https://doi.org/10.2352/ISSN.2470-1173.2017.11.IMSE-195

IN THE QUEST OF VISION-SENSORS-ON-CHIP: PRE-PROCESSING SENSORS FOR DATA REDUCTION

A. Rodríguez-Vázquez [1,2], *R. Carmona-Galán* [1], *J. Fernández-Berni* [1], *V. Brea* [3] and *J.A. Leñero-Bardallo* [4]
[1] IMSE-CNM (Universidad de Sevilla - CSIC)
[2] AnaFocus e2v, C/ Isaac Newton 4, Seville (Spain)
[3] Universidad de Santiago de Compostela; [4] Universidad de Cádiz
emails: - arodri-vazquez@us.es; angel@imse-cnm.csic.es; angel.rodriguez@anafocus.com

ABSTRACT

This paper shows that the implementation of vision systems benefits from the usage of sensing front-end chips with embedded pre-processing capabilities – called CVIS. Such embedded pre-processors reduce the number of data to be delivered for ulterior processing. This strategy, which is also adopted by natural vision systems, relaxes system-level requirements regarding data storage and communications and enables highly compact and fast vision systems. The paper includes several proof-o-concept CVIS chips with embedded pre-processing and illustrate their potential advantages.

INTRODUCTION

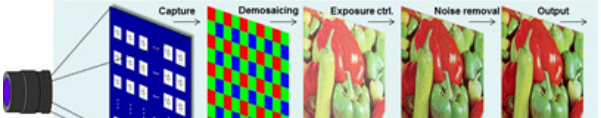
CMOS Image Sensors (CIS) market is dominated by smart phones, notebooks, tablets and other consumer equipment [1]. The design of CIS front-ends for these applications is mostly challenged by the necessity to reproduce and display captured images with larger possible details [2]. Emphasis of these CISs is on image data; they are *data-centric* front-ends. However, it is well known that images data are largely redundant, and that information contained into images can de extracted from reduced subsets of the raw image data [3]. This motivates interest on *information-centric* CISs; i.e. CISs with embedded pre-processing and conceived to deliver information, instead of raw data [4]. Because handling data is costly in terms of circuit resources, area and power, these unconventional CISs may be crucial to implement *vision systems* with fast response and minimum Size, Weight and Power (SWaP), 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. Using conventional CISs in these applications may result into rather slow systems with prohibitive SWaP values due to necessity to readout, encode, transmit and store myriads of irrelevant data [5] [6].

Fig.1 illustrates differences between conventional data-centric CISs and information-centric CIS. The latter may run under different names, such as *computational image sensors*, *vision sensors*, *silicon retinas* and the like. We use CVISs (CMOS VIsion Sensors) to highlight their similarities with conventional CISs regarding physical implementation. As Fig.1 shows, while the inputs of both CIS and CVIS are images captured by photo-sensors placed in the *focal-plane,* their primary outputs are of different nature. In the case of CIS, basic outcomes are just images, commonly in digital format, either grey-scale images or colour images, raw or corrected. CISs may incorporate some degree of *intelligence*; however, their smartness features are basically aimed to calibration, error correction and other similar tasks [2]. On the contrary, the outcomes of CVIS may not be images but either image *features* [i] or even *decisions* based on the *spatial-temporal* analysis of the information contained into the scene [5][6]. To that purpose, CVISs must embed much larger intelligence than CISs. Actually, CVIS architectures capable to extracting and interpreting the information contained into images and prompting sub-sequent *reaction* commands have been explored for years at academia [7]-[12], and industrial applications are recently ramping up [13]. Challenges of these newer architectures are linked to the incorporation of *computer vision* concepts to the design flow. The endeavour is ambitious because imager architects and computer vision architects have traditionally been disjoint groups with even different languages. CVIS chips reported in this paper are examples of CVISs with computer vision capabilities.

CVIS CONCEPT

CVIS Versus CIS Front-ends

Both CIS and CVIS chips are front-end devices of complex hardware-software camera systems [14]. Roughly speaking the front-end captures images and delivers data to a digital processor. There is hence a *border* between the sensing front-end and the processors. Differences between CIS and CVIS can be linked to the positioning of this border. Fig.2 illustrates



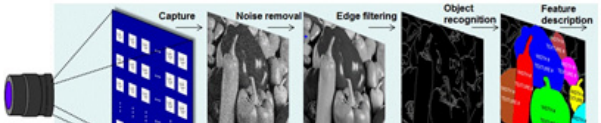


FIGURE 1 ILLUSTRATING DIFFERENCES BETWEEN SCIS (TOP) AND CVIS (BOTTOM)

i. Image features can be interpreted as characteristics of the information contained into images; for instance the number of objects included into an image, the location of maximum spots, etc.

