Form Factor Improvement of Smart-Pixels for Vision Sensors through 3-D Vertically-Integrated Technologies

Angel Rodríguez-Vázquez [1,2,3], Ricardo Carmona-Galán [2], Jorge Fernández Berni [1,2], Sonia Vargas [2], Juan A. Leñero [2], M. Suárez [4], Victor Brea [4], and Belén Pérez-Verdú [2]

[1] University of Seville
[2] Instituto de Microelectrónica de Sevilla, IMSE-CNM
[3] Innovaciones Microelectrónicas S.L. (ANAFOCUS)
[4] Centro de Investigación en Tecnologías de la Información (CITIUS), Universidad de Santiago de Compostela

Pabellón de Italia-Planta Ático, Parque Tecnológico Isla de la Cartuja, 41092-Sevilla (SPAIN)
arodri-vazquez@us.es, angel@imse-cnm.csic.es

Abstract − While conventional CMOS active pixel sensors embed only the circuitry required for photo-detection, pixel addressing and voltage buffering, smart pixels incorporate also circuitry for data processing, data storage and control of data interchange. This additional circuitry enables data processing be realized concurrently with the acquisition of images which is instrumental to reduce the number of data needed to carry to information contained into images. This way, more efficient vision systems can be built at the cost of larger pixel pitch. Vertically-integrated 3D technologies enable to keep the advnatges of smart pixels while improving the form factor of smart pixels.

I. INTRODUCTION

CMOS Image Sensors (CIS) are suited to *embed* sensing devices, signal conditioning, data conversion, communication and processing circuitry on a common semiconductor substrate [1]. This way sensing, signal conditioning and processing happen concurrently at the sensor, thus enabling camera functions be incorporated already at the image capture front-end and thereby paving the way towards camera systems with improved SWaP factors [2][3]. For illustration purposes, Fig.1 shows the block diagram of a smart CIS with on-chip, sensor-embedded image correction [4]. The embedded core handles digitized full *frames* to complete on-chip operations intended to either deliver *corrected* images (FPN-correction, PRNU-calibration, etc.) or to *analyse* images (filtering, convolutions,

morphology, etc.). In any case, the goal is reducing either the requirements or even the need of off-chip resources, thus simplifying the implementation of camera systems.

Circuit embedding can actually be made at different levels, namely: *per-pixel* [5]-[9], *per-column* [10], and *per-chip* [11]. The intensity of embedding at each level depend on the targeted applications. Thus, for instance, consumer applications seek for minimum possible in-pixel circuitry [12]-[14], while high-end machine vision applications may call for larger amounts of in-pixel circuitry to increase the speed in the extraction of image features and in the reaction thereof [15][16]. As a general rule of thumb, the incorporation of circuitry at pixel level enables images being processed as they are acquired thus increasing *speed* and reducing *power* consumption in the realization of vision tasks [17]-[21]. Sensors composed of these *smart pixels*, makes the next evolutionary step of CMOS pixels, following *passive* pixels and *active* pixels [1], by embedding within the pixel resources for mixed-signal processing, memory and the programming and control of information flows.

Smart pixels are typically arranged according to the paradigm of Single Instruction Multiple Data (SIMD) computer architectures, and the sensors formed by them may be called CVIS (CMOS VIsion Sensors) to stress the fact that they do not only acquire images (like CIS) but also complete *early-vision* tasks [22] to reduce the amount of data sent for ulterior processing. Cameras employing CVISs at the front-end are capable to analyse images at thousands frames per second rate and with very low power consumption. However, their pixels have large *pitch* (in the range of tens of μm´s) and reduced *fill factor*. These features make CVISs to have limited sensitivity and modest spatial resolution, thereby constraining their usage to applications with limited field-of-view and active illumination. These limitations can be overcame by resorting to 3-D integration technologies [23] and the subsequent improvement of the form factors and footprints of the functional structures embedded at the pixels and at the whole chip.

