

Optimisation de la Quantification par Modélisation Perceptuelle dans JPEG2000

Christophe Charrier
Université Laval
Dept. de génie électrique et de génie infomatique
LRTS
Québec city, Qc., Canada, G1K 7P4
Téléphone : (418) 656-2131 – Fax : (418) 656-3159
Courriel : charrier@gel.ulaval.ca

Hocine Cherifi
Université de Bourgogne
Faculté des Sciences Mirande
LIRSIA – EA1871
BP 400, 21011 Dijon cedex – France
Téléphone : 03 80 39 68 49 – Fax : 03 80 39 58 87
Courriel : cherifi@crld.u-bourgogne.fr

1. Introduction

Avec le développement fulgurant des applications multimédia, les besoins en technologies augmentent et doivent pouvoir s'adapter à cette nouvelle demande. Ainsi en mars 1997, le développement d'un nouveau standard de compression d'images fixes, JPEG2000, a officiellement été décidé. Ce projet, JTC¹ 1.29.14 (15444), devait aboutir à la création d'un nouveau système de compression adapté à des images de différente nature (noir et blanc, à niveau de gris, couleur, multi-composantes) et de diverses complexité (images naturelles, scientifiques, médicales, radar, etc.) tout en intégrant plusieurs modèles (client/serveur, transmission temps réel, etc.). Le processus de standardisation fut achevé en décembre 2000 et donna naissance au standard de compression JPEG2000.

La qualité des images compressées peut-être singulièrement améliorée dès lors que l'on prend en compte les exigences dictées par le système visuel humain. Il devient alors primordial d'intégrer les caractéristiques fonctionnelles de la vision humaine dans les schémas de compression.

Or de nombreuses recherches ont permis de mettre en exergue la relative similarité entre la transformation en ondelettes et la décomposition fréquentielle opérée par le système visuel humain [1,2]. Ceci permet alors de prendre en compte les effets de masquage de manière simple et efficace lors de l'utilisation de la transformation en ondelettes qui est présente dans le schéma de compression standard JPEG2000.

Ainsi une stratégie de quantification sur les coefficients d'ondelettes a été exploitée dans JPEG2000 afin d'exploiter au mieux les propriétés de masquage du système visuel humain.

1 Architecture du standard

La procédure d'encodage présente dans JPEG2000 est la suivante [3,4] :

- l'image est tout d'abord décomposée en composantes,

- ces dernières sont ensuite subdivisées en tuiles rectangulaires. Une tuile représente une unité de base de l'image, *i.e.* une composante,
- une transformation en ondelette est appliquée sur chacune de ces tuiles. La tuile est ainsi décomposée sur plusieurs niveaux de résolution,
- ces niveaux constituent des sous-bandes de coefficients décrivant les caractéristiques fréquentielles de zones locales de la tuile (plutôt que de la tuile entière),
- les sous-bandes sont ensuite quantifiées et regroupées en code-blocs, puis en précincts. Par défaut, tous les coefficients d'une même sous-bande forment un précinct,
- les plans de bits d'un code-bloc sont codés selon un codage entropique,
- les données sont ordonnées par couche de qualité permettant d'accroître la qualité de l'image reconstruite.
- La trame contient une en-tête principale permettant l'image originale ainsi que les diverses décomposition et procédures de codage utilisées pour localiser, extraire, decoder et reconstruire l'image en fonction d'une qualité désirée, de diverses régions d'intérêt, etc.

Il convient de noter que la procédure de compression est basée sur l'algorithme EBCOT—*Embedded Block Coding with Optimized Truncation of the embeded bitstream*—décrit en détails dans les références [5].

