

L'apport d'un bloc de segmentation d'erreur dans l'évaluation de la qualité d'images codées

Tiffany TWEED¹, Hakim SAADANE²

¹Laboratoire ERIC, Université Lumière Lyon 2

5 ave, Pierre Mendés France, 69676 BRON

ttweed@eric.univ-lyon2.fr

²IRESTE / IRCCyN / IVC, Site de la Chantierie

Rue Christian Pauc BP50609, 44306 Nantes

abdelhakim.saadane@polytech.univ-nantes.fr

Résumé – Ce papier présente un schéma d'évaluation de la qualité d'images où la conception de chaque bloc est basé sur les propriétés du système visuel humain. L'objectif principal est de déterminer l'apport de la prise en compte du contenu sémantique de l'image d'erreur dans le calcul de la note finale. Sur une batterie de 30 images codées JPEG, on montre que les notes fournies par l'évaluateur, quand le cumul spatial est effectué par une sommation de Minkowski, présentent un coefficient de corrélation avec le MOS de 0,79. Ce coefficient atteint la valeur de 0,91 quand le cumul spatial est effectué après une segmentation de l'image d'erreur.

Abstract – *This paper presents a quality assessment scheme where human visual system properties are considered in the design of each part of the scheme. The main goal is to specify how final objective measure varies with the semantic content of the error image. On a set of thirty JPEG coded images, we show that the linear correlation coefficient is of 0.79 when spatial pooling is performed by a Minkowski summation. This correlation coefficient grows to 0.91 when the spatial pooling is performed after a segmentation of error image.*

1. Introduction

L'apparition des nouvelles technologies de vidéo numérique, visant des niveaux de compression de plus en plus importants et entraînant la réalisation de multiples structures de codage, pose le problème de la qualité perceptuelle des images restituées.

Les tests subjectifs, bien que demeurant la référence dans le domaine de l'évaluation, sont toutefois coûteux, longs et fastidieux à mettre en œuvre.

Les mesures objectives classiques telles que le rapport signal à bruit crête et l'erreur quadratique moyenne restent de simples indicateurs et ne donnent pas d'idée sur comment les distorsions sont perçues par le Système Visuel Humain (SVH).

Les métriques perceptuelles représentent une alternative intéressante. Elles exploitent les différentes modélisations du SVH pour améliorer d'une manière significative la corrélation entre la note qu'elles fournissent et celle donnée par un ensemble d'observateurs. Les modèles considérés utilisent généralement les propriétés suivantes

- Le système visuel est plus sensible au contraste relatif qu'à la luminance absolue. Dans le cas des stimuli simples (sinusoïdes, incrément ou décrement ..), la variation de la perception de la luminance en fonction de la luminance locale du fond est bien modélisée par la loi de Weber-Fechner complétée par celle de Moon et Spencer pour les faibles luminances. Pour les stimuli complexes, différents modèles dépendant des différentes configurations considérées sont donnés dans [1,2,3,4].

- La sensibilité au contraste est une fonction de la fréquence. Différents modèles de cette fonction, appelée CSF (Contrast Sensitivity Function), sont décrits dans [5,6].
- L'information visuelle est analysée par le SVH par une batterie de canaux dont chacun est ajusté sur une bande de fréquences radiales et d'orientations données. Plusieurs études psychophysiques ont été menées pour caractériser ces canaux. Bien que les résultats soient assez disparates, la tendance générale s'oriente vers une largeur de bande radiale d'une octave et une sélectivité angulaire qui varie, en fonction de la fréquence, entre 20 et 60 degrés.
- Le masquage visuel traduit la capacité du système visuel à s'adapter au contraste local du fond. Différentes études se sont intéressées à la mesure de la variation, du seuil de visibilité d'un stimulus, induite par la présence d'un signal « masquant ». Plusieurs modèles dépendants de la nature, de l'orientation et de la phase des stimuli utilisés existent dans la littérature. Les plus connus sont largement discutés dans [7,8,9,10,11].

Dans ce papier nous nous intéressons également aux métriques perceptuelles. Un algorithme d'évaluation de la qualité, où les propriétés du SVH décrites ci dessus, sont exploitées. L'objectif principal est de quantifier l'apport, exprimé en terme de coefficient de corrélation, de la prise en compte du contenu sémantique de l'image dans le calcul de la note finale.