II. ARCHITECTURES FOR SMART-PIXEL CVISs

Camera architectures with conventional, frame-based, front-end sensors and per-chip centralized processor are not the most efficient regarding speed of the decision-making process,
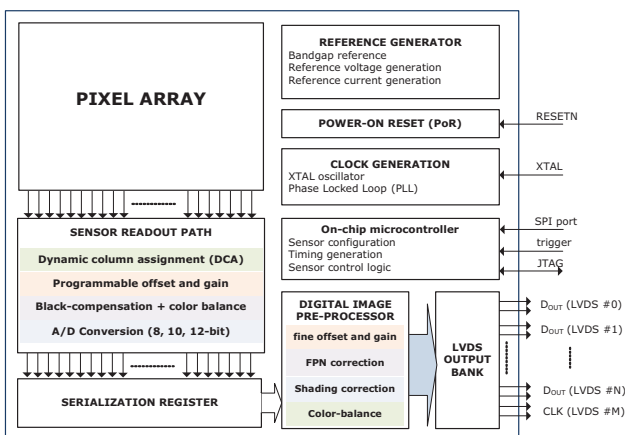

Figure 1. Block diagrams of a basic smart-CIS with embedded image correction

namely:

- On the one hand, they must transfer all pixel data from the sensor to the processor through the readout channel; despite these data carrying either relevant (such as pixels at object borders) or irrelevant information (such as pixels within an uniform background). This need to transfer all data creates communication bottlenecks.
- On the other hand, they require large power budget owing to the necessity to handle huge amount of data. Both features constraint or even preclude the usage of these architectures whenever either *on-line* analysis or *portability* are required.

For instance, vision-enabled wireless sensor networks employing this kind of architecture use multiple-of-ten μJ/pixel thus greatly constraining portability [18].

For increased efficiency, alternative architectures consisting of *distributed*, multi-core processor arrays are worth considering because they better fits to the peculiarities of the data processing through the *processing chain of vision* [15]. The quest for these alternative architectures becomes particularly necessary for *early vision* tasks due to the huge amount and large redundancy of the data involved in these tasks [22]. This fact has been highlighted in [16] which states that brute force pattern matching, the conventional approach adopted by many vision system developers, is not the right tool in many applications. Instead, sic, "*a majority of smart camera applications can be solved using only a small number of image processing algorithms that can be learned quickly and used very effectively*". Interestingly enough these simple algorithms (thresholds, blob analysis, edge detection, average intensity, binary operators, …) can be mapped down onto dedicated, processor architectures composed of simple processors with mostly *local interactions* [19].

The convenience to devise architectural concepts better suited to images is illustrated with the help of Fig.2, which shows the reduction of the amount of data as information progresses through the processing chain of vision [15]. It corresponds to an applica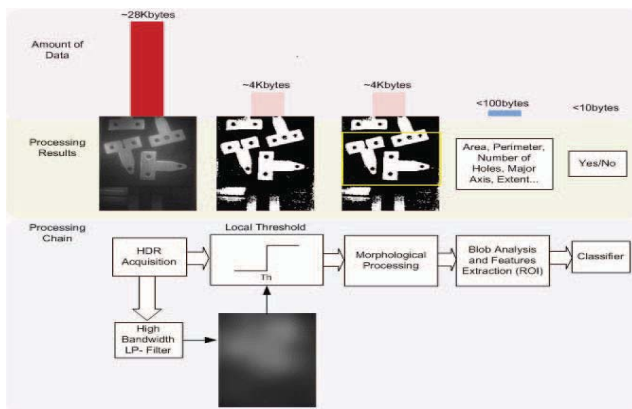tion where the target is detecting defective parts as they move on a conveyor belt. Images are acquired in asynchronous manner and analysed on-line to extract a number of features on the basis of which parts are classified as either defective or correct and a corresponding trigger signal is generated. The progressive data reduction highlighted by this example calls for corresponding progressive processing architectures.

The conception of architectures suitable for progressive processing benefits from the usage of concepts such as *multi-scale* representations, *feature* extraction, *sub-sampling*, and the like. Most efficient architectures employ *mixed-signal* smart-pixels for *parallel* completion of the computational-intensive early vision tasks, followed by sub-sampled topographic processor arrays (typically digital), processors-per-column and scalar processors. Unfortunately, there is not yet a standard, universally accepted set of functions to be incorporated at smart-pixels and most of the contributions are ad-hoc for specific tasks and requirements. Examples of state-of-the-art advanced smart-pixels are presented in the sections below.

