HAL
open science

Distributed Mixed-Signal Architecture for Programmable Smart Image Sensors

Juliette Le Hir, Anthony Kolar, Filipe Vinci dos Santos

To cite this version:

Juliette Le Hir, Anthony Kolar, Filipe Vinci dos Santos. Distributed Mixed-Signal Architecture for Programmable Smart Image Sensors. 2017 15th IEEE International New Circuits and Systems Conference (NEWCAS) , Jun 2017, Strasbourg, France. 10.1109/NEWCAS.2017.8010178 . hal-01672180

HAL Id: hal-01672180

https://centralesupelec.hal.science/hal-01672180v1

Submitted on 23 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed Mixed-Signal Architecture for Programmable Smart Image Sensors

Juliette LE HIR, Anthony KOLAR
GeePs | Group of electrical engineering – Paris,
UMR CNRS 8507, CentraleSupélec, Univ. Paris-Sud, Univ. Paris-Saclay, Sorbonne Univ., UPMC Univ Paris 06
Gif-sur-Yvette, France
juliette.lehir, anthony.kolar@centralesupelec.fr

Filipe VINCI DOS SANTOS
SANA | Advanced Analog Design Group
CentraleSupélec
Gif-sur-Yvette, France
filipe.vinci@centralesupelec.fr

***Abstract*— Smart vision systems on a chip are promising for embedded applications. Currently, flexibility in the choice of integrated pre-processing tools is obtained at the expense of total silicon area and fill factor, which are otherwise optimized provided that the sensor performs a specific task. We propose a new architecture based on macropixel-level processing to improve the trade-off by using the same processing elements (PEs) for a whole group of pixels. In this paper, we show through functional simulations the feasibility of using macropixel PEs, with analog operative parts to avoid analog to digital converter (ADC) bottlenecks and digital controls, distributed in and out of the matrix of pixels. PEs are designed to be suitable for coefficient-reconfigurable spatial filtering as well as temporal difference. Sharing electronics among several pixels and matching existing algorithms to the target architecture allow for such programmability without degrading pixel area nor fill factor.**

Keywords*— *Smart image sensor, vision system on a chip (VSoC), focal-plane array, algorithm-architecture matching, mixed analog-digital electronics.

I. INTRODUCTION

Smart Vision Systems-on-a-Chip (VSoCs) aim at outputting relevant information on the scene by performing low- and middle-level image processing, sometimes at the expense of image quality. Extracting image features such as edges or motion prior to transmitting it for further analysis can be a gain of speed and power consumption provided that the analog and digital processing units are co-designed and spatially distributed [1]. Such integrated imaging systems are becoming attractive for embedded applications such as drone vision thanks to their savings in area, power, weight and communication bandwidth [2]. They are also cost-effective, provided they are fabricated in standard (i.e. planar single-chip) CMOS image sensor (CIS) technology.

This paper is dedicated to the proposal and analysis of a new architecture for smart image sensors addressing important issues of smart VSoCs based on standard CIS processes, in particular their poor balance between reconfigurability and pixel optimization. The state of the art of VSoCs presented in section II shows an unavoidable trade-off between versatility and pixel pitch and fill factor. In section III, we propose a new design approach to reach an optimized solution regarding this trade-off, thanks to spatial distribution of processing elements. Section IV details the hardware architecture of a programmable sensor based on this approach, before section V concludes by the future work needed to implement a hardware prototype.

II. VSOCS STATE OF THE ART

During the last decade several smart vision sensors have been designed in standard CIS technology [1,3-15]. The increasing resolutions and frame rates result in a large data transfer between the imaging array and the processing unit. In order to avoid this highly energy consuming operation, image processing is moved as close as possible to the focal plane array. The straightforward approach is to implement in-pixel circuitry. Digital processing means that analog to digital converters (ADCs) are required in each pixel, which brings a tradeoff between resolution of conversion and used silicon area. On the other hand, analog processing needs no in-pixel ADC, and would consume less power. Therefore, and considering our application, we focus here on analog implementation of common processing tasks such as edge detection using spatial convolution [1,3], difference of averaged images [4] and neighbours comparison [5] ; motion detection using temporal difference [1, 5] ; or image enhancement [1,6].