2 Les modèles du système visuel

Malgré les connaissances acquises sur le système visuel humain (SVH) par les neurophysiciens, les psychologues, etc., sa complexité rend pratiquement impossible la construction d'un modèle physiologique complet. Néanmoins, quelques tentatives ont été réalisées [6,7,8], mais ont été restreintes à la modélisation de la rétine et ne permettent pas de prendre en compte la vision haut-niveau. Dès lors, le modèles de la vision humaine utilisés dans le traitement des images

¹ Joint Technical Committe

couleurs sont généralement basés sur le résultats d'études psychophysiques.

Ainsi ces divers modèle ne sont constituer que d'une suite d' « effets psychophysiques » implantés selon un processus séquentiel tel que décrit dans la Figure 1.

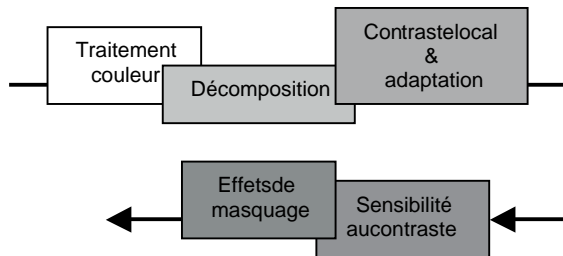


Figure 1: Diagramme du modèle typique du système visuel humain

2.1 La luminance et la couleur

La première étape dans la chaîne de traitement des modèles du SVH consiste en la transformation des données colorimétriques dans un espace couleur perceptuel, qui est généralement basé sur l'opposition des couleurs. Ceci permet d'obtenir une composante achromatique et deux composantes chromatiques. Ces dernières permettent de calculer les différences colorimétriques entre deux pixels.

Cette étape permet également de prendre en compte la perception non linéaire de la luminance par le SVH [9]. Cette particularité est inhérente à tout espace perceptuel tel que le CIE $L^*a^*b^*$, mais doit cependant être ajouté pour tous les autres espaces. Lors de compression, cette particularité peut-être prise en compte en fixant la précision de quantification des coefficients transformés [10].

2.2 Décomposition multi-canal

Il est communément admis que le SVH effectue une décomposition multi-canal des faisceaux lumineux reçus par les cônes et les bâtonnets de la rétine. Des mesures des champs récepteurs des cellules simples du cortex visuel primaire ont permis de démontrer que ces canaux exhibent une structure diadique [11,12]. Ce qui restituent relativement correctement les décomposition par ondelettes et les bancs de filtres.

2.3 Contraste et adptation

La réponse du SVH dépend bien moins d'une luminance absolue que d'une relation avec ses variations locales dans un voisinage, propriété connue sous le nom de « propriété de Weber-Fechner » [9]. Le contraste est une mesure de cette variation relative de luminance, communément utilisée dans les modèles de la vision humaine. Cependant la difficulté d'utilisation de cette loi prend toute sa dimension lors de l'application sur des images complexe puisque la perception du contraste varie avec le contenu local de l'image.

2.4 Sensibilité au contraste

L'une des propriétés les plus importantes utilisées dans les modèles du SVH concerne sa faible sensibilité aux hautes fréquences. Ce phénomène est reproduit par la fonction de sensibilité au contraste (FSC)—*contrast sensitivity function (CSF)*. Cependant son utilisation pour des images couleur est difficile. Afin de contourner cette difficulté, on suppose que la sensibilité aux couleurs et aux formes est différentes. Dès lors, une FSC doit être déterminée et implantée pour chacune des composantes de l'espace couleur. Les fonctions de sensibilité au contraste achromatique sont décrites par BARTEN [13] les mesures des FSC couleurs sont décrites par MULLEN [14] et une description détaillée d'une combinaison de ces trois FSC avec une décomposition en ondelette est donnée par NADENEAU *et al.* [15].

La sensibilité au contraste du SVH dépend également de la fréquence temporelle du stimuli. À l'instar de la FSC spatiale, la FSC temporelle est un filtre passe-bas. L'interaction entre les fréquences spatiales et temporelles peuvent être décrites par divers fonctions de sensibilités spatio-temporelles, qui sont largement repandue dans les modèles de vision appliqués aux séquences vidéo [1].

