HDR-ARtiSt: High Dynamic Range Advanced Real-time imaging System

Pierre-Jean Lapray*, Barthélémy Heyrman*, Matthieu Rossé* and Dominique Ginhac*†
*LE2I UMR 5158, Univ Burgundy, Dijon, France
Email: (Pierre-Jean.Lapray,heybar,mrosse,dginhac)@u-bourgogne.fr
†Corresponding author

***Abstract*—This paper describes the HDR-ARtiSt hardware platform, a FPGA-based architecture that can produce a real-time high dynamic range video from successive image acquisition. The hardware platform is built around a standard low dynamic range (LDR) CMOS sensor and a Virtex 5 FPGA board. The CMOS sensor is a EV76C560 provided by e2v. This 1.3 Megapixel device offers novel pixel integration/readout modes and embedded image pre-processing capabilities including multiframe acquisition with various exposure times. Our approach consists of a hardware architecture with different algorithms: double exposure control during image capture, building of an HDR image by combining the multiple frames, and final tone mapping for viewing on a LCD display. Our video camera system is able to achieve a real-time video rate of 30 frames per second for a full sensor resolution of $1,280 \times 1,024$ pixels.**

I. Introduction

Many camera sensors suffer from limited dynamic range (i.e. the ratio between the lightest and darkest pixel). The result is that there is a lack of clear details in displayed images and videos. This is specifically true for real scenes that simultaneously include areas of low and high illumination due to transitions between sunlit and shaded areas. When capturing such a scene, the camera is unable to store the full dynamic range resulting in low quality images where details are concealed in shadows or washed out by bright lights.

High Dynamic Range (HDR) imaging techniques appear as a solution to overcome this major issue by extending the precision of the digital images. HDR imaging encodes images with higher than standard 24-bit RGB format, increasing the range of luminance that can be digitally stored. HDR images can be created in two different ways. The first method requires the development of specific HDR sensors that can capture the entire scene dynamic range. Several sensors development have been done with techniques such as well-capacity adjusting, time-to-saturation, or self-reset (see [1] for a comparative analysis of these sensor architectures). To perform these functions, most of these sensors provide a processing unit at chip level or at column level [2], [3], and even at pixel level [4]–[6]. The second method relies on conventional Low Dynamic Range (LDR) sensors to capture HDR data by recording multiple exposures of the same scene [7]–[10]. By limiting the exposure time, the image loses low-light detail in exchange for improved details in areas of high illumination. A contrario, by increasing the exposure time, the resulting image contains the details in the dark areas but none of the details in the bright areas due to pixel saturation. Complex algorithms build a single HDR image (i.e. radiance map) that covers the full dynamic range by combining the details of the successive acquisitions. However, current display technology has a very limited dynamic range, so that HDR images need to be compressed by tone mapping operators [11]–[13] in such a way that the visual sensation of the real scene is faithfully reproduced.

This paper presents a complete hardware system dedicated to HDR imaging from multiple exposures acquisition, through radiance maps and tone mapping, to display. We present a simplified and practical hardware solution to produce real time HDR video. The remainder of the paper is as follows: in Section II, we briefly review the literature on real-time HDR imaging solutions. Section III describes HDR creating and tone mapping algorithms. We propose a dedicated hardware architecture in Section IV. Our results are presented in Section V. Conclusions are provided in Section VI.

II. Related work

The problems of capturing the complete dynamic range of a real scene and reducing this dynamic range to a viewable range have drawn the attention of many authors. However, the main part of the proposed algorithms have been developed without taking into account the specificities and the difficulties inherent to hardware implementations. Unfortunately, these works are not generally suitable for efficient real-time implementation on smart cameras. As a consequence, generating real-time HDR images still remains a interesting challenge.