A. Smart-pixel HDR CIS for 145dB Intra-Frame Capture

Fig.3 corresponds to a smart-pixel employed into a CIS conceived to capture High Dynamic Range (HDR) images using a content-aware compression law [24]. It employs a *tone mapping* algorithm [25] built into the pixel to achieve more than 151dB DR. Fig.3 depicts the pixel schematics, floorplan (where the architectural feature of heterogeneity is noticeable)
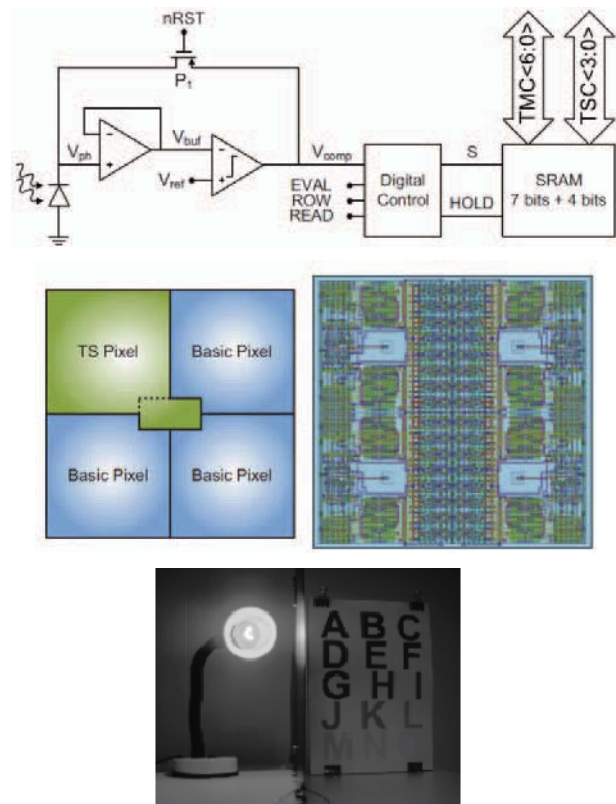

Figure 2. Illustrating the progressive reduction of data as images proceeds through the vision processing chain

Figure 3. Pixel and representative results of a 151dB DR smart-pixel-CIS - Technology CMOS 0,35μm [24].

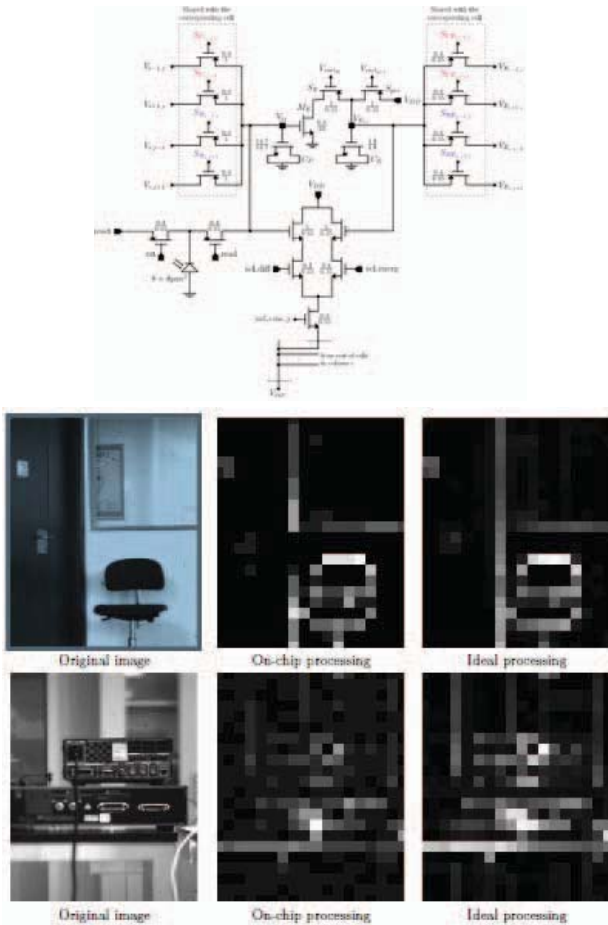Figure 4. Pixel and representative results of a smart-pixel-CVIS intended for programmable spatial-temporal filtering- Technology CMOS 0.35μm [9].

and representative HDR image captures.