In-pixel processing loosens data throughput requirements in exchange for decreased fill factor. Hence a trade-off has to be made with image quality. Moreover, image processing tasks have been proven to benefit from spatial distribution of processing circuits [7]. Therefore an improvement is to also integrate processing circuits once for the whole matrix [8] or at the bottom of each column. For example, one can take advantage of the column-wise correlated double sampling circuit to perform temporal difference [9].

On top of pixel-wise, column-wise and array-wise processing, one can consider the macropixel approach: blocks of several pixels (e.g. from 3x3 to 32x32 pixels) processed as a whole. Virtual macropixels are used for region-of-interest detection: pixels are processed together as a virtual cluster by out-of-the-matrix electronics or software. For example, this method is applied for spatial averaging [4, 10], computing of local integration time [11] or memory optimization by pixel interlacing [12]. On the other hand, the concept of macropixels can be implemented in hardware by mutualizing in-matrix circuitry for the block of pixels instead of repeating it in every pixel. Suárez *et al.* [13] proposed such a hardware macropixel: 4 photodetectors share an amplifier and an ADC. A solution for Gaussian filtering is also implemented in [13], but it relies on a

full resolution switched-capacitor network which does not really take advantage of the macropixel concept.

In short, smart vision sensors currently perform one or several simple tasks such as: edge and/or motion detection [5], edge detection, HDR and tracking [14], motion detection or low power imaging by programming pairs of pixels [15]. However, none of these systems grants real programmability in the choice neither of algorithms nor of the coefficients.

On the other hand, programmable VSoCs have been proposed in [1, 6, 10], but with only in-pixel processing circuits and thus they suffer from very low fill factor (e.g. 5,4% in [10]).

A key observation is that distributing analog processing in the matrix improves the area/programmability trade-off. Though a few programmable sensors do exist, there seems to be a lack of a tightly integrated solution. Therefore, in section III, we introduce a new design approach, furthering the macropixel concept, for a highly distributed fully configurable smart image sensor.

III. Algorithm-Architecture Matching for Distributed Electronics

The goal of this work is to develop a smart image sensor, embedding digitally controlled analog processing allowing for fully programmable image pre-processing tasks in the focal plane. This limits data transfers out of the system and thus energy consumption, by extracting relevant information as close as possible to the source. By distributing processing electronics between different levels - pixel, macropixel(s), column and whole matrix -, embedding more electronics for versatility purpose becomes possible without degrading significantly other characteristics such as fill factor or pixel size, so that smart high resolution sensors can be fabricated at low cost on standard CIS technology.

The idea is to map common image processing operations to processing circuits that are distributed all over the matrix. In particular, we consider moving away from pixel by pixel operations towards macropixel-level processing in both spatial and temporal image analysis tasks. Moreover, globalized programmable processing elements allow for electronic resources reuse for different tasks.

A. Spatial Convolution

Spatial convolution is widely used in pre-processing tasks such as edge detection or filtering, so efficiently implemented coefficient-programmable spatial convolution is of great interest. It has been done at pixel-level [1,3] but this implies high sensing surface loss in each pixel. Using macropixel-level implementation, another solution is proposed. The idea is to limit the number of processing elements (PEs) and interconnections inside the matrix. Therefore each pixel is linked to only one PE, and one PE manages as many pixels as the size of the mask (i.e. kernel), for ease of use of the control. Each PE is identical and performs the linear combination of the linked pixels weighted by the chosen coefficients of the mask. The result is then a down-sampled convolution since there is no superposition of the kernels (see Fig. 1). Hence drastic data and in-matrix circuitry reduction is obtained (division by the size of the mask) at the cost of quality loss due to downsampling.