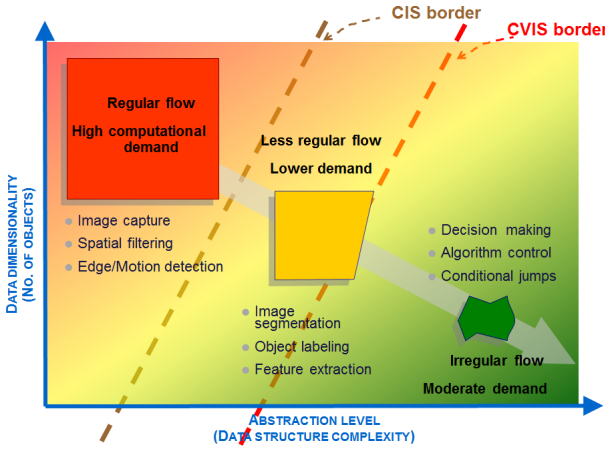


FIGURE 2 PROCESSING HIERARCHY, FROM LEFT-TOP TO BOTTOM-RIGHT, IN VISION.

this with reference to the sequence of operations of the vision processing chain. Data are assumed to evolve as indicated by the diagonal arrow; from sensors to decisions Left-top corresponds to input data captured by the sensor and bottom-right corresponds to output data on the basis of which decisions are made.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 some 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 second stage consisting of *feature extraction* tasks. Usually, feature extraction operations examine every pixel to verify if there is a feature present at that pixel considering its neighborhood. Interesting characteristics of images for subsequent image processing are *edges*, *corners* or *interest points*, *blobs* or *region of interest*, *ridges*, etc. Outputs of this second stage form irregular flow of data which are the inputs for high-level vision processing [15] [16].

Fig.3 illustrates he evolution of data throughout the vision processing chain. The example corresponds to an application 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. Data reduction and the increase of abstraction level of the progressing data are evident in this example.

CVSIs are aimed to place the border between front-ends and processors at a stage of the chain where data have been decimated. Hence vision systems built with CVISs front-ends may have better speed and SWaP metrics than those built with CISs. Actually, the strategy to reduce data at the front-end is smartly implemented in natural vision systems [17] and natural vision systems excel regarding power consumption, compactness and speed.

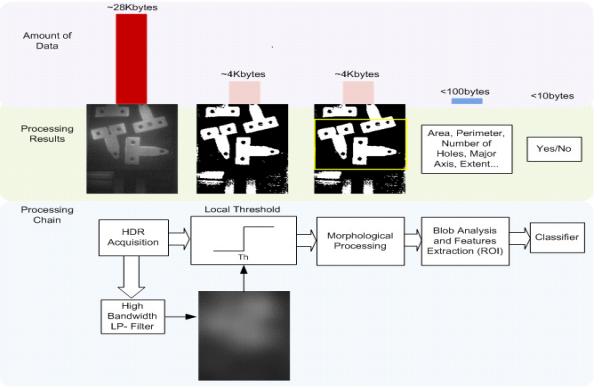


FIGURE 3 ILLUSTRATING THE PROGRESSIVE REDUCTION OF DATA AS IMAGES PROCEEDS THROUGH THE VISION PROCESSING CHAIN

On CVIS Architectures

Fig.4 shows the general concept of a vision system with CVIS front-end. Since the front-end senses and pre-processes the information, it sends and amount of data, represented by f, for ulterior processing, with $f \ll F$, where F denotes the number of raw sensor data. Indeed, in the architecture of Fig.4, processing is made progressively by distributing processing tasks between the front-end and the core processor sections.