Ce papier est organisé comme suit. Le paragraphe 2 donne une description de l'algorithme utilisé. Le paragraphe 3 rappelle brièvement l'algorithme de segmentation retenu et le paragraphe 4 présente quelques résultats obtenus.

2. Description de l'algorithme

Le schéma blocs de l'algorithme d'évaluation de la qualité est donné figure 1. L'image originale $I(m,n)$ et l'image dégradée $I_d(m,n)$ subissent d'abord une double transformation non linéaire. La première correspond à la prise en compte de la non linéarité de l'écran et permet donc le passage des niveaux de gris en luminances objectives. La deuxième transformation permet de passer des luminances objectives aux luminances subjectives.

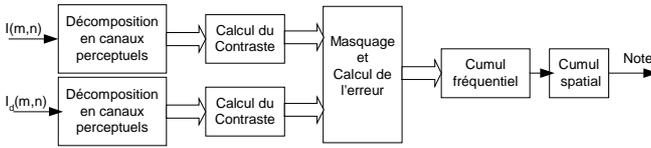


FIG. 1 : Algorithme d'évaluation

2.1 Décomposition en canaux perceptuels

Les deux images de luminance ainsi obtenues subissent ensuite une décomposition en canaux perceptuels. Cette décomposition (figure 2), modélisant le comportement du SVH, exploite un ensemble de 17 canaux fréquentiels répartis en un canal basses fréquences et quatre couronnes radiales. Chaque couronne est décomposée en secteurs angulaires.

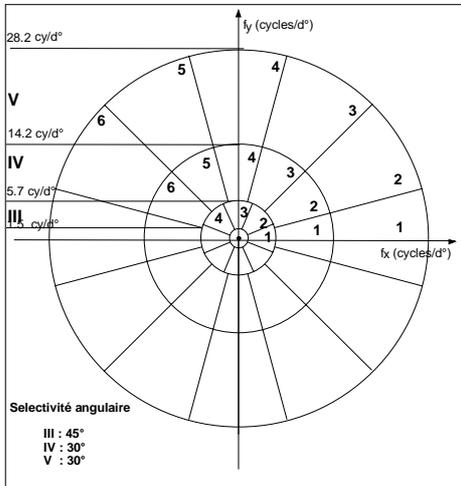


FIG. 2 : Décomposition en canaux perceptuels

Les filtres utilisés pour implémenter cette décomposition sont les filtres cortex de Watson [12]. Ces filtres sont définis comme étant le produit entre les filtres DOM, qui caractérisent la sélectivité radiale (couronne), et les filtres FAN qui déterminent la sélectivité angulaire.

2.2 Calcul du contraste

Le contraste est calculé pour chacune des 17 sous bandes de luminance issues de la décomposition. Ce calcul est nécessaire compte tenu de la sensibilité du SVH à cette

grandeur. Pour bien définir le contraste, il semble nécessaire que la résolution spatiale et fréquentielle soient liées. Une définition répondant à ce critère est celle du contraste local à bande limitée [3]. Pour une décomposition telle que celle de la figure 2, ce contraste s'écrit

$$c(m, n) = L_{i,j}(m, n) / \sum_{k=0}^{i-1} \sum_{l=0}^{\text{card}(l)} L_{k,l}^i(m, n)$$

où $L_{i,j}(m, n)$ et $c_{i,j}(m, n)$ représentent respectivement la luminance et le contraste au point (m, n) du $i^{\text{ème}}$ canal radial et $j^{\text{ème}}$ secteur angulaire. Card(l) est le nombre de secteurs angulaires dans le $k^{\text{ème}}$ canal radial.

Le dénominateur représente la luminance moyenne locale correspondante à tous les canaux de support spectral inférieur à celui du $i^{\text{ème}}$ canal.

2.3 Masquage

L'effet de masquage traduit la variation du seuil de visibilité d'un stimulus en fonction de la distribution de la luminance de son voisinage. En considérant la décomposition en canaux perceptuels de la figure 2, cet effet se manifeste entre signaux appartenant au même canal et entre signaux appartenant à des canaux différents.

