Points d’intérêt dans les vidéos HDR

Première évaluation différentielle de la quantité et de la qualité des points d’intérêt spatiaux et spatio-temporels obtenus sur des vidéos classiques et HDR

Alain Simac-Lejeune

Litii
15, rue Saint Exupéry
Alpespace
73800 Françin, France
alain.simaclejeune@litii.com

RÉSUMÉ. *La HDR (High Dynamic Range) permet de représenter des images d'une manière analogue à la représentation de notre système de vision et de manière plus complexe que dont nous disposons à l'heure actuelle sur notre ordinateur. Cet article propose d'analyser le comportement des extracteurs de points d'intérêt spatiaux et spatio-temporels lors de leur utilisation sur des vidéos HDR en les comparant avec leur utilisation sur des vidéos classique afin de déterminer l'influence des informations supplémentaires apportées par la HDR. Cette étude permet ainsi de vérifier si les détecteurs de points peuvent être utilisés directement dans leur expression habituelle ou s'il est nécessaire de les modifier pour s'adapter à cette nouvelle forme de représentation des images.*

ABSTRACT. *HDR (High Dynamic Range) is used to represent images in an analogous manner to the representation of our vision system and more complex than we have currently on our computer. This article proposes to analyze the behavior of extractors landmarks spatial and spatio-temporal in their use of HDR videos by comparing them with the use of classic videos to determine the influence of additional information provided by the HDR. This study allows to check whether the detector points can be used directly in their usual expression, or whether to change to adapt to this new form of representation of images.*

MOTS-CLÉS : *point d'intérêt, détection d'objet, vidéo, HDR, large gamme dynamique.*

KEYWORDS: *interest point, object detection, video, HDR, high dynamic range.*

Travaux effectués sous la direction de William Goubert – Litii
Thèse effectuée sous la direction de Michèle Rombaut et Patrick Lambert

JJC 2012, pp. 389–394, Bordeaux, 21-23 mars 2012

1. Introduction

Le système de vision humain et plus particulièrement l'œil humain permet de la visualisation dans une scène d'une très large gamme dynamique d'intensité. L'homme peut voir des détails très sombres ou très lumineux dans une même scène. Ce qui est impossible à obtenir avec des images numériques classiques. L'image numérique classique est codée sur 256 valeurs (entre 0 et 255) sur chaque plan rouge, vert et bleu, c'est-à-dire avec 24 bits par pixel (3 x 8 bits). L'écart d'intensité lumineuse entre le pixel le plus lumineux et le pixel le plus faible, non noir, n'est donc que de 255. Ce qui ne permet pas de représenter la même gamme dynamique que l'œil. Les images HDR utilisent plus de bits par pixel que les images classiques et permettent de stocker une dynamique largement supérieure. La technique la plus courante est de stocker les images avec un nombre flottant par couleur (96 bits par pixel). Les extracteurs de points d'intérêt spatiaux et spatio-temporels que nous avons étudiés dans (Simac-Lejeune *et al*., 2010) sont des extracteurs qui se basent sur l'intensité des pixels ainsi on peut légitimement se demander quel sera l'impact de cette nouvelle façon de stocker l'intensité et sur des plages de valeurs plus larges. Dans cette première étude, nous avons choisi d'utiliser uniquement les points d'intérêt spatiaux de Harris et spatio-temporels de Laptev. Après avoir présenté les points d'intérêt utilisés et la HDR, nous proposons une évaluation différentielle en termes de qualité et quantité de l'utilisation des extracteurs de points d'intérêt sur des données HDR par rapport à des données classiques

2. Les points d'intérêt

2.1. *Les points d'intérêt spatiaux : détecteur de Harris*

Les points d'intérêt spatiaux (Spatial Interest Points - SIP) sont définis comme des points où apparaît un changement significatif dans l'image. Par exemple, les coins, les intersections, les points isolés et les points spécifiques sur les textures sont des points d'intérêt. En pratique, ces points d'intérêt correspondent à un pixel présentant un fort rayon de courbure de l'intensité c'est à dire des variations du second ordre de celle-ci. Ils ont été proposés en 1988 par Harris (Harris *et al*., 1988) comme une extension du gradient 2D.