Regarding CVIS architectures, different solutions can be adopted either separately or in a combined way, such as using *per column* processors, using *topographic* array of processors, using Multi-Functional sensory-processing PixelS (MFPS), among others [5]-[12]. Most efficient architectures employ *mixed-signal* MFPS for *fully-parallel* completion of the computational-intensive early vision tasks, followed by sub-sampled topographic processor arrays (typically digital), processors-per-column and scalar processors [6]. MFPSs actually makes the next evolutionary step of CMOS pixels, following *passive* pixels (PPS) and *active* pixels (APS), by embedding within the pixel resources for analog processing, memory and programming and control of information flows [4].

Fig.5(a) illustrates the hardware concept of a MFPS by highlighting the functional structures included by pixel. These structures are aimed to complete a large variety of functions. It is illustrated at Fig.5(b) by showing the block diagram of the Q-Eye pixel [13]. The Q-Eye is a CVIS employed at the front-end

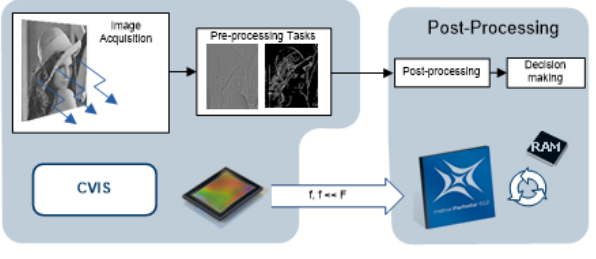


FIGURE 4 GENERAL ARCHITECTURE OF A VISION SYSTEM WITH CVIS FRONT-END.

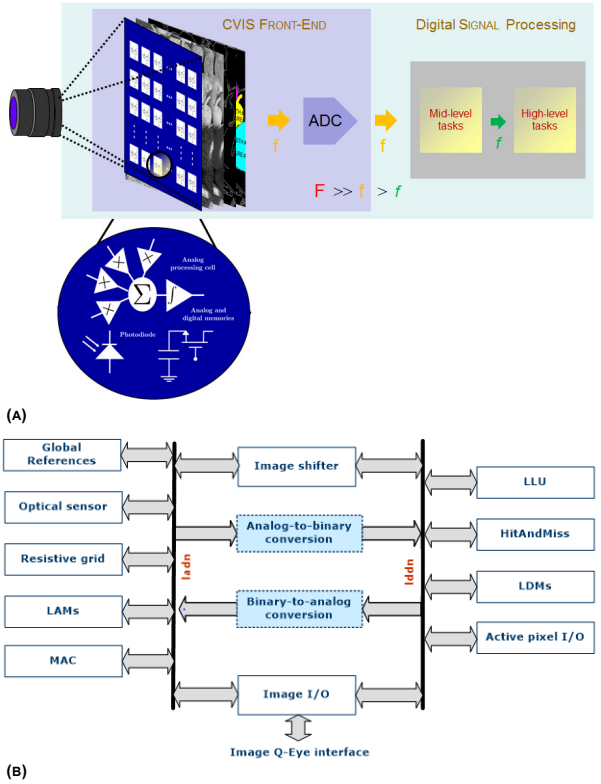


FIGURE 5 (A) CONCEPT OF VISION SENSOR WITH CVIS FRONT-END; (B) ARCHITECTURE OF A Q-EYE CELL [6] [13].

of the so-called Eye-RIS vision systems – a representative example of embedded, minimum SWaP, high-speed industrial vision system.

ILLUSTRATIVE CVIS CHIPS

Software-Programmable Visual Processor On-chip

Fig.6(a) shows the block diagram of the Eye-RIS vision system on a chip [13]. It embeds a CVIS front-end, a Digital Image Processor (DIP), a microprocessor, memories and I/O and communication ports. CVIS architecture follows the paradigm of Single Instruction Multiple Data (SIMD) processors, consisting of an array of interconnected mixed-signal processors, one per pixel, that operate in parallel – see Fig.5(a) and Fig.6(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