Le rôle de ce bloc est donc de préciser pour chaque sous bande et pour chaque point la variation du seuil de visibilité quand l'effet de masquage est pris en compte.

Le modèle utilisé ici est issu d'un très grand nombre d'expériences psychophysiques menées au laboratoire. Pour ce modèle le contraste seuil est donné par

$$\Delta c_{i,j}(m, n) = \frac{E_{i,j}}{\bar{L}_i(m, n)} \left(\frac{\Delta f_i}{f_{o,i}} \right)^{k_i}$$

où $\Delta c_{i,j}(m, n)$ est le contraste seuil de la sous bande (i, j) , $E_{i,j}$ est l'énergie de la sous bande (i, j) , $\bar{L}_i(m, n)$ est la luminance moyenne locale à la sous bande radiale i et au point (m, n) , Δf_i est la largeur de la sous bande radiale i , $f_{o,i}$ est la fréquence centrale de la sous bande radiale i et k_i est une constante qui dépend de la sous bande radiale i .

La connaissance de telles valeurs, rappelons le, permet de ne conserver que les erreurs situées au dessus de leur seuil et contribuant donc à l'élaboration de la qualité finale. Le calcul du contraste de ces erreurs perçues est donné par

$$c_{i,j}^p(m, n) = \begin{cases} \frac{c_{i,j}^{\text{err}}(m, n) - \Delta c_{i,j}}{\Delta c_{i,j}^0} + 1 & \text{si } c_{i,j}^{\text{err}}(m, n) \geq \Delta c_{i,j} \\ 0 & \text{sinon} \end{cases}$$

où le coefficient de normalisation $\Delta c_{i,j}^0$ correspond au seuil de détection sans masquage.

2.4 Cumul fréquentiel

Les erreurs visibles qui sont conservées dans chacune des sous bandes sont ensuite cumulées. Le premier cumul effectué est le cumul fréquentiel angulaire. Ce cumul consiste à

regrouper les images d'erreurs réparties sur toutes les sous bandes appartenant à une même bande radiale en une seule image d'erreurs par couronne. La sommation de Minkowski a été utilisée avec un exposant égal à l'infini. Ce qui correspond à appliquer l'opérateur Max

$$c_i^P(m, n) = \max_j (c_{i,j}^P(m, n))$$

Cet opérateur, qui ne considère que les erreurs les plus importantes, se justifie par le fait qu'une erreur locale de forte amplitude reste nettement plus gênante qu'un ensemble d'erreurs de moindre amplitude.

Le deuxième cumul effectué est le cumul radial. Ce cumul consiste à regrouper les quatre images d'erreur, correspondant aux quatre couronnes, en une seule et unique image d'erreur. Il est effectué moyennant une simple combinaison linéaire. L'image d'erreur finale est calculée en tout point (m,n) selon

$$c^P(m, n) = \left(\sum_{i=1}^4 \alpha_i c_i^P(m, n) \right)$$

où α_i représente le coefficient de pondération de la couronne radiale i .

2.5 Cumul spatial

Le but de ce dernier cumul est de combiner les erreurs spatiales $c^P(m,n)$ en une seule mesure finale représentant la note attribuée par l'algorithme à l'image dégradée présentée à l'entrée de l'évaluateur. Après plusieurs essais le modèle retenu est celui qui effectue le cumul par une sommation de Minkowski pondérée par les probabilités d'occurrences des erreurs. Pour ce modèle la note fournie Q est donnée par

$$Q = \sum_{n=1}^N \sum_{m=1}^M (c^P(m, n))^\alpha \left(\Pr(c^P(m, n)) \right)^\beta.$$