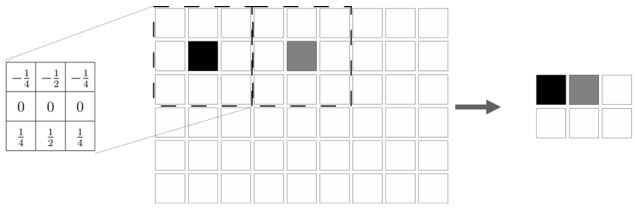

Fig. 1. Principle of the down-sampled convolution: linear combination of Sobel mask with pixels of the image is done once instead of nine times for a 3x3 kernel.

This theoretical adaptation of convolution has been functionally tested through Matlab simulations. An illustrative result is displayed on Fig. 2.

It has also been applied to the Histogram of Oriented Gradient (HOG) algorithm, which is widely used for pedestrian detection [16]. The first step of this algorithm is the gradient computation, which can be done by applying a {-1 0 1} mask or a Sobel mask, or else directly in polar coordinates [17] with

$$\begin{pmatrix} 0 & \sin(\theta_k) & 0 \\ -\cos(\theta_k) & 0 & \cos(\theta_k) \\ 0 & -\sin(\theta_k) & 0 \end{pmatrix} \quad (1),$$

for θ_k in {0°; 20°; 40°; 60°; 80°; 100°;120°; 140°; 160°}.

Results with SVMs (Support Vector Machines) trained by 600 positive images and 600 negative images from INRIA dataset, for each algorithm, are listed in Table I. Tests were conducted on 200 positive and 100 negative images from the INRIA dataset. Optimizing the training of the SVM is out of the scope of this paper. We simply used the same sets of images to qualitatively compare different low-level algorithms. Table I shows that using down-sampled convolution on cartesian or polar gradients gives comparable results to an implementation of classic HOG algorithm for false negative images. The down-sampled convolution shows a much higher rate of false positive detections. This can be explained by the fact that during training of positive images, edges can be lost and thus the SVM is considering non-pedestrian edges as pedestrian edges. For most applications, such as military detection of suspect person or pedestrian detection for automotive avoidance, false positive detections are not a critical issue. So for a negligible loss of quality, the proposed method divides by 9 the amount of electronic processing in the matrix for a 3x3 kernel convolution.

Besides, errors in the calculation of the downsampled convolution were simulated through addition of a normal law of chosen standard deviation. Simulations showed that to keep false positive and negative results comparable to those obtained with the ideal HOG algorithms, the standard deviation of the error must be kept below 0.25. Having an adapted algorithm still

TABLE I. Comparison of Implementations of HOG Algorithm

	Type of HOG				
	Classic HOG	***Down-sampling***	***Polar gradient***	***Down-sampling of polar gradient***	***Classic HOG - single large pixel***
False positive images (%)	0	14	0	4	13
False negative images (%)	6	8	5	8.5	10

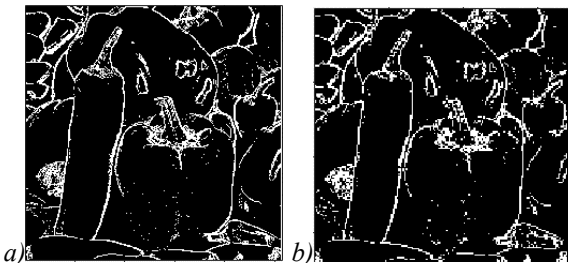

Fig. 2. Results of (a) classic convolution by a Sobel kernel (equivalent to 1 PE per pixel) and (b) down-sampled convolution with same kernel (equivalent to 1 PE per 9 pixels), performed on the classic image of peppers (512x512 pixels result in (b) 170x170 pixels).

effective with up to 25% error of cumulated computation suggests that an analog implementation is feasible.

B. Temporal Difference

The same methodology has been applied to temporal difference, which is a common technique for motion and Regions of Interest detection. Downsampling by 3x3 pixels seems to induce too much information loss, but downsampling by 2x2 appears to be a better trade-off between area saving and quality, as shown in Fig. 3.

