

Real-time single-exposure ROI-driven HDR adaptation based on focal-plane reconfiguration

J. Fernández-Berni^a, R. Carmona-Galán^a, R. del Río^a, R. Kleihorst^b,
W. Philips^b and Á. Rodríguez-Vázquez^a

^aInstitute of Microelectronics of Seville (IMSE-CNM), CSIC-Universidad de Sevilla, Spain

^bGhent University/iMinds/TELIN-IPI, St-Pietersnieuwstraat 41, Ghent, Belgium

ABSTRACT

This paper describes a prototype smart imager capable of adjusting the photo-integration time of multiple regions of interest concurrently, automatically and asynchronously with a single exposure period. The operation is supported by two intertwined photo-diodes at pixel level and two digital registers at the periphery of the pixel matrix. These registers divide the focal-plane into independent regions within which automatic concurrent adjustment of the integration time takes place. At pixel level, one of the photo-diodes senses the pixel value itself whereas the other, in collaboration with its counterparts in a particular ROI, senses the mean illumination of that ROI. Additional circuitry interconnecting both photo-diodes enables the asynchronous adjustment of the integration time for each ROI according to this sensed illumination. The sensor can be reconfigured on-the-fly according to the requirements of a vision algorithm.

Keywords: Viola-Jones algorithm, smart imaging, sensing-processing arrays, mixed-signal circuitry, Haar-like features, OpenCV library, integral images.

1. INTRODUCTION

When it comes to extracting meaningful information from a scene, vision algorithms have to cope with varying illumination conditions taking place at both intra-frame and inter-frame levels. Without a strategy to address this issue, important details can be missed due to saturation in over-exposed regions or noise and lack of contrast in under-exposed regions. The most usual technique to prevent this from happening consists of taking multiple captures per frame with different exposure periods.¹ The resulting images are combined into a single one featuring a much wider intra-frame dynamic range. Inter-frame changes of illumination are accommodated by adapting the range of exposure periods correspondingly. Unfortunately, this technique creates artifacts if motion happens to occur during multi-exposure. Specialized sensing architectures capable of extending the dynamic range through single exposure²⁻⁴ are thus highly demanded at present.⁵ Generally speaking, all these reported techniques targeting High Dynamic Range (HDR) deal globally with the image content. In other words, there is no special consideration for specific regions in the process of adjusting the capture according to the illumination conditions. However, vision algorithms usually focus their attention in particular so-called Regions Of Interest (ROI).^{6,7} Once a certain ROI is spotted, the algorithm tracks it across the scene while carrying out prescribed analytics. This tracking and corresponding analytics should not be affected by variations in the illumination over the ROI. Indeed, the priority should be to adapt the capture for that ROI while ensuring that new ROIs can still be detected in case they enter the scene.

All in all, this paper describes the operation of a prototype QVGA smart imager capable of adjusting the photo-integration time of multiple ROIs concurrently, automatically and asynchronously with a single exposure period. Each pixel of this prototype incorporates two photo-diodes. One of them senses the pixel value itself whereas the other, in collaboration with its counterparts in a particular ROI, senses the mean illumination of that ROI. Additional circuitry interconnecting both photo-diodes enables the automatic and asynchronous adjustment of the integration time for each ROI. Since the pixel matrix works in Single-Instruction Multiple-Data (SIMD) mode, the operation takes place in parallel for all the ROIs. The sensor can be reconfigured on-the-fly according

Further author information:

Jorge Fernández-Berni: C/ Américo Vespucio s/n, 41092, berni@imse-cnm.csic.es, telephone: +34 954 46 66 66

to the requirements of a vision algorithm. To this end, the array of pixels is surrounded by peripheral registers that can be modified on a frame basis. These registers encode the pixel interconnection patterns that set the different ROIs.

2. SENSING ARCHITECTURE

A simplified scheme of the proposed sensing architecture is depicted in Fig. 1 together with the mixed-signal circuitry to be included at pixel level. Two serial-in parallel-out digital registers for columns and rows, respectively, are required at the periphery of the pixel matrix. Each bit stored in these registers enables (logic ‘1’) or disables (logic ‘0’) the connection through switches between neighboring columns and rows across the matrix. After loading prescribed interconnection patterns into them, the focal plane gets divided into different rectangular-shaped regions. These patterns are meant to change on a frame basis according to the scene content and the analytics performed by the vision algorithm. Once the focal plane is properly divided, the two photo-diodes and corresponding sensing capacitances at pixel level are reset to V_{rst} by asserting the control signals RST and PLEN. After reset, RST is switched back to ‘0’ and photo-integration starts. The pixel value, $V_{px_{ij}}$, will be given by:

