High dynamic range adaptation for ROI tracking based on reconfigurable concurrent dual-sensing

J. Fernández-Berni, R. Carmona-Galán, R. del Río and Á. Rodríguez-Vázquez

A single-exposure technique to extend the dynamic range of vision sensors is presented. It is particularly suitable for vision algorithms requiring region-of-interest (ROI) tracking under varying illumination conditions. The operation is supported by two intertwined photodiodes 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 for each frame. At pixel level, one of the photodiodes senses the pixel value itself, whereas the other, in collaboration with its counterparts in every prescribed ROI, senses the mean illumination of that specific ROI. An additional circuitry interconnecting both photodiodes asynchronously determines the integration period for each ROI according to its mean illumination. The experimental results for a quarter video graphics array prototype CMOS vision sensor are reported.

Introduction: The most usual technique for imagers to deal with scenes featuring high dynamic range (HDR) consists in taking multiple captures per frame with different exposure periods and subsequently combining them [1]. Although this technique performs well for still images, it creates artefacts if motion occurs during multi-exposure. Specialised sensing architectures capable of extending the dynamic range through single exposure [2–4] are thus highly demanded at present [5]. This is also the case for vision sensors that, unlike imagers, are intended not to simply provide high-quality images but to support the automatic extraction of meaningful information from the activity, i.e. motion, taking place in a scene. Despite this fundamental difference between the targeted functionality of imagers and vision sensors, the latter commonly makes use of the HDR techniques devised for the former. The development of specific HDR techniques tailored for the requirements of vision algorithms has received little attention. In this Letter, we describe a sensing architecture particularly suitable for one of the basic tasks implemented by vision algorithms: region-of-interest (ROI) tracking. Once a certain ROI is spotted, a vision algorithm typically tracks it across the scene while carrying out the prescribed analytics. This tracking and the corresponding analytics must 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 and adapted in case they enter the scene. This is exactly the functionality provided by the proposed architecture.

Reconfigurable concurrent dual-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 neighbouring columns and rows across the matrix. After loading the 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 photodiodes and corresponding sensing capacitances at pixel level are reset to V_{rst} by asserting the control signals RST and PI_EN. After reset, RST is switched back to '0' and photo-integration starts. The pixel value, $V_{\mathrm{px}_{ij}}$, will be given by

$$V_{\mathrm{px}_{ij}} = V_{\mathrm{rst}} - \frac{I_{\mathrm{ph}_{ij}}}{C} T_k \tag{1}$$

where $I_{\mathrm{ph}_{ij}}$ 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. To obtain the expression for T_k, we must take the additional photodiode and sensing capacitance into account. These elements are scaled down by a factor m with respect to the main photodiode and sensing capacitance.