Concerning spatial convolution, for a similar fill factor and final resolution, one could suggest using classic 3x3 convolution, with a single large pixel instead of a group of 9. This scheme is evaluated through HOG algorithm in the last column of Table I. Results are comparable with other implementations, but a 2x2 temporal difference could not be implemented for instance.

C. Resources Reuse

The presence of a coefficient-programmable processing element allows for reusing it for different tasks. For instance, temporal difference can be computed with the same PE as spatial filtering using the appropriate set of coefficients. Note that if one PE is assigned to 3x3 pixels while temporal difference is to be computed on a 2x2 basis, a certain sequence of operations must be carried out. It takes longer than having a PE devoted to each temporal difference in parallel. But this is acceptable given the versatility gained with few added electronics.

Besides, analog memory of a frame is usually obtained through storing capacitors. An available capacitor linked to a pixel permits high dynamic range (HDR) imaging, since this storing capacitor can receive the charge surplus from the photodiode under high illumination. Therefore, this overflow capacitor allows for complete information of illumination at high light level as well as at low light level [18]. Here, using the memory capacitor as an overflow one would result in a downsampled HDR image.

IV. A Distributed Architecture

The proposed architecture consists of macropixels of a defined size. Most kernels are 3x3 pixels (e.g. Sobel kernel) so we propose to fit this size to the hardware architecture of PEs. Temporal difference is on a 2x2 basis. Considering both, we define a 6x6-pixel scheme (Fig. 4), which can be easily repeated to build the whole matrix. Note that the middle capacitor of this

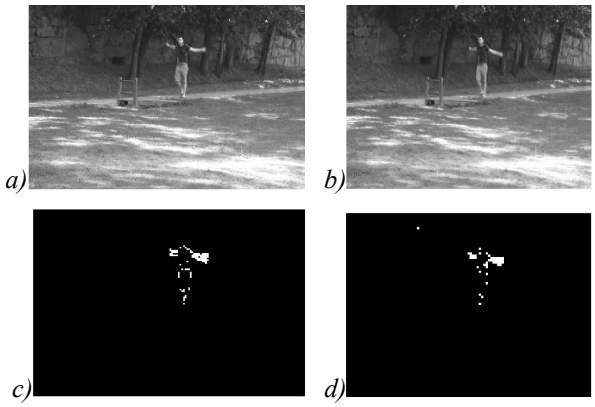

Fig. 3. Temporal difference: (a) image from a video of a walking and waving man, (b) following image of the video, and thresholded differences of the two previous images with (c) a 1/(2x2) down-sampling or (d) a 1/(3x3) down-sampling. (a) and (b) have a resolution of 240x320 pixels, and so (c) is 120x160 and (d) is 80x107 pixels.

pattern is to be connected to the bottom right PE so that each PE manages 3 capacitors at the most. This limits the computation time of temporal difference since each PE computes it for one of its dedicated capacitor after another.

Since PEs must be able to perform convolutions, multiply and add operators are required. A parallelized implementation would imply as many multipliers as mask coefficients. The huge area cost takes it out of consideration.

Moreover, it would be wasteful since masks containing zero coefficients would leave several multipliers unused. So we choose to have one multiplier and one accumulator for the sake of area and efficiency, at the cost of some velocity due to the sequential flow of operations. Note that those multiplier and accumulator are implemented in the analog domain, so that there is no ADC, which ensures fast calculation and area savings.

The PEs are digitally controlled, and that can be done from the out-of-matrix digital logic. This would mean numerous control buses crossing the whole matrix. Instead, we propose to distribute digital control over the matrix as well as analog operative circuits, since this implies less buses at the cost of only a few logic elements in each macropixel.

