

Tone reproduction and physically based spectral rendering

State of the art report, Eurographics 2002

K. Delvin, A. Chalmers, A. Wilkie, W. Purgathofer

Introduction

L'objectif des images de synthèse réalistes est de provoquer les mêmes réactions et les mêmes sensations que les scènes représentées. Nous n'en sommes pas là, et les principaux problèmes se situent aux deux bouts de la chaîne de production des images : La manière dont on représente la « lumière » dans les calculs d'image, et les procédés de restitution des images calculées sur écran. Cet *état de l'art* expose les évolutions récentes dans ces deux domaines.

Tone mapping

Le problème que le *tone mapping* tente de résoudre, c'est que la dynamique d'un écran est de $[1 ; 100]$ cd/m², alors que celle de l'œil humain est de $[10^{-6} ; 10^8]$ (mais pas simultanément). L'adaptation à l'obscurité est lente (en minutes), tandis que l'adaptation à la lumière est rapide (en secondes). La vision scotopique concerne le domaine $[10^{-6} ; 10^{-2}]$ cd/m², la vision mésopique le domaine $[10^{-2} ; 10]$ cd/m², et la vision photopique le domaine $[10 ; 10^8]$ cd/m². L'insuffisance des connaissances psychophysiques n'a permis jusqu'ici, que le développement d'algorithmes de *tone mapping* partiels, centrés sur une fonction visuelle particulière. **Cela dit, en parcourant la bibliographie, les seules références en « vision » sont le rapport de CIE 19/2.1 (1981), un livre de Hunt sur la reproduction des couleurs (1995), un article de Stevens & Stevens sur la luminosité (1960), et le « color science » de Wyszecki et Stiles (1982).**

Le *tone mapping* s'est développé, historiquement, pour la télévision et la photographie. L'idée de départ était que l'œil est plus sensible aux contrastes qu'aux niveaux absolus de luminance. Les premières tentatives ont consisté à ramener la luminance moyenne à la luminance moyenne disponible, ou la luminance maximale à la luminance maximale disponible. Ces procédés ne conservent pas la visibilité des fortes (ni des faibles) luminances.

Sur un écran CRT, la relation entre les valeurs RGB et les luminances affichées peut être modélisée par une courbe en γ . Une « correction γ » peut permettre de modifier l'exposant de cette courbe, sur un écran donné. Une relation linéaire entre RGB et L n'est pas nécessairement une « bonne » solution. **On aimerait avoir une explication psycho-physique précise de ce que signifie une « bonne » solution.**

Rendu spectral

L'apparence des objets peut être calculée, en général, à partir des lois de l'optique géométrique et d'une représentation de la lumière en 3 composantes RGB. Certains effets, cependant, demandent une description plus fine : la fluorescence, la diffraction, la dispersion et la polarisation. Néanmoins, la communauté en synthèse d'images reste sur des représentations à 3 composantes, en essayant de choisir au mieux l'espace colorimétrique. **L'idée selon laquelle une représentation de la lumière à 3 composantes permet de faire des calculs de rendu satisfaisant est fautive, même en lumière du jour, et même en s'intéressant seulement aux luminances.**

La sensibilité spectrale de l'œil humain est limitée à [380 ; 780] nm. Elle est très variable d'une personne à l'autre. L'espace standard de représentation des couleurs est XYZ, qui a été défini à partir de nombreux travaux de la CIE. **Les espaces colorimétriques postérieurs à 1932 ne sont pas mentionnés.**

La conversion XYZ→RGB est linéaire, la matrice de passage dépendant du système de restitution utilisé. **Ceci n'est vrai que pour un écran « linéaire », c'est à dire avec $\gamma=1$ pour les trois photophores RGB. La CIE recommande le modèle *Gain-Offset-Gamma*.** Réciproquement, à partir d'un triplet RGB ou XYZ, une infinité de spectres sont possibles (ce sont des métamères). Glassner et Smits proposent des méthodes pour passer de XYZ à un spectre « plausible ».

On rencontre essentiellement deux types de représentations des données spectrales : un échantillonnage (en général tous les 10 nm.) ou un vecteur défini à partir d'une base de fonction. La première approche ne permet pas de rendre compte correctement de la fluorescence. La seconde devient lourde si la base est trop importante. Il y a deux cas où la représentation spectrale est nécessaire : les effets optiques complexes, et le cas où on s'intéresse vraiment à un problème de perception. La discussion ci-dessous porte sur le premier cas.