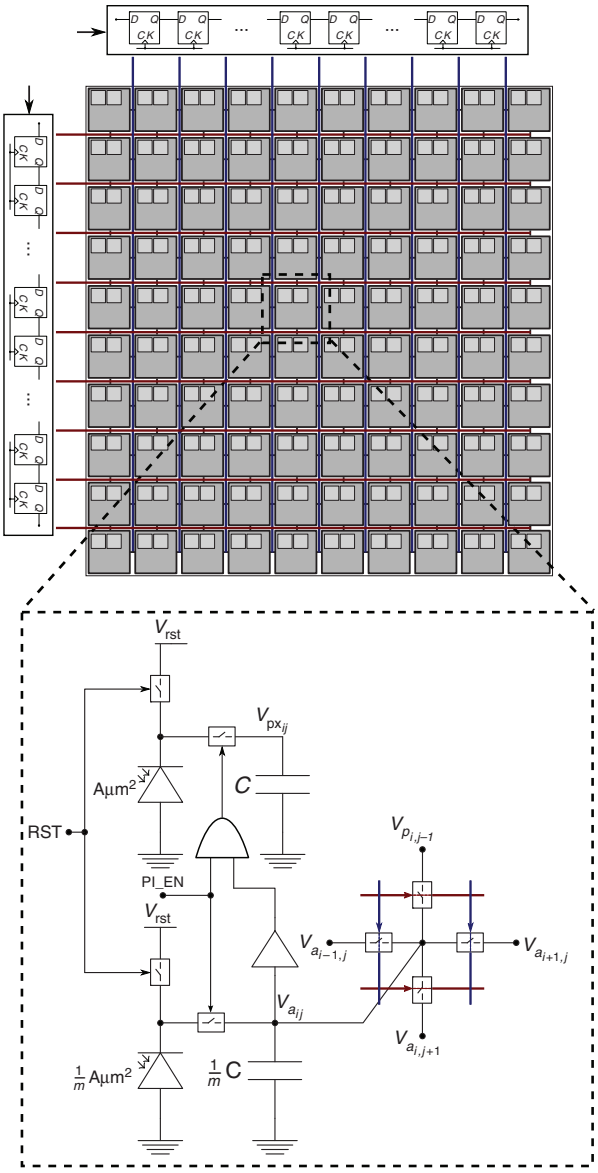

Fig. 1 *Simplified scheme of proposed sensing architecture (above) and mixed-signal circuitry (below) to be included at pixel level*

The specific value of m in a physical realisation 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 photocurrents generated in their associated photodiodes. The voltage $V_{a_{ij}}$ will therefore be the same for all the pixels of the generic region k, expressed as

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

where $W \times H$ are the dimensions of the considered region, in pixels. We are assuming that both pixel photodiodes 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 $V_{\mathrm{px}_{ij}}$. 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. $(V_{\mathrm{rst}} + V_{\mathrm{min}})/2$, where V_{min} is the lower limit of the signal range. Overall, (2) can be re-written as

$$\frac{V_{\mathrm{rst}} + V_{\mathrm{min}}}{2} = V_{\mathrm{rst}} - \frac{\bar{I}_{\mathrm{ph}_k}}{C} T_k \tag{3}$$

where $\bar{I}_{\mathrm{ph}_k}$ is the average current photo-generated in the region k,

directly proportional to its mean illumination. Solving (3) for T_k

$$T_k = \frac{C}{2}\frac{\Delta V_{\mathrm{px}_{\mathrm{MAX}}}}{\bar{I}_{\mathrm{ph}_k}} \tag{4}$$

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

$$V_{\mathrm{px}_{ij}} = V_{\mathrm{rst}} - \frac{\Delta V_{\mathrm{px}_{\mathrm{MAX}}}}{2}\frac{I_{\mathrm{ph}_{ij}}}{\bar{I}_{\mathrm{ph}_k}} \tag{5}$$

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 PI_EN switches back to '0'. This instant will in turn depend on the minimum frame rate affordable by the targeted application. Finally, note that (1)–(5) will be valid as long as the scale factor m can be applied accurately. If the photodiode and the capacitance sensing the mean illumination are scaled down too much, nonlinear terms and mismatch can lead to a great deviation with respect to the ideal linear operation just described.

Experimental results: We have implemented the sensing architecture sketched in Fig. 1 for a quarter video graphics array (QVGA) prototype CMOS vision sensor. A photograph of the chip together with a microphotograph of part of the pixel matrix and the pixel layout is depicted in Fig. 2. The photodiodes 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. The switches and capacitors are implemented by single MOS transistors. The main characteristics of the chip are summarised in Table 1.

