Automatic DR and Spatial Sampling Rate Adaptation for Secure and Privacy-Aware ROI Tracking Based on Focal-Plane Image Processing

Ricardo Carmona-Galán, Jorge Fernández-Berni, Ángel Rodríguez-Vázquez
Institute of Microelectronics of Seville (IMSE-CNM), CSIC-University of Seville (Spain)
E-mail: rcarmona@imse-cnm.csic.es

***Abstract*— 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 enough intelligence to implement security and privacy-protection directives. The incorporation of image signal processing at the focal-plane can help reducing the necessary resources to implement tasks like DR adaptation and privacy-aware ROI tracking. In this paper we present a vision sensor that is able to perform single-exposure HDR imaging and ROI obfuscation on-chip, with the help of a reduced set of focal-plane processing elements.**

Keywords*— Mobile/wearable applications, vision chips, single-exposure HDR, on-chip privacy-protection, focal-plane processing

I. INTRODUCTION

The pervasive use of networked cameras is introducing severe concerns on privacy and the subsequent social rejection [1]. On the other extreme, life-logging cameras and the like [2] are becoming usual in recording sport practices, professional activities and consumer/user behavior [3]. The smart embedded camera systems dedicated to these tasks need to be pro-active, run autonomously, easy to deploy, sometimes wearable and work in a low power budget. In these conditions, the conventional approach to vision fails to meet restrictions on latency and power. One possible approach is to convey elementary functions required to the implementation of context awareness to the sensor chip itself, converting it into a smart sensor and thus reducing the need for data transmission and storage off-chip. Although this has an incidence on the size of the image sensor and image quality, some high-level decisions can be triggered on the base of a reduced number of pixels [4].

For instance, in recording natural scenes, the sensor will have to deal with scenes featuring a high dynamic range (HDR). The most widely extended method for this is using multiple captures, what requires an appreciable amount of power to be implemented in real-time [5]. This technique is usually implemented in digital still cameras. Artifacts can be very noticeable when motion occurs during multi-exposure [6], and they require a considerable amount of computation to be eliminated or, at least, attenuated [7]. In order to avoid these problems, and to reduce the amount of processing necessary to provide artifact-free HDR images, a single-exposure solution is demanded. Of course, incorporating in-pixel circuitry to perform DR adaptation conveys a reduced fill-factor and a larger pixel size. However, power savings associated with this approach are well worth it. There are some reported works about the implementation of single-exposure DR extension. They aim at multilayered vertically integrated structures [8] [9]. Our proposed circuit has been implemented in planar technology as a proof of concept, while the architecture is easily mappable to a 3D-IC stack. In order to implement global compression of the illumination range into the available signal range in a single image capture, exposure needs to be guided by on-line local and regional averaging. The circuits employed to evaluate these magnitudes can provide support to realize different functionalities. One of them is the delimitation of the ROIs, and another one can be the content-aware regulation of the spatial sampling rate. This can be employed for the implementation of image compression algorithms or, as in the examples that we will be developed later, to prevent sensitive information to be delivered off-chip, thus realizing privacy-awareness at sensor level.

The organization of the paper is as follows: next section briefly describes the architecture of the sensor and the processing elements that are incorporated at pixel-level. The third section explains the mechanism for single-exposure DR adaptation. The fourth section is concentrated on the implementation of privacy-protection on-chip. And the last section is dedicated to conclusions.

II. IN-PIXEL PROCESSING ELEMENTS

The vision sensor chip that we will employ to illustrate the efficient implementation of DR adaptation and privacy-aware ROI tracking has the floorplan depicted in Fig. 1. 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, V_{ij}, by integrating the photocurrent in a sensing capacitance. The other photodiode generates a replica of this voltage value, that is initially stored as $V_{S_{ij}}$. The 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 [10], 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.

