Gaussian Pyramid Extraction with a CMOS Vision Sensor

M. Suárez*, V.M. Brea*, J. Fernández-Berni†‡, R. Carmona-Galán†, D. Cabello* and A. Rodríguez-Vázquez†‡
*Centro de Investigación en Tecnoloxías da Información (CITIUS)
University of Santiago de Compostela, Santiago de Compostela, Spain
Email: victor.brea@usc.es
†University of Seville, Instituto de Microelectrónica de Sevilla (IMSE-CNM), Seville, Spain
‡CSIC, Instituto de Microelectrónica de Sevilla (IMSE-CNM), Seville, Spain

***Abstract*—This paper addresses a CMOS vision sensor with 176 × 120 pixels in standard 0.18 μm CMOS technology that computes the Gaussian pyramid. The Gaussian pyramid is extracted with a double-Euler switched-capacitor network, giving RMSE errors below 1.2% of full-scale value. The chip provides a Gaussian pyramid of 3 octaves with 6 scales each with an energy cost of 26.5 nJ at 2.64 Mpx/s.**

I. Introduction

The Gaussian pyramid is a set of images extracted from the input scene that provides computer vision algorithms with scale robustness, i.e. the ability of an algorithm to give the same response regardless the distance of the object to the camera. Scale Invariant Feature Transform (SIFT) is an example of feature detector that includes the Gaussian pyramid to this end [1]. The Gaussian pyramid comprises octaves. Every new octave comes from downscaling by 1/2× the former octave. An octave is a set of images which are the result of applying Gaussian filters with increasing widths or sigma (σ) values. Usually, 3 octaves with 6 scales each suffice [1].

As expected, the Gaussian pyramid generation is a very time-consuming task, which, as reported in [2], might take up to 90% of computation time of SIFT. Massive parallelism with close to the sensors processing and pixel per processor assignment is a good approach to alleviate this bottleneck. Also, both the information redundancy conveyed in an image and the inherent uncertainty throughout the different stages of a modern computer vision algorithm as SIFT make hardware inaccuracies less relevant to the metrics of a given task [3]. In this context, the analog domain is very suitable for Gaussian pyramid. In particular, RC or switched-capacitor networks naturally compute the Gaussian kernel [4], [5]. The circuit addressed in this paper implements a switched-capacitor network. This minimizes the non-linearity of a conventional RC network, and it allows for a more accurate control of σ.

II. Chip Design

The chip comprises an array of 88 × 60 processing elements (PEs). Every PE contains 4 nwell/p-sub photodiodes configured as 3T structures, along with the circuitry for in-PE A/D conversion, in-PE CDS and a double-Euler switched-capacitor configuration along the 4 cardinal directions. The chip is fabricated within an area of 5 × 5 mm^2 in 0.18 μm CMOS technology. The chip gives the scales of the Gaussian pyramid or the input scene as 8-bit digital words. The image is read out through two frame buffers outside the PE array. Each PE is shorted to two 8-bit registers in their assigned frame buffer. This permits to read out pixels outside the chip as they are being read in from the PE array [6].

Fig. 1 shows the micrograph of the chip along with a schematic of a PE. Every PE occupies 44 × 44 μm^2. The area of the nwell/p-sub photodiode is 7.4 × 6.7 μm^2. The 4 3T pixels share the same current source drawing 1 μA. Both the A/D conversion of the input image (176 × 120 pixels) and the CDS operation are time-multiplexed throughout 4 cycles. The gain stages $-K$ for CDS and the one for the offset-compensated comparator labeled EoC in Fig. 1 during A/D conversion are double-cascode inverters with 65 dB gain and 1 μA of bias current. The capacitance C in Fig. 1 is used during CDS and A/D conversion. The output of CDS is stored at every capacitor C_{pij} in every PE. Subsequently, the Gaussian pyramid is computed with the double-Euler network displayed in Fig. 1 [7]. Capacitors C_{pij} are laid down as MiM structures on Metal5 and Metal6, being sized to 200 fF for CDS. Capacitors C_E are implemented with an MOS transistor, being set to 28.5 fF. The value of C_{pij} is increased up to 330 fF with an MOS capacitor in parallel with the aforementioned MiM structures during Gaussian kernel computation in order to minimize feedthrough and injection errors. The Gaussian pyramid computation is controled by the two non-overlapped signals ϕ_1 and ϕ_2 shown in Fig. 1, which combined define a clock cycle.

III. Demo Setup and Experimental Results

Fig. 2 is a picture of the experimental setup. Fig. 3 shows the chip with the lens, the carrier board, and an FPGA. The chip has a PGA120 package. It rests on a carrier board of 15 × 6 cm^2. A DE0 Terasic FPGA provides the control signals

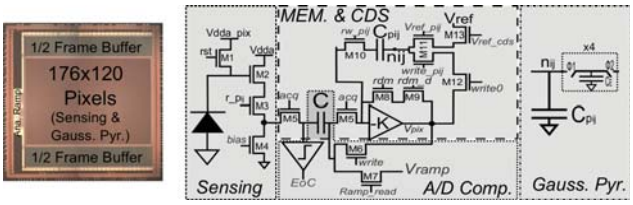

Fig. 1. Micrograph of the chip along with a schematic of a PE.