3. Segmentation de l'erreur

Une deuxième approche est considérée ici pour effectuer le cumul spatial. Elle consiste à segmenter d'abord l'image d'erreur $c^P(m,n)$ et à effectuer ensuite un cumul sur chacune des composantes générées. La méthode de segmentation, voulue simple pour ne pas alourdir les calculs, est celle proposée par Xu et al [13]. Son principe consiste à exploiter l'activité $A_o(p)$ d'un pixel p au niveau de l'image de référence et l'activité $A_d(p)$ de ce même pixel au niveau de l'image dégradée pour générer trois composantes : les vrais contours S_{vrai} , les faux contours S_{faux} et le reste de l'image S_{reste} (image de bruit). L'algorithme est simple et s'écrit

si $|A_d(p) - k_0 A_o(p)| > T_{faux}$ pour tout $p \in S$ alors $p \in S_{faux}$

si $|A_d(p)| > T_{vrai}$ pour tout $p \in S - S_{faux}$ alors $p \in S_{vrai}$

pour tout $p \in S - (S_{faux} + S_{vrai})$, $p \in S_{reste}$,

où T_{vrai} et T_{faux} représentent les seuils de décision des vrais et faux contours respectivement.

Trois facteurs de distorsions, notés x_{vrai} , x_{faux} et x_{reste} sont ensuite générés par application de la norme L1, ou de son logarithme, aux trois composantes précédemment calculées. La note finale Q_i est obtenue par combinaison de ces trois facteurs. Les quatre modèles retenus sont :

- En utilisant la norme L1

$$Q1 = 4.317 - 0.104 x_{vrai} - 0.696 x_{faux} - 0.556 x_{reste}$$

$$Q2 = \frac{4}{1 + \exp(-(1.914 - 0.191x_{vrai} - 0.946x_{faux} - 0.854x_{reste}))} + 1$$

- En utilisant le logarithme de la norme L1

$$Q3 = 4.809 - 0.761 x_{vrai} - 1.534 x_{faux} - 1.172 x_{reste}$$

$$Q4 = \frac{4}{1 + \exp(-(2.60 - 1.158x_{vrai} - 2.0246x_{faux} - 1.827x_{reste}))} + 1$$

4. Résultats

4.1 La base de données

Un ensemble de 6 images originales bien connues a servi de base à l'étude (figure 3). A chacune de ces images, cinq dégradations JPEG ont été générées pour couvrir l'ensemble de l'échelle de qualité définie dans la recommandation 500 du CCIR [14]. La corrélation entre les notes subjectives (MOS variant entre 1 et 5) de l'ensemble des trente images et les notes fournies par l'évaluateur a été utilisée pour optimiser les différents paramètres des modèles.



FIG. 3 : Base d'images utilisées

4.2 Sans segmentation

Dans le cadre de l'approche sans segmentation, le coefficient de pondération utilisé pour effectuer le cumul fréquentiel radial a été optimisé. Les résultats montrent qu'il suffit de pondérer fortement la couronne radiale IV car le système visuel humain y est plus sensible. La valeur 8 pour cette couronne et la valeur 1 pour toutes les autres couronnes ont été retenues.

Les coefficients α et β du cumul spatial ont également été optimisés. Le meilleur coefficient de corrélation obtenu avec

la batterie des 30 images test est de 0.799 pour $\alpha = 0.1$ et $\beta=0.5$.

4.3 Avec segmentation

Les performances de la méthode de segmentation dépendent du choix des seuils de détection des vrais et faux contours. Après plusieurs essais (histogramme, valeur moyenne, écart type...), le choix simple s'est fait en tenant compte de la dynamique de l'activité de l'image originale $A_0(p)$ pour le seuil de détection des vrais contours, et de la dynamique $|A_d(p)| - k_0|A_0(p)|$ pour les seuils de détection des faux contours. Dans les dynamiques considérées, les fortes valeurs de l'un ou l'autre des seuils sont à éviter car la segmentation dans ce cas est mal adaptée puisque elle privilégie l'une ou l'autre des composantes. Basés sur ces considérations, les meilleurs résultats obtenus, avec l'ensemble des trente images de la base, sont donnés tableau 1. Rappelons que ces résultats expriment le coefficient de corrélation entre les mesures subjectives (MOS) et les notes fournies par les quatre modèles ci dessus. Les seuils utilisés (faux contour, vrai contour) sont également donnés à titre indicatif.

TAB. 1 : Coefficients de corrélation obtenus avec les quatre modèles retenus.