In 2007, Hassan [14] described an FPGA-based architecture for local tone mapping of gray scale HDR images, able to generate $1,024 \times 768$ pixel image at 60 frames per second. The architecture is based on a modification of the nine-scale Reinhard's operator [15]. Recently, Vytla et al. [16] present another hardware implementation of tone mapping based on a simplified Poisson solver for Fattal's local operator [17]. Ching-Te et al. [18] developed an integrated tone mapping processor using an ARM SOC platform that can can process $1,024 \times 768$ images at 60 fps, runs at 100 MHz clock and consumes a core area of $8.1mm^2$ under TSMC 0.13 μm technology. However, these works focus only on the tone mapping process and do not care about the HDR capture, using a set of high dynamic range images extracted from the Debevec library [7] for example. Kang et al. [10] describe an algorithmic solution performing automatic exposure control, HDR stitching across neighboring frames, and tone mapping for viewing, while handling moving parts in the scene. However, the implementation on a 2GHz Pentium 4 machine does not reach the real-time constraints because the processing time for each video frame $(768 \times 1,024)$ is about 10 seconds. Based on Kang's algorithms, Youm et al. [19] create an HDR video by merging two images from different exposures acquired by a stationary video camera system. Their methodology mainly relies on the simple strategy of automatically controlling exposure times and effectively combines bright and dark areas in short and long exposure frames. Unfortunately, they do not reach real-time processing with approximately 2.5 seconds for each (640×480) pixel frame on a 1.53 GHz AMD Athlon XP 1800+ machine. In 2011, Tocci et al. [20] presented an high-cost HDR system built around a special optical architecture. The camera is composed of three sensors ("low", "medium" and "high" exposure sensors) using a beam splitter between lens and sensors to break up the light into three parts. This method captures perfectly synchronized video frames that are combined into a software application. Unfortunately, this system requires a lot of expensive equipment.

III. Efficient algorithms for HDR video

A. HDR creating

The pixel values of an image are the result of a non linear function of the exposure. This function is linked to the characteristic curve (i.e. the response to the variations in exposure) of the sensor. So, creating an HDR image from multiple LDR images, i.e producing a radiance map requires two major stages:

- Recovering the sensor response function;
- Reconstructing the HDR radiance map.

From literature, three popular algorithms for recovering the response curve can be extracted: Debevec et al. [7], Mitsunaga et al. [8], and Robertson et al. [21]. A technical paper [22] compares the first two algorithms in a real-time implementation in software. The results of time calculation for an image are significantly identical. Based on a detailed description of these methodologies [23] we decided to use the original algorithm developed by Debevec et al. [7].

From the system response curve stored in LUT, the radiance value of each pixel can be evaluated by a weighted combination of the pixels from each LDR frame.

B. Tone mapping

The HDR pixels are represented by a high bit-depth conflicting with the standard display devices, requiring a high to low bit-depth tone mapping. Cadik et al. [13] show that the global part of a tone mapping operator is most essential to obtain good results. Moreover, a global tone mapper is the easiest way to implement the algorithm in real time, because local operators require complex computations, and also may generate halo artifacts. The choice of a candidate tone mapping operator has been done after comparing some global algorithms applied to a radiance map constructed from two images. Several global algorithms have been extensively tested in C++. According to our temporal and hardware constraints, the best compromise is the global tone mapper introduced by Duan et al. [24].

IV. Proposed FPGA architecture for HDR video

A. Global hardware architecture

Our global pipeline architecture dedicated to HDR video creating is shown in Fig. 1. This architecture relies on a specific FPGA daughter board embedding a 1.3 million pixel CMOS imager from e2v [25]. With a dynamic range of about 60dB, this sensor can be viewed as a low dynamic range imaging system. It embeds dedicated image pre-processing such as image histogram. Each frame is delivered with its own histogram encoded in the image data stream footer. Based on these histograms, a double auto exposure algorithm has been implemented to evaluate the optimum integration times. The sensor also includes a bi-frame acquisition mode that is fundamental for HDR imaging. With such a mode, the sensor is able to successively acquire 2 images with different integration times, the second acquisition simultaneously occurring with the first frame readout. The sensor sends alternatively low exposure (LE) frame and high-exposure (HE) frame. The LE frame is first stored in DDR2 memory. Simultaneously with the acquisition of the HE frame, the LE frame is read in order to create the HDR image. The sensor sends full resolution images at 60 frame/s, which means that we have an output rate of 30 frame/s for the HDR image generation. Then, the HDR image is tone mapped as described in Subsection III-B. The 8-bit resulting image is displayable on a LCD screen through the DVI controller.