This architecture is illustrated in Fig. 5. The exterior digital part controls the pixels (*reset_pix* and *TX* for a 4T-pixel) and starts the PE with *proc_enable* (00 for idle, 01 for classic 3x3

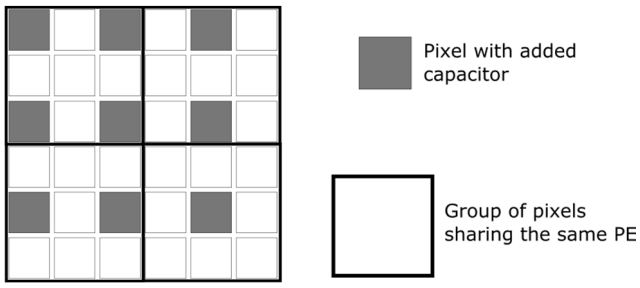

Fig. 4. Schematic representation of the 6x6-pixel basic unit of the proposed matrix: 1 PE for 3x3 pixels, and 1 capacitor for 2x2 pixels. This should be replicated as much as necessary to build the whole matrix.

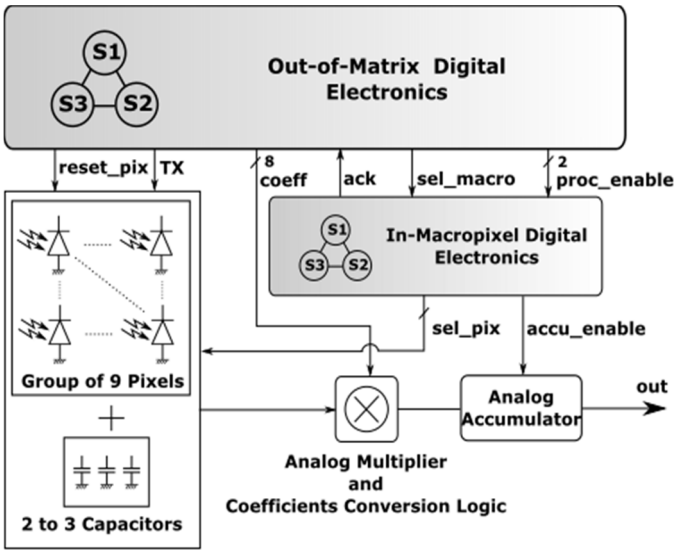

Fig. 5. Schematic representation of the proposed sensor: out-of-matrix digital electronics control a macropixel PE, composed of digital control interfaced with the general out-of-matrix control, and an analog operative part: 1 multiplier and 1 accumulator per macropixel (3x3 pixels).

mask, 10 for temporal difference). Then the PE sequentially selects the needed pixels or capacitors (*sel_pix*), multipliying their output by the corresponding coefficient (*coeff*) coming from the external digital control. Once the accumulation is done, the macropixel acknowledges (*ack*) and waits to be selected and read (*sel_macro*). If macropixels have digital outputs (e.g. simple 1-bit thresholding), they can be read all at once, and thus the system would be much faster.

The architecture has been simulated using functional VHDL-AMS model. The model can be easily modified to also permit classic convolution or 5x5-mask convolution, at the cost of more complex sequences for the control of the analog operative parts. Output rate of the sensor might thus be lowered in those choices, but very few electronics have to be added.

V. Conclusion

This paper has described a new approach to designing smart image sensors. To bring versatility while keeping reasonable fill factor and pixel area, the concept of hardware macropixel is used. Thanks to algorithm-architecture matching, spatial filtering and temporal difference are adapted to be computed using digitally controlled analog processing elements distributed in each macropixel. We show through simulations that the loss of quality is inconsequential for a subsequent high-level image processing such as pedestrian detection, whose circuit implementation is out of the scope of this work. An architecture of such a sensor has been presented, using analog multiplier and accumulator along with part of the digital control in the macropixel, and general digital control out of the matrix. Future work will focus on defining the requirements on the analog domain operators, followed by transistor-level design to fabricate a prototype. Our architecture is aimed at focal plane arrays, but would find even greater benefit when used with 3D stacking technology, which are likely to fundamentally transform VSoCs.

