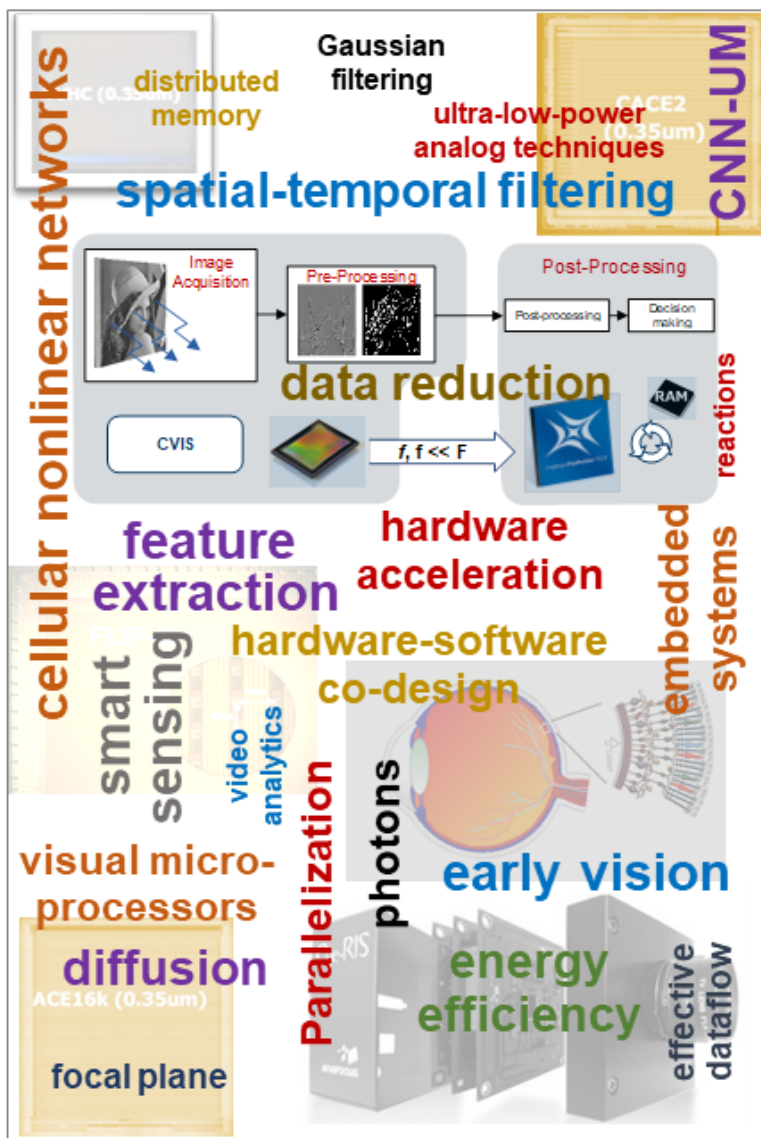# CMOS Vision Sensors: Embedding Computer Vision at Imaging Front-Ends

A. Rodríguez-Vázquez, J. Fernández-Berni, J.A. Leñero-Bardallo,I. Vornicu and R. Carmona-Galán
IMSE-CNM (Universidad de Sevilla - CSIC)
arodri-vazquez@us.es; angel@imse-cnm.csic.es

Abstract

CMOS Image Sensors (CIS) are key for imaging technologies. These chips are conceived for capturing optical scenes focused on their surface, and for delivering electrical images, commonly in digital format. CISs may incorporate *intelligence*; however, their smartness basically concerns calibration, error correction and other similar tasks. The term CVISs (CMOS VIsion Sensors) defines other class of sensor front-ends which are aimed at performing *vision tasks* right at the *focal plane*. They have been running under names such as *computational image sensors*, *vision sensors* and *silicon retinas*, among others.

CVIS and CISs are similar regarding physical implementation. However, while inputs of both CIS and CVIS are images captured by photo-sensors placed at the focal-plane*,* CVISs primary outputs may not be images but either image *features* or even *decisions* based on the *spatial-temporal* analysis of the scenes. We may hence state that CVISs are more "intelligent" than CISs as they focus on information instead of on raw data. Actually, CVIS architectures capable of extracting and interpreting the information contained in images, and prompting *reaction* commands thereof, have been explored for years in academia, and industrial applications are recently ramping up.

One of the challenges of CVISs architects is incorporating *computer vision* concepts into the design flow. The endeavor is ambitious because imaging and computer vision communities are rather disjoint groups talking different languages. The Cellular Nonlinear Network Universal Machine (CNNUM) paradigm, proposed by Profs. Chua and Roska, defined an adequate framework for such conciliation as it is particularly well suited for *hardware-software co-design* [1]-[4]. This paper overviews CVISs chips that were conceived and prototyped at IMSE Vision Lab over the past twenty years. Some of them fit the CNNUM paradigm while others are tangential to it. All them employ per-pixel *mixed-signal* processing circuitry to achieve sensor-processing concurrency in the quest of fast operation with reduced energy budget.

I. Introduction

A picture is worth a thousand words. The scope and meaning of this sentence are evident for human beings. Images carry the largest percentage of data involved in our interaction with the environment, and we employ more than 50% of our brain processing capabilities for handling visual scenes [5]. The same happens for many animals, and we do not need many arguments to get convinced of the advantages of conferring vision capabilities to *artificial sensory systems*.

However, large-scale deployment of imaging technologies was traditionally limited by cost, Size, Weight and Power (SWaP) constraints. Sensors employed mostly CCD technologies and delivered *analog* images. *Camera systems* built with these sensors were costly, bulky and power hungry. These drawbacks were particularly notorious for systems with *visual analysis* capabilities, thus rendering vision unfeasible for many applications. This scenario has recently changed owing to advances on CMOS *pixels* and CMOS Image Sensors (CIS) architectures, semiconductor technologies, packaging technologies, heterogeneous integration, and system-on-chip architectures, among others [6]-[9]. All-in-all, these advances have enabled imaging systems with reduced SWaP, low cost, large speed and large functional capabilities and flexibility. Given the relevance of the visual sense, it is not surprising that such increased availability had resulted in imaging technologies flooding practically all application territories. The growth rate of inventions and revenues concerning these technologies have been impressive; the number of IP assets and the revenue have been multiplied roughly by five in the last decade [6][10]. The imager market is dominated by smartphones, notebooks, tablets and other consumer equipment, but other sectors are rapidly growing [6]. For instance, most modern automobiles include several cameras to continuously monitor the outside and the inside, and many are capable of detecting pedestrians, classifying road-signs, and other tasks.

CISs have progressed towards ever smaller *pixel pitch*, and an ever larger image resolution (number of pixels). Besides pixel scaling, other CIS challenges include [6]-[10]:
- enhancing the image quality by improved readout, signal conditioning, and image enhancement circuitry;
- boosting the image downloading speed by improved communication techniques and;
- reducing the area, power, and cost by on-chip circuit embedding.

Recent milestones include multi-million-pixel sensors with pixel-pitch around 1μm, data rates above 10Gpx/s [11], reconfigurable A/D conversion and readout architectures [12][13], image correction, thermal and energy management, etc. [9].