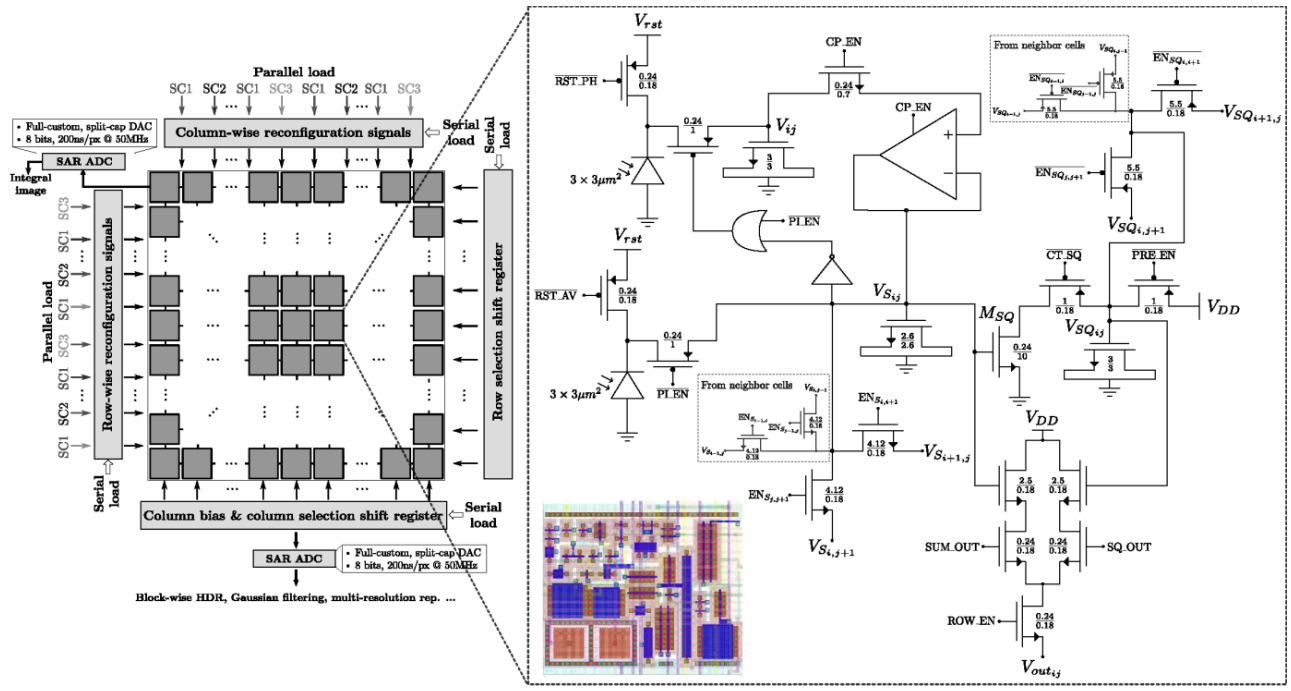

Fig. 1. Functional diagram of the chip architecture and schematic of the processing element.

In the schematics of Fig. 1, signals $\mathrm{EN}_{S_{i,i+1}}$, $\mathrm{EN}_{S_{j,j+1}}$, $\overline{\mathrm{EN}_{SQ_{i,i+1}}}$, $\overline{\mathrm{EN}_{SQ_{J,J+1}}}$ are the reconfiguration signals coming from the periphery map. The coordinates (i, j) denote the location of that specific cell in the array. Basically, these signals drive the gate of MOS switches employed to enable charge redistribution, thus voltage averaging, between the capacitors holding the voltages $V_{S_{ij}}$ and $V_{SQ_{ij}}$, respectively. Each time a new region is configured and a new average is required, $V_{S_{ij}}$ is reset to the value of V_{ij} by means of a voltage buffer. Then, it is squared with the help of transistor M_{SQ}, and then, the new average of the pixel value and its squared version is calculated with the capacitor network defined by the values of signals $\mathrm{EN}_{S_{i,i+1}}$, $\mathrm{EN}_{S_{j,j+1}}$, $\overline{\mathrm{EN}_{SQ_{i,i+1}}}$, $\overline{\mathrm{EN}_{SQ_{J,J+1}}}$. The main processing primitive upon which all the chip functionalities are built is charge redistribution, i. e. the averaging of voltages of the pixels belonging to the same region/subimage. The remaining in-pixel transistors are employed to read out $V_{S_{ij}}$ and $V_{SQ_{ij}}$.

