

Une carte de saillance adaptée aux images HDR

J. Petit^{1,2}, R. Brémond¹ & J.-P. Tarel¹

¹Université Paris Est, LEPSIS, UMR INRETS-LCPC, Paris, France

²Laboratoire d'InfoRmatique en Image et Système d'information, Lyon, France

Abstract

A ce jour, l'évaluation de la qualité des algorithmes de reproduction de tons (tone mapping operators, ou TMO) n'a été faite que pour des situations statiques, alors que l'essentiel des applications des images de synthèse (vidéo, télévision, cinéma, jeux vidéos) propose des images dynamiques. Nous pensons que le problème de l'évaluation dynamique des TMO peut être résolu si l'on dispose d'un modèle capable de prédire l'orientation du regard dans des images dynamiques, qui conduirait à une métrique sur les images. Le modèle proposé par Itti et al. [IDP03] correspond à ce besoin, mais fonctionne sur des images 8 bits classiques. Nous montrons qu'il n'est pas adapté aux images à haute dynamique (High Dynamic Range ou HDR), et donc à l'évaluation du tone mapping en situation dynamique. Face à cette situation, nous proposons plusieurs modèles pour calculer la saillance visuelle dans des images HDR. Finalement, nous proposons une méthode générale pour évaluer la qualité d'un TMO dynamique sur un critère d'attention visuelle, complémentaire d'autres critères, en proposant d'utiliser un pre-processing de type TMO avant le calcul de la saillance dynamique par Itti et al.

Keywords: tone mapping, images HDR, vision dynamique, qualité vidéo.

1. Introduction

Les images à haute dynamique de luminance (HDR images, pour *High Dynamic Range*) se sont imposées ces dernières années comme une manière de dépasser les limites des images 8 bits que nous manipulons sur les ordinateurs [RWPD05]. Cette évolution touche à la fois l'acquisition des images et leur restitution visuelle, ce dernier problème ayant donné lieu à de nombreuses publications sur les opérateurs de *tone mapping* (TMO). Tumblin et Rushmeier [TR93] ont introduit ces opérateurs il y a 15 ans dans la communauté des images de synthèse, et on observe depuis le début du siècle le développement d'une nouvelle problématique : l'évaluation de la qualité des TMO, du point de vue de la vision humaine. Cette question a été abordée avec des méthodes psycho-physiques, c'est-à-dire les méthodes de la psychologie expérimentale [KYL*07]. La question centrale est de comparer le rendu visuel d'une image HDR avec la même image en visualisation basse dynamique (LDR, pour *Low Dynamic Range*) comme peut l'être un écran d'ordinateur. Ledda et al. [LCTS05] utilisent un écran HDR [SHS*04] pour montrer à des observateurs humains des images HDR, tandis que Ashikhmin et Goyal [AG06] utilise une scène physique HDR.

Les métriques entre images représentent une alternative plus simple à mettre en oeuvre, par rapport aux expériences

psychophysiques, pour évaluer des différences entre images. Par exemple, Mantiuk et al. [MDMS05] ont utilisé cette approche dans le cas statique pour évaluer la qualité d'un TMO, à partir du modèle *Visual Difference Predictor* (VDP) de Daly [Dal93].

Dans cet article, nous proposons d'utiliser l'attention visuelle comme critère d'évaluation de la qualité des TMO dynamiques, ce qui est complémentaire des critères classiques. Nous proposons plusieurs implémentations de cette approche, à partir du calcul de la saillance visuelle dynamique proposé par Itti *et al.* pour des vidéos [IDP03], mais en incluant un *preprocessing* qui peut être lui-même un TMO. Ce *preprocessing* a pour but de pallier aux limites de l'algorithme de Itti, validé pour les images LDR, mais dont nous montrons les insuffisances pour les images HDR.

1.1. Comment évaluer un TMO en vision dynamique ?

La vision dynamique (explorer interactivement une scène virtuelle, regarder une vidéo) a conduit à de nouvelles questions autour des images HDR. De nouveaux opérateurs de *tone mapping* ont été proposés, prenant en compte certains aspects dynamiques de la vision, notamment l'adaptation temporelle à la lumière et à l'obscurité [PTYG00, IFM05]. Parallèlement, des TMO « statiques » ont été implémentés en temps réel, par exemple celui de Reinhard *et al.* [RSSF02] sur un FPGA [HC07]. Toutefois, aucun résultat n'a été publié à ce jour concernant l'évaluation de la qualité des images produites dans des situations dynamiques, à l'ex-

ception de Ledda et al. [LSC04] qui traite une situation très simplifiée (image d'un test carré sur fond uniforme), d'une manière qui ne peut être étendue à des images «normales». Une des raisons de ce manque, que l'on peut trouver surprenant étant donné que l'essentiel des applications (télévision, cinéma, jeux vidéo, interactivité, etc.) comporte une composante dynamique importante, est que les méthodes d'évaluation des TMO utilisées pour les situations statiques ne peuvent pas être directement transposées à des situations dynamiques (vidéos ou scènes interactives).

Concernant la méthodologie utilisée, Ledda *et al.* [LCTS05], Yoshida *et al.* [YBMS05] et McNamara [McN06] comparent des images HDR et LDR avec des critères tels que la luminosité, la reproduction des contrastes et des détails, d'une manière qui suppose une inspection fine des images, tandis que Grave et Brémond [GB08] enregistrent une performance visuelle (visibilité), qu'il n'est pas possible de mesurer de la même manière en dynamique. D'autres approches, plus subjectives, demandent à des observateurs d'évaluer les images en termes de «réalisme» ou de «naturalness» [DMAC03, KYL*07]. Ce type de jugement pourrait éventuellement s'appliquer à des vidéos, mais ils portent autant sur le TMO que sur la modélisation des scènes.