The last few years have also witnessed ever-increasing activities towards adding the estimation of *depth*, i.e. 3-D information, to 2-D scenes. One main driver is human-machine interfaces for the entertainment industry [14], but these technologies are also applicable to surveillance, automotive, industrial inspection and medicine, among other sectors with huge development potentials. Besides techniques based on stereoscopy, triangulation and the like, significant efforts are being made towards modifying CMOS pixels for capturing time information and estimate depth through Time-of-Flight (ToF) techniques. Imaging arrays consisting of Single Photon Avalanche Diodes (SPADs) pixels are receiving significant interest [15]-[18]. However, ToF measurements require active illumination that complicates system implementation. Also, 3-D sensors are still lagging behind mainstream CISs regarding the incorporation of on-chip processing circuitry.
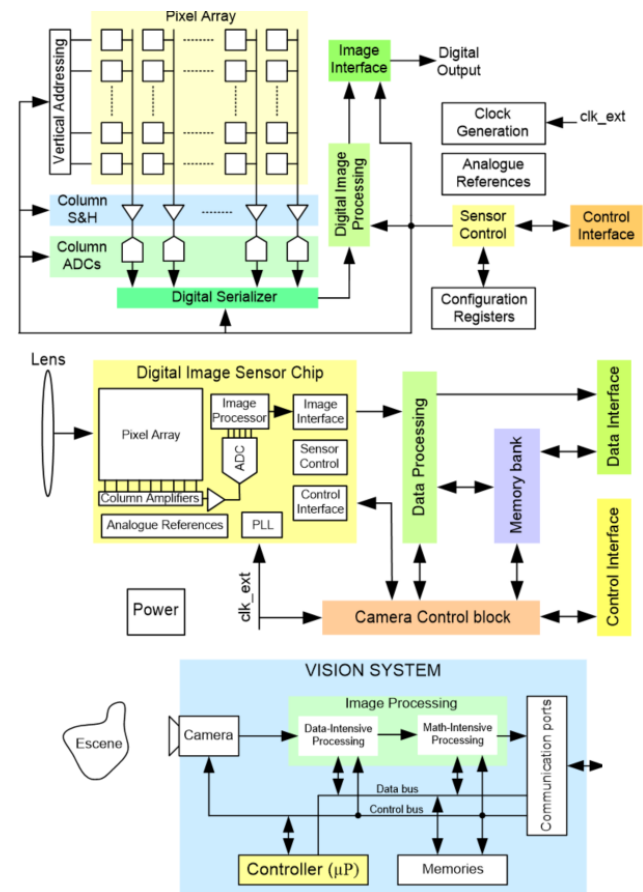
Figure 1. Conceptual block diagrams for an imager (top), a camera (mid) and an embedded vision system (bottom). The trend is towards full integration of these systems. Progresses in semiconductor technologies, heterogeneous integration and packaging enable compact implementations of these systems in the form of Systems-on-Chip and/or System-in-Package.

CVISs, the main characters of this paper, are similar to CIS regarding physical implementation; they both include photo-sensors and CMOS processing primitives on a common silicon substrate. Also, both CIS and CVIS chips are similar in that they can be used as *front-end* devices of complex *hardware-software* vision systems [19]-[21]. Roughly speaking the front-end captures images and delivers data for subsequent processing by digital processors. However, it is well known that raw pixel data are largely redundant, and that information contained into images can be extracted from reduced subsets of the spatial samples [22]. Consequently, vision systems architectures employing CISs at the front-end must read, encode, transmit and store myriads of irrelevant data.

However, CVISs are *information-centric* front-ends conceived to deliver information, instead of raw data [23]-[33]. Therefore, using CVISs at the front-end of visions systems yields smaller SWaP and larger throughput, as required for wireless sensor networks, unattended surveillance networks, automotive, low payload UAVs, visual prosthesis and internet-of-the-things, and in general whenever portable vision is required.

While CISs are solid industrial assets, CVISs are still lagging behind regarding industrial exploitation. Among other obstacles, they are lacking standardization. However, their concept has been already demonstrated through many silicon implementations and their industrial use is starting to ramp up [20].

II. CNNUM-based Visual Microprocessors

The Vision Processing Chain

Retinas are considered key for the outstanding performance of natural vision systems [34][35]. It is hence arguable that artificial vision systems will largely benefit from using front-ends with architecture and operation similar to retinas. Actually, many CVISs are named *silicon retinas* in the literature [36]-[38]. Retinas contain photo-receptors and *dynamically-coupled* processing cells of different types. They are able to complete complex spatial/temporal processing tasks to extract relevant information from the incoming sensory data, thus reducing the amount of data before transmitting them for subsequent processing. This data reduction strategy is justified by the specifics of the processing chain of vision – illustrated in Fig.3.

This figure highlights the steps to go from sensor raw data
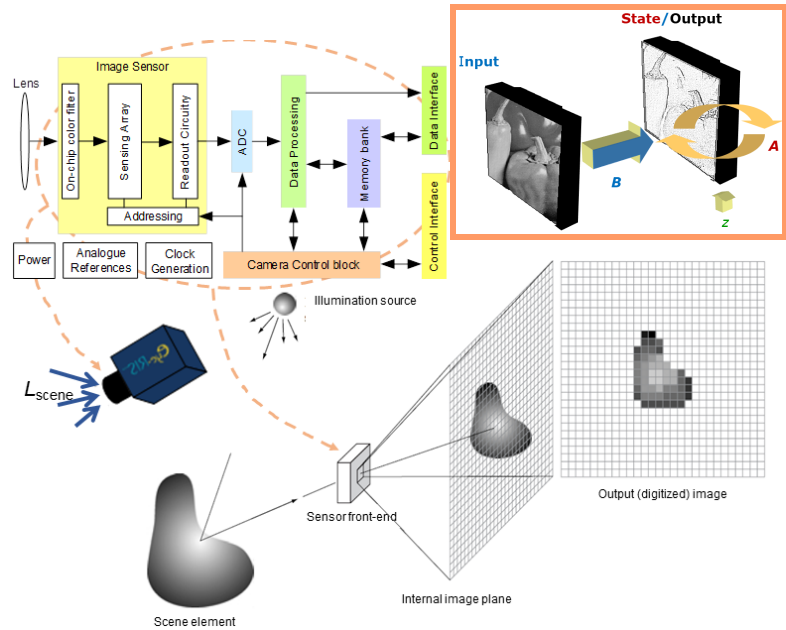
Figure 2. CISs and CVISs are front-end chips that embed photo-sensor arrays to capture scenes focused on the chip surface.
CISs produce electrical images consisting of digitized raw pixel data (bottom-right side in the figure – adapted from [39]).
CVISs, include processing structures in the focal plane to perform spatial/temporal tasks concurrently with acquisition, thus producing pre-processed, compressed images (top-right inset).
Eye-RIS^{TM} is a Trademark of AnaFocus Ltd [20].

to vision outcomes. The vertical axis represents data dimensions while the horizontal one represents the abstraction level of the data. The processing chain follows the diagonal arrow in Fig.3(a). Top-left corresponds to input data captured by the sensor and bottom-right corresponds to output data which support system actions. The first stage of the vision processing chain is usually devoted to *image enhancement* and *restoration*. During this stage, non-idealities of the sensing process are compensated and the quality of captured images is improved in relation to selected image features. This is achieved by applying several filters (*convolution* masks, *diffusion* process, etc.) and by performing *point-to-point* transformations. The output data provided by enhancement and restoration tasks is still a matrix of real numbers, which are the input of a second stage where *feature extraction* tasks are performed. These operations typically examine every pixel to verify if there is a feature present at that pixel considering its neighborhood. Of interest for subsequent processing are *edges*, *corners* or *interest points*, *blobs* or *region of interest*, *ridges*, etc. Outputs of this second stage form an irregular flow of data which are the inputs for the *high-level* vision processing tasks [39] [40].

