

Single-Exposure HDR Technique Based on Tunable Balance between Local and Global Adaptation

Jorge Fernández-Berni, Ricardo Carmona-Galán, *Member, IEEE*, and Ángel Rodríguez-Vázquez, *Fellow, IEEE*

Abstract—This paper describes a high dynamic range technique that compresses wide ranges of illuminations into the available signal range with a single exposure. An on-line analysis of the image histogram provides the sensor with the necessary feedback to dynamically accommodate changing illumination conditions. This adaptation is accomplished by properly weighing the influence of local and global illumination on each pixel response. The main advantages of this technique with respect to similar approaches previously reported are: i) standard active pixel sensor circuitry can be used to render the pixel values; ii) the resulting compressed image representation is ready either for readout or for early vision processing at the very focal plane without requiring any additional peripheral circuit block. Experimental results from a prototype smart image sensor achieving a dynamic range of 102dB are presented.

Index Terms—high dynamic range, single exposure, smart image sensor, focal-plane circuitry, tone mapping, histogram evaluation, split-diode.

I. INTRODUCTION

DESPITE the great deal of techniques reported in the literature [1], High Dynamic Range (HDR) continues to be a hot research topic within the image sensor community [2]–[5]. The primary reason for this diversity is that no specific approach is capable of optimally compensating all the trade-offs involved—fixed pattern noise, fill factor, motion artifacts, computational load, etc—for every possible application framework considered. Smart image sensors [6] constitute a good example where particular HDR specifications are demanded. For these sensors, whose targeted functionality entails image processing in addition to image capture, a faithful representation of a HDR scene must be available as soon as possible within the signal processing chain, ideally at the very focal plane. Designers do not thus have to face any constraints when it comes to distributing processing tasks in the sensor chip, including the choice about the analog or digital nature of their implementation. Common techniques like multi-capture [7] or multi-reset [8] require heavy digital post-processing to render each image. They are not therefore suitable for computationally-efficient smart image sensors. That digital post-processing precludes any kind of strategies to make the most of the inherent parallelism of many vision tasks, especially at early stages, from the very beginning of the

processing chain. Such strategies may include mixed-signal processing at pixel level, column level, Region of Interest (ROI) etc.

In this context, we propose a single-capture HDR technique that fits wide ranges of illumination into the available analog signal swing by balancing the influence of local and global illumination on the particular integration time of each pixel. This technique arises as an extended functionality of the circuit architecture described in [9]. This architecture was originally intended for ROI tracking algorithms to cope with HDR scenes by independently adjusting the integration time of ROIs according to their mean illumination. While fully functional for this task, artifacts are unavoidably generated due to the need of adapting the capture for specific regions. Moreover, when no focus of attention is provided—in other words, when the ROI is the entire image—, this approach boils down to simple adaptation to the global mean illumination. Such adaptation is not good enough to prevent details from being missed even in scenes featuring moderate illumination variation. In this paper, we explain the procedure we have devised to carry out artifact-free HDR compression by exploiting the same circuitry of [9] in a different way.

Our contribution takes inspiration from several concepts and ideas previously reported. First, it makes use of two photodiodes per pixel. This is used in the industry for dual concurrent exposure [10], with a large photodiode for high sensitivity and a small one for low sensitivity. In our case, a single exposure suffices, with the large photodiode sensing the pixel value itself and the small one enabling the tunable balance between local and global adaptation. The function of the small photodiode in this scheme resembles, generally speaking, that of the so-called *intrinsically photosensitive retinal ganglion cells* as additional non-image-forming photosensitive devices in the mammalian retina. These cells, whose exact role in vision remains under debate, seem to take part in different natural mechanisms for light adaptation [11]. In our case, the small photodiode also constitutes a non-image-forming photosensitive device taking part in a particular process of light adaptation. The proposed technique can also be described in terms of a global operator for HDR tone-mapping compression based on an on-line evaluation of the image histogram [2]. In our case, this on-line evaluation as well as the required circuitry at pixel level are simpler, allowing the use of standard Active Pixel Sensor (APS) structures for the image-forming photodiode. Finally, we combine the local adaptation suggested in [4] with the gain reduction based on the global average illumination experimentally measured in natural vision systems [12]. This is achieved without introducing a peripheral

The authors are with the Instituto de Microelectrónica de Sevilla (CSIC-Univ. de Sevilla), 41092 Seville, Spain (email: berni@imse-cnm.csic.es).