B. Smart-pixel CVIS for Programmable Spatial-Temporal Processing

Fig.4 shows the pixel schematics and illustrative processing for a smart-CIS which employs mixed-signal MFPS to realize programmable spatial-temporal filtering [9]. It employs 22nJ/cycle to complete *binomial filtering* operations at nsec rate which makes it very well suited for the front-end of portable vision-enabled wireless sensors. The target of the processing illustrated in the figure is segmenting the zones of the image with the largest changes of intensity, that is, the relative values among the blocks of the scene representation are the key point here. It is done by the chip in [9] by in-pixel energy computation to guide subsequent foveation.

C. High-Speed Vision System-on-Chip

FIG.5 shows the functional diagram of a smart-pixel employed at the front-end of an industrial CVIS that combines on-chip per-pixel and per-chip (32-bit RISC digital processor for post-processing) circuitry to go from image acquisition to decision making at 1,000F/s rate with around 60nW per pixel required.
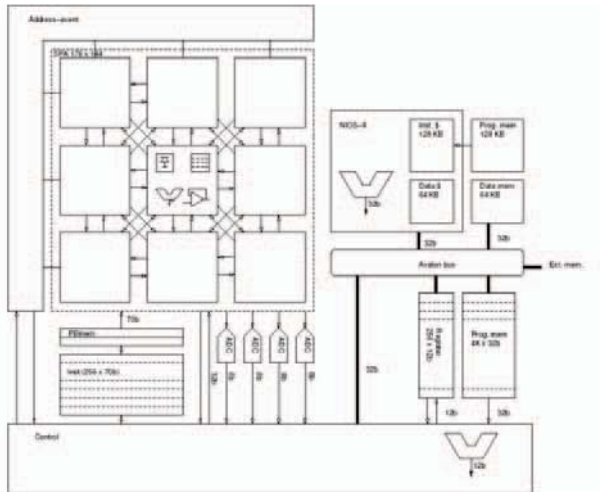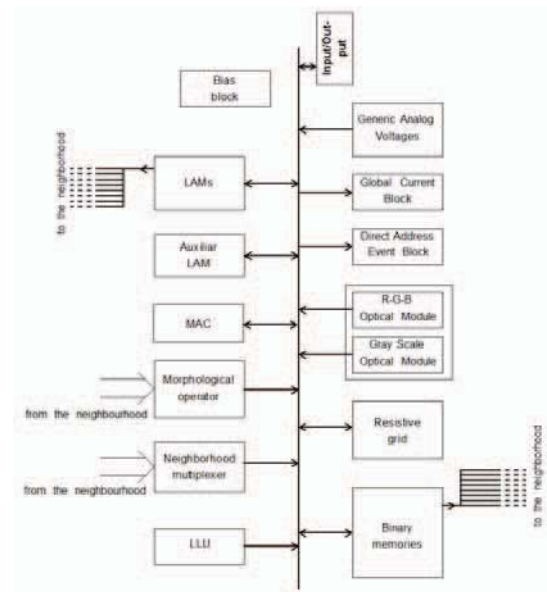

Figure 5. Block diagram of an advanced smart-pixel for machine vision (top figure) and complete chip architecture including a per-chip digital post-processor (bottom figure) - Technology is 0.18μm [17]

For instance, the processing chain of Fig.3 is entirely realized at the sensory plane and the central processor is provided with just 10bytes information indicating if pieces are either deffective or correct.

III. MAPPING SMART-PIXEL CVIS ONTO 3-D ARCHITECTURES

The schematics and/or bock diagrams of the smart-pixels in Figs. 3, 4 and 5 show significant overhead, non-sensing circuitry that penalizes sensor spatial resolution and photo-responsiveness. t is clearly highlighted at the pixel layout of the HDR smart-pixel depicted in Fig.3, where the photosensor occupies 3×3 μm^2 for a pixel pitch of 33×33 μm^2. Vertical stacking of the different functions embedded within pixels can improve the form factor of smart-pixels thus keeping their advantages while overcoming the drawbacks of large pitch. The form factor improvement yielded by 3D technologies is

particularly notorious when complex processing algorithms such as object detection and recognition [26], image retrieval, image registration, or tracking must be implemented. Since these algorithms rely on local properties, *diffusions* and *multiscale* representations play a prominent role for their implementation. This is actually the very domain of analog and mixed-signal circuits [8][27][27].