L'argument principal qui nous conduit à nous éloigner des méthodes statiques dans l'évaluation de la vision dynamique des images est toutefois d'un autre ordre : les notions de visibilité des détails, ou d'apparence visuelle, nous semblent moins importantes pour de nombreuses applications que les notions dynamiques de capture du regard et de saillance visuelle. Les métriques statiques sont toutes fondées sur l'hypothèse de vision fovéale, c'est-à-dire que la distance entre deux images, en un pixel donné, est évaluée en supposant qu'un observateur regarde attentivement ce pixel, ce qui ne paraît pas pertinent en vision dynamique. Notre parti pris consiste au contraire à privilégier la dimension dynamique de l'observation d'une scène (interactive ou séquence vidéo) et de choisir des critères de qualité à partir de cette aspect dynamique. Concrètement, cette approche nous a amenés à nous intéresser aux modèles d'attention visuelle, qui cherchent à prédire l'orientation du regard d'un observateur face à une image. Ce choix d'un critère, en l'occurrence l'attention visuelle, n'est pas anodin, car les différents types de modèles de vision ne sont pas équivalents [SF96]. Choisir un critère, c'est donc choisir à partir de quoi on jugera de la qualité d'un algorithme dans des applications dynamiques. En outre, s'intéresser spécifiquement à la situation dynamique rend non pertinentes la plupart des métriques sur les images, d'une part à cause de l'hypothèse de vision centrale, mais aussi parce que des différences qui sont nettement visibles entre deux images que l'on observe attentivement pendant plusieurs secondes peuvent passer complètement inaperçues dans un flux vidéo à 60 Hz.

1.2. L'attention visuelle, un critère pour juger les aspects dynamiques des images

Par rapport à une expérimentation psychophysique HDR en dynamique, l'utilisation d'une métrique a les avantages de la simplicité de mise en oeuvre, et surtout de la généralité (on peut l'utiliser sur n'importe quelle séquence vidéo HDR). Ceci nous amène aux modèles de vision capables de prédire, dans un flux vidéo, le focus de l'attention, autrement dit l'orientation du regard. La littérature sur l'attention visuelle s'est fortement développée depuis 1998 et la publication par Itti *et al.* d'un premier modèle de «carte de saillance» [IKN98]. En particulier, une version dynamique de cet algorithme a été proposée en 2003 [IDP03]. Le principe de cet algorithme est de calculer, à partir d'un ensemble de caractéristiques élémentaires (intensité, orientation, couleur, *flicker*, mouvement) à plusieurs échelles spatiales, une unique carte de saillance associée à l'image. Cette *saliency map* est interprétée, en l'absence de biais (Cf. section 2.2), comme prédictive du comportement oculaire, c'est-à-dire des orientations du regard. Malheureusement, nous montrons que cet algorithme donne des résultats décevants avec des images HDR, ce qui nous conduit à proposer une alternative qui produit une carte de saillance plus plausible pour des images HDR, tout en restant cohérente avec [IDP03] dans le cas des images LDR.

2. Etat de l'art

2.1. Métriques sur les images et les vidéos

Si l'on souhaite éviter la complexité que représentent les expérimentations psycho-physiques, l'alternative consiste à utiliser une métrique vidéo basée sur un modèle de vision, c'est-à-dire sur des connaissances portant sur le fonctionnement du système visuel humain. Toutefois, les métriques entre images [EB98] et entre vidéos [Win05] ont été développées dans le but d'évaluer la qualité d'algorithmes de compression, ce qui limite leur utilisation aux situations dans lesquelles l'image de référence et l'image transformée sont difficiles à distinguer. D'autre part, les critères de qualité de ces métriques, s'ils sont variés (objectifs comme dans le VDP [Dal93], subjectifs comme dans CIECAM02 [MFH*02]), font l'hypothèse que l'on est en vision centrale : ils évaluent une distance entre les valeurs d'un pixel dans deux images, en supposant que ce pixel est dans l'axe du regard. Or pour juger de la qualité d'images dynamiques, nous avons besoin d'une métrique capable de prédire des caractéristiques dynamiques, la première d'entre elles étant l'orientation du regard dans l'image. Ces critères ne sont pas contradictoires, mais il convient, pour une application donnée, de les hiérarchiser.

Certains auteurs ont utilisé des métriques entre images comme alternative à la psychophysique pour évaluer la qualité des TMO dans des situations statiques. Mantiuk *et al.* et McNamara [MDMS05, McN06] utilisent le VDP et la

« similarité structurelle » [WBSS04], tandis que Smith *et al.* [SKMS06] proposent leur propre métrique à partir des *transducers* de Mantiuk *et al.* [MMS06]. Notre objectif est d'identifier une telle métrique, adaptée à une situation de vision dynamique.

2.2. Attention visuelle

Les recherches actuelles en neurosciences cognitives permettent de décrire les processus attentionnels selon trois composantes [Knu07] :

1. La sélection ascendante (*bottom-up*) de caractéristiques saillantes parmi les données sensorielles (pour le canal visuel, la saillance visuelle).
2. Les biais descendants (*top-down*) qui dépendent de nombreux facteurs, comme la tâche en cours, l'état émotionnel, les connaissances *a priori*, etc.
3. La compétition entre les différentes caractéristiques pour accéder aux ressources limitées, comme la mémoire à court terme, et la sélection de stimuli pertinents.

Parmi ces facteurs, celui qui permet de juger de la qualité des images dynamiques est la saillance visuelle [IK01], dans la mesure où c'est la partie du système attentionnel qui dépend des données, par opposition avec les biais descendants (*top-down*) qui dépendent de l'observateur et de la tâche en cours. Le jugement que l'on porte sur la qualité d'une vidéo doit être lié aux propriétés intrinsèques des données.

Parmi les théories de l'attention visuelle, la *Feature Integration Theory* (FIT) [TG80] s'est imposée grâce aux travaux de Itti [IKN98] qui a proposé un modèle efficace pour calculer la saillance visuelle à partir d'une image. Ce modèle est fondé sur des analogies avec le fonctionnement physiologique du système visuel (sélection ascendante et compétition entre les caractéristiques). Le modèle initial de 1998 [IKN98], ainsi que des variantes ultérieures [IK00, LLBT06], ont pour but de prédire la composante ascendante de l'attention visuelle, qui est la contribution de l'image d'entrée du système visuel à la sélection de l'orientation du regard. Le modèle utilise l'hypothèse d'une unique carte de saillance pour l'ensemble du processus attentionnel [KU85].