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

Manuscript received XXX XX, XXXX; revised XXX XX, XXXX.

balance block as needed in [4]. Massively parallel early vision processing can thus be carried out right at the focal plane [13] on the resulting HDR image representation.

II. TECHNIQUE DESCRIPTION

Consider the pixel circuitry depicted in Fig. 1. The pixel value itself is represented by $V_{px_{i,j}}$. The voltage $V_{a_{i,j}}$ encodes the signal adjusting the integration time according to the prescribed balance between local and global adaptation explained later on. The scale factor m , equal to 1 in our prototype, permits to save area while keeping the operation unaffected as long as proportional linear photo-response is ensured for the lower sensing structure [9]. Note that a single global signal GL_EN controls the switches interconnecting neighboring pixels. This signal simultaneously sets ON or OFF all the switches across the whole image plane, turning $V_{a_{i,j}}$ into a common averaged voltage \bar{V}_a when set ON. In such a case, it can be demonstrated [9] that the temporal dynamics of \bar{V}_a during photointegration is governed by:

$$\bar{V}_a(t) = V_{rst} - \frac{\bar{I}_{ph}}{C} t \quad (1)$$

where V_{rst} is the reset voltage and \bar{I}_{ph} is the average current photogenerated across the image, directly proportional to the average illumination impinging on the sensor. The integration time determining the pixel values can thus be adjusted to the available signal swing. To this end, the input threshold voltage of the digital buffers connected to \bar{V}_a must be designed to coincide with V_{mid} , the middle point of the signal range. As mentioned in the introduction and proved in [9], this approach works well for the adjustment of the integration time in specific ROIs. There, large deviations with respect to the average illumination are not usual. However, such deviations do occur across HDR scenes, leading to saturated pixels in bright areas and poor contrast in dark regions. An example considering four arbitrary pixels is shown in Fig. 2. Pixels $V_{px_{c,d}}$ and $V_{px_{e,f}}$ feature illuminations close to the global mean illumination represented by \bar{V}_a . Consequently, the integration time derived from this mean illumination is adequate for them. On the contrary, the pixel $V_{px_{a,b}}$ belonging to a dark region could take advantage of a longer integration interval until the limit T_{max} ultimately established by the targeted frame rate. Likewise, the saturated pixel $V_{px_{g,h}}$ would need to reduce its gain to accommodate its response to the available signal range.

In order to overcome the drawbacks of an adaptation exclusively based on the global illumination, we propose to balance the influence of the local and global illuminations over the integration time commensurate with the content of the scene. This means that each pixel determines its own integration time according to such balance, which can be seen as a global tone-mapping operator since a given pixel luminance will always lead to the same output pixel intensity for a particular image. An illustrative example of this strategy applied to the pixels of Fig. 2 is depicted in Fig. 3. We have removed the temporal dynamics of $V_{px_{c,d}}$ and $V_{px_{e,f}}$ to avoid clutter since we now have to consider each $V_{a_{i,j}}$ individually. Let us start by explaining $V_{px_{g,h}}$ and its companion $V_{a_{g,h}}$. As GL_EN is initially