- Le premier effet est la diffraction de la lumière dans des milieux diélectriques (le verre, en général) qui dépend de la longueur d'onde de la lumière, et forme des franges ou des arcs-en-ciel. Des modèles analytiques et des données expérimentales (selon le matériau) sont disponibles.
- Le second effet est la polarisation de la lumière, qui est « visible à l'œil nu », contrairement à un opinion répandue. Dans une onde électro-magnétique, on décrit le champs électrique \mathbf{E} avec ses composantes E_{\perp} et E_{\parallel} . la relation entre eux peut être modifiée lors de la réflexion sur une surface, d'une manière qui dépend de la nature de la surface. On exprime cela en terme de « retard de phase » qui se calcule avec les équations de Fresnel.
- Le 3^e effet est la phosphorescence, nom donné au fait qu'un rayonnement peut être réfléchi par une surface à une longueur d'onde différente de la longueur d'onde d'origine. Plutôt que par un spectre de réflexion, de tels matériaux doivent être décrits par une matrice de ré-émission. De telles matrices ne sont pas simples à mesurer.

Bibliographie sur le tone mapping

Les algorithmes de *tone mapping* de la littérature sont classés en 3 catégories, selon qu'ils sont uniformes, qu'ils varient dans l'espace ou qu'ils varient dans le temps.

Algorithmes fondamentaux

L'article de **Tumblin et Rushmeier** (1993) est détaillé. L'objectif de l'algorithme est de conserver les luminosités. Ils s'appuient sur Stevens & Stevens pour définir la relation entre luminosité et luminance. La relation entre L et B n'est normalement valable que pour un fond uniforme, mais elle est utilisée ici avec un fond complexe. **L'expression de la luminance à afficher en fonction de la luminance calculée utilise la notion de luminance d'adaptation, à la fois pour la scène d'origine et pour la scène affichée. Le calcul de ces luminances d'adaptation n'est pas abordé.** L'idée centrale de l'algorithme est d'égaliser la luminosité affichée et la luminosité calculée, ce qui permet d'exprimer la luminance affichée en fonction de la luminance calculée (et de la luminance maximum affichable). Cet algorithme est réputé bon dans les cas extrêmes.

L'objectif du modèle de **Ward** est de préserver la perception des contrastes, plutôt que la luminosité. **Il s'intéresse à la performance visuelle, alors que le précédent article est centré sur l'apparence visuelle.** Ward s'appuie sur des résultats de Blackwell qui décrivent l'évolution de la performance visuelle en fonction de la luminance d'adaptation. Il fait l'hypothèse que la luminance d'adaptation sur écran est la moitié de la luminance maximum affichable. A partir des courbes de Blackwell de la « *just noticeable difference* » (JND) en fonction de L_a , il écrit que ce JND est égal dans la scène calculée et dans la scène affichée. L'algorithme est linéaire. Son avantage est de préserver la visibilité, mais les résultats de Blackwell correspondent à un fond uniforme. **De plus, l'hypothèse permettant de fixer la luminance d'adaptation sur écran est arbitraire.**

Spencer, Shirly, Zimmerman et Greenberg (1995) ont proposé un modèle fondé sur des données physiologiques, permettant de visualiser dans l'image les effets de l'éblouissement que devrait causer une source lumineuse, lorsque le système de restitution visuelle ne permet pas d'éblouir l'observateur.

Algorithmes à variation spatiale

Les algorithmes de *tone mapping* à variations spatiales locales ont un défaut commun. Ils ont pour objectif de conserver la « qualité » de l'image dans la région fovéale, ce qui les conduit à dégrader la « qualité » de l'image dans les régions périphériques. Cet inconvénient peut devenir critique si on s'intéresse à des tâches visuelles impliquant la détection d'objets susceptibles d'apparaître dans tout le champ visuel (notamment la conduite automobile). Le seul à échapper, semble-t-il, à cette critique, est l'algorithme de Pattanaik (1998) qui se fonde sur des arguments physiologiques.

Chiu, Herf, Shirley, Swang, Wang et Zimmerman (1993) prétendent que l'œil est « plus sensible à la réflectance qu'à la luminance ». Pour avoir un meilleur rendu global, ils considèrent que l'observateur ne s'apercevra pas d'une modification de l'image pourvu que cela se passe dans les basses fréquences spatiales.

Schlick (1994) a proposé des modèles très rapides en temps de calcul, mais sans fondement psychophysique.

La démarche de **Jobson, Rahman et Woodell** (1997) s'appuie sur la « *retinex theory* ». Ils construisent une image multi-échelle, à partir de laquelle ils calculent la compression des luminances. Ils parviennent à une « *colour consistency* » et une « *lightness rendition* ». Ces notions sont obscures. On suppose que les images traitées ne sont pas exclusivement des images en luminance, puisque le rendu des couleurs fait partie des points forts de l'algorithme. Une validation de l'algorithme a été réalisée à partir de scènes réelles. Un des défauts de cet algorithme est de faire l'hypothèse systématique que la « couleur moyenne » est achromatique. Il semble qu'il s'agit d'un algorithme adapté aux appareils photo numériques.