In the periphery of the array, there are the above referred controls for the array subdivision, the registers for the selection signals and four 8b SAR ADCs. These converters have a conversion time of 200ns when clocked at 50MHz. Two of them are connected to the first pixel of the array, because they provide the values of the integral image and the integral image squared, which are globally computed in a network like this. These magnitudes are very useful for the extraction of Haar-like features [11]. The other two ADCs convert the pixel voltage $V_{out_{ij}}$, which corresponds to the selected output of the nodes $V_{S_{ij}}$ and $V_{SQ_{ij}}$. The chip has been fabricated in a standard 0.18µm CMOS process. Its power consumption ranges from 42.6mW for high dynamic range operation to 55.2mW for integral image computation at 30fps. When compared to other focal-plane processors reported (see Table I), the conclusion is that this chip embeds a larger amount of functionality, with the largest resolution and the smallest pitch in the state-of-the-art.

Reference	**[12]**	**[13]**	**[14]**	**This chip**
Function	edge filtering, tracking, HDR	Gaussian filtering	2D optic flow estimation	HDR, Gaussian filter, integral image, multiresolution
Technology (µm)	0.18	0.18	0.18	0.18
Supply voltage (V)	0.5	1.8	3.3	1.8
Array size	64 × 64	176 × 120	64 × 64	320 × 240
Pixel pitch (µm)	20	44	28.8	19.6
Fill factor (%)	32.4	10.25	18.32	5.4
Dynamic range (dB)	105	—	—	102
Power (nW/px·frame)	1.25	26.5	0.89	23.9

Table I. Comparison with state-of-the-art focal-plane sensor/processors.

Fig. 2. Experimental results showing image capture with global adaptation (top) and ROI-driven adaptation (bottom).

III. SINGLE-EXPOSURE DR ADAPTATION

In order to have an on-line estimation of local or regional average illumination, each processing element counts on two photodiodes and two separate sensing capacitances (Fig. 1). Once they have been reset to V_{rst}, photocurrent integration starts concurrently in both the pixel capacitance —holding voltage V_{ij}— and the averaging capacitance —holding $V_{S_{ij}}$. However, while in the former photocurrent integration is carried out in an isolated way, in the latter charge redistribution takes place in parallel among the set of averaging capacitances that are interconnected through the switches controlled by $\mathrm{EN}_{S_{i,i+1}}$ and $\mathrm{EN}_{S_{j,j+1}}$. Photocurrent integration is thus stopped at a certain time instant depending on the input threshold voltage of the inverter connected to $V_{S_{ij}}$. If this threshold voltage is designed to be at the middle point of the signal range, it can be demonstrated [15] that the voltage excursion due to integration of the photogenerated current for each pixel within a certain block k —blocks can be pre-set or dynamically set by a vision algorithm according to the scene content— is given by:

$$\Delta V_{ij_k} = \frac{V_{rst}-V_{min}}{2} \cdot \frac{I_{ph_{ij_k}}}{\overline{I_{ph_k}}} \qquad (1)$$

where $V_{rst} - V_{min}$ represents the maximum pixel excursion, $I_{ph_{ij_k}}$ denotes the pixel photogenerated current and $\overline{I_{ph_k}}$ is the block average photocurrent generated during the integration period. We can see from Eq. (1) that the maximum pixel illumination to be detected without saturation is double of the average illumination of the block. It is this property, together with the possibility of confining its application to any particular rectangular-shaped image region, what endows our array with the capability of retrieving information, otherwise missed, from scenes with a high dynamic range.

The proposed sensing architecture, based in the twin-photodiode scheme, is then able to concurrently elaborate a reference frame containing the average illumination of regions defined by the user or by the program itself. At each pixel in the array, once the threshold —designed to fall in the middle of the signal range— is reached, photocurrent integration stops. The resulting image contains several regions that are balanced around their respective average illumination (Fig. 2). A tracking algorithm running in the host system, for instance a PC (Fig. 3), defines the ROI and the rest of regions to be considered for adaptation. It generates a representation of the scene that is balanced in the ROI and still can be used to keep track of any important changes in the surroundings.