Seuils	(-10,10)	(-10,20)	(-5,5)	(-5,10)
Modèle 1	0.903	0.890	0.886	0.903
Modèle 2	0.887	0.878	0.855	0.880
Modèle 3	0.911	0.884	0.901	0.901
Modèle 4	0.906	0.884	0.880	0.886

Pour les seuils (-10,10) et (-5,5), l'examen visuel des images montre que pour une dégradation donnée, une bonne répartition se fait entre l'image des erreurs sur les vrais contours, l'image des erreurs sur les faux contours et l'image du bruit.

Une comparaison des différents modèles montre que le modèle 3 présente, dans les deux cas, le meilleur coefficient de corrélation.

5. Conclusion

Une métrique perceptuelle de la qualité d'image a été présentée. Cette métrique exploite une modélisation multicanaux du système visuel humain pour décomposer l'image originale et l'image dégradée en un ensemble de composantes visuelles. Un modèle de masquage dépendant des caractéristiques de chacune de ces composantes est ensuite appliqué pour générer des images d'erreurs perceptuelles. Un cumul fréquentiel angulaire suivi d'un cumul radial permet de regrouper les différentes erreurs réparties fréquemment en une seule image d'erreur. Sur cette image d'erreur un dernier cumul est effectué pour construire la note finale. Pour ce dernier cumul, deux approches ont été considérées. La première utilise la sommation de Minkowski pondérée par les probabilités d'occurrences des erreurs, et la deuxième exploite une

segmentation de l'image d'erreur en trois composantes. Les résultats montrent que le coefficient de corrélation, qui est de 0.799 dans le premier cas, atteint la valeur de 0.911 dans le deuxième cas. Cette amélioration, exclusivement due à la prise en compte du contenu sémantique, ouvre de nouvelles perspectives dans l'élaboration des critères objectifs.

Références

- [1] A.B.Watson. *DCTune : A technique for visual optimization of DCT quantization matrices for individual images*. SID Digest of technical paper, XXIV, pp. 949-964, 1993.
- [2] A.J.Ahumada et H.A.Peterson. *Luminance-model based DCT quantization for color image compression*. Proc. SPIE 1666, pp.365-374, 1992.
- [3] E.Peli. *Contrast in complex images*. JOSA, A7, N°10, pp. 2032-2040, 1990.
- [4] S.Daly. *The visible difference predictor : an algorithm for the assessment of image fidelity*. Digital images and human vision, Chap. 14, pp.179-206, 1993.
- [5] A.B.Watson, G.Y.Yang, J.A.Solomon et J.Villasenor. *Visibility of wavelet quantisation noise*. IEEE Trans. Image Processing, 6 (8), pp. 1164-1175, 1997.
- [6] E.Peli, L.E.Arend, G.M.Young et R.B.Goldstein. *Contrast sensitivity to patch stimuli : Effects of spatial bandwidth and temporal presentation*. Spatial Vision, 7 (1), pp.1-14, 1993.
- [7] G.E. Legge et J.M. Foley. *Contrast masking in human vision*. JOSA, Vol.70, pp. 1458-1471, 1980.
- [8] J.A.Solomon, A.B.Watson et A.J.Ahumada. *Visibility of DCT basis functions : Effects of contrast masking*. Proc. Of the data compression conf., IEEE Computer Society Press, pp. 361-371, 1994.
- [9] J.M. Foley. *Human luminance pattern mechanisms : Masking experiments require a new model*. JOSA, A11 (6), pp. 1710-1719, 1994.
- [10] J.M. Foley et G.M. Boynton. *A new model of human luminance pattern vision mechanisms : Analysis of the effects of pattern orientation, spatial phase and temporal frequency*. Proc. SPIE, Vol. 2054, pp.32-42, 1993.
- [11] P.C.Teo et D.J.Heeger. *Perceptual image distortion*. Proc. SPIE, Vol. 2179, pp. 127-141, 1994.
- [12] A.B.Watson. *The cortex transform : rapid computation of simulated neural images*. Computer Vision, Graphics and image Processing, N°39, pp. 311-327, 1987.
- [13] W.Xu et G.Hauske. *Picture quality evaluation based on error segmentation*. Proc. SPIE, Vol. 2308, pp. 1454-1465, 1994.
- [14] CCIR. *Projet de révision de la recommandation 500-4 : Méthode d'évaluation subjective de la qualité des images de télévision*. Document commissions d'études du CCIR 11/BL/51-F, 1992.