Ce modèle standard a été très fructueux, sous sa forme statique, dans des applications de synthèse d'images (e.g. [Red01, CCM03, PO03, LVJ05]). Toutefois ce modèle ne considère que des caractéristiques statiques dans les images : luminance, couleur et orientation. Yee *et al.* [YPG01] utilisent une extension de ce modèle qui inclut une caractéristique de mouvement, et par la suite Itti lui-même a proposé un modèle dynamique [IDP03], permettant de calculer la saillance visuelle dans une scène dynamique en ajoutant deux caractéristiques aux trois précédentes : le mouvement (estimé à partir des détecteurs d'orientation dans des images successives) et le *flicker* (variation temporelle de l'intensité). Ce modèle a été validé par comparaison avec des fixations

oculaires sur des vidéos projetées à l'écran [Itt04a, LLBT06], et il ressort que les caractéristiques dynamiques de la vision ont un impact plus important que les caractéristiques statiques sur les fixations oculaires.

2.3. Calcul de la carte de saillance dynamique

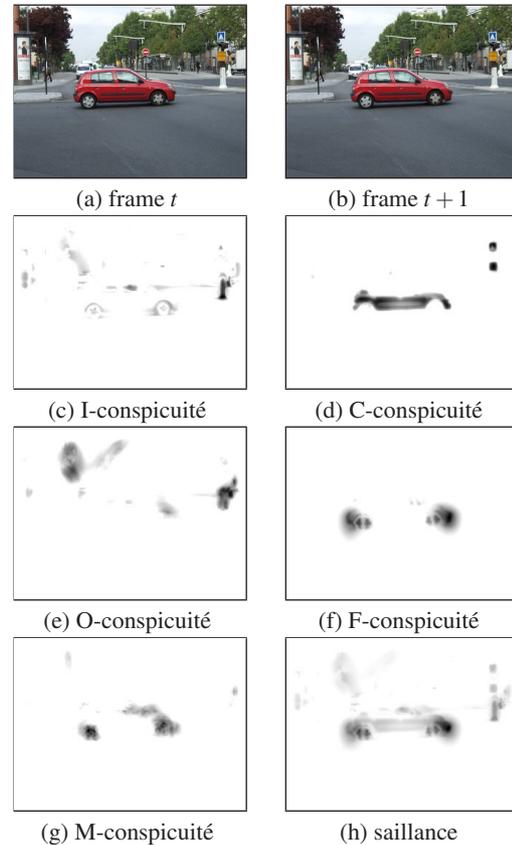


Figure 1: (c-g) Cartes de conspicuité calculées par [IDP03] à partir des frames (a-b). (h) carte de saillance calculée à partir de (c-g). Les régions saillantes apparaissent en noir.

L'algorithme dynamique de Itti *et al.* [IDP03] extrait donc cinq caractéristiques visuelles élémentaires d'une image, à plusieurs échelles spatiales. Ce calcul est suivi par des opérateurs centre-périphérie, d'inspiration biologique (implémentés par des différences de gaussiennes, DoG), puis par une étape de normalisation entre les différentes échelles. Finalement, une nouvelle normalisation conduit à une carte dite de « conspicuité » pour chaque caractéristique. L'étape de normalisation est reprise de Itti et Koch [IK00] : une carte de conspicuité est convoluée par un filtre DoG, la carte est ajoutée au résultat, les valeurs négatives sont mises à zéro, et un terme d'inhibition est ajouté (cette normalisation améliore la méthode antérieure proposée dans [IKN98]). Finalement, les cinq cartes de conspicuité (Intensité, Orientation, Couleur,

Flicker et Mouvement : I, O, C, F & M) sont ajoutées pour produire la carte de saillance. La Fig. 1 donne un exemple, avec les 5 cartes de conspécuité et la carte de saillance, sur une paire d'images consécutives dans une séquence vidéo.

Le maximum de la carte de saillance est interprété comme le focus de l'attention en l'absence de biais descendants. Dans le cas général, il est plus prudent de considérer cette carte comme une distribution d'entrée pour les processus descendants, mais dans la suite de cet article, nous ne considérons pas les biais descendants (qui ne dépendent pas de l'image elle-même, mais de la tâche, des états internes de l'observateurs, etc.) et nous considérons la carte de saillance comme un prédicteur de l'attention visuelle.

3. Limites du modèle de Itti pour les images HDR

Dans cette section, nous utilisons l'implémentation de l'algorithme dynamique de Itti *et al.* [IDP03] disponible sous Matlab [Itt04b], et nous mettons en évidence ses limites dans le cas HDR. Ces limites sont visibles aussi bien dans le cas statique que dans le cas dynamique, ce qui nous conduit, dans cette section et dans la suivante, à choisir des exemples statiques pour faciliter la lecture : les défauts de l'algorithme y sont déjà visibles.

Pour mettre en évidence la différence entre les cartes de saillance calculées à partir d'images LDR et HDR, nous utilisons dans un premier temps une version LDR (8 bits) d'images HDR (32 bits) provenant du site de P. Debevec [Deb] au *Graphic Lab* de l'USC : *Grace New* et *Memorial Church*. Les images HDR sont d'abord traitées par le TMO de Reinhard *et al.* [RSSF02]. La Fig. 2 montre des cartes de saillance calculées sur la version LDR de ces images. On observe la finalité de ces cartes de saillance : identifier les régions de l'image qui font sens, pour les niveaux les plus élémentaires du système visuel.