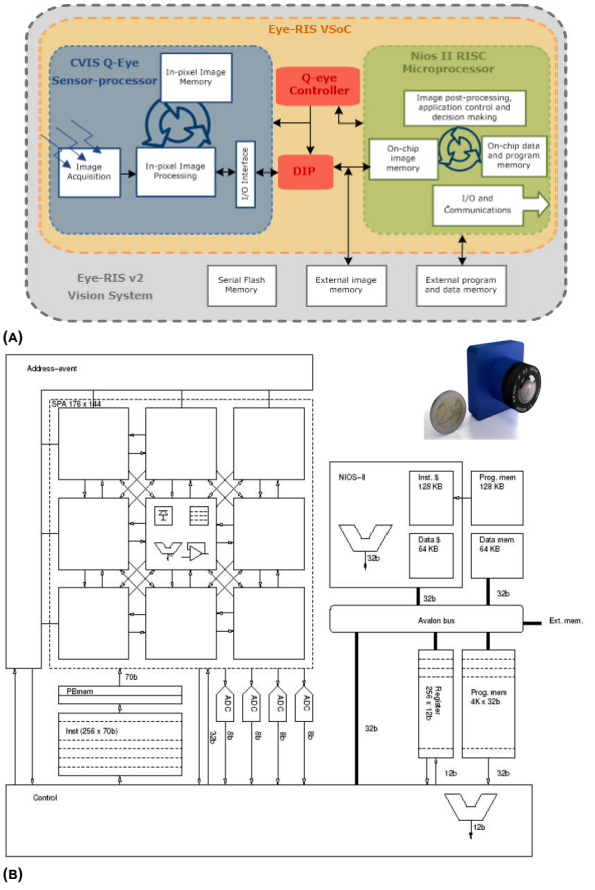


FIGURE 6 EYE-RIS VIVION SYSTEM: (A) BLOCK DIAGRAM; (B) ARCHITECTURE [13].

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 sus of mixed-signal circuitry enable going from sensing to actuation at rates about 1,000F/s rate with around 60nW per pixel required. Also, software programming of the front-end feature 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 [19]. 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 [20] or unmanned aerial vehicles [21].

Fig.7(a) shows the microphotograph of a CVIS to extract the Gaussian pyramid consisting of an arrangement of 88×60 Processing Elements (PEs) which captures images of 176×120 resolution and performs concurrent parallel processing right at pixel level [18]. The Gaussian pyramid is generated by using a *switched-capacitor* network embedded per PE. In order to shorten routing length and speed I/O operations up, the image is read out through two frame buffers outside the PE array. Each PE 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 PE is shown in Fig.7(b). The scene is acquired with 4 3T-APS per PE with nwell/p-subs. photodiodes. Every PE contains the local circuitry of an 8-bit single-slope ADC and one CDS circuit. Also, the PE 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.

The 4 3T-APS structures are biased with only one current source drawing 1 µA. The design of the source follower aims at the largest possible operating range, which is met with low threshold voltage transistors, reaching 1 V of operating range with a gain error spread inferior to 0.4%.

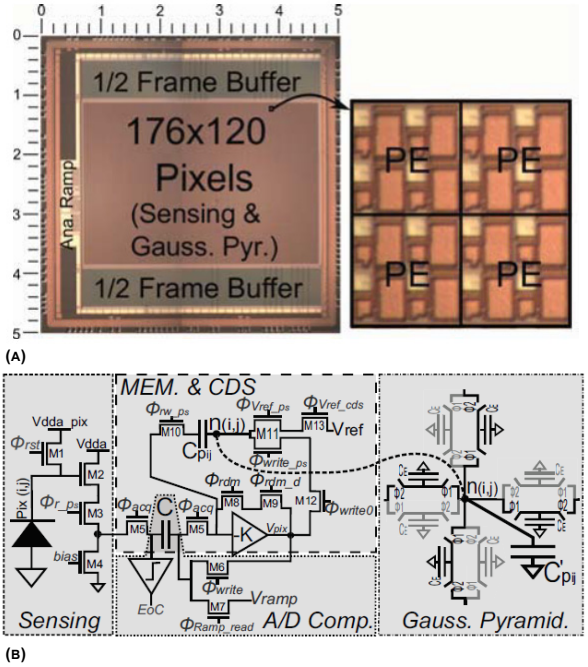


FIGURE 7 (A) GAUSIIAN PYRAMID CVIS MICROGRAPH; (B) PROCESSING ELEMENT (PE) OF THE CHIP. [18]