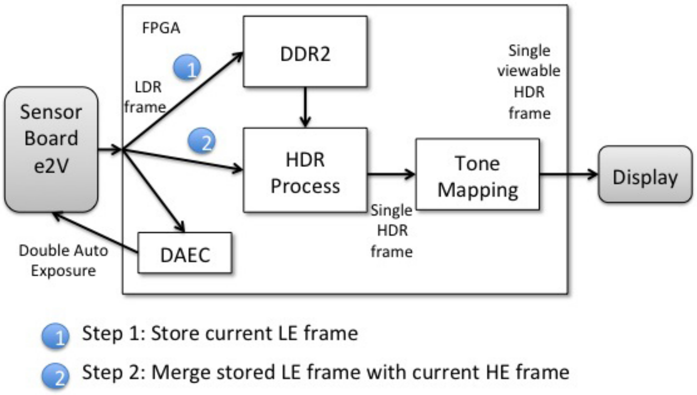
FPGA
DDR2
Single viewable HDR frame
1
LDR frame
2
Sensor Board e2V
HDR Process
Tone Mapping
Display
Single HDR frame
DAEC
Double Auto Exposure
1 Step 1: Store current LE frame
2 Step 2: Merge stored LE frame with current HE frame

Fig. 1. HDR imaging system architecture

B. Double Exposure Control

The auto-bracketing system implemented in many cameras is a good way to expand the dynamic range captured by the sensor. Traditionally, the camera estimates the good exposure and takes several pictures with many multiples of f-stops up and down. Our way is not to set fixed exposures, but rather to make the exposures varying from frame to frame to get the more pixels well exposed. Our exposure control algorithm is based on pixels statistics on two successive images. The sensor is able to automatically computes a 64 16-bit categories histogram, the number of dark pixels (S_{low}), and the number of saturated pixels (S_{high}) for each acquired frame. Our aim is to calculate the number of pixels in the upper and lower part of the histograms of the two images as:

$$P_{t',s} = \sum_{h=1}^{h=32} \frac{H_{t',s}(h)}{N - S_{low}} \text{ and } P_{t',l} = \sum_{h=33}^{h=64} \frac{H_{t',l}(h)}{N - S_{high}} \quad (1)$$

P is the proportion of pixels located in a part of the histogram. s means the short exposure, l means the long exposure, H is the histogram of the image, h is the category position (1 - 64), S is the number of saturated pixels (dark pixel at level 0 or bright pixel at level 1023) and N is the total number of pixels. The better exposure time for each image can be evaluated as follows:

$$\Delta t_{t',s} = \Delta t_{t'-1,s} \cdot (\frac{\Delta t_{opt,s}}{P_{t',s}}) \text{ and } \Delta t_{t',l} = \Delta t_{t'-1,l} \cdot (\frac{\Delta t_{opt,l}}{P_{t',l}}) \quad (2)$$

$\Delta t_{t'}$ is the exposure time at time t'. Δt_{opt} depends on the number of saturated pixels. It is not possible to precisely evaluate the best exposure for a real scene because the tone mapped HDR image rendering depends on many factors (number of images, type of the scene, sensor sensitivity, etc.). In order to evaluate as precisely as possible the best values of the integration times, the following assumption was made: 80% of the pixels in the lower part of the histogram are present in the low exposure image, and 80% of the pixels in the upper part of the histogram are present in the high exposure image.

To prevent an useless updating system for every frames, we introduce a threshold so that the system is not too slow for our architecture:

$$\Delta t_{sensor} = \begin{cases} \Delta t_{t'} \text{ for } \Delta t_{t'} - \Delta t_{t'-1} \geq \text{threshold} \\ \Delta t_{t'-1} \text{ for } \Delta t_{t'} - \Delta t_{t'-1} < \text{threshold} \end{cases} \quad (3)$$

V. Results and experiments

A. Results of hardware synthesis

Device:	xc5vfx70t-1ff1136
Number of Slice LUTs:	14168/44800 (31%)
Number of Slice Registers:	8132/44800 (18%)
Number of Block RAM/FIFO:	40/148 (27%)
Number of DSP48Es:	4/128 (3%)
Maximum frequency:	94.733MHz

TABLE I
Summary of hardware synthesis report

The proposed architecture has been checked using Modelsim and implemented on our Virtex 5 platform. The architecture has been adapted to deliver a real-time 1280×1024 resolution HDR video at 30 frames/s. Table I summarizes the synthesis report. The clock speed shows the simplicity of hardware and the low utilization of logic cells. The internal memory used is quite large, which justifies the choice of the device of the Virtex 5. Most of the operations are simple, since the use of DSP blocks remains modest. Usually FPGA image processing requires specific device with extended memory capabilities. Here, only the DDR2 memory is used to store an image, and to manage the flow of images delivered by the sensor. The radiance values and the arithmetic operations are performed using the IEEE754 floating-point representation with 32 bits of depth (single precision).