Lorsque l'on calcule la carte de saillance d'une image HDR avec le même algorithme, le contenu est fortement appauvri (Cf. Fig. 3). Pour l'image *GraceNew*, par exemple, l'information des zones autres que les fenêtres et les lampes, qui sont les régions les plus lumineuses, est pratiquement perdue. Or l'interprétation de cette carte de saillance concerne le comportement visuel : on devrait avoir une prédiction compatible avec l'orientation du regard d'un observateur. Il est douteux que seules la lampe au plafond et les fenêtres attirent l'attention. Pour *Memorial Church*, la perte d'information est encore plus radicale, puisqu'il ne reste qu'une fenêtre !

Sur les images HDR, l'algorithme de [IDP03] tend à sélectionner un unique point saillant, et à faire disparaître les autres, ou au moins à les gommer fortement (comparer les versions HDR et LDR des cartes de saillance Fig. 2 et 3). Les cartes de saillance calculées à partir d'images HDR se focalisent ainsi sur un point lumineux, au détriment d'autres régions qui peuvent pourtant contenir des objets saillants.

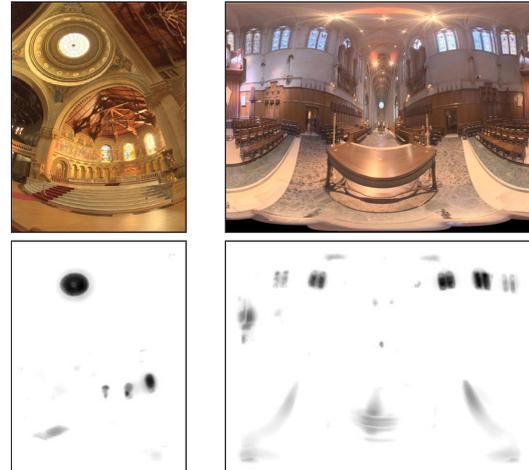


Figure 2: Images Memorial et Grace New : images LDR et cartes de saillance associées, calculée selon [IDP03].

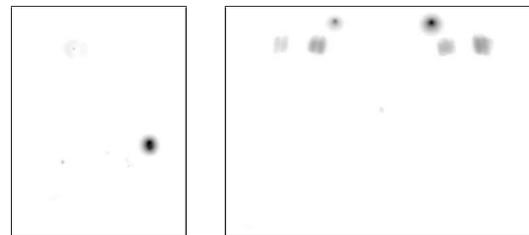


Figure 3: Carte de saillance des images HDR Memorial Church et Grace New, calculées avec [IDP03]. Les images HDR ne sont pas affichables directement.

Elles sont nettement moins prédictives de l'attention que celles calculées à partir d'images LDR. Autrement dit, l'algorithme d'Itti n'est pas adapté aux images HDR, *a fortiori* à l'évaluation des TMO dynamiques.

4. Cartes de saillance HDR

On propose dans cette section des méthodes permettant de calculer la saillance dans une image HDR, tout en restant valables pour les images LDR.

Le principe de notre approche consiste à faire précéder le calcul de [IDP03] par une étape de *preprocessing* (Cf. Fig. 4), qui peut être un TMO (Cf. section 4.2). Cette approche nous permet de garder partiellement le bénéfice de l'algorithme de Itti, qui a été validé sur des images et des vidéos LDR [IK00, Itt04a, LLBT06].

4.1. Analogies biologiques

Une première approche consiste à poursuivre l'analogie biologique que fait Itti, en assimilant les différentes étapes

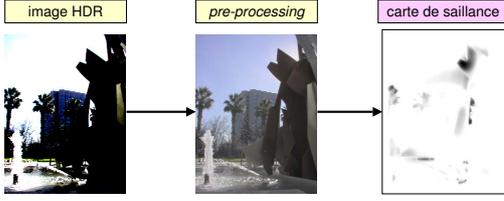


Figure 4: La carte de saillance des images HDR est calculée par [IDP03] après une première transformation (pré-processing).

du traitement aux calculs qui sont « câblés » dans la rétine et dans les niveaux ultérieurs de traitement dans les voies visuelles [Wan95].

4.1.1. Luminance ou luminosité ?

Les différents algorithmes de calcul de la saillance visuelle assimilent les canaux RGB d'une image à une « intensité » lumineuse. Si on considère, ce qui est vrai dans de nombreux domaines, que le système visuel est plus sensible à la luminosité qu'à la luminance, on est conduit à alimenter l'algorithme de calcul, non pas avec l'image d'intensité (assimilée à la luminance, en cd/m^2), mais avec l'image de luminosité, qui peut s'exprimer à un facteur près comme le log de la luminance [SS63].

Nous proposons donc une première approche, d'inspiration biologique, qui modifie le calcul de la carte de saillance des images HDR au moyen d'un pré-traitement : le calcul de l'image log (assimilée à la luminosité, ou *brightness*), qui est ensuite traitée par [IDP03]. On calcule ainsi :

$$\begin{cases} R' = \log_{10}(1 + R) \\ G' = \log_{10}(1 + G) \\ B' = \log_{10}(1 + B) \end{cases} \quad (1)$$

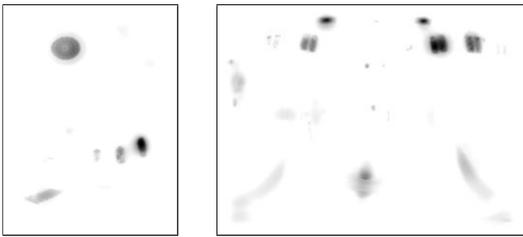


Figure 5: Cartes de saillance d'images HDR, calculées à partir de l'image des luminosités (équation 1).

Ce pré-traitement comprime la dynamique de l'image d'entrée, sans imposer une dynamique pré-définie à l'image de sortie comme les TMO usuels. S'il n'a pas des conséquences très marquées pour des images 8 bits (Cf. Section

4.3), il modifie de manière importante la carte de saillance des images HDR. La Fig. 5 donne des exemples de résultats (à comparer avec la Fig. 3).