$$V_{px_{ij}} = V_{rst} - \frac{I_{ph_{ij}}}{C} T_k \quad (1)$$

where $I_{ph_{ij}}$ represents the average current photo-generated during the integration period denoted as T_k . This period will be the same for all the pixels composing a particular image region k . It directly depends on the mean illumination of that region, as demonstrated next. In order to obtain the expression for T_k , we must take the additional photo-diode and sensing capacitance into account. These elements are scaled down by a factor m with respect to the main photo-diode and sensing capacitance. The specific value of m in a physical realization will depend on a trade-off between operation accuracy and area limitations, as further explained below. The second sensing capacitance will be interconnected through switches with its counterparts within the considered region k . The resulting larger capacitance will integrate all the photo-currents generated in their associated photo-diodes. The voltage $V_{a_{ij}}$ will therefore be the same for all the pixels of the generic region k , expressed as:

$$V_{a_{ij}} = V_{rst} - \frac{\frac{1}{m} \sum_{\forall i,j \in k} I_{ph_{ij}}}{W \cdot H \cdot \frac{1}{m} C} T_k \quad (2)$$

where $W \times H$ are the dimensions of the considered region, in pixels. We are assuming that both pixel photo-diodes are close enough to proportionally sense the same amount of light. Note that T_k will be determined by the time instant at which $V_{a_{ij}}$ reaches the input threshold voltage of the digital buffer. At that instant, the output of the buffer will switch to ‘0’, stopping the photo-integration associated with the pixel value $V_{px_{ij}}$. In order to extend the dynamic range as much as possible, the input threshold voltage of the buffer must be designed to coincide with the middle point of the signal range, i.e. $(V_{rst} + V_{min})/2$, where V_{min} is the lower limit of the signal range. All in all, equation (2) can be re-written as:

$$\frac{V_{rst} + V_{min}}{2} = V_{rst} - \frac{\bar{I}_{ph_k}}{C} T_k \quad (3)$$

where \bar{I}_{ph_k} is the average current photo-generated in the region k , directly proportional to its mean illumination. Solving equation (3) for T_k :

$$T_k = \frac{C}{2} \frac{\Delta V_{px_{MAX}}}{\bar{I}_{ph_k}}, \quad (4)$$

where $\Delta V_{px_{MAX}} = V_{rst} - V_{min}$ represents the maximum pixel excursion. Substituting equation (4) in equation (1), we obtain that:

$$V_{px_{ij}} = V_{rst} - \frac{\Delta V_{px_{MAX}}}{2} \frac{I_{ph_{ij}}}{\bar{I}_{ph_k}} \quad (5)$$

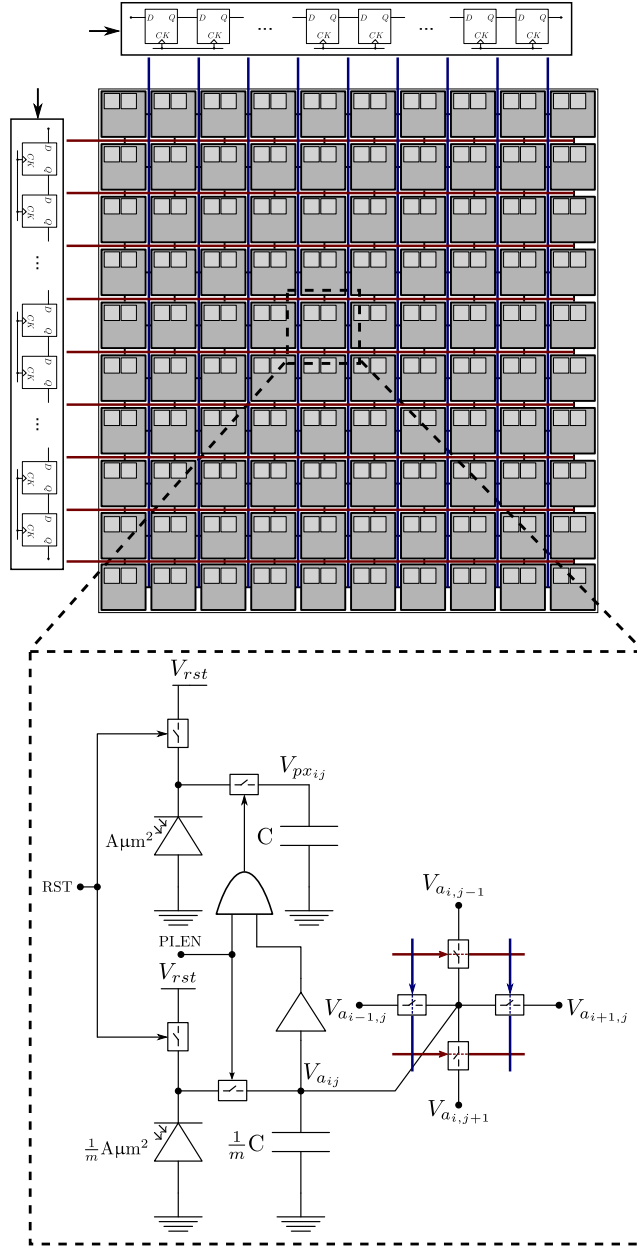


Figure 1. Simplified scheme of the proposed sensing architecture (above) and mixed-signal circuitry (below) to be included at pixel level.