Fig.3(b) illustrates vision-chain data evolution by means of an application example where the target is detecting
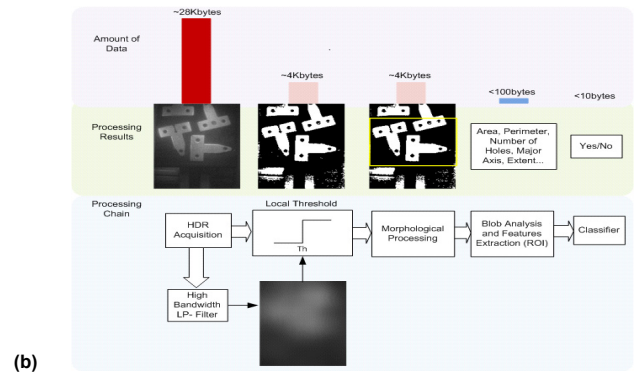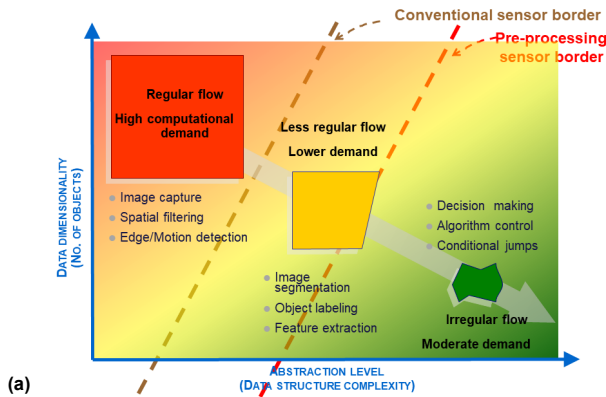
Figure 3. (a) Processing hierarchy, from left-top to bottom-right, in vision; **(b)** Illustrative example of vision processing chain [19].

defective parts as they move on a conveyor belt. This application is entirely executed by the EyeRIS™ system-of chip visual microprocessor of AnaFocus Ltd. [20]; the only data downloaded from the chip are those codifying the classification decision. Images are acquired in an asynchronous manner and analyzed *on-line* to extract several features on the basis of which the pieces are classified as either defective or correct and a corresponding trigger signal is generated. Data reduction and increased abstraction levels as data progress across the chain are highlighted at the figure lettering.

Dashed oblique lines in Fig.3(a) mark where front-end borders are located in different vision system architectures. The left-side line corresponds to systems that employ CISs at the front-end; the right-side corresponds to systems that employ CVISs. Operations at the left of each borderline are performed by the corresponding system front-end while operations at the right are completed by the remaining hardware components using as inputs the front-end outputs. Note that CVISs enable placing the border at a stage of the chain where data have been reduced through *early processing* − similar to what retinas do [35]. Hence vision systems built with CVISs front-ends have the potentials for larger speed and better SWaP than those built with CISs.

Fig.4 highlights differences between architectures with CIS and CVIS front-ends respectively. Note that CVISs sense and pre-process in a concurrent manner, thus sending for subsequent processing an amount of data, represented by f that is much smaller $(f \ll F)$ than the number F of raw sensor data. Indeed, in the architecture of Fig.4(b), processing is performed progressively by distributing processing tasks between the front-end and the core processor sections.

Concept of CNN-Based Visual Microprocessors
There are two general classes of CVIS architectures:
- Specific-purpose ones pick up a specific task and implement it on silicon. This is quite common also for other neuro-morphic systems [36][37][41][42].

- "General-purpose" mixed-signal *visual microprocessors* [3]. That is, processors which combine optical sensing with analog cellular spatial-temporal dynamic circuits and some form of logic. These processors have elementary instructions mapped onto *receptive fields* [4], and embed the possibility of storing and executing user-selectable sequences of instructions.

Visual microprocessors architectures based on the CNNUM paradigm aim at combining the best of analog and digital worlds. On the one hand, analog circuits are known to excel concerning SWaP; they are smaller, faster and require less energy than digital ones for tasks with limited *signal-to-noise-ratio* requirements [43][44]. Among other advantages, analog techniques fully exploit the functional capabilities offered by basic VLSI design primitives, and
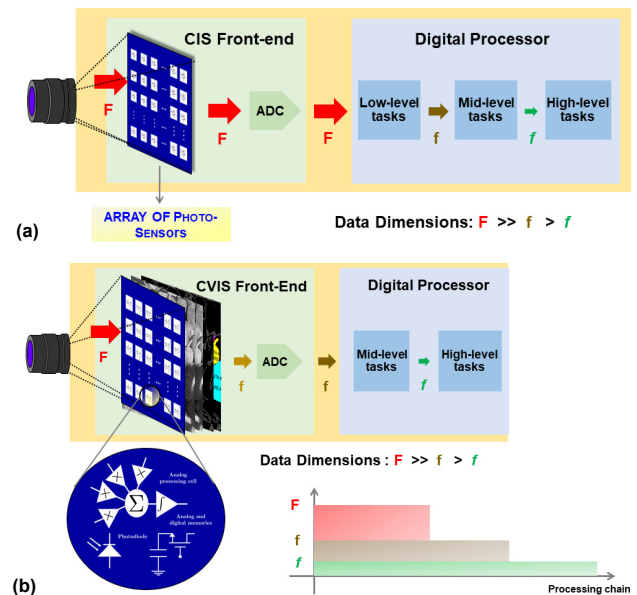

Figure 4. Conceptual vision system architectures with: **(a)** a CIS front-end; **(b)** a CVIS front-end.
CVISs perform early vision tasks right at the front-end, thereby bringing much less number of data into play and hence significantly reducing memory, bandwidth and computation payloads [19].

particular by MOS transistors. to implement a large variety of circuit blocks with a minimum number of devices [3][42]. On the other hand, digital circuits excel regarding controllability, flexibility, and robustness.

Back in the 1960s, the building blocks for logic design had been the various logic circuits (micro-modules) implementing different "smart" logic tasks. These had also been used to make digital computers. The digital computer has a key attribute due to J. Von Neumann, namely *stored programmability*. It means that the same core architecture, via *algorithms* coded in *software*, can be used for a myriad of tasks. Or, to put it in another way, the architecture is open to the human intellect for millions of algorithmic innovations. This is the functional secret behind the success of the digital microprocessor, first made in the early 1970s.