2.5 Masquage

L'effet de masquage est un phénomène perceptuel permettant à des artéfacts d'être masqués localement par le contenu de l'image. Quelquefois, l'effet inverse, *i.e.* la facilitation peut se produire : un stimulus qui n'est pas nécessairement visible lors qu'il est considéré tout seul, peut être détecté grâce à son voisinage.

Plusieurs types de masquage peuvent être distingués, mais les critères permettant de les séparer ne sont pas francs et nets. Les expressions *masquage de contraste*, *masquage de contours* et *masquage de texture* sont couramment utilisés pour décrire des masquages dus à un contraste local, de contour et d'activité locale. Le masquage temporel représente une petite élévation du seuil de visibilité due à des discontinuités temporelles présentes lors de busques variations dans les images. [16]

3 Exploitation du masquage

Une optimisation visuelle largement répandue est l'utilisation de la fonction de sensibilité au contraste caractérisant la variation de la sensibilité du système visuel humain aux fréquences spatiales 2D. Cette fonction est somme toute très efficace lors d'applications hautes résolutions ou lorsque la distance de visualisation de l'image est importante. Cependant, cette efficacité tend à disparaître lorsque l'affichage est effectué en basse résolution ou lorsque l'observateur se situe proche de l'image. Sous ces conditions, la propriété de masquage semble pouvoir améliorer la qualité des images.

Cette propriété de la vision humaine se produit principalement dans les canaux de fréquence spatiale limités en fréquence radiale et en orientation. A l'instar de la vision humaine, la transformation en ondelettes permet d'obtenir une décomposition spatio-fréquentielle de l'image qui se traduit par une série d'images spatiales dont le contenu fréquentiel diffère d'une image à l'autre. D'où l'idée d'exploiter la propriété de masquage afin d'améliorer la qualité des images.

Une première idée est d'utiliser une quantification non uniforme qui serait plus grossière en fonction de l'activité en fréquence spatiale ainsi qu'en fonction de la localisation spatiale et qui serait appliquée sur les coefficients. Puisque ces effets de masquage sont à peu près identiques dans chacun des canaux, il est alors possible d'utiliser la même procédure de masquage. Une possibilité permettant d'exploiter cet effet de masquage est d'appliquer une transformation (basée sur une quantification non linéaire) sur les coefficients obtenus *via* la fonction de sensibilité au contraste. Ceci permet d'avoir une quantification non-uniforme des coefficients originaux dans laquelle ceux ayant une grande amplitude sont quantifiés de façon plus grossière.

Une seconde possibilité permettant d'exploiter la propriété de masquage est de contrôler la contribution de chacun des blocs sur le masquage d'une région ou d'un artefact. Actuellement, le codage de chaque bloc s'effectue indépendamment les uns des autres. Cependant l'effet de masquage est pris en compte dans la mesure de distorsion intégrée à JPEG2000. Ceci est réalisé en pondérant chacun des coefficients de distorsion par un facteur de masquage visuel qui est généralement fonction du voisinage. Cette approche est cependant limitative car elle n'affecte que la mesure de distorsion présente dans l'encodeur et ne permet de définir que les points de troncature de chacun des blocs.

Afin d'obtenir une meilleure prise en compte de la propriété de masquage, nous faisons coopérer les deux approches citées précédemment, à l'instar des travaux de ZENG *et al.* [17]. L'idée principale est de prendre également en compte la propriété de masquage mesurable dans le voisinage spatial des coefficients. Ceci permet entre autre de distinguer les coefficients de grande amplitude appartenant à des contours de ceux appartenant à des régions de grande activité telles que les zones texturées d'une image. Ainsi cette distinction permet de conserver une bonne restitution des contours dans un environnement lissé. En effet, le système visuel humain est extrêmement sensible à toute dégradation précoce des contours, ce qui est souvent fatal lors du jugement de la qualité globale de l'image.