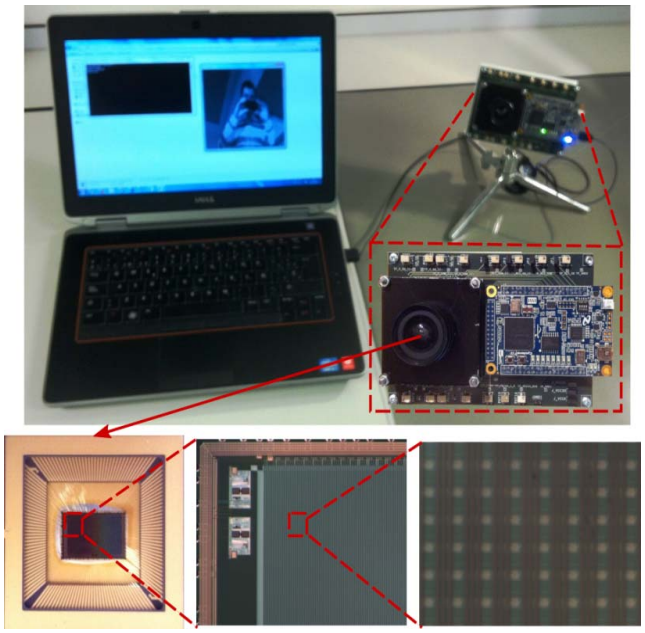

Fig. 3. PC-controlled testboard, chip photograph and two microphotograpohs showing close-ups of the array of pixels

IV. PRIVACY-PROTECTION AT CHIP-LEVEL

Still further processing can be done at the focal-plane which permits secure and privacy-aware monitoring of human activity. One of the major concerns in networked cameras is the video stream meddling on the part of unfaithful users [16]. Implementing privacy protection measures right at the sensor chip reduces the opportunities of tampering. The most elementary technique for privacy-protection on images is blanking [17]. It consists in completely removing sensitive regions from the captured images. In the case of monitoring of human behavior, this technique precludes any behavioral analysis. Alternatives to this that still permit this analysis are obfuscation and scrambling [18]. Concerning obfuscation, the pixelation of sensitive regions provides the best performance in balancing privacy-protection and intelligibility of the surveyed scene when compared to blurring and masking filters [19] [20].

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 V_{ij}. These values can be copied into $V_{S_{ij}}$ in parallel, what takes only 150ns and is non-destructive. This is going to be very important to avoid artifacts due to obfuscation. Once the voltages $V_{S_{ij}}$ 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. 4). The rest remains the same, so privacy-protection is implemented at chip level. No sensitive information is delivered by the sensor.

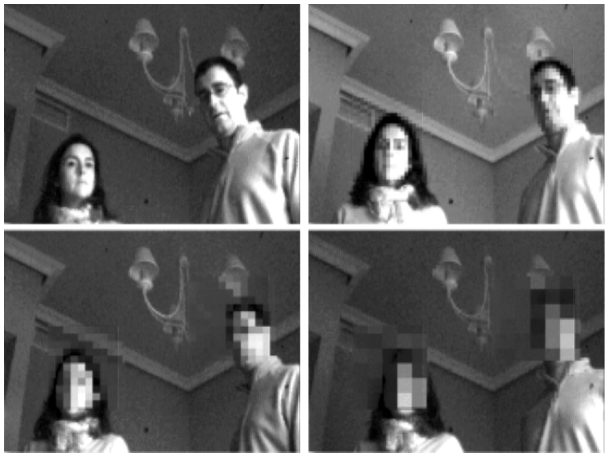

Fig. 4. On-chip pixelation by selective adaptation of the spatial sampling rate. A face-detection algorithm defines the regions that need to be obfuscated for privacy protection.

V. CONCLUSIONS

Vision sensors can play a key role in boosting the performance of embedded vision systems for the mobile consumer application market at affordable power consumption. In this paper, we present a reconfigurable focal-plane vision sensor intended to efficiently provide useful low-level processing capabilities to vision algorithms. The most interesting functionalities implemented and successfully tested are region-wise HDR and obfuscation of sensitive information. This permits unsupervised privacy-aware ROI tracking.

ACKNOWLEDGMENT

This work has been funded by Office of Naval Research (USA) ONR, grant No. N000141410355, the Spanish Government through projects TEC2012-38921-C02 MINECO (ERDF/FEDER), IPT-2011-1625-430000 MINECO, IPC- 20111009 CDTI (ERDF/FEDER) and Junta de Andalucía, Consejería de Economía, Innovación, Ciencia y Empleo (CEICE) TIC 2012-2338.

REFERENCES

[1] T. Winkler and B. Rinner, "Security and Privacy Protection in Visual Sensor Networks: A Survey," *ACM Computing Surveys*, Vol. 47, No. 1, paper no. 2, March 2015.