CNNUM-based visual microprocessors aim at mimicking this functional secret. However, they are mixed-signal devices which realize analog-and-logic spatial/temporal processing tasks and hence require quite different building blocks [3]. One key aspect of visual microprocessors is the integration of sensing and *stored programmable processing* (SPP) at the analog signal array level. Among other things, this allows us to tune the sensors dynamically, pixel by pixel, depending on the content and even on the context of the changing scene.

CNNUM-based visual microprocessors belong to the general class of *topographic* smart sensors processors formed by an array of processing cores. Some features which make CNNUM different from other topographic processors include the following:

- They elementary processors (*cells*) are intrinsically mixed-signal processors which mutually interact with tunable interaction weight patterns.
- Data memories are embedded per-pixel to locally store partial processing outcomes that are further employed to either generate global processing outcomes or control the sequence of processing steps.
- This programmable and reconfigurable array is embedded in a computer architecture resulting in CNNUM general-purpose architecture.
- The CNNUM is stored programmable and capable of implementing mixed-signal spatial/temporal algorithms through the smart synergy of hardware and software.

All the signal variables are *continuous*, except for the discreteness in space (pixels). At the same time, visual microprocessors retain the extraordinary strength of digital computers, their unconstrained variability via programming or software. Obviously, such software and related algorithms are different from conventional ones.
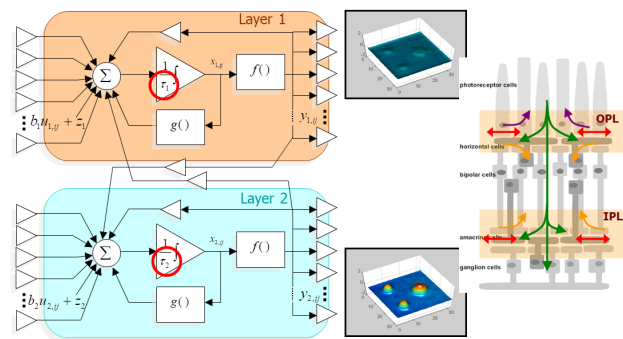
Figure 5. Coupled processing layers and waves generated by the CACE CNNUM chip emulating the behavior of Inner and Outer Plexiform Layers in mammal retinas [45].

Functional mechanisms underlying CNNUM processing capabilities are briefly summarized at the top-right inset of Fig.2. It shows that processing pixel cells include circuits structures to control the dynamic evolution of a dynamic state driven by:

- two-dimensional *gain transformations of the* inputs (matrix B);
- two-dimensional *offset factors* (parameter z); and
- subjected to dynamic interactions among the states of neighbour cells (matrix A).

Complex spatial/temporal tasks can be performed by proper setting of the parameters. Also, programs can be executed by executing algorithms in a process where parameter values are changed by software, thus covering a very large variety of vision tasks [4].

Also, as Fig.5 illustrates, bio-inspired models that mimic the way in which images are processed at retina visual pathways can be implemented by extending the CNNUM concept to include two state variables per cell. Complex spatial-temporal dynamics can be generated in this way to achieve computation through waves. The outcome of such processing can be used to develop control feedback actions to adapt the response of photo-receptors to local image features. Besides simple resistive grid filtering, it is possible to program other spatial-temporal processing operators into the model core, such as non-linear and anisotropic diffusion, among others.

On CVIS Architectural Choices

Most efficient CVISs architectures employ *mixed-signal* Multi-Functional sensory-processing PixelS (MFPS) for *fully-parallel* completion of computational-intensive early vision tasks, followed by sub-sampled topographic processor arrays (typically digital), *processors-per-column* and *scalar* processors [33]. MFPSs actually make the next evolutionary step of CMOS pixels, after *passive* pixels (PPS) and *active* pixels (APS), by embedding within the pixel
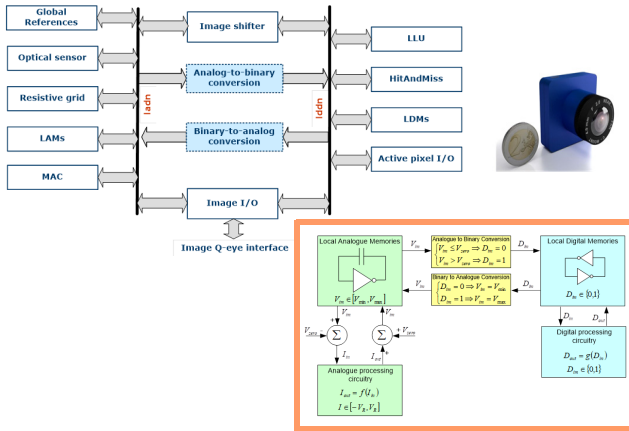
Figure 6. (a) Concept of vision sensor with CVIS front-end; (b) Architecture of a Q-Eye™ cell [19][20][27].

resources for analog processing, memory and programming and control of information flows [27]. However, embedding circuitry per-pixel enlarges the pixel pitch and may result in spatial sampling *aliasing* artifacts. Despite considerations concerning the number of pixels required for vision tasks [22], design trade-offs arise which may require alternative architectural solutions. A sound strategy is resorting to 3-D, *vertically integrated* technologies for improving the pixel foot-print by distributing the different circuit types across different physical layers [46][47]. This is already happening in CIS-APSs and we are convinced that CVIS roadmap will evolve towards 3-D architectures. Other alternatives include using *per column* processors and a sub-sampled *topographic* array of processors, among others.

Fig.6 illustrates the functional structures encountered within an industrial MFPS, namely the Q-Eye™ pixel. This CVIS is the front-end of the Eye-RIS™ vision system from AnaFocus Ltd [20], which picture is included at the top-right in the figure. The figure inset at the bottom-right highlights the different signal modalities included per pixel.

III. Illustrative CVIS Chips

ACE and CACE Chips

ACE (Analogic Cellular Engines) and CACE (Complex ACEs) were devised and designed by the IMSE vision lab over around ten years, following the proposal of improved mixed-signal circuits for analog processing and memory [3][44][48][49]. These chips employed the CNNUM paradigm and were designed for robust analog behaviour owing to the extensive use of *dynamic biasing*, *error correction* and *calibration* loops. ACE and CACE chip milestones are shown in the roadmap of Fig.7, which epitomes were the CACE2 chip [45] and the ACE16k-v2 [50] chip.

These chips demonstrated the concept of CNNUM and the

viability of ultra-fast vision front-ends with small SWaP. Both were fabricated using standard 0.35μm CMOS technology. The ACE16k-v2 displayed peak computing figures of 330GOPs with 3.6GOPs/mm2 and 82.5GOPs/W. It performed linear convolutions on 3x3−neighborhoods in less than 1.5μs, image-wise Boolean combinations in less than 200ns, image-wise arithmetic operations in about 5μs, and CNN-like temporal evolutions with a time constant of about 0.5μs. Regarding CACE2, this chip opened vistas for application of the CNNUM paradigm to the emulation of the dynamic phenomena observed in mammalian retinas.