Cette coopération s'effectue en commençant par une transformation des coefficients $(x_i)_{i \in [1, \dots, N]}$ via une quantification non linéaire f telle que :

$$f : x_i \rightarrow x'_i = f(x_i)$$

Cette étape suppose que chaque signal auquel est associé un coefficient se situe sur un arrière plan de même amplitude. Sous cette hypothèse, les coefficients $(x'_i)_{i \in [1, \dots, N]}$ sont perceptuellement quantifiés. Dans ce cas, chaque signal recouvre une partiellement les autres signaux voisins. Ce phénomène engendre alors des effets de masquage issus du voisinage. Afin d'exploiter cette propriété, les valeurs x'_i sont normalisées par un facteur de voisinage α_i qui est fonction de l'amplitude des coefficients présents.

4 Mesure de la qualité

De façon à évaluer l'impact de la technique utilisée, nous utilisons également des tests psychophysiques.

4.1 Environnement expérimental

Nous avons utilisés 15 images différentes dont la complexité varie d'une image de synthèse à une image fortement texturée. Chacune de ces images ont été compressées en utilisant la méthode proposée, et celle décrite par ZENG *et al.* [17].

Les images ont été présentées à chaque sujet en alternant aléatoirement la position de la première image (à droite ou à gauche de l'image originale) sur l'écran. Cependant, l'image doit avoir été présentée 80 fois à droite et 80 fois à gauche. Ainsi 160 présentations ont été réalisées pour chaque triplet d'images afin d'éviter un biais lié à l'accoutumance du sujet. Nous obtenons alors un meilleur estimé de sensibilité de l'observateur.

4.2 Mesures psychophysiques

Afin de comparer les deux méthodes, nous avons utilisé la technique du choix forcé qui permet de quantifier la sensibilité de l'observateur.

Cette sensibilité, appelée *mesure de sensibilité*, est dans ce cas définie à l'aide de l'estimé deux probabilités $p(D)$ (Décision Correcte) et $p(FA)$ (Fausse Alarme). Cette mesure de sensibilité d' , représente simplement la distance entre la moyenne de la distribution du <stimulus 1> et de la distribution du <stimulus 2>. Il est toutefois utile de préciser que la sensibilité d' caractérise uniquement la paire de stimuli et non la méthode qui est à l'origine de chacun des stimuli. Sous l'hypothèse d'avoir une réponse non biaisée, la mesure de la sensibilité d' s'exprime sous la forme suivante :

$$d' = \frac{1}{\sqrt{2}} [z(p(D)) - z(p(FA))]$$

où $z[\cdot]$ est la fonction inverse de la fonction de distribution normale.

4.3 Résultats

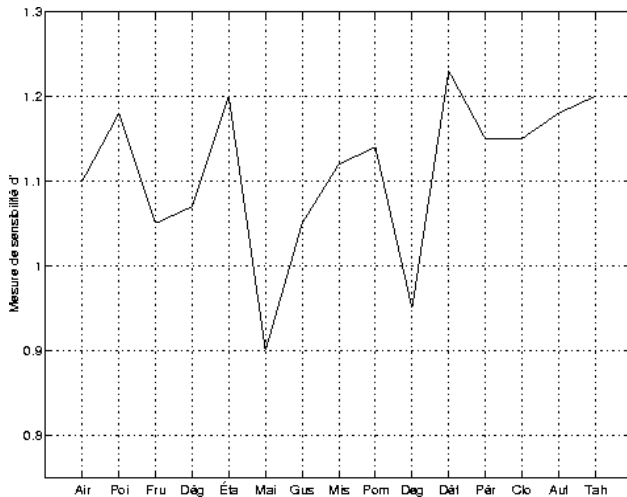


Figure 2 : Mesures de sensibilité pour les images testées

La figure \ref{fig:tout} présente les mesures de sensibilité obtenues pour toutes les images testées. Plusieurs remarques peuvent alors être formulées :