Fig. 5 shows several snapshots of the Gaussian pyramid along with the image acquired by the chip. Fig. 6 plots both the expected and the actual on-chip σ as a function of the number of clock cycles *n*. The upper curve is the experimental σ. The lower curve is the theoretical σ. The on-chip σ levels are found by comparing the different on-chip Gaussian-filtered images, known as *scales*, with the acquired image filtered by using a conventional computer within a given range of sigmas [σ1, σ2] around the expected σ value. The minimum RMSE sets the on chip σ level. Fig. 6 also shows RMSE levels with 255 as Full Scale Value (FSV). The RMSE slightly changes across octaves, being inferior to 1.2% of FSV. This method accounts for the errors of the on-chip Gaussian pyramid generation and the A/D conversion. The effect of such error levels in terms of an application is addressed in the next section. The 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 A/D conversion, and 150 ns 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 in around three

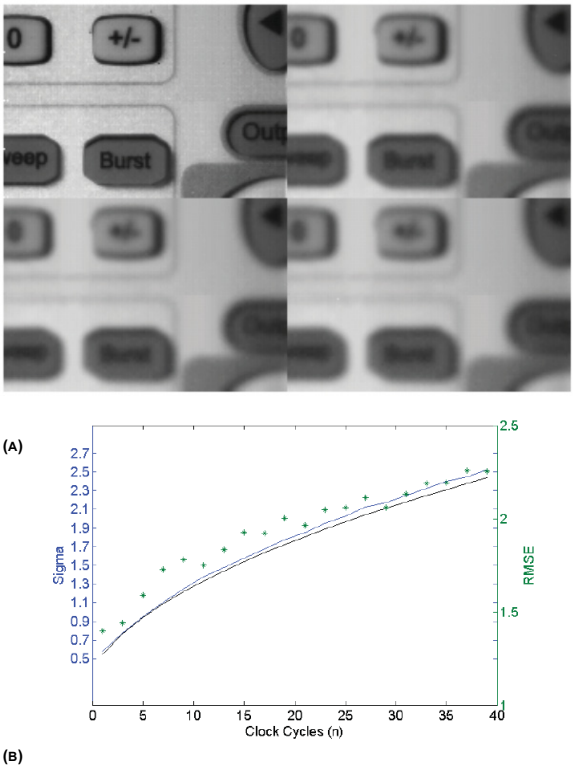


FIGURE 8 (A) IMAGE ACQUISITION AND DIFFERENT SNAPSHOTS OF THE ON-CHIP GAUSSIAN PYRAMID. THE UPPER LEFT IMAGE IS THE INPUT SCENE, THE REST OF THE IMAGES FROM LEFT TO RIGHT AND TOP TO DOWN CORRESPOND TO σ=1,77 (CLOCK CYLES *N*=19), σ=2,17 (*N*=29), AND σ=2,51 (*N*=39);
(B) EXPECTED AND ACTUAL σ VS. CLOCK CYCLES (N) PLOTS ALONG WITH THE RMSE VALUES WHEN COMPARING ACTUAL AND IDEAL GAUSSIAN-FILTERED IMAGES

orders of magnitude energy consumption reduction while having similar or faster processing speed. It leads to a combined speed-power figure of merit from two to five orders of magnitude superior to that of conventional solutions.

Multifunctional Feature Extraction Sensor

Embedded camera systems for the consumer mobile and wearable application market need to operate in a tight power budget. They need to cope with a vast range of illumination conditions, and at the same time, they need to incorporate intelligent features dictated by security and privacy-protection directives. This can be achieved by using CVISs with MFPSs conceived specifically for Dynamic Range (DR) adaptation the tracking of Region-of-Interests (RoI) selected on the basi of privacy-aware considerations.

Fig.9 shows the architecture and the pixel of a CVIS conceived specifically for DR adaptation and privacy-aware RoI tracking. The central element is an array of 4-connected mixed-signal processing elements (PE). Each PE contains two photodiodes. One of them is responsible for generating the pixel value by integrating the photocurrent in a sensing capacitance. The other photodiode 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 neighborhoods. The array can be divided into different regions by means of control lines distributed along the horizontal and vertical edges of the array [22], 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 photocurrent 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 – Fig.10(a). The rest remains the same, so privacy-protection is implemented at chip level. No sensitive information is delivered by the sensor. Fig.10(b) further illustrates the operation of this CVIS by displaying a capture of a scene with a high DR (102dB) using a single exposure with optimized exposition time. Despite a spatial resolution of only 320×240 pixels, details can be appreciated both outside of the window and in the picture on the wall inside the room.