ACE architectures prompted the launching of the start-up company AnaFocus Ltd. in Sevilla-Spain; they were also transferred to the hungarian start-up AnaLogic Ltd.

Eye-RIS™ Visual Processor On-Chip

Fig.8(a) shows the block diagram of the Eye-RIS™ vision system on a chip [20]. It embeds a CVIS front-end, a Digital Image Processor (DIP), a microprocessor, memories and I/O and communication ports. CVIS architecture follows a modified version of the CNNUM paradigm, similar to Single Instruction Multiple Data (SIMD) processors, consisting of
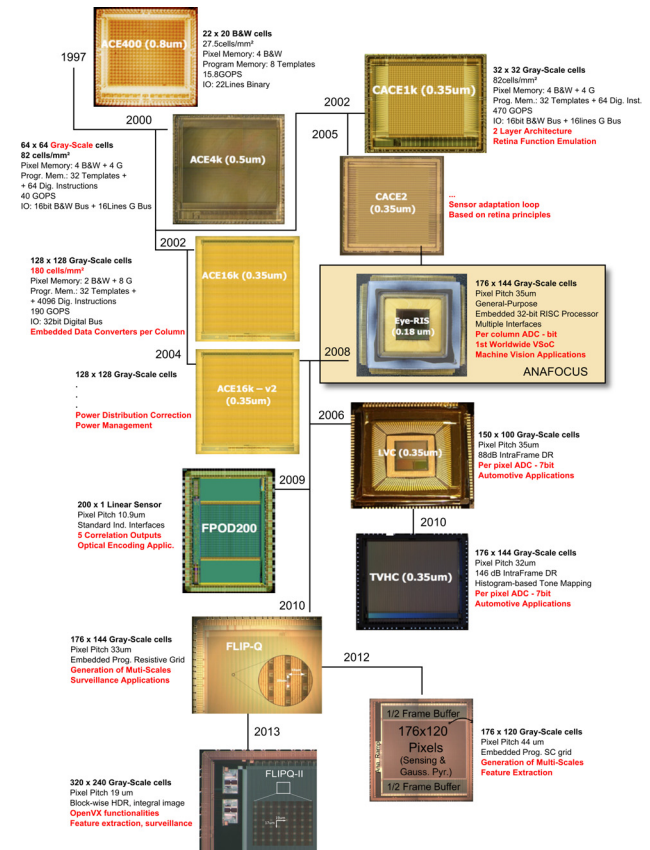

Figure 7. Roadmap of vision chips with CNNUM architecture designed at the vision lab of Institute of Microelectronics of Seville (The bottom-right one was designed in collaboration with CITIUS-Universidad de Santiago de Compostela).
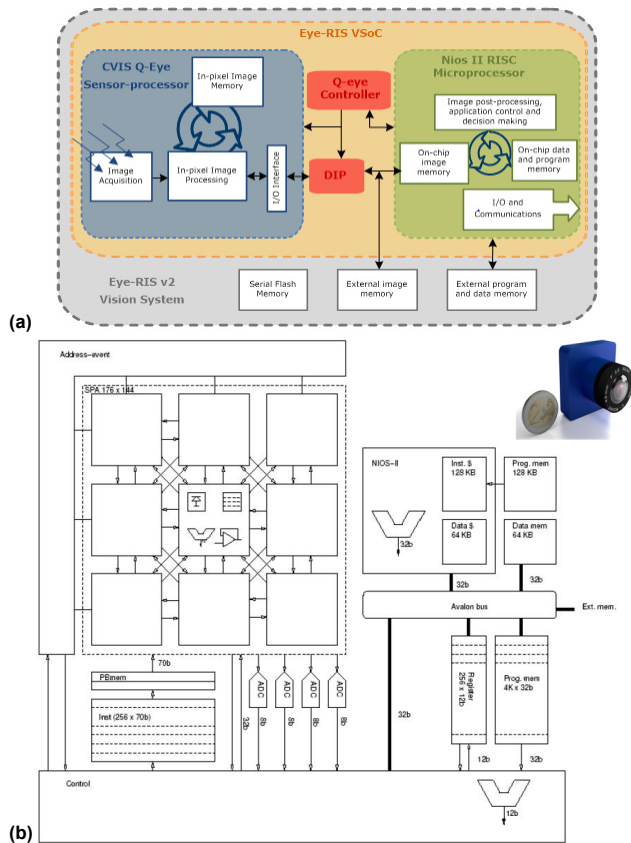
Figure 8. Eye-RIS^{TM} vision system: **(a)** Block diagram; **(b)** Architecture [19][20][27].

an array of interconnected mixed-signal processors, one per pixel, that operate in parallel – see Fig.8(b). Since the CVIS is software-controllable, the systems must include a dedicated microprocessor to control and configure its operation. Users can defined a particular algorithm or sequence of operations through the NIOS microprocessor, and the microprocessor of the CVIS controller sends the microinstructions through the control interface.

Architecture and parameters of this CVIS are conceived for efficient completion of pre-processing vision tasks. The implementation of regular algorithms in hardware involves mapping of operations onto dedicated processing elements and representation of data dependencies by hardware interconnections or intermediate memories. For regular algorithms of image processing, array processors are typically derived as appropriated hardware structures. Favourable properties of array structures are the incorporation of parallel processing and pipelining and the locality of connections between processing elements. Thus, high performance and throughput are obtained at moderate hardware expense.

Parallelism and the use of mixed-signal circuitry enable going from sensing to actuation at rates about 1kF/s rate with around 60nW per pixel. Also, software programming of the

front-end features large flexibility to cope with a wide range of machine vision applications.

Low Power CVIS for Gaussian Pyramid Extraction

Compatibility with computer vision tools is cornerstone for CVIS adoption and can be achieved by focusing on the embedding of pre-processing functions customarily used by computer vision system engineers. This is actually the case of *image pyramids*, such as the *Gaussian pyramid* [51]. 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, calculating them with CVIS-SIMDs may represent a first step towards embedding complete computer vision on a single die with vision capabilities into SWaP sensitive systems such as vision-enabled wireless sensor networks [52] or unmanned aerial vehicles [53].

Fig.9 shows the block diagram of a MFPS micro-photograph of a 176x120 resolution CVIS designed to extract the Gaussian pyramid [26]. This pyramid is generated by using a *switched-capacitor* network embedded per MFPS. In order to shorten routing length and speed I/O operations up, the image is read out through two frame buffers outside the MFPS array. Each MFPS is connected to two 8-bit registers in the corresponding frame buffer, allowing for reading out pixels outside the chip as they are being A/D converted.

The scene is acquired with 4 3T-APS per MFPS with nwell/p-subs. photo-diodes. Every MFPS contains the local circuitry of an 8-bit single-slope ADC and one circuit to perform correlated double sampling. Also, the MFPS comprises 4 state capacitors with their corresponding switches along the four cardinal directions to configure a double-Euler switched-capacitor network that yields the Gaussian pyramid.