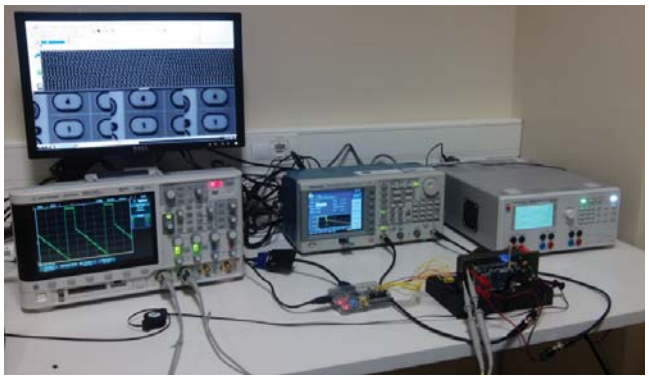

Fig. 2. Photograph of the chip setup.

for the chip. A Raspberry-Pi with an ARM processor is used for display purposes.

Fig. 3 shows different Gaussian-filtered images throughout the pyramid for the first three octaves ($O1$, $O2$, and $O3$). RMSE errors by comparing our chip measurements with pure software implementations are below 1.2% with respect to the full-scale value. The chip consumes 70 mW with scene acquisition and the Gaussian pyramid of 3 octaves and 6 scales each. The Gaussian pyramid takes 8 ms (A/D conversions included), with 200 μs per A/D conversion, and 150 ns as the clock cycle for the switched-capacitor network, rendering 26.5 nJ/px at 2.64 Mpx/s. Real-time tests will be performed during the conference.

ACKNOWLEDGMENT

This work has been funded by ONR N000141410355, Spanish government projects TEC2009-12686 MICINN, TEC2012-38921-C02 MINECO (European Region Development Fund, ERDF/FEDER), IPT-2011-1625-430000 MINECO, IPC-20111009 CDTI ((ERDF(FEDER)), Junta de Andalucía with TIC 2338-2013, Xunta de Galicia with EM2013/038 (ERDF(FEDER)), AE CITIUS (CN2012/151, ERDF(FEDER)), and GPC2013/040 ERDF(FEDER).

REFERENCES

[1] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints". International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[2] K. Mizuno et al., "Fast and Low-Memory-Bandwidth Architecture of SIFT Descriptor Generation with Scalability on Speed and Accuracy for VGA Video". FPL 2010, pp. 608-611, 2010.

[3] S. Gauglitz et al., "Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking". Int. J. of Computer Vision, vol. 94, pp. 335-360, 2011.

Fig. 3. Photograph of the chip with the carrier board, the lens, and the FPGA for control signals.

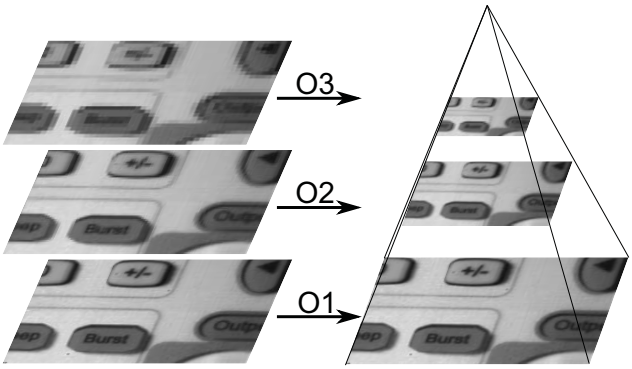

Fig. 4. Gaussian-filtered images throughout the pyramid for the first ($O1$, 176 $\times$ 120 px), second ($O2$, 88 $\times$ 60 px), and third ($O3$, 44 $\times$ 30 px) octaves.

[4] J. Fernández-Berni et al., "FLIP-Q: A QCIF Resolution Focal-Plane Array for Low-Power Image Processing". IEEE J. of Solid-State Circuits, vol. 46, No. 3, pp. 669-680, March 2011.

[5] M. Suárez et al., "CMOS-3D Smart Imager Architectures for Feature Detection", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol.2, no.4, pp.723-736, Dec. 2012.

[6] M. Suárez et al., "A 176×120 Pixel CMOS Vision Chip for Gaussian Filtering with Massivelly Parallel CDS and A/D-Conversion", ECCTD 2013.

[7] M. Suárez et al., "Switched-capacitor networks for scale-space generation", ECCTD 2011, pp. 190-193, 29-31 Aug. 2011.