4.1.2. Equation de Naka-Rushton

Une deuxième approche, toujours d'inspiration neuro-biologique, consiste à s'inspirer, comme certains auteurs [PTYG00, RSSF02, RD05] de l'équation de Naka-Rushton [NR66] qui décrit la réponse d'un photo-récepteur en fonction de l'intensité du signal d'entrée. On considère que le système visuel traite le signal d'entrée avec une courbe de réponse du type :

$$S = \frac{I^n}{I^n + \sigma^n} \quad (2)$$

où S est le signal de sortie d'un photo-récepteur (compris entre 0 et 1), et I l'intensité au sens de [IKN98] : $I = R + V + B$. σ est habituellement interprété comme la luminance d'adaptation, ce qui permet d'écrire $\sigma = \langle I \rangle$, la luminance moyenne de l'image. n est un facteur qui vaut 0,73 (Cf. [PTYG00]). Nous adaptions cette équation pour un signal sur 3 canaux :

$$\begin{cases} R' = \frac{R^n}{I^n + \sigma^n} \\ V' = \frac{V^n}{I^n + \sigma^n} \\ B' = \frac{B^n}{I^n + \sigma^n} \end{cases} \quad (3)$$

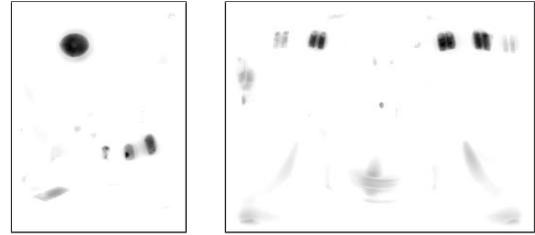


Figure 6: Cartes de saillance d'images HDR, calculées avec [IDP03], après preprocessing (équation 3, Naka-Rushton).

La Fig. 6 donne des exemples de résultats, à comparer avec la Fig. 5. Les résultats sont cohérents, mais les zones sélectionnées comme saillantes sont un peu différentes. Cet algorithme est un TMO, dans le sens où il force la dynamique de toute image dans [0,1].

4.2. Pré-processing par un TMO

On distingue deux caractéristiques des images HDR, qui ne sont pas nécessairement liées :

1. la dynamique de luminance, que l'on exprime généralement par le rapport entre les valeurs max et min de l'intensité (par exemple, dans les caractéristiques d'un écran de visualisation) ;

- le nombre de bits de codage des images (e.g. 8 bits pour les images usuelles, 12 bits pour les images raw, 32 bits pour les formats HDR, etc.).

Les opérateurs que nous avons proposés jusqu'ici modifient la dynamique des images au premier sens (I_{max}/I_{min}), mais pas au second (les calculs sont faits en valeurs flottantes).

Cette remarque nous conduit à une autre approche, plus informatique. Les *preprocessing* que nous avons appliqué jusqu'ici (Cf. Fig. 4) réduisent la dynamique de l'image, mais sans faire de quantification. Des TMO plus classiques, qui compriment la dynamique dans une plage fixe et font également une quantification, peuvent aboutir à des résultats comparables, concernant le critère qui nous intéresse ici : ne pas perdre d'information pertinente dans la carte de saillance. La Fig. 7 présente quelques résultats à partir de ce principe, en utilisant à titre d'exemple l'opérateur de Irawan *et al.* [IFM05], qui dans sa version statique est assez proche de celui de Ward *et al.* [WRP97] ; d'autres opérateurs donneraient des résultats comparables, bien que distincts, concernant l'aspect de la carte de saillance.

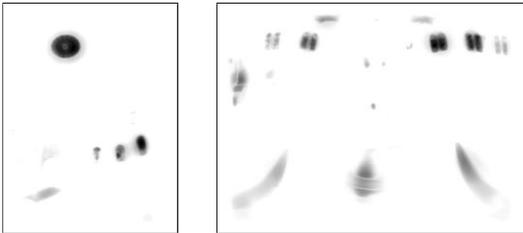


Figure 7: Cartes de saillance calculées avec [IDP03] à partir d'images HDR, après tone mapping par [IFM05].

4.3. Cohérence entre HDR et LDR

Nous avons proposé plusieurs méthodes de calcul de la carte de saillance d'images HDR. Il est nécessaire de regarder leur effet sur des images LDR : idéalement, on devrait retrouver la carte de saillance d'origine. Autrement dit, la transformation (pré-traitement) devrait être proche de l'identité lorsque la dynamique devient LDR. La Fig. 8 montre des cartes de saillance calculées à partir d'une image LDR (Lena) avec les méthodes proposées : $\log(1+I)$, Naka-Rushton et Irawan *et al.* [IFM05]. Les résultats après pré-traitement (c-e) sont en effet assez proches de l'identité (b).

On peut ainsi examiner, à partir des opérateurs d'inspiration biologique proposés en section 4.1, l'effet prévisible du pré-traitement sur le calcul de la carte de saillance pour une image LDR. Qualitativement, on dira que les fonctions utilisées sont d'autant plus linéaires que la dynamique de l'image est faible. Cette propriété est évidente pour la fonction de Naka-Rushton, puisque le point d'inversion de la courbure correspond à la moyenne de l'image. Dans le cas de l'Eq. 1,

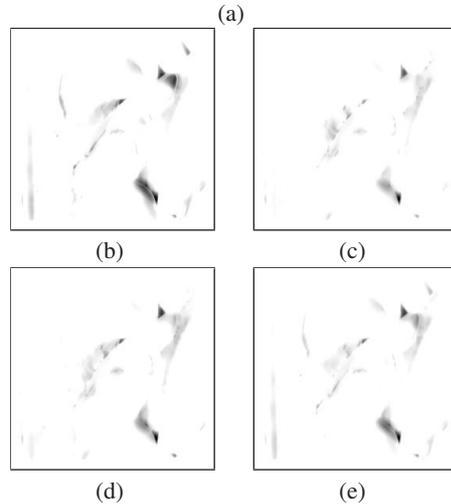
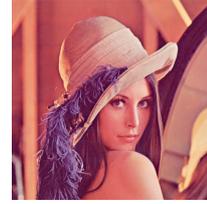


Figure 8: Carte de saillance calculée à partir d'une image LDR (Lena, a), avec [IDP03] précédé par quatre pré-traitements : (b) l'identité, (c) $\log(1+I)$, (d) Naka-Rushton, et (e) un TMO [IFM05].

il est également manifeste que plus l'histogramme est dispersé, moins on peut approximer la fonction log de manière linéaire. Or l'algorithme de Itti *et al.* [IDP03] est invariant à une multiplication de l'image par une constante, ce qui explique que les cartes de saillance des images LDR ne sont que faiblement modifiées par le pré-traitement.