Fig.10 shows the image acquired by the chip and several snapshots of the Gaussian pyramid for different values of the
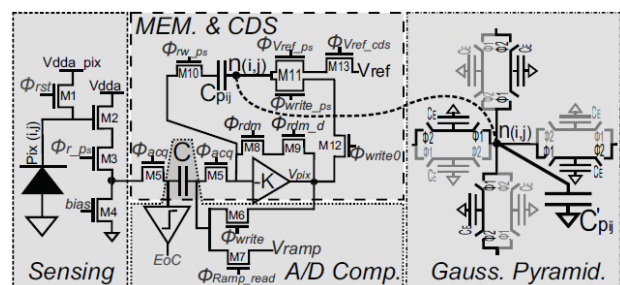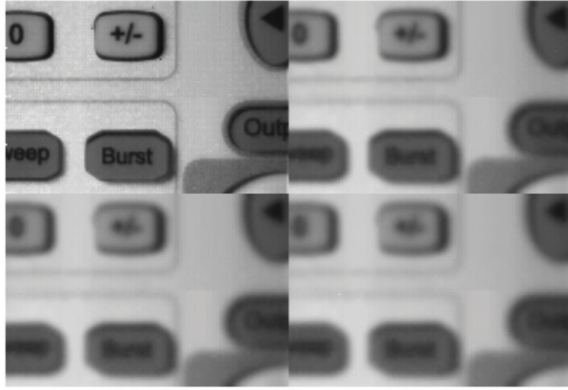

Figure 9. MFPS block diagram of Gaussian pyramid CVIS chip [19][26].

Figure 10. Image acquisition and different snapshots of the on-chip Gaussian pyramid. The upper left image is the input scene, the rest of the images from left to right and top to down correspond to $\sigma=1,77$ (clock cyles $n=19$), $\sigma=2,17$ ($n=29$), and $\sigma=2,51$ ($n=39$) [19].

diffusion width, σ, and the number of clock cycles n. Deviations between images filtered on-chip, on the one hand, and filtered by a conventional computer, on the other hand are tolerated by the SIFT algorithm [26].

This chip consumes 70mW with scene acquisition and the Gaussian pyramid of 3 octaves with 6 scales each. The Gaussian pyramid is executed in 8ms (A/D conversions included), with 200μs per AD conversion, and 150ns as the clock cycle for the switched-capacitor network. This leads to 26.5nJ/px at 2.64Mpx/s. As compared to conventional architectures consisting of a CIS front-end and a conventional MPU (even a low-power MPU), this CVIS chip features around three orders of magnitude energy consumption reduction while having similar or faster processing speed.

Multifunctional Feature Extraction Sensor

Embedded camera systems for markets like smart surveillance or wearable devices need to operate with a tight power budget. They also need to cope with a vast range of illumination conditions, and at the same time, they need to incorporate intelligent features dictated by high-level specifications. This can be achieved by using CVISs with MFPSs tailored for meeting these stringent requirements. An example developed at IMSE vision lab for environmental monitoring is shown in Fig.11.

There is also a growing demand for *privacy-aware* visual systems. MFPS-based CVISs can implement privacy policies from the very beginning of the signal processing chain. Fig.12 shows the architecture and the pixel of a CVIS conceived specifically for dynamic range adaptation and privacy-aware Region-of-Interest (RoI) tracking. The central
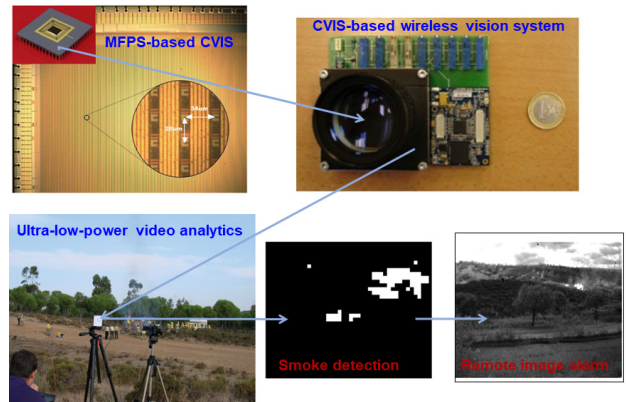
Figure 11. Ultra-low-power smart surveillance for forest fire detection as an application scenario of CVISs [54].
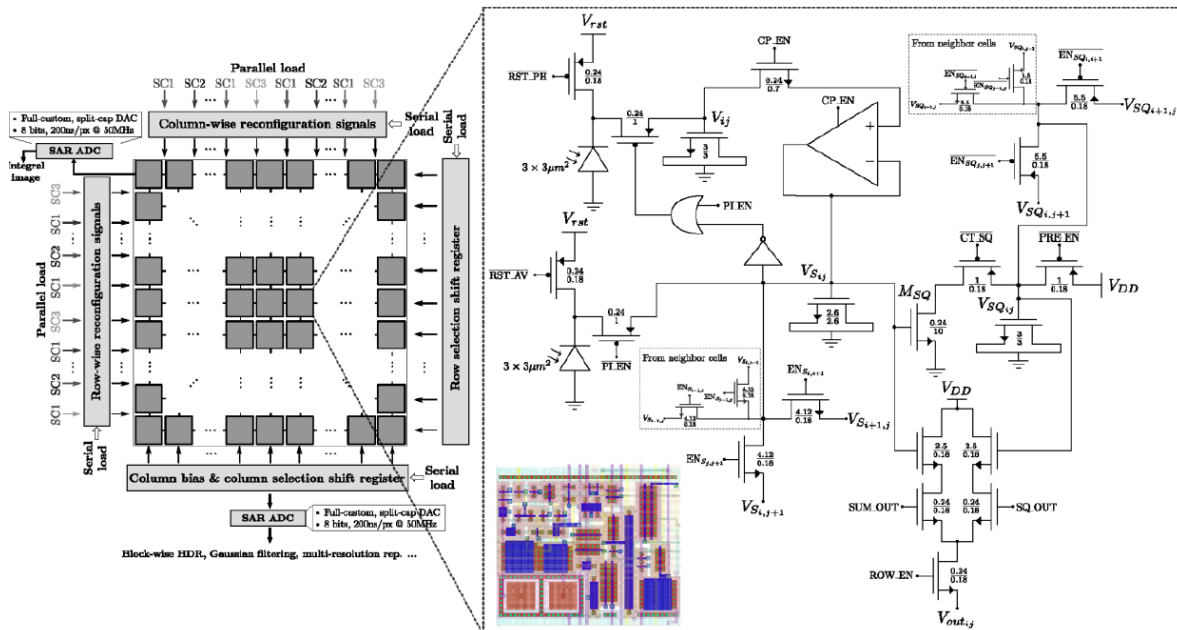
Figure 12. Functional diagram of the chip architecture and schematic of the pixel of a CVIS for privacy-aware applications [19] [55].

element is an array of 4-connected mixed-signal MFPSs. Each MFPS contains two photo-diodes. One of them is responsible for generating the pixel value by integrating the photo-current in a sensing capacitance. The other photo-diode generates a replica of this voltage value that is initially stored. This stored voltage at this node will be employed later to evaluate the average value of different neighbourhoods. The array can be divided into different regions by means of control lines distributed along the horizontal and vertical edges of the array [55], which are operated by peripheral control blocks and selection registers. These registers can be serially updated with different interconnection patterns. There is also the possibility of setting up six different successive pixelation scales, with patterns that can be loaded in parallel for fast reconfiguration.