2.2. *Les points d'intérêt spatio-temporels : détecteur de Laptev*

Laptev and al. (Laptev *et al*., 2003) proposent une extension spatio-temporelle du détecteur de SIP de Harris (Harris *et al*. 1988) pour détecter les "Space-Time Interest Points", noté STIP dans la suite de l'article. Dans une séquence d'images,

ces points d'intérêt spatio-temporels sont des points présentant des changements significatifs dans le temps et dans l'espace. Ces points sont particulièrement intéressants car ils concentrent l'information initialement contenue dans toute l'image en quelques points spécifiques. L'intégration de la composante temporelle permet d'effectuer un filtrage sur les points d'intérêt spatiaux (SIP) et de donner plus d'importance à ceux qui présentent également une évolution temporelle non régulière. La détection des STIP est réalisée en utilisant une matrice Hessienne-Laplace $\boldsymbol{H}$ (Laptev, 2005). De manière analogue au détecteur de Harris, un filtre gaussien est appliqué dans le domaine spatial et dans le domaine temporel. Les deux paramètres σ_s et σ_t, contrôlent les échelles spatiale et temporelle pour les coins détectés.

Pour extraire les STIP, différents critères ont été proposés. Comme dans (Laptev, 2005), nous avons choisi d'utiliser l'extension spatio-temporelle de la fonction d'extraction $\boldsymbol{R}$ permettant de créer une image représentant la carte d'intérêt, appelée "*salience function*", définie par :

$$R(x,y,t) = \det\left(H(x,y,t)\right) - k \times trace(H(x,y,t))^3 \qquad [1]$$

où le paramètre k est ajusté de manière empirique à 0.04 comme pour la détection des SIP. Les STIP correspondent aux plus grandes valeurs de la fonction d'extraction $\boldsymbol{R}$ en utilisant une valeur de seuil. Une valeur typique pour ce seuil est 150.

3. La HDR et les vidéos HDR

3.1. *Définition : High Dynamic Range*

Les images HDR (Schneider, 2007) sont codées en utilisant plus de bits par pixel que les images classiques afin de permettre le stockage de dynamique plus grande. De manière générale, les images HDR sont stockées en utilisant un nombre flottant (32 bits) par couleur soit 96 bits par pixel.3.2.

3.2. *Tonemapping*

Nos écrans n'étant pas prévu pour afficher des pixels autrement qu'en 24 bits (8 bits par couleurs), il est nécessaire d'appliquer des corrections à l'image HDR si on souhaite la visualiser. Le Tonemapping est un procédé qui consiste à passer d'une image HDR à une image à faible dynamique, en gardant tous les détails que l'on souhaite : il s'agit de compresser la gamme dynamique à quelque chose de plus facile à percevoir, en attribuant à certaines couleurs une autre couleur. Pour cela, il existe une multitude d'algorithmes différents (Mantuik *et al*., 2008, Qiu *et al*., Krawczyk *et al.*, 2005, Fattal *et al*., 2002).

3.3. *Les différents modes de production de données HDR*

On distingue plusieurs méthodes pour produire des données HDR. Les **données réelles** sont obtenues via l'appareil d'acquisition qui prend simultanément plusieurs données et qui effectue la fusion directement. Les **données calculées** sont obtenues directement en appliquant un algorithme de fusion (Mertens *et al*., 2007) sur des images prises à des expositions différentes (généralement par la technique du 'braketing' – (Efros *et al*., 2009)) à l'instar des **données simulées** qui sont obtenues de manière analogue sur des images dont on a modifié l'exposition (sous ou sur exposition). L'image sous-exposée va permettre de récupérer des détails dans les ombres et les couleurs les plus sombres, tandis que l'image surexposée[1] permettra de récupérer des détails dans les zones lumineuses et les couleurs claires. Enfin, les **données tonemappées** sont obtenues en appliquant un algorithme de tonemapping sur des données HDR ce qui permet de recoder chaque pixel de son codage HDR (96 bits par exemple) vers un codage classique RGB (24 bits). Le tonemapping permet de repasser en 24 bits tout en conservant au maximum la dynamique de contraste obtenue dans la HDR.