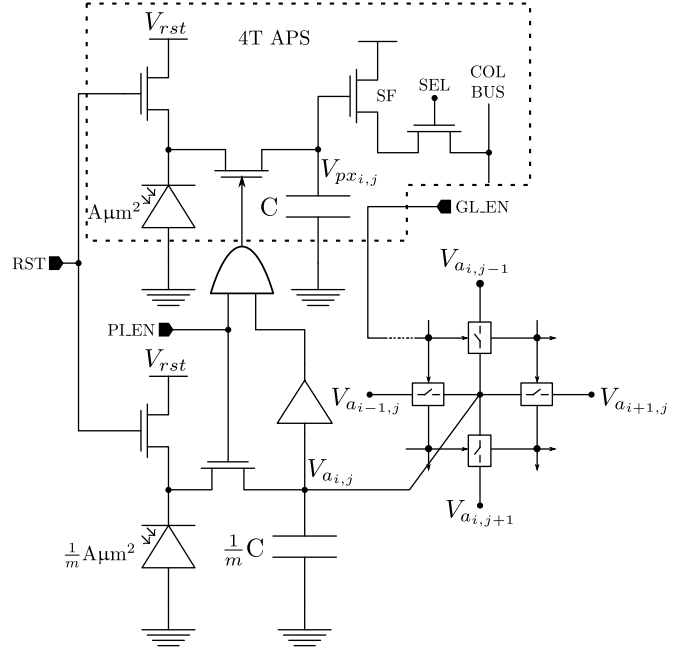


Fig. 1. Circuitry required at pixel level for the proposed HDR technique.

set to '0', both voltages follow the same evolution until $V_{a_{g,h}}$ reaches V_{mid} . At that time instant, photointegration stops for $V_{px_{g,h}}$, continuing for $V_{a_{g,h}}$ until a short pulse of GL_EN arrives. All the charge redistribution that was constantly taking place when GL_EN was held at '1' in Fig. 2, giving rise to \bar{V}_a , occurs now temporally concentrated in this pulse. The interconnecting switches must therefore be properly sized for this redistribution to complete as fast as possible, especially when high image resolutions are considered. $V_{a_{g,h}}$ —and the rest of $V_{a_{i,j}}$ —is thus reset to a voltage depending on the global illumination. Local adaptation is then enabled again when GL_EN switches back to '0'. Since the value of $V_{a_{g,h}}$ after this reset exceeds V_{mid} , photointegration is also activated for $V_{px_{g,h}}$ until $V_{a_{g,h}}$ reaches V_{mid} once again. This new photointegration interval is shorter than the previous one. The next pulse of GL_EN leads to an even shorter interval, the last one for $V_{px_{g,h}}$ during the current capture. This process of local adaptation modulated by periodic sampling of the global illumination succeeds in accommodating the pixel value to the available signal range by reducing its net gain. Concerning $V_{px_{a,b}}$, the effect is also beneficial. We achieve for this pixel a longer integration interval than for the global adaptation in Fig. 2, resulting in better contrast. The extra integration time, highlighted in Fig. 3, ends when $V_{a_{a,b}}$ crosses V_{mid} during its third reset to \bar{V}_a .

This technique therefore adjusts the temporal evolution of all the pixels for a better fitting of the illumination conditions of the scene into the signal range. Thus, unlike in Fig. 2, \bar{V}_a is not the only factor determining the final value of each pixel but a tunable balance between that voltage, proportional to the global illumination, and the voltage $V_{a_{i,j}}$, proportional to the local illumination. A single parameter regulates this balance: the sampling period T_{adj} determined by the distance

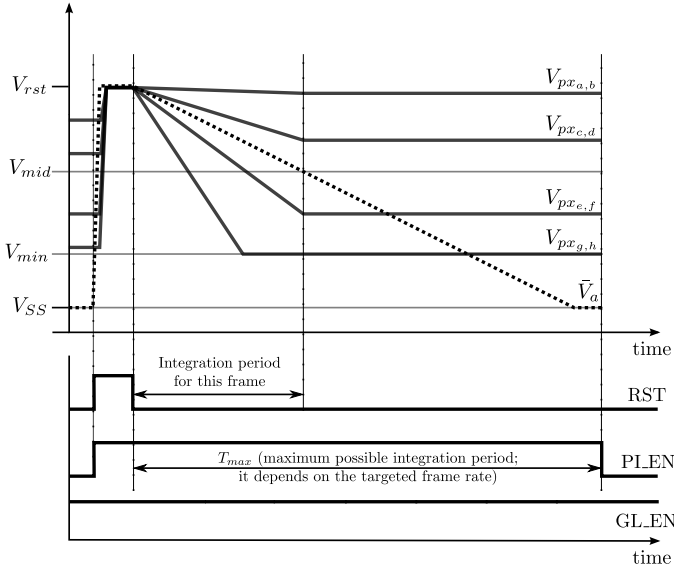


Fig. 2. Adaptation exclusively based on the global mean illumination leads to saturated pixels in bright areas and poor contrast in dark regions.