5. Résultats

5.1. Cartes de saillance HDR dynamiques

On donne ici un exemple de cartes de saillance calculées sur une vidéo HDR mise en ligne sur le site du Max Planck Institute [MPI]. La Fig. 9 montre le résultat du calcul avec [IDP03] après pré-traitement par l'équation 3, qui semble cohérent avec le contenu de la séquence.

5.2. Evaluation de TMO en dynamique

Nous pouvons maintenant revenir à notre question de départ : l'évaluation de la qualité des TMO sur un critère de vision dynamique. Ayant proposé un modèle de calcul de la saillance visuelle prédictif de l'attention visuelle aussi bien en HDR et en LDR, nous avons du même coup un outil

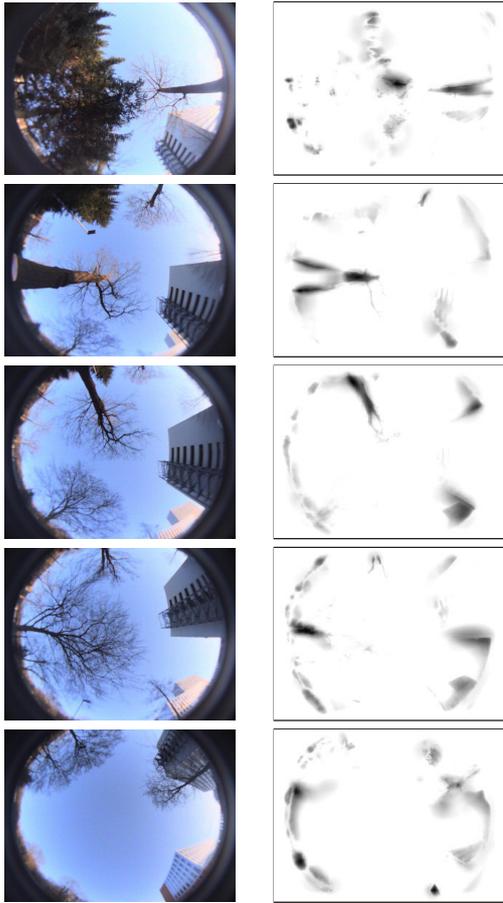


Figure 9: Exemple de cartes de saillance à partir d'une vidéo HDR calculées par un pré-traitement (Eq. 3) puis par [IDP03].

pour comparer l'effet d'un TMO particulier sur les cartes de saillances, selon le schéma de la Fig. 10.

On peut donc sélectionner, du point de vue de la vision dynamique, entre plusieurs TMO, celui qui modifie le moins possible ces cartes de saillances. A titre d'exemple, nous utilisons comme preprocessing notre opérateur basé sur l'équation de Naka-Rushton (Cf. Eq. 3), à partir de la vidéo du MPI ; on pourrait multiplier les exemples. On peut comparer, Fig. 11, la différence entre les cartes de saillance calculées avec et sans *tone mapping*. On utilise à titre d'exemple, le TMO dynamique de Pattanaik *et al.* [PTYG00].

Il peut sembler paradoxal d'introduire, dans le calcul de la carte de saillance HDR, un pré-traitement qui a certaines caractéristiques des TMO (Cf Eq. 1 et 3) ou qui est un TMO (section 4.2), et de se servir de cette carte pour évaluer un TMO. Nous pensons au contraire que le fait que les *pre-processing* proposés compriment la dynamique des images peut justement être utilisée pour sélectionner des opérateurs

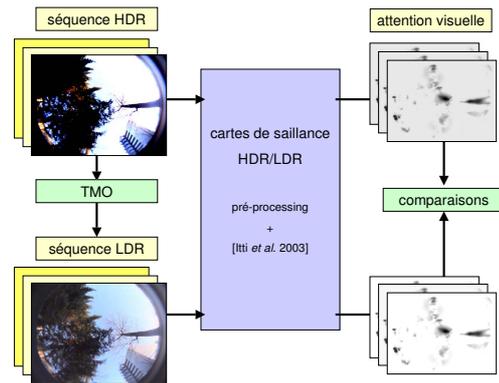


Figure 10: Principe de comparaison d'une séquence d'images HDR et de la séquence tone-mappée, en utilisant un modèle de calcul de la saillance proposé dans cet article.

de TMO de bonne qualité sur le plan de la vision dynamique. En effet, les pré-traitements qui permettront de prédire correctement l'attention visuelle nous donneront non seulement un modèle calculatoire de la saillance visuelle pour des images LDR et HDR, mais aussi des informations importantes sur les types de TMO qui sont les plus compatibles avec la perception humaine du point de vue de la vision dynamique. Nous espérons que cela nous permettra à l'avenir de concevoir des TMO plus fidèles au comportement visuel humain sur des vidéos, aussi bien LDR que HDR.

6. Conclusion

Dans cet article, nous nous sommes intéressés à l'évaluation de la qualité des opérateurs de *tone mapping* sur un critère de vision dynamique, qui nous paraît complémentaire des critères statiques pour un grand nombre d'applications des images de synthèse. Pour cela, nous nous sommes tournés vers les modèles de calcul de la saillance visuelle, et nous avons mis en évidence les défauts des modèles existants dans le cas des images à haute dynamique (HDR) : focalisation sur un seul point saillant, perte d'information dans la carte de saillance.

Partant de ce constat, nous avons proposé une méthode pour calculer la saillance visuelle sur des images HDR. Nous utilisons un modèle de calcul reconnu et validé pour des séquences dynamiques en LDR [IDP03], que nous faisons précéder d'un pré-traitement, pour lequel nous proposons différentes solutions, que l'on peut assimiler à différents degrés à des TMO. Les exemples présentés montrent que qualitativement, les cartes de saillance calculées ainsi n'ont plus les défauts pointés précédemment.