4. Evaluation

4.1. Les données classiques et HDR

Pour effectuer notre évaluation, nous avons utilisé une base de données composées de trois types de données en noir et blanc (classique 8 bits, simulée 24 bits et tonemappé 8 bits - algorithme de (Mantiuk *et al*., 2008)). Les données réelles n'ont pas été utilisées car elles ne peuvent actuellement être produite car pas une seule caméra spécifique très onéreuse à laquelle nous n'avons pas eu accès.

Les données sont issues d'une partie des données de la base la base UCF Sports Action 50[1] initialement composée de 5000 séquences collectées sur Youtube et réparties sur 50 catégories. Chaque séquence dure 4 secondes à raison de 25 images par seconde (100 images), dans une résolution de 320 pixels par 240 pixels.

4.2. Les paramètres des algorithmes

On dispose de deux groupes de paramètres : ceux concernant les extracteurs et ceux concernant la HDR (fusion et tonemapping). Pour ce qu'il est des extracteurs, nous avons utilisé des paramètres classiques à savoir un sigma spatial de 1.5 pour l'extracteur de Harris et de Laptev, un sigma temporel de 1.5 pour l'extracteur de Laptev ainsi qu'un seuil de 150 pour les images 8 bits (0-255) pour les 2 extracteurs.

---

[1] http://server.cs.ucf.edu/~vision/data.html

Le seuil pour les images 24 bits (0-65535) a été choisi empiriquement (par règle de 3) à 38500 pour les deux extracteurs également. En ce qui concerne les algorithmes de HDR, nous avons utilisé le triplet (2,1.5,1) pour les facteurs (contraste, saturation, détail) concernant l'algorithme de Mantiuk utilisé pour l'étape de tonemapping.

4.3. Quantité et qualité des SIP et des STIP générés

Pour évaluer la quantité, nous avons appliqué les extracteurs avec les paramètres cités sur les données classiques, simulées et tonemappées. Les résultats (tableau 1) obtenus montrent que l'augmentation de la dynamique d'intensité provoque une légère augmentation du nombre de points d'intérêt spatiaux et spatio-temporels et que le tonemapping augmente sensiblement cette quantité. La HDR semble ne rien apporter de plus alors que le tonemapping génère d'avantage de points.

	Classique		HDR simulée		HDR tonemappée	
	SIP	STIP	SIP	STIP	SIP	STIP
Par séquence	8947	3741	8971	3756	9153	4013
Par image	101	42	101	42	104	45
Minimum	47	11	47	9	48	13
Maximum	165	89	460	91	172	93
Ecart-type	27	18	27	18	28	20

Tableau 1. *Nombre de SIP et de STIP extraits sur des vidéos en format classique et en format HDT simulée/tonemappée.*

Dans le cadre de cette étude, nous n'avons pas souhaité utiliser de descripteurs associés aux points d'intérêt comme SIFT (Lowe, 2003) ce qui sera effectué dans une prochaine étude. Ainsi, l'évaluation de la qualité proposée est basée sur la position des points générés. Pour cela, on propose de découper chaque image en une grille 5x5 et d'analyser la quantité de points d'intérêt dans chacune des 25 zones formées en comparant le contenu de ces zones sur les différentes formes HDR proposées. On prend comme référence la séquence classique et on calcule le pourcentage de variation. Les résultats (tableau 2) montre que la qualité des points d'intérêt spatiaux et spatio-temporels est conservée lors du passage à des données HDR ou tonemappées puisque le pourcentage de variation du nombre d'intérêt par zone est plutôt faible.