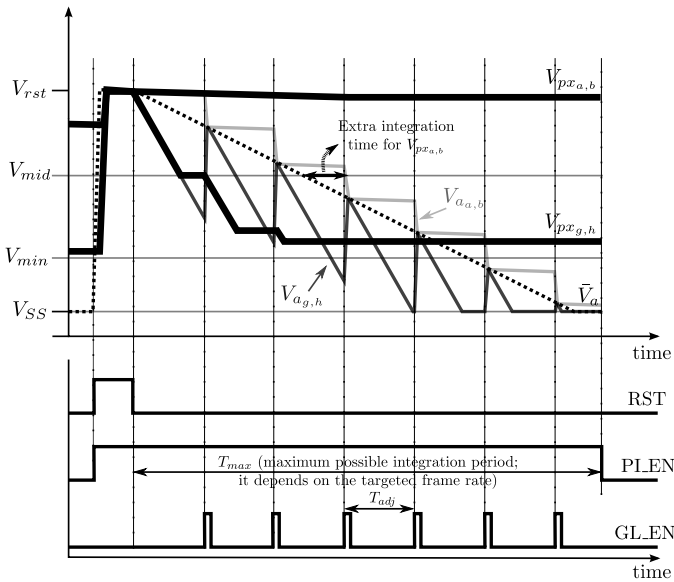


Fig. 3. Local adaptation modulated by periodic sampling of the global illumination improves the parameters of the capture.

between pulses of GL_EN. We must emphasize the importance of a good adjustment of T_{adj} . The drawbacks of a purely global adaptation encoded by $T_{adj} = 0$ have been already remarked. But the other extreme, purely local adaptation encoded by $T_{adj} \rightarrow \infty$, is even worse since all the pixels will be reaching the same value after photointegration, V_{mid} , removing all image contrast. So this parameter demands tuning in agreement with the content of the scene. It will become smaller—global adaptation gains more influence—for those scenes featuring low or moderate variation of illuminations with respect to the global mean illumination. Conversely, T_{adj} will increase when large variations of intra-frame illuminations take place, that is, when a greater degree of local adaptation

is required. In order to dynamically calibrate this parameter, we suggest an on-line analysis of the image histogram as explained next.

III. IMAGE HISTOGRAM ANALYSIS

A flowchart showing the different steps involved in the adjustment of T_{adj} is depicted in Fig. 4. The strategy is straightforward: we aim at reducing the number of saturated pixels as long as such reduction encompasses a better correlation between the image histogram and a perfectly equalized histogram that we are considering as the ideal benchmark. Said another way, we try to retrieve information from bright regions while it does not mean to deteriorate the image contrast. To this end, several variables and parameters are required in addition to T_{adj} . The starting point is an adaptation exclusively based on the global mean illumination, what is equivalent to keeping GL_EN unchanged at ‘1’ during the photocurrent integration. This implies initially setting $T_{adj} = 0$. The number of saturated pixels and the correlation factor are calculated for every input image. They are respectively represented by the variables *sat_pixels* and *corr*. We keep track of the value of T_{adj} set for the previous frame in $T_{adj_{prev}}$. Likewise, the previous number of saturated pixels and the correlation factor with the ideal equalized histogram are also stored in *prev_sat_pixels* and *prev_corr* respectively. Note that *prev_sat_pixels* is initially set to $W \times H + 1$, one pixel over the maximum possible number of saturated pixels for an input image size of $W \times H$ pixels. This enables the immediate trigger of the tuning of T_{adj} as soon as saturated pixels, no matter how many, are detected. The tuning of T_{adj} simply consists in increasing its current value by a prescribed offset denoted by ΔT_{adj} . That is, local adaptation progressively gains more weight in the joint balance with global adaptation. Note that T_{adj} is never decreased in practice since this procedure starts the search for the optimum from an extreme case, namely purely global adaptation encoded by $T_{adj} = 0$. It continues until we are not capable of improving the image representation according to the aforementioned criteria. In such a case, the last value established for T_{adj} is considered the optimum. From that frame on, T_{adj} remains unchanged unless a change in the illumination conditions of the scene occurs. This change is determined by a concurrent increase of saturated pixels and decrease of correlation factor. The parameters k_1 and k_2 , being $k_1 > 1$ and $0 < k_2 < 1$, encode the sensitivity to these changes triggering a new tuning process from scratch. These parameters are highly dependent on the application requirements. Values of k_1 and k_2 close to 1 will make the algorithm very sensitive to illumination changes, better accommodating such changes but also increasing the number of frames where adaptation can render visual artifacts until reaching a stable point. Conversely, values of k_1 and k_2 departing from 1 will delay the adaptation to new conditions but will produce a more stable sequence. An alternative to starting over this process is to search a new T_{adj} by varying the previous optimal value. However, we have found from the experimental tests realized that the former approach attains better performance. Keep in mind that the changes in the scene forcing the search of a new optimum are