La validité de la méthode proposée pour évaluer les TMO dépend de la validité du modèle de calcul de saillance HDR. Autant celui de [IDP03] a été validé sur des images LDR,

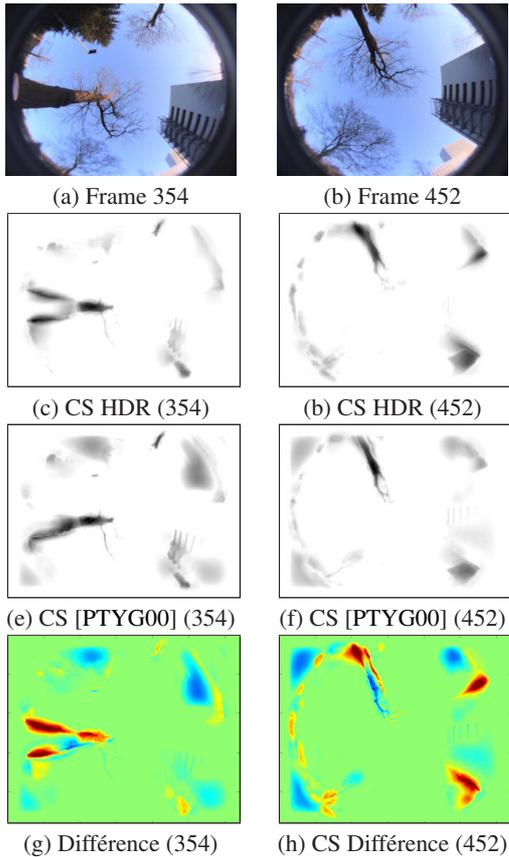


Figure 11: Comparaison des cartes de saillance (CS) dynamiques calculées à partir d'une vidéo HDR (frames 354 et 452), sans tone mapping (CS HDR) et avec tone mapping (CS [PTYG00]). La différence entre ces deux types de CS apparaît en fausses couleurs (g-h), avec les valeurs négatives en bleu et les valeurs positives en jaune-rouge.

autant il n'existe pas à notre connaissance de données oculométriques sur des images HDR qui permettraient de valider l'une ou l'autre des variantes que nous proposons. Une telle validation est pourtant nécessaire, pour faire confiance au caractère prédictif de l'attention visuelle telle qu'elle est définie dans la carte de saillance visuelle. Si les différentes approches que nous proposons pour calculer des cartes de saillance HDR sont toutes « plausibles », elles ne sont pas équivalentes (Cf. Fig. 8). Il y a donc besoin d'une méthode objective de comparaison des résultats, ce qui doit se faire selon la méthodologie classiquement utilisée pour valider les cartes de saillance : l'oculométrie.

C'est pourquoi une expérimentation psycho-visuelle est en cours de réalisation sur une scène HDR, au cours de laquelle nous enregistrons les fixations de différents observateurs à l'aide d'un *eye tracker*. Parallèlement à cette expérimentation, nous avons réalisé des prises de vue pour obtenir

l'image HDR correspondante à partir de plusieurs temps de pose (*bracketing*) [DM97]. Finalement, nous allons pouvoir comparer les prédictions de différents modèles de calcul de la saillance appliqués aux images HDR, par rapport aux fixations oculaires enregistrées par le système de suivi du regard.

Bien entendu, un TMO peut être performant du point de vue de la vision dynamique (mesuré avec un critère d'attention visuelle), cela ne garantit pas qu'il sera bon pour d'autres critères, puisque les modèles de vision correspondants ne sont *a priori* pas corrélés. Il est donc utile de compléter l'évaluation dynamique des TMO par des critères statiques, ce qui donnerait des indicateurs de qualité complémentaires (e.g. apparence visuelle, qualité subjective du rendu ou performance visuelle).

References

- [AG06] ASHIKHMINE M., GOYAL J. : A reality check for tone mapping operators. *ACM Transactions on Applied Perception* 3, 4 (2006), 399–411.
- [CCM03] CHALMERS A., CATER K., MAFLIOLI D. : Visual attention models for producing high fidelity graphics efficiently. In *Proceedings of SCCG* (2003), ACM, pp. 39–45.
- [Dal93] DALY S. : *The visible differences predictor : an algorithm for the assessment of image fidelity*. A. B. Watson Ed., Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993, pp. 179–206.
- [Deb] DEBEVEC P. : <http://gl.ict.usc.edu/data/highresprobes/>.
- [DM97] DEBEVEC P., MALIK J. : Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH* (1997), pp. 369–378.
- [DMAC03] DRAGO F., MYRSZKOWSKI K., ANNEN T., CHIBA N. : Adaptive logarithmic mapping for displaying high contrast scenes. In *Proceedings of Eurographics* (2003).
- [EB98] ECKERT M. P., BRADLEY A. P. : Perceptual quality metrics applied to still image compression. *Signal Processing* 70 (1998), 177–200.
- [GB08] GRAVE J., BRÉMOND R. : A tone mapping algorithm for road visibility experiments. *ACM Transactions on Applied Perception* 5, 2 (2008), art. 12.
- [HC07] HASSAN F., CARLETTA J. E. : A real time fpga-based architecture for a reinhard-like tone mapping operator. In *Proceedings of Graphics Hardware* (2007), pp. 65–72.
- [IDP03] ITTI L., DHAVALA N., PIGHIN F. : Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proc. SPIE 48th Annual International Symposium on Optical Science and Technology* (2003), vol. 5200, SPIE Press, pp. 64–78.

