

# HDR-ARtiSt : une caméra intelligente dédiée à la vidéo à grande dynamique en temps réel

Pierre-Jean LAPRAY<sup>1</sup>, Barthélémy HEYRMAN<sup>1</sup>, Dominique GINHAC<sup>1</sup>

<sup>1</sup>Le2i - Laboratoire Electronique, d'Informatique et Image - UMR CNRS 6306, Dijon, France  
Pierre-Jean.Lapray@u-bourgogne.fr, Barthelemy.heyrman@u-bourgogne.fr,  
Dominique.Ginhac@u-bourgogne.fr

**Résumé** – Une caméra vidéo standard n'est capable de capturer qu'une partie de l'information contenue dans une scène de gamme dynamique élevée. La même scène peut être presque entièrement perçue par le système visuel humain. Cela est particulièrement vrai pour les scènes réelles où la différence d'intensité lumineuse entre les zones sombres et les zones claires est élevée. La technique d'imagerie qui peut surmonter ce problème est appelé HDRi (High Dynamic Range imaging). Elle produit des images à partir d'un ensemble de plusieurs images LDRs (Low Dynamic Range), capturées avec différents temps d'exposition. Cette technique apparaît comme l'une des solutions la plus appropriée et la moins coûteuse afin d'améliorer la plage dynamique acquise par le capteur. Nous avons développé un prototype de caméra intelligente à base de FPGA, qui peut produire un flux vidéo de haute gamme dynamique en live, et ce à partir de trois acquisitions successives. Notre plate-forme matérielle est construit autour d'un capteur CMOS LDR, et d'un FPGA Virtex 6. Notre approche consiste en une architecture matérielle mélangant en oeuvre les différents algorithmes HDR suivants : un contrôle de l'exposition multiple, une unité de gestion de mémoire, la création HDR et le tone mapping. Notre caméra vidéo permet une vidéo en temps réel à 60 images par seconde pour une grande résolution du capteur de  $1280 \times 1024$  pixels.

**Abstract** – A camera or a video camera is able to capture only a part of a high dynamic range scene information. The same scene can be almost totally perceived by the human visual system. This is true especially for real scenes where the difference in light intensity between the dark areas and bright areas is high. The imaging technique which can overcome this problem is called HDR (High Dynamic Range). It produces images from a set of multiple LDR images (Low Dynamic Range), captured with different exposure times. This technique appears as one of the most appropriate and a cheap solution to enhance the dynamic range of captured environments. We developed an FPGA-based smart camera prototype that can produce a high dynamic range live video colour stream from three successive sensor acquisitions. Our hardware platform is build around a standard LDR CMOS sensor and a Virtex 6 FPGA board. Our approach consists of a hardware architecture with different algorithms : a multiple exposure control, a memory management unit, the HDR creating, and the tone mapping. Our video camera enables a real-time video at 60 frames per second for a full sensor resolution of  $1,280 \times 1,024$  pixels.

## 1 Introduction

Les techniques d'acquisition d'images, que ce soit la photographie ou la vidéo, sont intrinsèquement limitées par la dynamique des capteurs utilisés, c'est-à-dire leur capacité à acquérir et restituer l'étendue des intensités lumineuses présentes dans une scène. Or, ceux-ci ne sont capables d'enregistrer qu'une gamme limitée d'intensités lumineuses en une seule capture. Cela se traduit par des images qui peuvent contenir des zones très sombres, saturées dans les noirs, et des zones sur-exposées saturées dans les blancs. Cela entraîne inévitablement une perte d'information significative de part et d'autre de ces niveaux de saturation. De nombreuses techniques d'optimisation du temps d'exposition ont vu le jour permettant ainsi de maximiser l'information capturée. Seulement ces techniques n'augmentent pas la dynamique du capteur, mais essaye plutôt de l'exploiter au mieux afin d'obtenir une exposition de la scène la plus adéquate possible. Pour augmenter la dynamique et ainsi fournir des images dites HDR (High Dynamic Range), la technique

la plus efficace est la reconstruction HDR à partir de la capture multiple d'images avec des temps d'exposition différents de la même scène. Ainsi, la création d'une image HDR affichable sur un écran nécessite 3 étapes : 1) la capture multiple, 2) la génération de l'image HDR ou carte de radiance, et 3) le rendu sur écran nommé "Tone mapping".