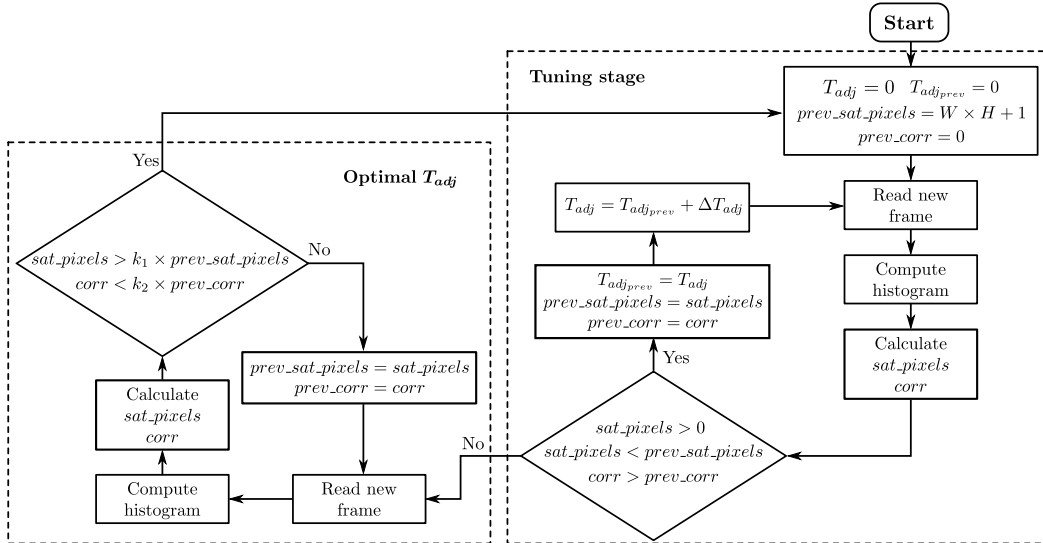


Fig. 4. Flowchart illustrating the tuning loop until reaching the optimal value for T_{adj} .

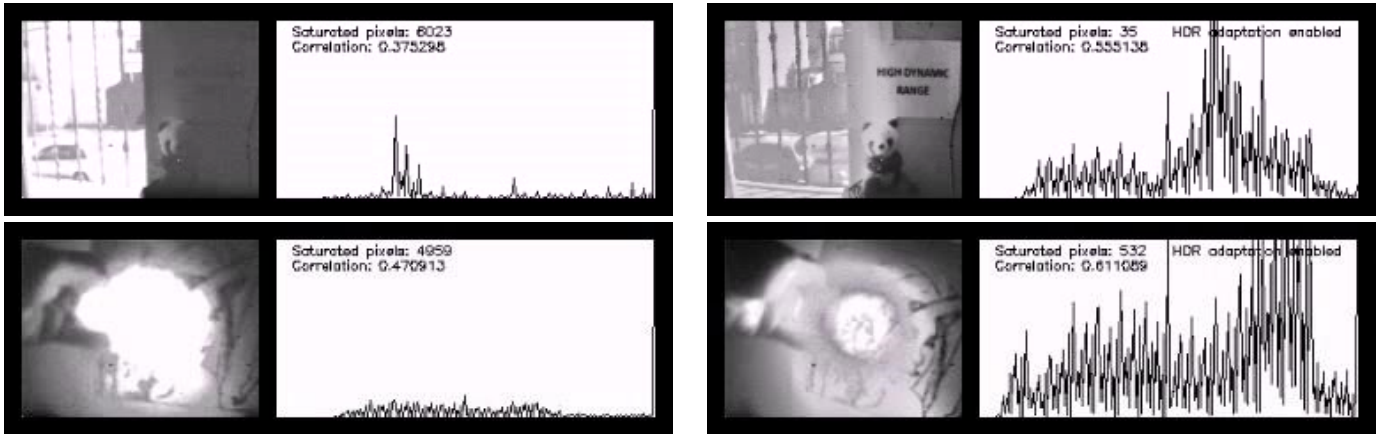


Fig. 5. Experimental results from our prototype sensor: on the left, adaptation exclusively based on the global mean illumination; on the right, resulting scene representations after balancing local and global adaptation from the on-line histogram evaluation. The complete sequences can be downloaded from [14].