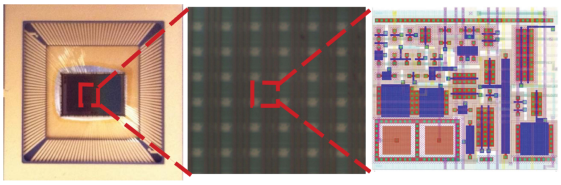

Fig. 2 *Prototype vision sensor along with microphotograph of part of pixel matrix and pixel layout*

Table 1: Summary of main chip characteristics

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

The prototype has been embedded into a field-programmable gate array-based system for testing purposes. The captured images are sent to a personal computer where we make use of the Open CV library to run ROI tracking vision algorithms on them. When a certain ROI is detected, the coordinates of its bounding rectangle are transmitted on-the-fly to the test board for the sensor to adapt the next capture correspondingly. Two examples are shown in Fig. 3: face tracking (Fig. 3*a*) and pedestrian tracking (Fig. 3*b*). 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 the details of the detected ROIs previously missed. Furthermore, the details on other regions are also retrieved, thanks to the focal-plane division required to adapt the capture for those ROIs. The whole sequences can be downloaded from [6]. Intra-frame dynamic ranges of up to 102 dB have been experimentally achieved for different image regions by applying the technique described in this Letter.

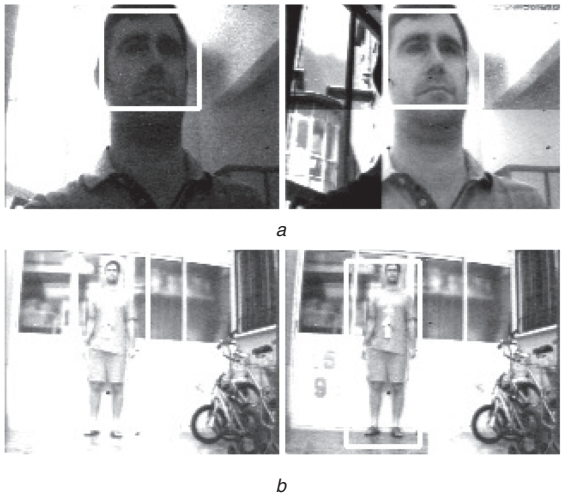

Fig. 3 *Experimental results from prototype chip*

a Face tracking
b Pedestrian tracking
With global adaptation (left) and ROI-driven adaptation (right). Whole sequences can be downloaded from [6]

Conclusion: When it comes to extracting meaningful information from a scene, vision algorithms have to cope with changing illumination conditions. Generally, all the reported techniques targeting HDR 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 Letter presents a specialised sensing architecture suitable for HDR ROI tracking. It permits adapting and reconfiguring 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.

Acknowledgment: This work was 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 the Junta de Andalucía through project TIC 2338-2013 CEICE and by the Office of Naval Research (USA) through grant N000141410355.

28 August 2014
doi: 10.1049/el.2014.3136
One or more of the Figures in this Letter are available in colour online.

J. Fernández-Berni, R. Carmona-Galán, R. del Río and Á. Rodríguez-Vázquez (*Institute of Microelectronics of Seville, CSIC-Universidad de Sevilla, Seville, Spain*)

E-mail: berni@imse-cnm.csic.es

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 J. Solid-State Circuits*, 2005, **40**, (12), pp. 2787–2795
2 Teixeira, E.C., Santos, F.V., and Mesquita, A.C.: 'High fill factor CMOS APS sensor with extended output range', *Electron. Lett.*, 2010, **46**, (25), pp. 1658–1659
3 Ma, C., San Segundo Bello, D., Hoof, C., and Theuwissen, A.: 'High dynamic range hybrid pixel sensor', *Electron. Lett.*, 2011, **47**, (12), pp. 695–696
4 Xhakoni, A., and Gielen, G.: 'A 132-dB dynamic-range global-shutter stacked architecture for high-performance imagers', *IEEE Trans. Circuits Syst. II*, 2014, **61**, (6), pp. 398–402
5 Int. Solid-State Circuit Conference 2014 Trends. Available http://www.isscc.org/trends/, accessed 28 August 2014
6 MONDEGO Project Web Site. Available at http://www.imse-cnm.csic.es/mondego/Elect_Letters/, accessed 28 August 2014