree in information and telecommunication technologies from the University of Santiago de Compostela, Santiago de Compostela, Spain, in 2008 and 2015, respectively, where the research was done with the Centro Singular de Investigación en Tecnologías da Información in collaboration with the Institute of Microelectronics of Seville (IMSE-CNMC-SIC), Seville, Spain.

Currently, he is an Analog Designer with Atomos GmbH, Montabaur, Germany. He is the first inventor of a U.S. patent.

Dr. Suárez was a recipient of the third Best Student Paper Award at ECCTD 2013.



**Víctor Manuel Brea** received the Ph.D. (Hons.) degree in physics from the University of Santiago de Compostela, Santiago de Compostela, Spain, in 2003.

Currently, he is an Associate Professor with the Centro Singular de Investigación en Tecnologías da Información, University of Santiago de Compostela, Santiago de Compostela, Spain. He has authored or co-authored over 100 papers in refereed journals, conferences, and workshops. His current research interests include the design of hardware architectures and CMOS solutions for computer vision, especially in early vision and microenergy harvesting.

Dr. Brea was a recipient of the Best Paper Award at ECCTD 2003 and the third Best Student Paper Award at ECCTD 2013. He has served on Technical and Steering Committees of international conferences such as ICDSC and CNNA.



**Jorge Fernández-Berni** received the B.Eng. degree in electronics and telecommunication, the M.Sc. degree in microelectronics, and the Ph.D. (Hons.) degree from the University of Seville, Seville, Spain, in 2004, 2008, and 2011, respectively.

From 2005 to 2006, he was working in the telecommunication industry. He has been a Visiting Researcher with the Computer and Automation Research Institute, Budapest, Hungary, Ghent University, Ghent, Belgium, and the University of Notre Dame, Notre Dame, IN, USA. He has authored or co-authored over 50 papers in refereed journals, conferences, and workshops. He has also authored a book and two book chapters and is the first inventor of two licensed patents. His current research interests include smart image sensors, vision chips, and embedded vision systems.

Dr. Fernández-Berni was a recipient of the Best Paper Award at the Image Sensors and Imaging Systems, SPIE Electronic Imaging 2014, San Francisco, CA, USA, and the Third Prize of the Student Paper Award at IEEE CNNA 2010: 12th International Workshop on Cellular Nanoscale Networks and their Applications, Berkeley, CA, USA.



**Ricardo Carmona-Galán** (M'98–SM'16) received the B.Sc. degree in physics and the Ph.D. degree in microelectronics from the University of Seville, Seville, Spain, in 1993 and 2002, respectively.

He was a Research Assistant with Prof. Chua's Laboratory, Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA, USA. He has been an Assistant Professor with the Department of Electronics, University of Seville. His research focuses on the very large scale integration implementation of concurrent sensor/processor arrays for real time image processing and vision. He also held a post-doctoral position with the University of Notre Dame, Notre Dame, IN, USA, from 2006 to 2007, where he was involved in interfaces for CMOS compatible nanostructures for multispectral light sensing. He has collaborated with startup companies such as AnaFocus, Seville, and Eutecus, Berkeley. He has designed several vision chips implementing different focal plane operators for early vision processing. Since 2005, he has been a Tenured Scientist with the Institute of Microelectronics of Seville (IMSE-CNM-CSIC), Seville. He has co-authored over 120 journal and conference papers and a book on low-power vision sensors for vision-enabled sensor networks. He is the Co-Inventor of several patents. His current research interests include the design of low-power smart image sensors and 3-D integrated circuits for autonomous vision systems.

Dr. Carmona-Galán was a recipient of the Certificate of Teaching Excellence from the University of Seville. He was also a recipient of the Best Paper Award of the IEEE-CASS Technical Committee on Sensory Systems at ISCAS 2015, together with Dr. Vornicu and Prof. Rodríguez-Vázquez. He has been an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART I and the Springer Journal on Real-Time Image Processing.



**Diego Cabello** (M'96) received the B.Sc. degree in physics from the University of Granada, Granada, Spain, in 1978, and the Ph.D. degree in physics from the University of Santiago de Compostela, Santiago de Compostela, Spain, in 1984.

From 1997 to 2002, he was the Dean with the Faculty of Physics, University of Santiago de Compostela, where he was also the Head of the Department of Electronics and Computer Science from 2002 to 2006. Currently, he is a Professor of Electronics with the Centro Singular de Investigación en Tecnologías da Información, University of Santiago de Compostela. His current research interests include the design of efficient architectures and CMOS solutions for computer vision, especially in early vision.

Prof. Cabello was co-recipient of the Best Paper Award at ECCTD 2003 and the third Best Student Paper Award at ECCTD 2013.



**Ángel Rodríguez-Vázquez** (F'99) received the B.Sc. and Ph.D. degrees in physics electronics from the University of Seville, Seville, Spain, in 1976 and 1983, respectively, with several national and international awards, including the IEEE Rogelio Segovia Torres Award (1981).

After different research stays at the University of California at Berkeley, Berkeley, CA, USA, and Texas A&M University, College Station, TX, USA, he became a Full Professor of Electronics with the University of Seville, Seville, Spain, in 1995. He co-founded the Institute of Microelectronics of Sevilla, Seville, under the umbrella of the Spanish Council Research, Barcelona, Spain, and the University of Sevilla, and started a research group on analog and mixed-signal circuits for sensors and communications. In 2001, he was the Main Promotor and the Co-Founder of the startup company AnaFocus Ltd., Seville, and served as the CEO, on leave from the University, until 2009. AnaFocus Ltd. was founded on the basis of his patents on vision chips and he participated in the foundation of the Hungarian startup company AnaLogic Ltd. His research work has been cited about 6700 times. He has an h-index of 43 and an i10-index of 133. He has authored over 11 books, 36 additional book chapters, and 150 journal articles in peer-reviewed specialized publications. He holds eight patents, three of which have been licensed to companies. His current research interests include the design of analog and mixed-signal front-ends for sensing and communication, including smart imagers, vision chips, and low-power sensory-processing microsystems.

Prof. Rodríguez-Vázquez has presented invited plenary lectures at different international conferences and was a recipient of several awards for his research such as the IEEE Guillemin-Cauer Best Paper Award, two Wiley's IJCTA Best Paper Awards, two IEEE ECCTD Best Paper Award, one SPIE-IST Electronic Imaging Best Paper Award, the IEEE ISCAS Best Demo-Paper Award, and the IEEE ICECS Best Demo-Paper Award. He has served as an Editor, an Associate Editor, and a Guest Editor for different IEEE and non-IEEE journals, and is on the committee of several international journals and conferences. He has chaired several international IEEE and SPIE conferences. He served as the Vice President Region 8 of the IEEE Circuits and Systems Society from 2009 to 2012 and as the Chair of the IEEE CASS Fellow Evaluation Committee in 2010, 2012, 2013, 2014, and 2015.