unknown a priori. This implies that a divergent variation of T_{adj} towards the new optimum could be mistakenly applied, slowing down the tuning phase. Another parameter greatly impacting the performance of the adjustment is ΔT_{adj} . The larger its value, the faster the tuning stage but also the coarser the adaptation balance. This trade-off must be solved from the specifications of the particular application scenario considered. Even a strategy of dynamic variation of ΔT_{adj} aligned with the characteristics of the scene to be surveyed could provide better performance than a fixed value. In any case, we would like to emphasize the lightweight nature of the proposed technique in terms of computational demand. In addition to only requiring a single exposure, which is inherently faster than the usual multi-exposure procedure, a single histogram and some comparisons between few variables and parameters are simply needed per image.

IV. EXPERIMENTAL RESULTS

We have used the same prototype image sensor, test board and OpenCV-based software environment as in [9] for the experimental demonstration of the technique just described. All these elements were fully reported in [15]. Each image captured by the sensor is grabbed by the FPGA on the test board and sent to a PC for the on-line adjustment of T_{adj} . Specifically, the analysis of the histogram makes use of the functions `calcHist` for the calculation of the histogram and `compareHist` for its comparison with the ideal equalized histogram. Both functions are included in OpenCV, release 2.4.0. A value of T_{adj} is generated for each image coming from the sensor. It takes about $120\mu\text{s}$ for an Intel Core i7-2640M at 2.8GHz to calculate it per frame. This value is sent back in real time to the FPGA that reconfigures the control signals of the sensor chip for the next frame to be captured accordingly. Two examples of the results obtained are shown in Fig. 5. On the left, we depict the representation of two different HDR scenes rendered by an adaptation exclusively based on



Fig. 6. First scene of Fig. 5 captured by an iPhone 4: (a) single-exposure mode; (b) HDR mode. These images can also be downloaded from [14].

the global mean illumination along with the corresponding histogram. On the right, the same scenes after applying the tuned balance between local and global illuminations together with its histogram. No aesthetic post-processing is employed on these images. They are the raw outcome of the HDR focal-plane adaptation and subsequent A-to-D conversion stage. The number of saturated pixels and the correlation factor, sat_pixels and $corr$ respectively in Fig. 4, are registered on the histograms. The whole sequences from which the components of Fig. 5 have been extracted can be downloaded from [14]. Qualitatively, we can see how most of the details from the areas saturated for global adaptation are retrieved after tuning T_{adj} while a much better contrast is accomplished for the dark region. The goodness of the adaptation can also be quantitatively gauged from the reduction in the number of saturated pixels and the greater correlation factor with an equalized histogram. We set ΔT_{adj} to $800\mu s$ as this value proved to adequately trade speed for balance accuracy in most of the tests performed. The optimal values found for T_{adj} were respectively 5.6ms and 12.8ms, with $T_{max} = 100ms$, $k_1 = 1.2$ and $k_2 = 0.8$. For the sake of comparison, we have taken two snapshots of the first scene of Fig. 5 with the rear camera of an iPhone 4. These snapshots, shown in Fig. 6, can also be downloaded from [14]. The left image was captured with the single-exposure mode whereas the right one made use of the HDR mode. In this mode, three images with progressive exposure times are captured and merged to create the HDR representation. Leaving apart the obvious difference of resolution—the iPhone 4 features a backside-illuminated 5-megapixel image sensor—our approach is capable of achieving a better compression of the illumination map, specially in terms of contrast in the dark zones. Finally, Table I summarizes most of the aspects already mentioned for this work in relation with techniques previously reported. Dynamic ranges of up to 102dB have been measured from our prototype sensor, fabricated in a standard $0.18\mu m$ CMOS process and conceived not as an image sensor but as a vision sensor. This means that its targeted processing capabilities were given priority over image quality—specifically concerning noise—during the design. It also affected the pixel pitch. Hence, we consider that both dynamic range and pixel pitch can definitely be improved in future realizations paying careful attention to the HDR functionality. Manufacturing in CMOS Image Sensor (CIS) technologies will significantly contribute to succeed in this objective too.