On-chip programmable pixelation can be implemented in this chip by combining focal-plane reconfigurability, charge redistribution and distributed memory. Right after photo-current integration, all the pixels in the image are represented by their respective voltages; then these values are copied and stored in parallel, what takes only 150ns and is non-destructive. This is important to avoid artifacts due to obfuscation. Once the stored voltages are set, the adequate interconnection pattern must be established. Parameters like RoI address and the required degree of obfuscation are provided by the algorithm. These patterns, activated by the corresponding control signals, enable charge redistribution among the connected capacitors, thus averaging selected areas of the image. The rest remains the same, so privacy-protection is implemented at chip level. No sensitive information is delivered by the sensor.

IV. Conclusions

Applications targeting image analysis instead of just capture are gaining relevance within the ecosystem of imaging technologies and are expected to grow at rapid pace in the near future. Advances in sensor technologies, heterogeneous packaging and processor embedding help to reduce the SWaP of *vision systems*. However, architectural changes may also be required for large-scale deployment of vision systems in applications demanding minimum SWaP and large speed. It is well known that the front-end is key for the outstanding performance and energy efficiency of natural vision systems. CVIS are meant to emulate functional attributes of these natural front-ends, namely parallelism, sensor-processor concurrency, and data reduction. CVISs embed computer vision techniques at the sensor focal plane, and are suitable candidates for replacing conventional imagers (CIS) as front-ends of efficient vision systems. The CNNUM paradigm, devised by Profs. Chua and Roska, provided a unifying framework for conciliating CVIS chip design with digital processor concepts, on the one hand, and computer vision concepts, on the other hand. CVISs chip prototypes devised at IMSE Vision Lab and AnaFocus Ltd. demonstrate the suitability of compact, fast and energy efficient vision systems.

References

[1] L. O. Chua and T. Roska, "The CNN Paradigm". IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, Vol. 40, No.3, pp. 147-156, March 1993.

[2] T. Roska and L.O. Chua, "The CNN Universal Machine: An Analogic Array Computer". *IEEE Tran. on Circuits and Systems-II: Analog and Digital Signal Processing*, Vol.40, pp. 163-173, March 1993.

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

[4] L.O. Chua and T. Roska, Cellular Neural Networks and Visual Computing. Cambridge University Press, Cambridge-UK 2002.

[5] L.M. Chalupa and J.S. Werner, The Visual Neurosciences. MIT Press 2004.

[6] Yole Development, 2016. http://www.yole.fr

[7] *2015 International Technology Roadmap for Semiconductors*. https://www.semiconductors.org/main/2015_international_technology_roadmap_for_semiconductors_itrs/

[8] Smithers Apex's Image Sensors events Image Sensors Europe and USA Conferences. https://www.image-sensors.com/

[9] *Proc. 2017 Int. Image Sensing Workshop*. Hiroshima, June 2017.

[10] R. Fontaine, "The State of the Art of Mainstream CMOS Image Sensors". *Proc. 2015 Int. Image Sensor Workshop*, Vaals, June 2015.

[11] I. Takayanagi and J. Nakamura, "High-Resolution CMOS Video Image Sensors". *Proceedings of the IEEE*, Vol. 101, pp. 61-73, Jan. 2013.

[12] S. Kawahito et al., "A CMOS Image Sensor Integrating Column-Parallel Cyclic ADCs with On-Chip Digital Error Correction Circuits". *IEEE Int. Solid-State Circuits Conf.*, pp 56-595, Feb. 2008.

[13] J.A. Leñero-Bardallo and A. Rodríguez-Vázquez, "Review of ADCs for Imaging". *Proc. Image Sensors and Imaging Systems 2014-SPIE Electronic Imaging 2014,* January 2014.

[14] A. Paybe et al., "A 512×424 CMOS 3D Time-of-Flight Image Sensor with Multi-Frequency Photo-Demodulation up to 130MHz and 2GS/s ADC". *Proc. IEEE Int. Solid-State Circuit Conference*, February 2014.

[15] P. Seitz and A.J.P. Theuwisen (Ed.), *Single Photon Imaging*. Springer 2011.

[16] A. Tosi and F. Zappa, "MiSPiA: Microelectronic Single-Photon 3D Imaging Arrays for Low-light High-speed Safety and Security Applications". *Proc. SPIE*, vol. 8899, p. 88990D, Nov. 2013.

[17] G.F. DallaBetta et al., "Avalanche Photodiodes in Submicron CMOS Technologies for High-Sensitivity Imaging". *Advances in Photodiodes*. Rijeka, Croatia: InTech, 2011, pp. 225–248.

[18] I. Vornicu et al., "Real-Time Inter-Frame Histogram Builder for SPAD Image Sensors". IEEE Sensors Journal, Vol. 18, pp. 1576-1584, February 2018.

[19] A. Rodríguez-Vázquez et al., "In the Quest of Vision-Sensors-on-Chip: Pre-processing Sensors for Data Reduction". *Proc. IS&T Electronic Imaging: Image Sensors and Imaging Systems 201*, pages 96-101, IS&T, Springfield, VA, 2017.

[20] Anafocus Ltd. http://www.anafocus.com/

[21] A.N. Belbachir, *Smart Cameras*. Springer, ISBN:978-1-4419-4419-0952-7, 2009.

[22] A. Torralba, "How Many Pixels Make an Image?". *Visual Neuroscience*, vol. 26, n. 01, pp. 123-131.

[23] A. Rodríguez-Vázquez et al., "ACE16k: The Third Generation of Mixed-Signal SIMD-CNN ACE Chips Toward VSoCs". *IEEE Tran. on Circuits and Systems-I*, vol. 51, no. 5, pp. 851-863, May 2004.

[24] 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, no. 3, pp. 669-680, March 2011.

[25] S. Vargas-Sierra et al. "A 151dB High Dynamic Range CMOS Image Sensor Chip Architecture with Tone Mapping Compression Embedded in-Pixel". *IEEE Sensors Journal*, Vol. 15, pp. 180-195, January 2015.

[26] M. Suárez et al., "Low Power CMOS Vision Sensor for Gaussian Pyramid Extraction". *IEEE Journal of Solid-State Circuits*, Vol. 52, pp. 483-495, February 2017.

[27] A. Rodríguez-Vázquez et al., "A CMOS Vision System On-Chip withMulti-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.

[28] S.J. Carey et al., "A 100,000 fps Vision Sensor with Embedded 535GOPS/W 256 x 256 SIMD Processor Array". *Proc. 2013 Symposium on VLSI Circuits (VLSIC)*, pp. C182-C183, 2013.

[29] S. Park et al., "243.3 pJ/Pixel Bio-Inspired Time-Stamp-Based 2D Optic Flow Sensor for Artificial Compound Eyes". *Proc. 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 126-127, 2014.