where we can see that the voltage excursion for all the pixels belonging to a certain region k will depend on the illumination conditions of that particular region, specifically on its mean illumination. This asynchronous adaptation of the integration period takes place concurrently for each region previously set from the peripheral registers. For regions poorly illuminated, the maximum integration period will be given by the time instant at which PLEN switches back to '0'. This instant will in turn depend on the minimum frame rate affordable by the targeted application. Finally, note that equations (1)-(5) will be valid as long as the scale factor m can be applied accurately. If the photo-diode and the capacitance sensing the mean illumination are scaled down too much, non-linear terms and mismatch can lead to a great deviation with respect to the ideal linear operation just described.

3. EXPERIMENTAL RESULTS

We have implemented the sensing architecture sketched in Fig. 1 in a QVGA prototype CMOS vision sensor. A photograph of the chip together with a microphotograph of part of the pixel matrix and the pixel layout are depicted in Fig. 2. The photo-diodes can easily be identified in the lower left corner of the layout. The scale factor is $m = 1$ since this chip incorporates additional functionalities at pixel level that require sensing capacitances with the same nominal value.⁸ The switches and capacitors are implemented by single MOS transistors. The main characteristics of the chip are summarized in Table 1.

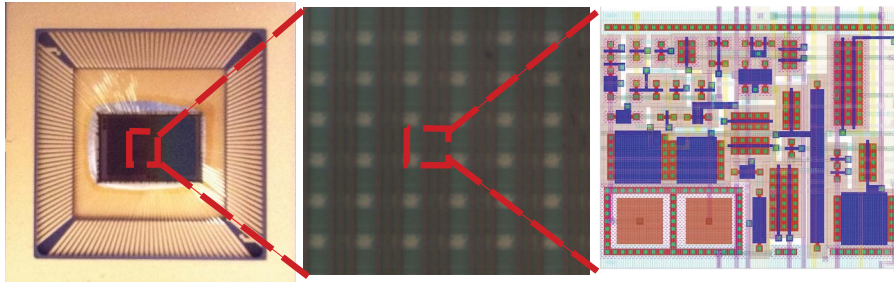


Figure 2. Photograph of the prototype vision sensor along with a microphotograph of part of the pixel matrix and the pixel layout.

Technology	Std 0.18 μm 1.8V 1P6M CMOS process
Die size (with pads)	7.5mm \times 5mm
Pixel size	19.59 μm \times 17 μm
Fill factor	5.4%
Photodiode type	n-well/p-substrate
Power supply	3.3 (pads), 1.8V (core)
DSNU	1.7%
PRNU (50% signal range)	3.5%
ADC throughput	5MSa/s (200ns/Sa)
Power consumption@30fps	42.6mW

Table 1. Summary of the main chip characteristics.

The prototype has been embedded into the FPGA-based system shown in Fig. 3 for test purposes. The captured images are sent to a PC where we make use of the OpenCV library⁹ to run ROI tracking vision algorithms on them. When a certain ROI is detected, the coordinates of its bounding rectangle is transmitted on-the-fly to the test board for the sensor to adapt the next capture correspondingly. Two examples are shown in Fig. 4: (a) face tracking, and (b) pedestrian tracking. The left images correspond to an adaptation based on the global mean illumination of the scene. This leads to a noisy capture of poorly illuminated regions as well as to saturated pixels in regions featuring very high illumination. ROI-driven HDR adaptation was activated for the right images. In this case, we retrieve details of the detected ROIs previously missed. Furthermore, details from other regions are also retrieved thanks to the focal-plane division required to adapt the capture for those ROIs. The whole sequences can be freely downloaded.¹⁰ Intra-frame dynamic ranges of up to 102dB have been experimentally achieved for different image regions by applying the technique described in this paper.