Dans le cas de la vidéo HDR, on trouve dans la littérature des méthodes hardware dédiées à la capture multiple d'images [1, 2], à la génération HDR[3, 4] ou au tone mapping[5, 6, 7]. Malheureusement, de tels systèmes sont souvent incomplets et aucun d'entre eux ne propose une plate-forme matérielle intégrant simultanément la capture, la génération et l'affichage de contenu HDR en temps réel. Néanmoins, un travail de Mann et al. [8] portant sur une architecture complète, traite de l'implémentation du pipeline HDR en temps réel. Toutefois, c'est une architecture à deux caméras à première vue assez coûteuse, utilisable uniquement dans des environnements lumineux fixes ou limités, puisqu'ils limitent les captures à des temps d'exposition définis au préalable. D'autre part, une caméra industrielle

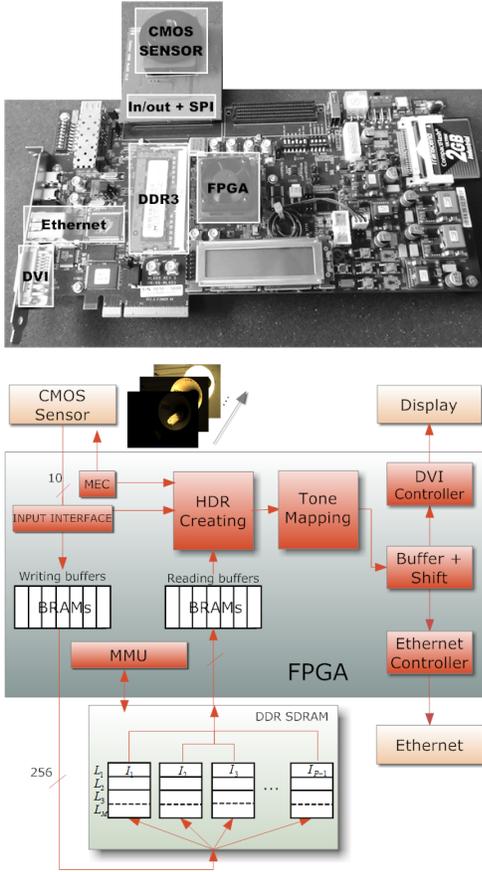


FIGURE 1 – Vue globale de la plate-forme HDR-ARtiSt

complète ("HDR-60 video Camera development kit") est commercialisée, mais nous ne disposons pas suffisamment d'informations sur ce produit pour effectuer une comparaison avec notre architecture. Partant de ce constat, nos travaux se sont donc focalisés sur le développement d'une caméra intelligente dédiée à la vidéo HDR en temps réel, construite autour d'un capteur CMOS et d'un FPGA. Cette plateforme nommée HDR-ARtiSt (HDR Adaptive Real-time Smart camera) est présentée dans la suite de cette communication.

## 2 Description de la plate-forme HDR-ARtiSt

Notre architecture est conçue autour du capteur e2v 1.3 MegaPixels sur 10 bits à la cadence de 60 images/s (cf. Fig. 1). Le capteur CMOS est capable d'envoyer de façon séquentielle 3 images  $I_L$ ,  $I_M$  et  $I_H$  (Low, Medium et High) de la même scène avec 3 temps d'exposition différents. Une mémoire DDR3 SDRAM de 8 bits par composante RVB des images HDR dont la dynamique est utilisée pour stocker la séquence d'image ligne par ligne ( $\Lambda_1 \dots \Lambda_\lambda$ ) et la restituer à la cadence du capteur. Avec chaque image, le capteur nous fournit un histogramme codé sur 64 catégories. Ces histogrammes sont utilisés par un algorithme d'auto exposition capable d'évaluer de manière optimale les 3

temps d'exposition.