- si l'on considère les valeurs d' inférieures à 1, on note que les images correspondantes (« Mai » et « Deg ») sont peu complexes. Dans ce cas, les deux méthodes de compression fournissent des résultats identiques.
- pour des valeurs de la sensibilité comprises entre 1 et 1.15, on observe que les images associées présentent plus de zones homogènes que précédemment. La technique développée conserve quelque peu la qualité des images.
- pour des valeurs de d' supérieures à 1.15, les images correspondantes présentent des plages de dégradé de couleur. Dans ce cas, la technique développée considère, dans une moindre mesure, un dégradé de couleurs en plusieurs régions distincte, et introduit alors un minimum des contours fictifs.

5 Conclusion

Les résultats obtenus précédemment ont permis de démontrer que la coopération mise en œuvre tout au long de cette étude améliore sensiblement la qualité des images compressées lors d'un affichage en basse résolution ou lorsque la distance de l'observateur à l'écran est faible.

Références

[1] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity", in *Digital Images and Human Vision*, ed. A. B. Watson, pp. 179-206, MIT Press, 1993

[2] P. W. Jones, S. Daly, R. Gaborski and M. Rabbani, "Comparative study of wavelet and DCT decomposition with equivalent quantization and encoding strategies for

medical images", in *SPIE Medical Imaging: Image Display*, vol. 2431, (San-Diego, CA), pp.571-582, Feb. 1995

[3] C. Christopoulos, "JPEG2000 Verification Model 8.0 (technical description)", ISO/ EIC JTC1/SC29/WG1 N1822, July 21, 2000.

[4] M. Boliek, C. Christopoulos and E. Majani, "JPEG2000 Part I Final Draft International Standard", ISO/SC29/WG1 N1855, August 18, 2000

[5] D. Taubman, "High Performance Scalable Image Compression with EBCOT", in *Proc. IEEE Int. Conference Image Processing*, Vol. III, pp. 344-348, (Kobe, Japan), October 1999

[6] J. Atick et al., "Understanding retinal color coding from first principles", in *Neural computation*, Vol. 4, pp 559-572, 1992

[7] S. Guth, "Model for color vision and light adaptation", in *J. Opt. Soc. Am.*, Vol. 8, n° 6, pp 976-993, 1991

[8] R. D. Valois and K. D. Valois, "A multi-stage color vision", in *Vision Res.*, Vol. 33, n° 8, pp 1053-1065, 1993

[9] W. Schreiber, *Fundamentals of Electronic Imaging Systems*, Springer Verlag, NY, 1993

[10] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston; 1992

[11] J. G. Daugman, "Two dimensional spectral analysis of cortical receptive field profile", in *Vision Rech.*, Vol. 20, n° 10, pp. 847-546, 1990

[12] M. A. García-Pérez, "The perceived image : efficient modelling of visual inhomogeneity", in *Spatial Vis.*, Vol. 6, n° 42, pp. 89-99, 1992

[13] P. G. J. Barten, *Contrast sensitivity of the human eye and its effects on image quality*, SPIE, Bellingham, Washington, 1999

[14] K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings", *J. Physiol.* Vol. 359, pp. 381-400, 1985

[15] M. Nadeneau and J. Reichel, "Image compression related contrast masking measurements", in *proc. SPIE Human Vision and Electronic Imaging*, Vol. 3959, (San-José, CA), 2000

[16] A. J. Seyler and Z. L. Budrikis, "Detail perception after scene changes in television presentations", in *IEEE Trans. Inform. Theory*, Vol. 11, n° 1, pp. 31-43, 1965

[17] W. Zeng, S. Daly and S. Lei, "Visual optimisation tools in JPEG2000", in *ICIP (Vancouver, BC)*, Sept. 2000