[2] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia, "Privacy behaviors of lifeloggers using wearable cameras". *Proc. of the 2014 ACM Int. Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14)*, pp-571-582. New York, NY, USA, 2014.

[3] A. W. Senior et al., "Video analytics for retail". *IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 423-428, Sept. 2007.

[4] A. Torralba, "How many pixels make an image?". *Visual Neuroscience*, Vol. 26, No. 1, pp. 123-131, Jan.-Feb. 2009.

[5] M. Mase, S. Kawahito, M. Sasaki, Y. Wakamori and M. Furuta, "A wide dynamic range CMOS image sensor with multiple exposure-time signal outputs and 12 bit column-parallel cyclic A/D converters', *IEEE J. Solid-State Circuits*, Vol. 40, No. 12, pp. 2787-2795, 2005.

[6] J. An, S. Ha, and N. Cho, "Probabilistic motion pixel detection for the reduction of ghost artifacts in high dynamic range images from multiple exposures," *EURASIP Journal on Image and Video Processing*, Vol. 42, No. 1, pp. 1-15, 2014.

[7] O. Gallo, N. Gelfandz, W. C. Chenz, M. Tico, K. Pulli, "Artifact-free high dynamic range imaging". *Proc. of IEEE International Conference on Computational Photography*, pp. 1-7, 2009.

[8] C. Ma, D. San Segundo Bello, C. Hoof and A. Theuwissen, "High dynamic range hybrid pixel sensor", *Electronic Letters*, Vol.47, No. 12, pp. 695-696, June 2011.

[9] A. Xhakoni and G. Gielen, "A 132-dB dynamic-range global-shutter stacked architecture for high-performance imagers", *IEEE Transactions on Circuits and Systems II*, Vol. 61, No. 6, pp. 398-402, June 2014.

[10] J. Fernández-Berni, R. Carmona-Galán, R. del Río, A. Rodríguez-Vázquez, "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

[11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features". *Proc. IEEE Computer Vision and Pattern Recognition (CVPR'01)*, Vol. 1, pp. I-511-I-518, 2011.

[12] C. Yin and C. Hsieh, "A 0.5V 34.4μW 14.28kfps 105dB smart image sensor with array-level analog signal processing". *Proc. Asian Solid-State Circuits Conference (A-SSCC)*, pp. 97-100, 2013.

[13] M. Suárez et al., "A 26.5nJ/px 2.64Mpx/s CMOS vision sensor for Gaussian pyramid extraction". *40th European Solid-State Circuits Conference (ESSCIRC)*, pp. 311-314, Venice (Italy), September 2014.

[14] S. Park et al., "243.3pJ/pixel bio-inspired time-stamp-based 2D optic flow sensor for artificial compound eyes". *Int. Solid-State CircuitsConf. (ISSCC)*, pp. 126–127, San Francisco (CA), Feb. 2014.

[15] J. Fernández-Berni, R. Carmona-Galán, and A. Rodríguez-Vázquez, "Reconfigurable focal-plane hardware for block-wise intra-frame HDR imaging". *Int. Image Sensor Workshop (IISW 2013)*, pp. 289-292, Snowbird Resort, Utah (USA), June 2013.

[16] D. N. Serpanos and A. Papalambrou, "Security and privacy in distributed smart cameras". *Proceedings of the IEEE*, Vol. 96, No. 10, pp. 1678–1687, 2008.

[17] S. Cheung, J. Zhao, M. Vijay, "Efficient Object-Based Video Inpainting". Proceedings of the IEEE International Conference on Image Processing, pp. 705-708, Atlanta (GA) October 2006.

[18] D. Dufaux, T. Ebrahimi, "Scrambling for Privacy Protection in Video Surveillance Systems". *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 18, No. 8, pp. 1168-1174, Aug. 2008.

[19] P. Korshunov et al. "Subjective Study of Privacy Filters in Video Surveillance". *Proc. of the IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, pp. 378-382, Banff, AB, Canada, Sept. 2012.

[20] P. Korshunov, S. Cai, T. Ebrahimi, "Crowdsourcing Approach for Evaluation of Privacy Filters in Video Surveillance". *Proc. of the ACM Int. Workshop on Crowdsourcing for Multimedia*, pp. 35-40, Nara, Japan, Oct.-Nov. 2012.