4. CONCLUSIONS

When it comes to extracting meaningful information from a scene, vision algorithms have to cope with changing illumination conditions. Generally speaking, all the reported techniques targeting high dynamic range deal globally with the image content. There is no special consideration for specific regions in the process of adjusting the capture according to their illumination conditions. However, vision algorithms usually focus their attention in particular ROIs. This paper presents a specialized sensing architecture suitable for HDR ROI tracking. It permits

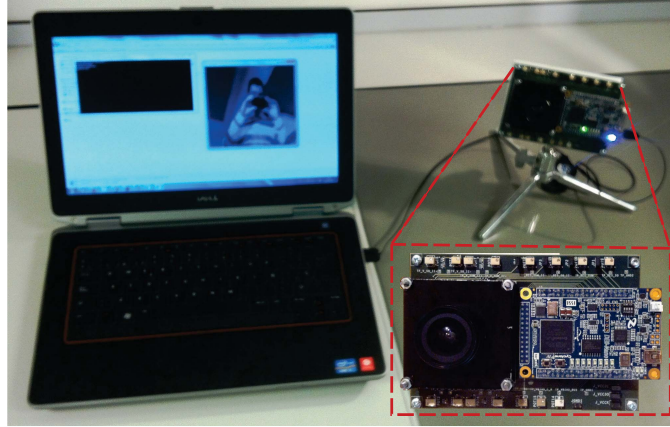
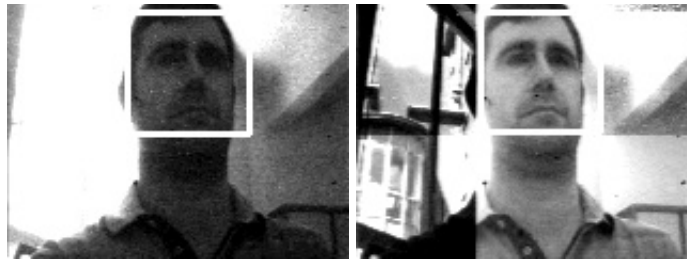


Figure 3. FPGA-based system for test purposes. The FPGA controls the chip and also serves as the interface to communicate with a computer through a USB port.



(a) Face tracking



(b) Pedestrian tracking

Figure 4. Experimental results from the prototype chip:¹⁰ (a) face tracking, and (b) pedestrian tracking, with global adaptation (left) and ROI-driven adaptation (right).

to adapt and reconfigure the image capture on a frame basis according to the scene content. This functionality has been experimentally proved by a prototype vision chip implementing the proposed architecture.

ACKNOWLEDGMENTS

This work has been funded by the Spanish Government through projects TEC2012-38921-C02 MINECO (European Region Development Fund, ERDF/FEDER), IPT-2011-1625-430000 MINECO and IPC-20111009 CDTI (ERDF/FEDER), by Junta de Andalucía through project TIC 2338-2013 CEICE and by the Office of Naval Research (USA) through grant N000141410355

REFERENCES

- [1] Mase, M., Kawahito, S., Sasaki, M., Wakamori, Y., and Furuta, M., "A wide dynamic range CMOS image sensor with multiple exposure-time signal outputs and 12-bit column-parallel cyclic A/D converters," *IEEE*

Journal of Solid-State Circuits **40**(12), 2787–2795 (2005).

- [2] Teixeira, E., Santos, F., and Mesquita, A., “High fill factor CMOS APS sensor with extended output range,” *Electronics Letters* **46**(25), 1658–1659 (2010).
- [3] Ma, C., San-Segundo-Bello, D., Hoof, C., and Theuwissen, A., “High dynamic range hybrid pixel sensor,” *Electronics Letters* **47**(12), 695–696 (2011).
- [4] Xhakoni, A. and Gielen, G., “A 132-dB dynamic-range global-shutter stacked architecture for high-performance imagers,” *IEEE Trans. on Circuits and Systems-II* **61**(6), 398–402 (2014).
- [5] “Int. Solid-State Circuit Conference 2014 Trends.” <http://isscc.org/trends/> (accessed on 28 October 2014).
- [6] Boltz, S., Debreuve, E., and Barlaud, M., “High-dimensional statistical measure for Region-of-Interest tracking,” *IEEE Trans. on Image Processing* **18**(6), 1266–1283 (2009).
- [7] Chiverton, J., Xianghua, X., and Mirmehdi, M., “Automatic bootstrapping and tracking of object contours,” *IEEE Trans. on Image Processing* **21**(3), 1231–1245 (2012).
- [8] Fernández-Berni, J., Carmona-Galán, R., del Río, R., Kleihorst, R., Philips, W., and Rodríguez-Vázquez, A., “Focal-plane sensing-processing: A power-efficient approach for the implementation of privacy-aware networked visual sensors,” *Sensors* **14**(8), 15203–15226 (2014).
- [9] Bradski, G., “The OpenCV library,” *Dr. Dobb’s Journal of Software Tools* (2000). <http://opencv.org/> (accessed on 28 October 2014).
- [10] “MONDEGO Project Web Site.” <http://www.imse-cnm.csic.es/mondego/EI2015/> (accessed on 28 October 2014).