With a video frame rate of 30 frames per second, the computational frame-rate is $30 \times (1280 + 307) \times (1024 + 20) = 49.70$ megapixels per second. As we can see in Table I, our hardware system has a maximum operating frequency of 94.733 MHz, so our system is able to process each pixel at each clock tick. The sensor has an horizontal blanking period of 307 pixels, and a vertical blanking period of 20 rows. The entire design introduces a latency at the end of each row of 65 extra- clock ticks (whether $570ns$ for a clock pixel of $114Mhz$). This constant latency appears but not alters the frame rate because of the horizontal blanking period that is larger than our system latency. Finally, our architecture embeds all the algorithmic operators to produce a single tone-mapped output pixel in real-time.

B. Visual results and experiments

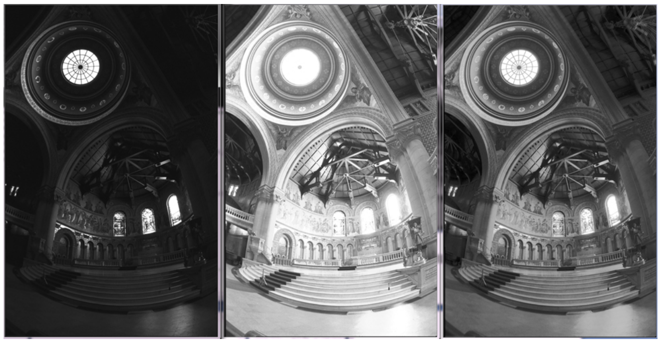

Fig. 2. Evaluation of the hardware architecture using image samples from the Debevec library.

We bechmarked our architecture using multiple image sources. First, we chose a database from the Debevec library [7], which includes images with variable exposure times. In our specific case, set of two images have been used to evaluate our architecture. An example of a resulting HDR image is shown on the right part of the Fig. 2 obtained from a LE frame (left part) and a HE frame (central part). Details in dark areas and bright are represented in the tone mapped image without artifacts. Secondly, we also evaluated our platform with realistic scenes acquired by our sensor board (see Figure 3). Results are similar to those obtained from Debevec library with a good reproduction of fine details in dark and bright areas. We can both distinguish the word "HDR" overexposed in the lampshade, and the R2D2 figurine underexposed in the cup.

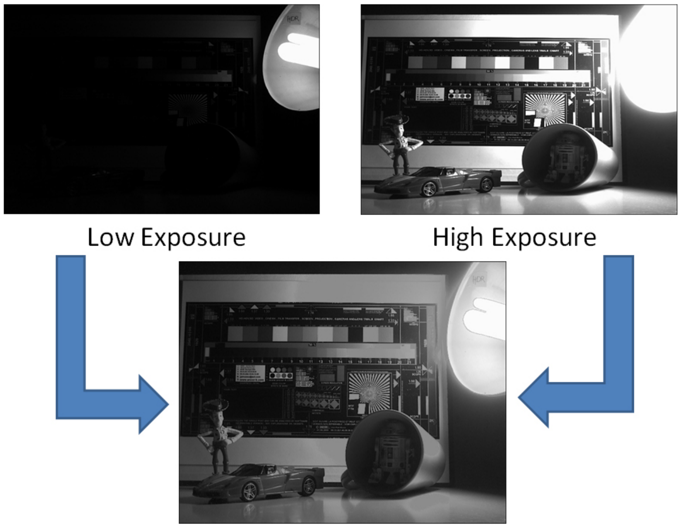

Fig. 3. The HDR image (tone mapped) reconstructed from two exposures captured by our sensor board ($0.062ms$ and $2.254ms$).

VI. Conclusion

In this paper, we present a complete hardware architecture dedicated to HDR video that can fulfill drastic real-time constraints while satisfying image quality requirements. We propose a hardware vision system based on a standard image sensor associated with a FPGA development board. We obtain good results with the conventional technique of HDR creating and a global tone mapping. We are also planning to develop and benchmark other several state-of-art algorithms both for computing radiance maps, calibrating HDR images, and local/global tone mapping. Development of motion correction and alignment of images are also considered.

Acknowledgment

The authors would like to thank the DGCIS (French Ministry for Industry) for financial support within the framework of the project HiDRaLoN.

References

[1] S. Kavusi and A. El Gamal, "A quantitative study of high dynamic range image sensor architectures," in *Proceedings of the SPIE Electronic Imaging '04 Conference*, vol. 5301, Jan 2004, pp. 264–275.