### 2.1 Auto exposition multiple

L'auto exposition est une partie cruciale lors de la capture d'une scène à forts contrastes. En effet, si celle-ci est mal effectuée, seulement une partie de la dynamique de la scène sera capturée. Notre algorithme évalue à chaque nouvelle acquisition les prochains temps d'exposition haut et bas ( $T_L$  et  $T_H$ ) à partir des histogrammes fournis par le capteur selon le principe suivant : moins de 10% de pixels saturés blanc (respectivement noirs) doivent être présents dans l'histogramme de l'image  $I_L$  (respectivement  $I_H$ ). Le temps d'exposition de l'image  $I_M$  est automatiquement déduit par une combinaison linéaire de  $T_L$  et  $T_H$ .

### 2.2 Choix algorithmique pour le HDR

Nous avons développé *HDR-Lab*, un logiciel dédié aux tests d'algorithmes HDR sur Matlab. Ce logiciel a pour but de regrouper les techniques de génération de contenu à grande dynamique les plus répandues, dans l'optique de pouvoir quantifier la qualité des résultats, effectuer des comparaisons entre algorithmes, analyser statistiquement les images HDR, quantifier la complexité des solutions etc. Le programme propose de nombreuses fonctionnalités, autour de la mise en place du pipeline HDR complet, partant du choix des images en allant vers la visualisation du résultats. Nous ne détaillerons pas ici le logiciel.

Pour la génération de la carte de radiance, les techniques possibles ont été principalement développées pour la génération logicielle de photos HDR. Nous avons sélectionné, parmi les techniques existantes, l'algorithme de Debevec et al. [9] en raison de ses bonnes performances. Il présente l'avantage de pouvoir être adapté à notre architecture matérielle :

$$\ln R_{ij} = \frac{\sum_{p=1}^{p=3} \omega(Z_{p,ij}) [g(Z_{p,ij}) - \ln T_p]}{\sum_{p=1}^{p=3} \omega(Z_{p,ij})} \quad (1)$$

où  $R$  est la valeur de la radiance,  $p$  est l'indice d'image,  $Z_{ij}$  est la valeur du pixel à la position  $ij$ ,  $\omega$  est une fonction de pondération,  $T$  est la valeur du temps d'exposition du pixel concerné et  $g()$  est la courbe de réponse inverse du système optique qui a été au préalable évaluée et stockée sous forme de LUTs à l'intérieur du FPGA. La fonction de pondération permet de donner plus ou moins d'importance aux pixels pendant la somme des pixels, suivant si ils sont trop saturés, trop bruités, ne contenant que peu d'information etc.

En ce qui concerne l'opération finale de tone mapping, celle-ci permet d'afficher sur un écran standard (avec une dynamique atteint plusieurs dizaines de bits. La difficulté du tone mapping est donc d'obtenir une image affichable sur un écran tout en essayant de préserver l'intégralité de l'information qui est contenue dans la carte de radiance de l'image HDR. Pour cela, nous avons adapté un algorithme de tone mapping global

proposé par Reinhard et al. [10], choisi pour sa simplicité et sa restitution naturelle des contrastes d'images HDR (voir l'étude de Akyüs et al. [11]). On calcul d'abord la radiance moyenne de l'image HDR comme ceci :

$$\bar{R}_{ij} = \exp \left( \frac{1}{N} \sum_{i,j} \ln (R_{ij}) \right) \quad (2)$$

Où  $R_{ij}$  est la luminance du pixel à la position  $(i,j)$ ,  $N$  est le nombre total de pixel dans l'image. Ensuite, nous faisons correspondre la radiance moyenne de la scène avec la valeur grise de l'image affichable :

$$\begin{aligned} D_{ij} &= 255 \cdot \frac{a \frac{R_{ij}}{\bar{R}_{ij}}}{1 + a \frac{R_{ij}}{\bar{R}_{ij}}} \\ &= 255 \cdot \frac{1}{1 + \frac{R_{ij}}{a \bar{R}_{ij}}} \end{aligned} \quad (3)$$

### 2.3 Mémoire

La technique envisagée de création de vidéo HDR a besoin de quatre étapes successives :

1. Capture de  $P$  images uniques sélectionnées par l'auto exposition multiple
2. Stockage de  $P - 1$  images en mémoire
3. Calcul de la carte des radiances pendant l'acquisition de la  $P^{ieme}$  image.
4. Exécution de l'opération de tone mapping directement pendant la reception de la  $P^{ieme}$  image.