[30] A. Dupret et al., "A DSP-like Analogue Processing Unit for Smart Image Sensors". *Int. J. of Circuit Theory and Applications*, Vol. 30, pp. 595-609, 2002.

[31] A. Paasio et al., "A 176 x 144 Processor Binary I/O CNN-UM Chip Design". *Proc. 1999 European Conference on Circuit Theory and Design-ECCTD*, 1999.

[32] M. Laiho et al., "MIPA4k: Mixed-Mode Cellular Processor Array". *Focal-Plabe Sensor-Processor Chips*. Springer 2011.

[33] A. Zarandy (editor). *Focal-Plabe Sensor-Processor Chips*. Springer 2011.

[34] B. Roska and F. S.Werblin, "Vertical Interactions Across Ten Parallel, Stacked Representations in the Mammalian Retina". *Nature*, Vol. 410, pp. 583–587, Mar. 2001.

[35] F. Werblin et al., "The Analogic Cellular Neural Network as a Bionic Eye," *Int. J. Circuit Theory Applicat.*, Vol. 23, pp. 541–549, 1995.

[36] C. Koch and H. Li, Eds., *Vision Chips, Implementing Vision Algorithms with Analog VLSI Circuits*. IEEE Press 1995.

[37] A. Moini, *Vision Chips*. Kluwer 2000.

[38] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 × 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor". *IEEE J. Solid- State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008.

[39] R.C. González and R.E. Woods, Digital Image Processing. Prentice Hall 2002.

[40] R.C. González et al., *Digital Image Processing Using MATLAB - 2nd Ed*. Gatesmark Publishing 2015.

[41] A. Rodríguez-Vázquez et al., "A Modular Programmable CMOS Analog Fuzzy Controller Chip". *IEEE Tran. Circuits and Systems - II*, Vol. 46, pp. 251-265, IEEE March 1999.

[42] R. Saspeshkar, *Ultra Low-Power Bioelectronics: Fundamentals, Biomedical Applications and Bio-Inspired Systems*. Cambridge 2010.

[43] B. J. Hosticka, "Performance Comparison of Analog and Digital Circuits". *Proceedings of the IEEE*. Vol. 73, pp. 25-29, January 1985.

[44] A. Rodríguez-Vázquez et al., "MOST-based Design and Scaling of Synaptic Interconnections in VLSI Analog Array Processing Chips," *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, Vol. 23, pp. 239–266, Nov./Dec. 1999.

[45] R. Carmona et al., "A Bio-Inspired 2-Layer Mixed-Signal Mixed-Signal Flexible Programmable Chip for Early Vision". *IEEE Tran. on Neural Networks*, Vol. 14, pp. 1313-1336, September 2003.

[46] A. Rodríguez-Vázquez et al, "A 3-D Chip Architecture for Optical Sensing and Concurrent Processing". SPIE Photonics Europe 2010 Symposium - Conf. on CMOS and Detector Technology, *Proc. of SPIE,* Vol. 7726, CCC 0277-786X, pp. 772613-1-12, April 2010.

[47] 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.

[48] A. Rodríguez-Vázquez, et al., "Current Mode Techniques for the Implementation of Continuous and Discrete-Time Cellular Neural Networks". *IEEE Tran. on Circuits and Systems*, Vol. 40, pp. 132-146, IEEE March 1993.

[49] R. Carmona et al., "A 0.5μm CMOS Random Access Analog Memory Chip for TeraOPS Speed Muitimedia Video Processing". *IEEE Tran. on Multimedia*, Vol.1, pp. 121-136, IEEE June 1999.

[50] G. Liñán et al., "A 1000 FPS at 128 x 128 Vision Processor with 8-Bit Digitized I/O". *IEEE Journal of Solid-State Circuits*, Vol. 39, pp. 1044-1055, IEEE July 2004.

[51] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints". International Journal of Computer Vision, Vol. 60(2): 91-110, 2004.

[52] J. Fernández-Berni et al., *Low-Power Smart Imagers for Vision-Enabled Sensor Networks*. Springer Science & Business Media, 2012.

[53] M. Nathan et al. "The Grasp Multiple Micro-UAV Testbed," *IEEE Robotics & Automation Magazine*, Vol. 17, no. 3, pp. 56-65, 2010.

[54] J. Fernández-Berni et al., "Early Forest fire Detection by Vision-Enabled Wireless Sensor Networks". *Int. J. of Wildland Fire*,Vol. 21, pp. 938–949, July 2012

[55] J. Fernández-Berni et al.,"Bottom-up Performance Analysis of Focal-Plane Mixed-Signal Hardware for Viola-Jones Early Vision Tasks". *Int. Journal of Circuit Theory and Applications*. Vol. 43, pp. 1063-1079, 2015.

Ángel Rodríguez-Vázquez (IEEE Fellow, 1999) received the bachelor's (Univ. de Sevilla, 1976) and Ph.D. degrees in physics-electronics (Univ. de Sevilla, 1982) with several national and international awards, including the IEEE Rogelio Segovia Torres Award (1981). After research stays in University of California-Berkeley and Texas A&M University he became a Full Professor of Electronics at the University of Sevilla in 1995.

He co-founded the Instituto de Microelectrónica de Sevilla, a joint undertaken of the Consejo Superior de Investigaciones Científicas (CSIC) and the Universidad de Sevilla and started a Research Lab on Analog and Mixed-Signal Circuits for Sensors and Communications. He was the main promotor of the start-up company AnaFocus Ltd. and served as CEO, on leave from the University, from 2004 until 2009, when the company reached maturity as a worldwide provider of smart CMOS imagers and vision systems-on-chip. AnaFocus was founded on the basis of his early patents. Dr. Rodríguez-Vázquez holds eight patents.

While in Academia, he conducted R&D activities on mixed-signal microelectronics for massive sensory data, including vision chips and neuro-fuzzy controllers. He also pioneered the application of chaos to instrumentation and communications. His team designed the first, world-wide, integrated circuits with controllable chaotic behavior and the design and prototyping

of the first world-wide chaos-based communication MoDem chips. His team made also significant contributions to the area of structured analog and mixed-signal design and the area of data converter design, including the elaboration of advanced teaching materials on this topic for different industrial courses and the production of two widely quoted books on the design of high-performance CMOS sigma-delta converters.

His research work has received some 8,900 citations; he has an h-index of 48 and an i10-index of 173 according to Google Scholar. Dr. Rodríguez-Vázquez has received a number of awards for his research: the IEEE Guillemin-Cauer Best Paper Award, two Wiley's IJCTA Best Paper Awards, two IEEE ECCTD Best Paper Awards, one IEEE-ISCAS 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 Editor, Associate Editor, and Guest Editor for IEEE and non-IEEE journals, is on the committee of several international journals and conferences, and has chaired several international IEEE (NDES 1996, CNNA 1996, ECCTD 2007, ESSCIRC 2010, ICECS 2013) and SPIE conferences. He served as VP Region 8 of the IEEE Circuits 1087 and Systems Society (2009-2012) and as Chair of the IEEE CASS Fellow Evaluation Committee (2010, 2012, 2013, 2014, and 2015). He has been appointed General Chairman for IEEE ISCAS 2020.