Larson, Rushmeier et Piatko (1997) proposent de travailler directement sur l'histogramme. Ils calculent un histogramme des luminosités (avec une relation entre L et B du type Stevens & Stevens) puis prennent des moyennes dans des cercles de 1°, ce qui est censé correspondre à une « adaptation fovéale ». Cet histogramme est utilisé pour comprimer l'histogramme de l'image. L'algorithme est spatialement uniforme. Des procédures supplémentaires modélisent l'éblouissement, l'acuité visuelle et la sensibilité à la couleur.

Pattanaik, Ferwerda, Fairchild et Greenberg (1998) ont développé un modèle en deux parties : tout d'abord un modèle de vision, puis un modèle de restitution. L'image est décrite dans une représentation multi-échelle, et les auteurs cherchent à prendre en compte le niveau local d'adaptation visuelle. Les canaux chromatique et achromatique sont séparés dans la représentation de l'image. L'intérêt de ce modèle est de s'appuyer sur des données physiologiques du fonctionnement local du traitement de l'information visuelle.

Tumblin, Hodgkins et Guenter (1999) proposent 2 algorithmes prétendument basés sur le fonctionnement du système visuel humain. Le premier est limité aux images de synthèses. L'idée est de ne comprimer la dynamique de luminance que pour les sources lumineuses, et pas (ou moins) pour les surfaces réfléchissantes. Le second est interactif en fonction de la région pointée par la souris, l'algorithme étant modifié de manière à être optimum dans la « région d'intérêt ». Cet algorithme est très parlant pour mettre en évidence les problèmes soulevés par les *tone mapping* à variation spatiale.

Tumblin et Turk (1999) proposent la méthode « LCIS ». Elle reprend les idées développées par Tumblin, Hodgkins et Guenter, en ajoutant la notion d'adaptation locale. Un des objectifs est de conserver les hautes fréquences spatiales, et de compresser les basses fréquences.

Ashikhmin (2002) propose un modèle permettant de préserver les détails d'une image de grande dynamique. Il est fondé sur un modèle de perception de la luminance. Il utilise un algorithme de transposition « localement linéaire ».

Fattal, Lischinski et Ferwerda (2002) s'appuient sur un procédé courant en photographie (*zone system*). L'image est découpée en 11 zones, allant de noir à blanc, la zone médiane étant affectée à la luminance moyenne. La forme du *tone mapping* est ensuite choisie en fonction des valeurs constatées des luminances moyennes des différentes zones. Les régions entourée de forts contrastes subissent ensuite des modifications locales.

Algorithmes à variation temporelle

Les modèles d'adaptation temporelle ont pour but de tenir compte de la variation du niveau de luminance d'une image à l'autre (notamment lors d'un parcours dans une scène). **Ferwerda, Pattanaik, Shirley et Greenberg** (1996) ont développé un modèle basé sur des données psycho-physiques. Ils utilisent comme Ward un modèle linéaire dont le coefficient est déterminé par le JND, qui tient compte des caractéristiques de cônes et des bâtonnets de la rétine, selon le niveau lumineux considéré. Les modifications de l'acuité et de la perception des couleurs qui interviennent dans les domaines mésopique et scotopique sont prises en compte. En ce qui concerne l'adaptation temporelle elle-même, les auteurs proposent d'ajouter un terme « constant » à la luminance : $L_a = \lambda L_c + b$ (L_a luminance affichée, L_c luminance calculée, b terme spatialement constant). Le terme spatialement constant varie dans le temps de manière à garantir que la luminance moyenne affichée reste constante dans le temps. **La justification psychophysique de ce procédé, qui évite les situation d'adaptation temporelles plutôt qu'il n'en tient compte, n'est pas claire. D'autre part ce procédé ne s'intéresse pas à la baisse des performances visuelles qui correspond à l'adaptation visuelle, pour laquelle des données psycho-physiques existent pourtant. Enfin on peut remarquer que le terme constant rend illusoire la procédure choisie pour calculer le terme λ à partir de données psycho-physiques.**

L'objectif de **Scheel, Stamminger et Seidel** (2000) est d'adapter en temps réel le *tone mapping* lors d'un parcours dans une scène virtuelle. Il ne s'agit pas de prendre en compte les phénomènes d'adaptation temporelle, mais de recalculer à tout instant la compression des luminances affichées en fonction de la distribution des luminances qui sont effectivement vues à cet instant. Les luminances pré-calculées sont représentées dans la base de données sous la forme de textures de luminance.