Le principal inconvénient est le débit de sortie limité. Comme illustré sur la Figure 2, avec 3 images uniques (basse, moyenne et haute expositions) capturées à 60 ips, la sortie d'une caméra HDR standard est considérablement limitée à 20 images par seconde : la première image HDR  $H_1$  est calculée à partir des cartes de radiances  $I_1, I_2$  et  $I_3$ , la deuxième  $H_2$  à partir de  $I_4, I_5$  et  $I_6$  etc. Pour contourner cette limite, nous proposons une gestion de mémoire spécifique, dans la mesure où l'on crée une nouvelle image HDR à la cadence du capteur et en temps réel. La méthode peut être apparentée à la technique de "memory swapping". Afin de suivre la cadence du capteur, il est nécessaire de capturer, stocker et restituer en temps réel trois images ( $I_L, I_M$  et  $I_H$ ) nécessaires à la production de contenu HDR dans le flux. Pour cela, la mémoire externe DDR3 SDRAM capture l'image courante à l'instant  $t$  ( $I_{L,t}$  par exemple), pendant qu'elle lit les 2 images précédemment stockées  $I_{M,t-1}$  et  $I_{H,t-1}$ . Ainsi, un système de FIFOs en sortie de la SDRAM permet de délivrer en parallèle les pixels des 2 images stockées ainsi que ceux de l'image courante, et ce à la cadence du capteur (60 images/s). Un tel système permet ainsi de générer en temps réel une nouvelle image HDR pour chaque nouvelle image acquise. De cette manière, 3 flux de pixels liés aux 3 images  $I_L, I_M$  et  $I_H$  sont distribués à l'intérieur de notre FPGA et alimentent à chaque instant le module de génération de la carte de radiance.

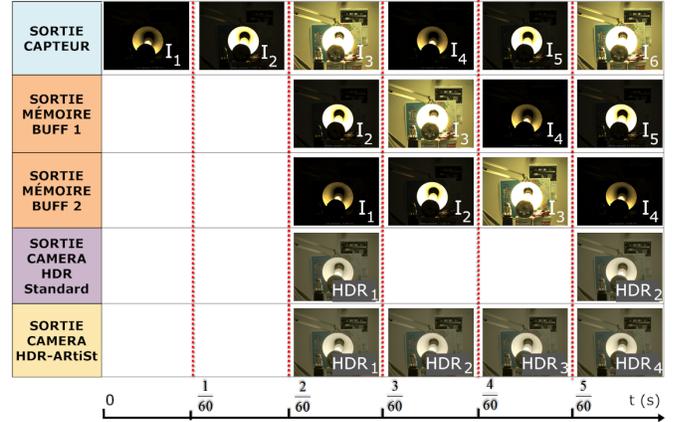


FIGURE 2 – Comparaison entre les framerates de la technique HDR standard et de notre technique. Une MMU dédiée de "memory swapping" est mise en place afin de délivrer à chaque instant plusieurs flux vidéos en parallèle.

## 3 Résultats

Un exemple de résultats d'images issues des flux obtenus est visible sur la Figure 1. Les Fig. 1(a), 1(b) et 1(c) représentent simultanément les 3 frames utilisées pour la génération HDR de la Fig. 1(d). On remarque bien l'effet de l'auto exposition multiple sur les trois images LDRs, lorsque la lampe est allumée. Quand la dynamique de la scène est faible, l'algorithme d'auto exposition a tendance à rapprocher les valeurs de temps d'exposition des images  $I_L, I_M$  et  $I_H$ . Dans le cas de la scène éclairée, nous obtenons au final une vidéo HDR présentant à la fois des détails dans les zones fortement éclairées (autour de l'ampoule) et dans les zones sombres (à l'intérieur du tube). La dynamique capturée ici est d'environ  $120dB$ ; que l'on peut comparer avec les 10 stops du capteur (soit  $60dB$ ).

Ce système a été implanté sur la plate-forme HDR-ARTiSt conçue autour d'un Virtex 6 de chez Xilinx. La description en VHDL a été réalisée sur PlanAhead. Les résultats de synthèse du Tableau 1 montrent que l'utilisation des ressources est relativement faible (17% des ressources disponibles). Les paramètres algorithmiques, tels que  $a$  ou la LUTs contenant la courbe de réponse du système ont une influence minim sur l'utilisation des ressources.