	Classique		HDR simulée		HDR tonemappée	
	SIP	STIP	SIP	STIP	SIP	STIP
Par séquence	3.74	1.26	3.72	1.27	3.74	1.26
Par image	17	15	16	15	18	16

Tableau 2. *Nombre de SIP et de STIP extraits par zone sur des vidéos en format classique et en format HDR simulée/tonemappée (minimum toujours à 0).*

5. Conclusion

Dans cet article, nous nous sommes intéressés à l'imagerie HDR et surtout aux opérateurs de Harris et de Laptev qui permettent l'extraction de points d'intérêt spatiaux et spatio-temporels. L'étude proposée permet de fournir une première évaluation quantitative et qualitative de ces extracteurs lors de leur utilisation sur des données à haute dynamique de valeurs. Ces premiers résultats nous indiquent que le changement de dynamique affecte de façon modérée la génération des points d'intérêt et laissent à penser que les applications utilisant les points d'intérêt sont applicables sur des données classiques comme sur des données HDR. Cependant, ces résultats sont à relativiser. Le paramétrage de la génération des points d'intérêt notamment celui du seuil de détection est à revoir. Les données utilisées peuvent être très variables et il conviendrait d'effectuer des tests sur de nouvelles données notamment avec des changements d'éclairage important (en sous comme en sur exposition) comme par exemple comparer le taux de génération par rapport à l'exposition de la zone considérée. Au final, le résultat le plus intéressant est que l'utilisation d'images HDR semble ne pas produire plus d'informations que les images classiques en utilisant ce type d'extracteur mais qu'on ajoute une difficulté supplémentaire lors des réglages qui sont initialement très empirique.

12. Bibliographie

Dollar P., Rabaud V., Cottrell G., Belongie S., « Behavior recognition via sparse spatio-temporal features », *Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, vol. 0, p. 65-72, October, 2005.

Efros A., Raskar R., Seitz S., « Next billion cameras », *ACM SIGGRAPH 2009 Courses*, SIGGRAPH '09, ACM, New York, NY, USA, p. 17 :1-17 :131, 2009.

Fattal R., Lischinski D., Werman M., « Gradient domain high dynamic range compression », *ACM Trans. Graph.*, vol. 21, p. 249-256, July, 2002.

Harris C., Stephens M., « A Combined Corner and Edge Detector », *Proceedings of the 4th Alvey Vision Conference*, p. 147-151, 1988.

Krawczyk G., Myszkowski K., Seidel H.-P., « Lightness Perception in Tone Reproduction for High Dynamic Range Images », *Computer Graphics Forum*, vol. 24, p. 635-645, 2005.

Laptev I., « On Space-Time Interest Points. », *International Journal of Computer Vision*, vol.64, n° 2-3, p. 107-123, 2005.

Laptev I., Lindeberg T., « Space-time Interest Points », *IN ICCV*, p. 432-439, 2003.

Lowe D., « Distinctive image features from scale-invariant keypoints », *International Journal of Computer Vision,* 60, 2, p. 91-110, 2003.

Mantiuk R., Seidel H.-P., « Modeling a Generic Tone-mapping Operator », *Computer Graphics Forum*, vol. 27, n° 2, p. 699-708, 2008.

Mertens T., Kautz J., Reeth F. V., « Exposure Fusion », *Pacific Conference on Computer Graphics and Applications*, p. 382-390, 2007.

Qiu G., Guan J., Duan J., Chen M., « Tone mapping for HDR image using optimization. A new closed form solution », *ICRP* 2006.

Schneider V., « The High-Dynamic-Range Sensor »,volume 26, p. 13-56, 2007.

Simac-Lejeune A., Rombaut M., Lambert P., « Points d'intérêt spatio-temporels pour la détection de mouvements dans les vidéos », *MajecSTIC'2010*, 2010.