3D sensor architectures for the extraction of features and the calculation of interest points within imagers have been previously proposed by the authors in [27] and [28]. Both architectures rely on the implementation of Gaussian filtering at the sensory plane and the calculation of Gaussian pyramids thereof. For illustration purposes, Fig.6(a) shows the functional splitting among consecutive layers for the detection of salient points into an image. Also, Fig.6(b) shows an architecture for interest point detection where sensors lie in a dedicated tier as in reference [27], so that different active pixels are possible. The last tier in this architecture is a DRAM memory block which is actually the implementation choice in [29]. These architectures represents a first move towards a general architectural solution where each layers captures a given data abstraction within the hierarchical processing chain of vision. The proposal of such a general architectural solution remains still an open problem.
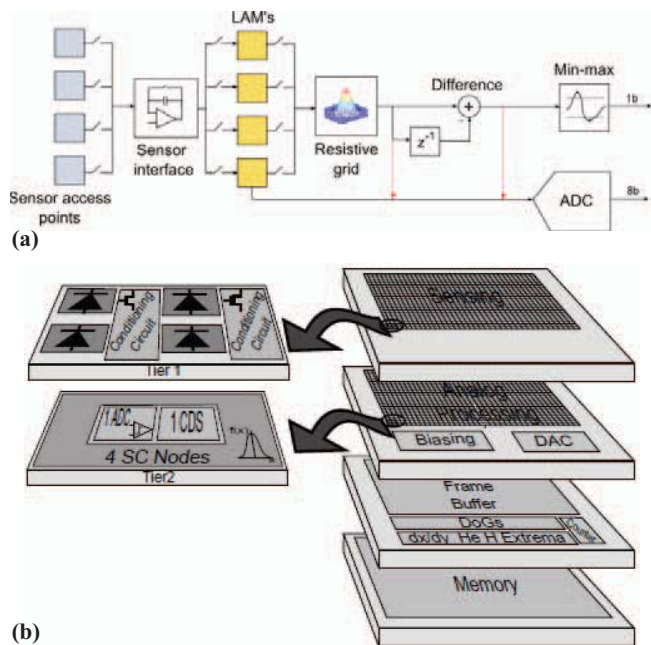
ACKNOWLEDGMENTS

(a)

(b)

Figure 6. (a) Mixed-signal circuit for salient-points location using a 3-D chip architecture [27]; (b) Functionality distribution across tiers on a 3D-IC technology for feature extraction.

REFERENCES

[1] A. El Gamal and H. Eltoukhy, "CMOS image sensors". *IEEE Circuits and Devices Magazine*, vol. 21, no. 3, pp. 6 – 20, May-June 2005.

[2] Teledyne DALSA. Genie HM640. [Online]. Available:http://www.teledynedalsa.com/mv/products/cameradetail.aspx?partNumber=CR-GM00-H640x

[3] L. Nicolosi et al., "A monitoring system for laser beam welding based on analgorithm for spatter detection". *Proc. of the 2011European Conference on Circuit Theory and Design* (ECCTD), pp. 25 –28, aug. 2011.

[4] F. Jiménez-Garrido et al.. "High-Speed Global Shutter CMOS Machine Vision Sensor with High Dynamic Range Image Acquisition and Embedded Intelligence". *Proc. of SPIE-IS&T Electronic Imaging - Sensors, Cameras, and Systems for Industrial and Scientific Applications XIII*, SPIE-IS&T/ Vol. 8298 829803-1-10, January 2012.

[5] J. Crooks et al., "A CMOS image sensor with in-pixel ADC, time stamp, and sparse readout". *IEEE Sensors Journal*, vol. 9, no. 1, pp. 20–28, 2009.

[6] L.-W. Lai et al., "A novel logarithmic response CMOS image sensor with high output voltage swing and in-pixel fixed pattern noise reduction". *IEEE Sensors Journal*, vol. 4, no. 1, pp. 122–126, 2004.

[7] R. Steadman, G. Vogtmeier, A. Kemna, S. Quossai, and B. J. Hosticka,"An in-pixel current-mode amplifier for computed tomography." *IEEE Sensors Journal*, vol. 6, no. 6, pp. 1372–1373, 2006.