TABLE 1 – Implémentation du système HDR à trois images sur un Virtex 6 XC6VLX240T.

Metric	System	Utilization
Puissance consommée	6.039 W	
Fréquence maximum	125.0 MHz	
Slices occupés	6,692 / 37,680	17%
LUTs	16,880 / 150,720	11%
Registres	20,192 / 301,440	6%
IOBs liés	196 / 600	32%
BRAMs 36K	17 / 416	4%



FIGURE 3 – Résultat du système complet. Notre contrôle d'exposition multiple peut choisir les trois temps d'exposition adéquate, et le MMU nous permet de diffuser en parallèle les trois flux vidéos au sein du FPGA. La frame HDR correspondante à la reconstruction des trois frames courante (a), (b) et (c) est directement visible en sortie de notre caméra après une latence de traitement de 1, 2 $\mu$ s.

## 4 Conclusion

Nous avons développé une nouvelle caméra intelligente enchaînant successivement acquisition, traitement et restitution vidéo de contenu HDR. La plate-forme HDR-ARtiSt est capable de fonctionner à une cadence de 60 images/s, quelle que soit la résolution, grâce à une gestion mémoire dédiée et à l'implantation temps réel des différents algorithmes nécessaires à la reconstruction HDR.

Nos travaux se concentrent actuellement sur différentes pistes permettant d'améliorer la qualité des vidéos HDR. Premièrement, nous focalisons nos efforts sur l'implantation d'autres algorithmes de tone mapping plus sophistiqués qui permettront à court terme d'obtenir un rendu HDR de meilleure qualité. Deuxièmement, même si le flou vidéo est très peu visible du fait de la cadence du capteur (60 images/s) et de la rapidité des traitements effectués, le problème de ghost bien connu dans le domaine HDR peut toutefois apparaître si la quantité de mouvements est trop importante dans la scène. Nous envisageons donc d'implanter sur notre architecture des algorithmes (voir [12]) pour palier à ces problèmes.

## Références

- [1] H. Wang, R. Raskar, and N. Ahuja, "High dynamic range video using split aperture camera," in *IEEE 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras*, 2005.
- [2] 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.
- [3] S. W. Sing Bing Kang, Matthew Uyttendaele and R. Szeliski, "High dynamic range video," Interactive Visual Media Group, Microsoft Research, Redmond, WA, Tech. Rep., 2003.
- [4] M. Ali and S. Mann, "Comparametric image compositing : Computationally efficient high dynamic range imaging," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 913–916.
- [5] R. Ureña, P. Martínez-Cañada, J. M. Gómez-López, C. A. Morillas, and F. J. Pelayo, "Real-time tone mapping on gpu and fpga," *EURASIP J. Image and Video Processing*, vol. 2012, p. 1, 2012.
- [6] F. Hassan and J. E. Carletta, "An fpga-based architecture for a local tone-mapping operator," *J Real-Time Image Proc*, vol. 2, pp. 293–308, 2007.
- [7] M. T. Chris Kiser, Erik Reinhard and N. Tocci, "Real-time automated tone mapping system for hdr video," in *IEEE International Conference on Image Processing*, 2012.
- [8] S. Mann, R. Lo, K. Ovtcharov, S. Gu, D. Dai, C. Ngan, and T. Ai, "Realtime hdr (high dynamic range) video for eyetap wearable computers, fpga-based seeing aids, and glasseyes (eyetaps)," in *Electrical Computer Engineering (CCECE), 2012 25th IEEE Canadian Conference on*, 29 2012-may 2 2012, pp. 1–6.
- [9] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *SIGGRAPH*, 1997, pp. 369–378.
- [10] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 267–276, 2002.
- [11] A. O. Akyüz and E. Reinhard, "Perceptual evaluation of tone-reproduction operators using the cornsweet–craik–o'brien illusion," *ACM Trans. Appl. Percept.*, vol. 4, pp. 1 :1–1 :29, February 2008. [Online]. Available : <http://doi.acm.org/10.1145/1278760.1278761>
- [12] A. Srikantha and D. Sidibé, "Ghost detection and removal for high dynamic range images : Recent advances," *Signal Processing : Image Communication*, vol. 27, no. 6, pp. 650–662, 2012.