Conclusions

Applications targeting image analysis instead of just displaying are gaining relevance within the CIS ecosystem and are expected to experience significant growing in near future. Advances on image sensor technologies, heterogeneous packaging and system embedding enable to reduce Size,

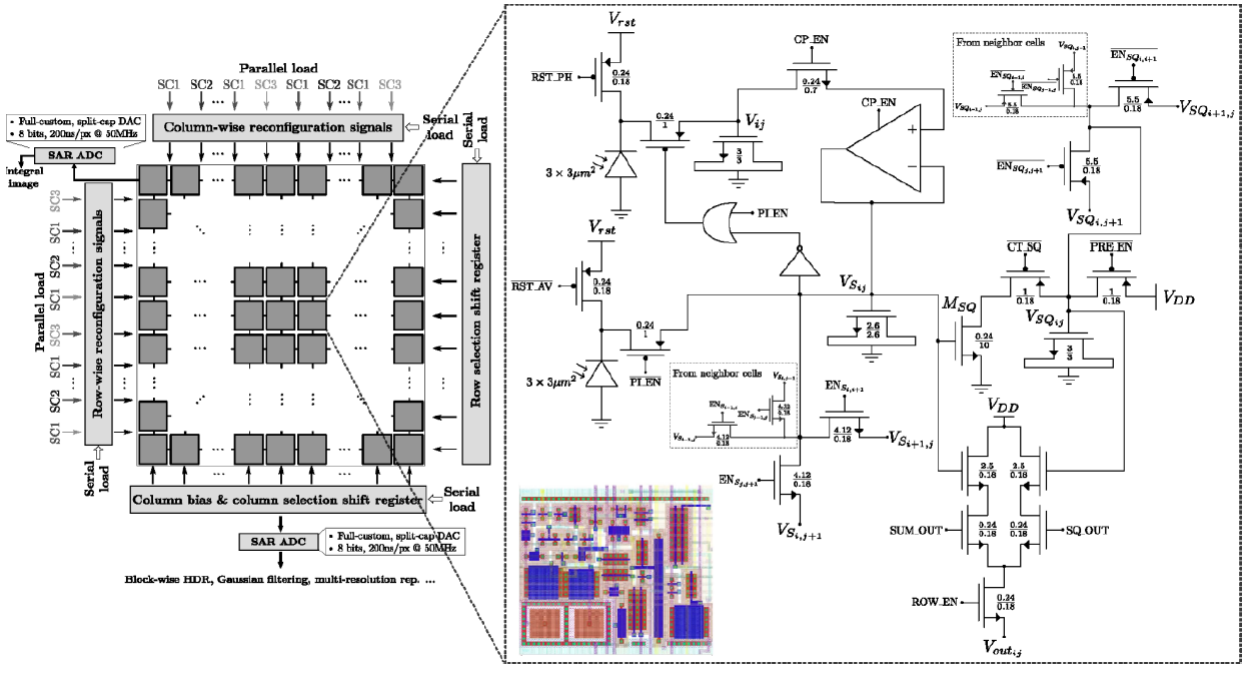
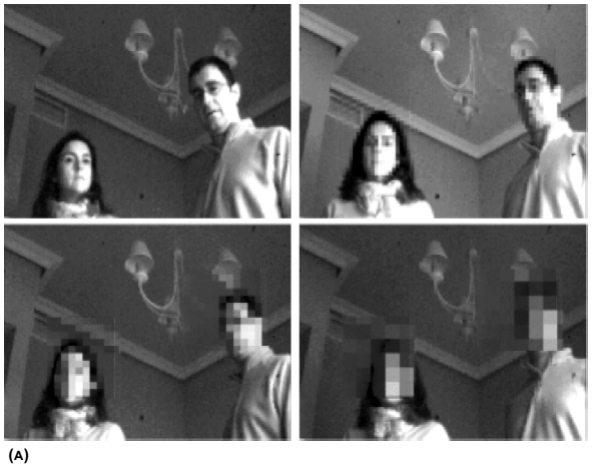


FIGURE 9 FUNCTIONAL DIAGRAM OF THE CHIP ARCHITECTURE AND SCHEMATIC OF THE PIXEL OF A CVIS FOR PRIVACY-AWARE APPLICATIONS.