- [IFM05] IRAWAN P., FERWERDA J. A., MARSCHNER S. R. : Perceptually based tone mapping of high dynamic range image streams. In *Proceedings of Eurographics Symposium on Rendering* (2005), pp. 231–242.
- [IK00] ITTI L., KOCH C. : A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research* 40, 10-12 (2000), 1489–1506.
- [IK01] ITTI L., KOCH C. : Computational modeling of visual attention. *Nature Reviews Neuroscience* 2, 3 (2001), 194–203.
- [IKN98] ITTI L., KOCH C., NIEBUR E. : A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (1998), 1254–1259.
- [Itt04a] ITTI L. : Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing* 13, 10 (Oct 2004), 1304–1318.
- [Itt04b] ITTI L. : <http://ilab.usc.edu/toolkit/>, 2004.
- [Knu07] KNUDSEN E. I. : Fundamental components of attention. *Annual Review Neuroscience* 30 (2007), 57–78.
- [KU85] KOCH C., ULLMAN S. : Shifts in selective visual attention : towards the underlying neural circuitry. *Human Neurobiology* 4 (1985), 219–227.
- [KYL*07] KUANG J., YAMAGUCHI H., LIU C., JOHNSON G. M., FAIRCHILD M. D. : Evaluating HDR rendering algorithms. *ACM Transactions on Applied Perception* 3, 3 (2007), 286–308.
- [LCTS05] LEDDA P., CHALMERS A., TROSCIANKO T., SEETZEN H. : Evaluation of tone mapping operators using a high dynamic range display. In *Proceedings of ACM SIGGRAPH* (2005), ACM Press, pp. 640–648.
- [LLBT06] LEMEUR O., LECALLET P., BARBA D., THOREAU D. : A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 5 (2006), 802–817.
- [LSC04] LEDDA P., SANTOS L. P., CHALMERS A. : A local model of eye adaptation for high dynamic range images. In *Proceedings of AFRIGRAPH* (2004), ACM, pp. 151–160.
- [LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W. : Mesh saliency. *ACM Transactions on Graphics* 24, 3 (2005), 659–666.
- [McN06] MCNAMARA A. : Exploring visual and automatic measures of perceptual fidelity in real and simulated imagery. *ACM Transactions on Applied Perception* 3, 3 (2006), 217–238.
- [MDMS05] MANTIUK R., DALY S., MYSZKOWSKI K., SEIDEL H.-P. : Predicting visible differences in high dynamic range images. model and its calibration. In *Proceedings of SPIE Human Vision and Electronic Imaging X* (2005), SPIE, pp. 204–214.
- [MFH*02] MORONEY N., FAIRCHILD M. D., HUNT R. W. G., LI C. J., LUO M. R., NEWMAN T. : The CIE-CAM02 color appearance model. In *Proceedings of IS&T-SID 10th color imaging conference* (2002), pp. 23–27.
- [MMS06] MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P. : A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception* 3, 3 (2006), 286–308.
- [MPI] www.mpi-inf.mpg.de/resources/hdr/vem/maps. Max Planck Institute.
- [NR66] NAKA K. I., RUSHTON W. A. H. : S-potential from colour units in the retina of fish (cyprinidae). *Journal of Physiology* 185 (1966), 536–555.
- [PO03] PETERS C., O’SULLIVAN C. : Bottom-up visual attention for virtual human animation. In *Proceedings of CASA* (2003), IEEE Computer Society, p. 111.
- [PTYG00] PATTANAİK S. N., TUMBLIN J., YEE H., GREENBERG D. P. : Time-dependent visual adaptation for fast realistic image display. In *Proceedings of SIGGRAPH* (2000), ACM Press, pp. 47–54.
- [RD05] REINHARD E., DEVLIN K. : Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics* 11, 1 (2005).
- [Red01] REDDY M. : Perceptually optimized 3d graphics. *IEEE Computer Graphics and Applications* 22, 5 (2001), 68–75.
- [RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J. : Photographic tone reproduction for digital images. In *Proceedings of SIGGRAPH* (2002).
- [RWPD05] REINHARD E., WARD G., PATTANAİK S. N., DEBEVEC P. : *High dynamic range imaging : acquisition, display, and image-based lighting*. Morgan Kaufmann, 2005.
- [SF96] SILVERSTEIN D. A., FARRELL J. E. : The relationship between image fidelity and image quality. In *Proceedings of ICIP* (1996), IEEE, pp. 881–884.
- [SHS*04] SEETZEN H., HEIDRICH W., STUERZLINGER W., WARD G., WHITEHEAD L., TRENTACOSTE M., GHOSH A., VOROZCOVS A. : High dynamic range display systems. *ACM Transactions on Graphics* 23, 3 (2004), 760–768.
- [SKMS06] SMITH K., KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P. : Beyond tone mapping : enhanced depiction of tone mapped HDR images. In *Proceedings of Eurographics* (2006), pp. 1–12.
- [SS63] STEVENS J., STEVENS S. : Brightness function :

- effects of adaptation. *Journal of the Optical Society of America* 53 (1963), 375–385.
- [TG80] TREISMAN A. M., GELADE G. : A feature-integration theory of attention. *Cognitive Psychology* 12 (1980), 97–136.
- [TR93] TUMBLIN J., RUSHMEIER H. : Tone reproduction for realistic images. *IEEE computer Graphics and Applications* 13, 6 (1993), 42–48.
- [Wan95] WANDELL B. : *Foundations of vision*. Sinauer associates, Sunderland, MA, USA, 1995.
- [WBSS04] WANG Z., BOVIK A., SHEIK H., SIMONCELLI E. : Image quality assessment : from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13 (2004), 600–612.
- [Win05] WINKLER S. : *Digital video quality : vision models and metrics*. Wiley, NY, 2005.
- [WRP97] WARD G., RUSHMEIER H., PIATKO C. : A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics* 3, 4 (1997), 291–306.
- [YBMS05] YOSHIDA A., BLANZ V., MYSZKOWSKI K., SEIDEL H.-P. : Perceptual evaluation of tone mapping operators with real world scenes. In *Proceedings of the SPIE* (2005), vol. 5666, pp. 192–203.
- [YPG01] YEE H., PATTANAİK S. N., GREENBERG D. P. : Spatio-temporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics* 20, 1 (2001), 39–65.