Reference	[2]	[4]	[10]	This work
Single exposure	Yes	Yes	No	Yes
Focal-plane HDR representation	Yes	No	No	Yes
APS afforded	No	Yes	Yes	Yes
Dynamic Range (dB)	151	n.a.	118	102
Pixel pitch (μm)	33	15	6	19.6

TABLE I
COMPARISON OF OUR TECHNIQUE WITH SIMILAR APPROACHES.

V. CONCLUSIONS

We have presented a HDR technique that simultaneously fulfills three characteristics not reported to coincide in any previous approach, to the best of our knowledge. First, a single exposure suffices to generate the HDR image representation. This significantly reduces motion artifacts when compared to multi-exposure techniques. Second, such representation is available for readout or further processing at the very focal plane. Vision sensors can clearly benefit from this feature by exploiting massively parallel early vision processing at pixel level. And third, it affords the use of standard APS structures for high-quality image capture. Future work will aim at reducing the area overhead of the non-image-forming photodiode and its related circuitry. Distributing and re-using these elements among several pixels seem to be a promising strategy to be explored.

REFERENCES

- [1] A. Spivak *et al.*, “Wide-dynamic-range CMOS image sensors—comparative performance analysis,” *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2446–2461, 2009.
- [2] S. Vargas-Sierra *et al.*, “A 151 dB high dynamic range CMOS image sensor chip architecture with tone mapping compression embedded in-pixel,” *IEEE Sensors J.*, vol. 15, no. 1, pp. 180–195, 2015.
- [3] A. Xhakoni and G. Gielen, “A 132-dB dynamic-range global-shutter stacked architecture for high-performance imagers,” *IEEE Trans. Circuits Syst. II*, vol. 61, no. 6, pp. 398–402, 2014.
- [4] G. Sicard *et al.*, “A CMOS HDR imager with an analog local adaptation,” in *Int. Image Sensor Workshop*, 2013.
- [5] D. Kim and M. Song, “An enhanced dynamic-range CMOS image sensor using a digital logarithmic single-slope ADC,” *IEEE Trans. Circuits Syst. II*, vol. 59, no. 10, pp. 653–657, 2012.
- [6] J. Ohta, *Smart CMOS Image Sensors and Applications*. CRC Press, 2007.
- [7] M. Mase *et al.*, “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.
- [8] A. Belenky *et al.*, “A snapshot CMOS image sensor with extended dynamic range,” *IEEE Sensors J.*, vol. 9, no. 2, pp. 103–111, 2009.
- [9] J. Fernández-Berni *et al.*, “High dynamic range adaptation for ROI tracking based on reconfigurable concurrent dual sensing,” *Electronics Letters*, vol. 50, no. 24, pp. 1832–1834, 2014.
- [10] J. Solhusvik *et al.*, “A comparison of high dynamic range CIS technologies for automotive,” in *Int. Image Sensor Workshop*, 2013.
- [11] K. Wong *et al.*, “Photoreceptor adaptation in intrinsically photosensitive retinal ganglion cells,” *Neuron*, vol. 48, no. 6, pp. 1001–1010, 2005.
- [12] R. Shapley and C. Enroth-Cugell, “Visual adaptation and retinal gain controls,” *Progress in Retinal Research*, vol. 3, pp. 263–346, 1984.
- [13] A. Zarándy, Ed., *Focal-plane Sensor-Processor Chips*. Springer, 2011.
- [14] MONDEGO project web site. [Online]. Available: <http://www.imsecnm.csic.es/mondego/TCASII/> (accessed on October 9, 2015)
- [15] J. Fernández-Berni *et al.*, “Focal-plane sensing-processing: A power-efficient approach for the implementation of privacy-aware networked visual sensors,” *Sensors*, vol. 14, no. 8, pp. 15,203–15,226, 2014.