[2] P. Acosta-Serafini, M. Ichiro, and C. Sodini, "A 1/3" VGA linear wide dynamic range CMOS image sensor implementing a predictive multiple sampling algorithm with overlapping integration intervals," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1487–1496, Sept 2004.

[3] M. Sakakibara, S. Kawahito, D. Handoko, N. Nakamura, M. Higashi, K. Mabuchi, and H. Sumi, "A high-sensitivity CMOS image sensor with gain-adaptative column amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 5, pp. 1147–1156, May 2005.

[4] G. Cembrano, A. Rodriguez-Vazquez, R. Galan, F. Jimenez-Garrido, S. Espejo, and R. Dominguez-Castro, "A 1000 FPS at 128 × 128 vision processor with 8-bit digitized I/O," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 7, pp. 1044–1055, Jul 2004.

[5] L. Lindgren, J. Melander, R. Johansson, and B. Mller, "A multiresolution 100-GOPS 4-Gpixels/s programmable smart vision sensor for multisense imaging," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 6, pp. 1350–1359, Jun 2005.

[6] J. Dubois, D. Ginhac, M. Paindavoine, and B. Heyrman, "A 10 000 fps CMOS sensor with massively parallel image processing," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 3, pp. 706–717, Mar 2008.

[7] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 1997, pp. 369–378.

[8] T. Mitsunaga and S. Nayar, "Radiometric Self Calibration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, Jun 1999, pp. 374–380.

[9] M. A. Robertson, S. Borman, and R. L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *Journal of Electronic Imaging*, vol. 12, no. 2, pp. 219–228, 2003.

[10] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Transactions on Graphics (TOG)*, vol. 22, pp. 319–325, July 2003.

[11] K. Devlin, A. Chalmers, A. Wilkie, and W. Purgathofer, "Tone reproduction and physically based spectral rendering," in *Eurographics 2002*. Eurographics Association, Sep 2002, pp. 101–123.

[12] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, pp. 13 – 24, 2005.

[13] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of HDR tone mapping methods using essential perceptual attributes," *Computers & Graphics*, vol. 32, pp. 330–349, 2008.

[14] F. Hassan and J. E. Carletta, "A real-time FPGA-based architecture for a Reinhard-like tone mapping operator," in *Proceedings of the 22nd ACM SIGGRAPH/EUROGRAPHICS symposium on Graphics hardware*, Aire-la-Ville, Switzerland, Switzerland, 2007, pp. 65–71.

[15] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 267–276, 2002.

[16] L. Vytla, F. Hassan, and J. Carletta, "A real-time implementation of gradient domain high dynamic range compression using a local poisson solver," *Journal of Real-Time Image Processing*, pp. 1–15, 2011, doi: 10.1007/s11554-011-0198-5.

[17] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics (TOG)*, vol. 21, pp. 249–256, July 2002.

[18] C.-T. Chiu, T.-H. Wang, W.-M. Ke, C.-Y. Chuang, J.-S. Huang, W.-S. Wong, R.-S. Tsay, and C.-J. Wu, "Real-time tone-mapping processor with integrated photographic and gradient compression using 0.13μm technology on an ARM SoC platform," *Journal of Signal Processing Systems*, vol. 64, no. 1, pp. 93–107, 2010.

[19] S.-J. Youm, W. ho Cho, and K.-S. Hong, "High dynamic range video through fusion of exposured-controlled frames," in *Proceedings of IAPR Conference on Machine Vision Applications*, 2005, pp. 546–549.

[20] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A Versatile HDR Video Production System," *ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH 2011)*, vol. 30, no. 4, p. 9, 2011.

[21] M. A. Robertson, S. Borman, and R. L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *Journal of Electronic Imaging*, vol. 12, no. 2, pp. 219–228, 2003.

[22] W. S. G. Yourganov, "Acquiring high dynamic range video at video rates," Acquiring High Dynamic Range Video at Video Rates, Tech. Rep., 2001.

[23] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. Seidel, and H. Lensch, "Optimal hdr reconstruction with linear digital cameras," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 215–222.

[24] J. Duan, M. Bressan, C. Dance, and G. Qiu, "Tone-mapping high dynamic range images by novel histogram adjustment," *Pattern Recognition*, vol. 43, pp. 1847–1862, May 2010.

[25] e2v technologies, "Ev76c560 BW and colour CMOS sensor," 2009. [Online]. Available: http://www.e2v.com/products-and-services/high-performance-imaging-solutions/industrial-imaging/industrial-imaging-sensors/datasheets—sensors/