Le modèle dynamique de **Pattanaik, Tumblin, Yee et Greenberg** (2000) s'appuie sur le modèle d'apparence colorée de Hunt. L'observateur est sensé déterminer, dans la scène, un blanc et un noir de référence par rapport auxquels il perçoit l'ensemble de la scène. L'adaptation temporelle est prise en compte, mais d'une manière qui n'est pas précisée dans le STAR. **D'après JP Farrugia, ce modèle est le seul, à ce jour, à intégrer la problématique de l'adaptation temporelle.**

Durand et Dorsey (2000) s'intéressent à la luminance d'adaptation, estimée selon des méthodes propres aux appareils photos. Le modèle de Ferwerda est enrichi (adaptation chromatique). L'algorithme est interactif, ce qui permet de calculer les paramètres du *tone mapping* à la volée lors du parcours d'une scène virtuelle.

On constate que la plupart des chercheurs qui se sont intéressés à l'adaptation temporelle se sont bornés, jusqu'ici, à essayer de tenir compte du niveau d'adaptation « de long terme » estimé à chaque instant, sans prendre en compte, justement, les questions d'adaptation temporelles. De plus, même dans ce domaine restreint, les fondements psycho-visuel ne sont pas convaincants. Il reste à lire plus précisément l'article de Pattanaik et al. (2000).

Bilan de la bibliographie sur le *tone mapping* :

- Tumblin et Rushmeier (1993) sur la luminosité.
- Ward (1994) sur la luminance.
- Spencer, Shirley, Zimmerman et Greenberg (1995) sur l'éblouissement.

- Pattanaik, Ferwerda, Fairchild et Greenberg (1998) sur les variations locales.
- Pattanaik, Tumblin, Yee et Greenberg (2000) sur l'adaptation temporelle.

Bibliographie sur le rendu spectral

Concernant la lumière polarisée, les développements en synthèses d'image portent sur les techniques de lancer de rayons. Un panorama des « systèmes » de rendu spectral est présenté, de préférence à une bibliographie exhaustive.

Les auteurs qui se sont intéressés à la lumière polarisée ont tenté de prendre en compte les équations électro-magnétiques de propagation de la lumière. Une onde incidente est décrite par une matrice de cohérence (dont les éléments peuvent être mesurés). Un milieu (ou une surface) est décrit comme un modificateur de la matrice de cohérence (CMM). Un mode de description plus simple est souvent utilisé : l'onde est décrite par un « vecteur de Stokes » et le milieu par une « matrice de Muller ».

La fluorescence peut être modélisée, si on considère n échantillons dans le spectre, par une matrice $[n \times n]$ de ré-émission. Dans le cas où la lumière est ré-émise, systématiquement, à des longueurs d'onde plus courtes, on peut se contenter d'une matrice triangulaire.

Curieusement, aucun des systèmes de rendu standard en synthèse d'image n'utilise une représentation spectrale de la lumière. Le plus important de ceux qui le font est celui de l'université de Cornell, qui en outre produit des images au format XYZ. La thèse de **Collins** (1996) sur le tracé de photons utilise un modèle spectral. **Peercy** (1993) a proposé une alternative à l'échantillonnage en utilisant une base de fonction bien choisies pour représenter le spectre visible. Des techniques spécifiques de rendu spectral ont été développées par SGI, puis abandonnées, faute de marché. **Iehl et Péroche** (2000) ont proposé un modèle de calcul spectral dont l'échantillonnage s'adapte, au cours du calcul, en fonction des objectifs de précision, afin d'optimiser le temps de calcul. **Sun, Fracchia, Drew et Calvert** (2001) utilisent une représentation mixte, comportant à la fois un échantillonnage et une base de fonction. L'université d'Utah a produit un certain nombre de papiers dans lesquels une représentation spectrale de la lumière était utilisée. **Raso et Fournier** (1991) ont proposé une représentation purement polynomiale. **Icart et Arquès** ont étudié à partir des équations physiques des cas particuliers de couches minces. **Evans et McCool** proposent une stratégie de factorisation des calculs dans des techniques de Monte Carlo, afin d'optimiser le temps de calcul. C'est également l'objectif de **Rougeron et Péroche**, qui cherchent en même temps à minimiser l'erreur colorimétrique finale. **Gondek, Meyer et Newman** (1994) ont proposé un modèle spectral de BRDF. Ils ont également développé un « gonio-photomètre virtuel » à partir d'un algorithme de lancer de rayons.

Conclusion

Il y a encore du travail à faire. Dans le domaine spectral, la difficulté est surtout celle du coût de calcul. En ce qui concerne le *tone mapping*, l'arrivée des écrans LCD et micro-miroirs doit être prise en compte. Les points importants concernant les LCD sont la dépendance angulaire, la température, le manque de contrôle du gamma, et les variations des 3 spectres d'émission pour les faibles niveaux. Le développement d'un système de restitution à grande dynamique serait le bienvenu pour tester les modèles existants. L'idée de fonder les modèles sur les propriétés du système visuel humain doit être poursuivie, mais il existe des zones d'ombres en science de la vision. La validation psychophysique des algorithmes doit être recherchée.