(A)

(B)

FIGURE 10 (A) ON-CHIP PIXELATION BY SELECTIVE ADAPTATION OF THE SPATIAL SAMPLING RATE. A FACE-DETECTION ALGORITHM DEFINES THE REGIONS THAT NEED TO BE OBFUSCATED FOR PRIVACY PROTECTION. (B) BALANCED IMAGE CAPTURED BY THE CVIS

Weight and Power (SWaP) of *vision systems*. Vision can hence be incorporated to applications that require minimum SWaP and large speed, thus outperforming conventional solutions employing an imager and a separate digital processor. New sensor front-end architectures embedding computer vision principles are required to that purpose. The paper shows examples of three of these CVIS front-ends. Other examples included in the oral presentation associated to this correspond to the adaptive, content-aware adaptation of the DR of images at video rates [23]. Possible extension to 3D implementation is also briefly outline in the oral presentation [24].

ACKNOWLEDGMENTS

This research has been partially funded by ONR N00014-14-1-0355 and Spanish government projects MINECO TEC2015-66878-C3-1-R&TEC2015-66878-C3-3-R and Junta de Andalucía, Proyectos Excelencia- Conv. 2012 TIC 2338.

REFERENCES

[1] Yole Development. http://www.yole.fr/.
[2] J. Nakamura (editor). *Image Sensors and Signal Processing for Digital Still Cameras*. Taylor & Francis 2006.
[3] A. Torralba, "How Many Pixels Make an Image?". Visual Neuroscience, vol. 26, n. 01, pp. 123-131.
[4] T. Roska and A. Rodríguez-Vázquez. *Towards the Analogic Visual Microprocessor*. John Wiley & Sons, Chichester 2001.
[5] A. Zarandy (editor). Focal-Plabe Sensor-Processor Chips. Springer 2011.
[6] 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.
[7] 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.
[8] C.L. Lee and C.C. Hsieh, "A 0.8-V 4096-Pixel CMOS Sense-and-Stimulus Imager for Retinal Prosthesis". *IEEE Transactions on Electron Devices,* vol. 60, no. 3, 1162-1168, 2013.
[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, no. 3, pp. 669-680, March 2011.
[10] S.J. Carey et al., "A 100,000 fps Vision Sensor with Embedded 535GOPS/W 256 x 256 SIMD Processor Array". *2013 Symposium on VLSI Circuits (VLSIC),* pp. C182-C183, 2013.
[11] S. Park et al., "243.3 pJ/Pixel Bio-Inspired Time-Stamp-Based 2D Optic Flow Sensor for Artificial Compound Eyes". *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 126-127, 2014.
[12] A. Rodríguez-Vázquez et al., "ACE16k: The Third Generation of Mixed-Signal SIMD-CNN ACE Chips Toward VSoCs". *IEEE Transactions on Circuits and Systems-I*, vol. 51, no. 5, pp. 851-863, May 2004.
[13] Anafocus Ltd. [Online]. Available. http://www.anafocus.com.
[14] A.N. Belbachir, *Smart Cameras*. Springer, ISBN:978-1-4419-4419-0952-7, 2009.
[15] R.C. González and R.E. Woods, *Digital Image Processing*. Prentice Hall.
[16] R.C. González et al., *Digital Image Processing Using MATLAB - 2nd Ed*. Gatesmark Publishing, 2015.
[17] B. Roska and F. Werblin, "Vertical Interactions Across Ten Parallel, Stacked Representations in the Mammalian Retina". *Nature*, 410, pp. 583-587, 2001.
[18] M. Suárez et al., "Low Power CMOS Vision Sensor for Gaussian Pyramid Extraction". *IEEE J. Solid-State Circuits*, doi:10.1109/JSSC.2016.2610580.
[19] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints". *International Journal of Computer Vision*, Vol. 60(2): 91-110, 2004.
[20] J. Fernández-Berni et al., *Low-Power Smart Imagers for Vision-Enabled Sensor Networks*. Springer Science & Business Media, 2012.
[21] M. Nathan et al. "The Grasp Multiple Micro-UAV Testbed," *IEEE Robotics & Automation Magazine*, Vol. 17, no. 3, pp. 56-65, 2010.
[22] 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*. April 2014, doi:10.1002/cta.1996.
[23] S. Vargas-Sierra et al. "A 151dB High Dynamic Range CMOS Image Sensor Chip Architecture with Tone Mapping Compresion Embedded in-Pixel". *IEEE Sensors Journal*, Vol. 15, pp. 180-195, January 2015.
[24] 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. (Invited Paper).