[8] A. Rodríguez-Vázquez et al., "ACE16k: the third generation of mixed-signal SIMD-CNN ACE chips toward VSoCs". *IEEE Trans. on Circuits and Systems I*, vol. 51, no. 5, pp. 851-863, May 2004.

[9] J. Fernández-Berni et al., "FLIP-Q: A QCIF Resolution Focal-Plane Array for Low-Power Image Processing". *IEEE Journal of Solid-State Circuits*, vol. 46, pp. 669-680, March 2011.

[10] M.W. Seo et al., "A low noise wide dynamic range CMOS image sensor with low-noise transistors and 17b column-parallel adcs". *IEEE Sensors Journal*, vol. 13, no. 8, pp. 2922–2929, 2013.

[11] M. Dadkhah et al., "Block-based compressive sensing in a CMOS image sensor". *IEEE Sensors Journal*, vol. 12, no. 99, pp. 1–1, 2012.

[12] [7] R. Fontaine, "Recent innovations in CMOS image sensors". *Proc. of the 2011 Conference in Advanced Semiconductor Manufacturing* (ASMC), pp. 1 –5, may 2011.

[13] R. Fontaine, "A Review of the 1.4um Pixel Generation". *Proc. IEEE 2011 Int. Image Sensor Workshop*, pp 5-8, Hokkaido-Japan, June 2011.

[14] K. Itonaga et al., "0.9μm pitch pixel CMOS image sensor design methodology". *Proc. of the 2009 IEEE International Electron Devices Meeting* (IEDM), pp. 1 –4, dec. 2009.

[15] J.C. Russ, *The Image Processing Handbook.* CRC Press 1992.

[16] G. Devaraj et al., "Applying Algorithms". *Vision System Design*, Vol. 13(11), pp. 17-20 and 85-87, 2008.

[17] A. Rodríguez-Vázquez et al., "A CMOS vision system on-chip with multi-core, cellular sensory-processing front-end". Chapter 6 in *Cellular Nanoscale Sensory Wave Computers* (edited by C. Baatar, W. Porod and T. Roska). Springer 2010.

[18] J. Fernández-Berni et al, *Low-Power Smart Imagers for Vision-Enabled Sensor Networks*. ISBN 978-1-4614-2391-1, Springer, May 2012.

[19] T. Roska and A. Rodríguez-Vázquez. *Towards the Analogic Visual Microprocessor.* John Wiley & Sons, Chichester 2001.

[20] T. Laforest et al., "Algorithm Architecture Co-Design for Uktra Low-Power Image Sensor". *IST&SPIE Electronic Imaging*, San Francisco, January 2012.

[21] W. Zhang et al., "A Programmable Vision Chip Based on Multiple Levels of Parallel Processors". *IEEE J. Solid-State Circuits*, vol.46, no.9, pp.2132-2147, Sept. 2011.

[22] C. Tomasi, "Early Vision". *Encyclopedia of Cognitive Sciences. Nature.* Pub. Group. McMillan 2002.

[23] P. Garrou et al., *Handbook of 3D Integration: Technology and Applications of 3D Integrated Circuits*, Wiley-VCH, 2008.

[24] S. Vargas-Sierra et al., "A 145dB Focal-Plane Tone-Mapping QCIF Imager". *Proc. of the 2012 IEEE International Symposium on Circuits and Systems* (ISCAS 2012), pp. 1616-1619, Seoul-Korea, May 2012.

[25] E. Reinhard et al., *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting.* Elsevier Morgan Kaufmann, 2006.

[26] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints". *International Journal of Computer Vision*, Vol. 60, pp. 91-110, 2004.

[27] R. Carmona-Galán et al., "A Hierarchical Vision Processing Architecture oriented to 3D Integration of Smart Camera Chips". *Journal of System Architectures* (special issue on *Smart Camera Architectures*), 2013. (http://dx.doi.org/10.1016/j.sysarc.2013.03.002).

[28] M. Suárez et al., "CMOS-3D Smart Imager Architectures for Feature Detection". *IEEE J. on Emerging and Selected Topics in Circuits and Systems*, Vol. 2, pp. 723-736, December 2012.

[29] Tezzaron Semiconductors. http://www.tezzaron.com/.