References

[1] N. Massari, M. Gottardi, L. Gonzo, D. Stoppa and A. Simoni, "A CMOS image sensor with programmable pixel-level analog processing", IEEE Trans. Neural Netw., Vol.16, pp. 1673-1684, 2005.

[2] E. Fossum, "CMOS image sensors: electronic camera-on-a-chip", IEEE Trans. Electron Devices, Vol. 44, pp.1689-1698, 1997.

[3] J. Dubois, D. Ginhac, M. Paindavoine and B. Heyrman, "A 10 000 fps CMOS sensor with massively parallel image processing", IEEE J. Solid-State Circuits, Vol. 43, pp. 706-717, 2008.

[4] N. Katic, A. Schmid and Y. Leblebici, "A retina-inspired robust on-focal-plane multi-band edge-detection scheme for CMOS image sensors", Midwest Symp. on Circuits and Systems, August 2014.

[5] D. Kim and E. Culurciello, "Tri-mode smart vision sensor with 11-transistors/pixel for wireless sensor networks", IEEE Sensors J., Vol. 13, pp. 2102-2108, 2013.

[6] J. Fernández-Berni, 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, pp. 669-680, 2011.

[7] H. Li, Z. Zhang, J. Yang, L. Liu and N. Wu, "A novel vision chip architecture for image recognition based on convolutional neural network", IEEE 11th International Conf. on ASIC, November 2015.

[8] B. Zhao, X. Zhang, S. Chen, K.-S. Low and H. Zhuang, "A 64x64 CMOS image sensor with on-chip moving object detection and localization", IEEE Trans. Circuits Syst. Video Technol., Vol. 22, pp. 581-588, 2012.

[9] N. Massari and M. Gottardi, "A 100dB dynamic-range CMOS vision sensor with programmable image processing and global feature extraction", IEEE J. Solid-State Circuits, Vol. 42, pp. 647-657, 2007.

[10] R. Carmona-Galán, J. Fernández-Berni and A. Rodriguez-Vázquez, "Automatic DR and spatial sampling rate sdaptation for secure and privacy-aware ROI tracking based on focal-plane image processing", Image Sensors Workshop, 2015. http://imagesensors.org/2015-paper/

[11] A. Peizerat *et al.*, "A 120dB DR and 5μm pixel pitch imager based on local integration time adaptation", Image Sensors Workshop, 2015. http://imagesensors.org/2015-paper/

[12] J. Martel, M. Chau, M. Cook and P. Dudek, "Pixel interlacing to trade off the resolution of a cellular processor array against more registers", European Conf. on Circuit Theory and Design (ECCTD), August 2015.

[13] M. Suárez *et al.*, "A 176 × 120 pixel CMOS vision chip for gaussian filtering with massivelly parallel CDS and A/D conversion", European Conf. on Circuit Theory and Design (ECCTD), September 2013.

[14] C.Yin, C.-F. Chiu and C.-C. Hsieh, "A 0.5V, 14.28kframes/s, 96.7dB smart image sensor with array-level image signal processing for IoT applications", IEEE Trans. Electron Devices, Vol.63, pp1134-1140, 2016.

[15] J. Choi, S. Park, J. Cho and E.Yoon, "A 3.4-μW object-adaptive CMOS image sensor with embedded feature extraction algorithm for motion-triggered object-of-interest imaging", IEEE J. Solid-State Circuits, Vol. 49, pp.289-300, 2014.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, June 2005.

[17] C. Bourrasset, L. Maggiani, C. Salvadori, J. Sérot, P. Pagano and F. Berry, "FPGA implementations of Histograms of Oriented Gradients in FPGA", WASC 2013. http://www.eunevis.org/wasc2013/index.php/en/programwasc.html

[18] S. Sugawa, N. Akahane, S. Adachi, K. Mori, T. Ishiuchi, K. Mizobuchi, "A 100dB dynamic range CMOS image sensor using a lateral overflow integration capacitor", IEEE International Solid-State Circuits Conf., February 2005.