

Critère objectif de qualité visuelle d'images vidéo de documents

Dominique BARBA, Pascale SIEFFERT

IRESTE – Laboratoire SEI, EP CNRS
rue Christian Pauc, La Chantrerie, BP 60601, Nantes, France

dbarba@ireste.fr

Résumé – Ce papier présente une méthode de construction d'un critère mesurable de qualité d'images vidéo de documents, critère devant correspondre à une mesure subjective de la lisibilité de texte contenus dans les images de documents. La méthode s'appuie sur une modélisation du système visuel humain (SVH) en terme de perception de dégradations et inclut une décomposition en sous-bandes visuelles. Elle s'appuie aussi sur une modélisation explicite du cumul inter sous-bandes et du cumul spatial des dégradations, fournissent une mesure globale de qualité. Le modèle global obtenu conduit à une bonne prédiction par le modèle de la qualité subjective donnée par un panel d'observateurs.

Abstract – This paper presents the design of an objective quality criterion of document video images. The measurements given by this objective criterion have to be in good correspondance with subjective measurements of lisibility of the documents. The proposed method uses a model of the human visual system (HVS) in term of degradation perception based on a psychovisual sub-band decomposition of the images. The method deals also with an explicit non linear modeling of the inter sub-band pooling and the spatial pooling of the perceived degradations and provides a unique measurement of the overall image quality. This objective criterion is able to perform well for predicting the subjective quality of quite different document images.

1 Introduction

Les tests subjectifs de qualité sont encore très largement utilisés pour évaluer la qualité d'images dégradées fournies par des systèmes de traitement d'images (codage ou autres). Ce sont des procédures complexes, lourdes à mettre en œuvre et fastidieuses. Il est de ce fait très difficile d'optimiser sérieusement un système (méthode) de traitement du point de vue de la qualité réelle des images qu'il produit. D'où l'idée apparue depuis plus de 20 ans de concevoir un critère mesurable de la qualité d'image comparable à celle que fournirait un observateur standard dans des conditions normalisées de visualisation. À partir d'une simple mesure bruitée (critère très mauvais) de la distorsion quadratique, un ensemble de critères, à complexité croissante, ont été proposés pour améliorer les performances du critère considéré (écart entre mesure objective et mesure subjective, conservation de la relation d'ordre, ...). Un premier exemple type, très récent, qui inclue les derniers perfectionnements est celui de Miyahara, Kotami et Algazi, appelé le PQS [1]. Il incorpore les non-linéarités d'entrée, la pondération en fréquences spatiales, et détermine 5 facteurs (F1 à F5) faisant intervenir différents aspects visuels (seuillage visuel, corrélation des distorsions, effets de masquage) qui servent à construire linéairement la mesure de qualité Q. On ne peut pas dire qu'il y ait vraiment une construction basée sur une modélisation systématique et logique du Système Visuel Humain dans la formation de la qualité. En particulier l'aspect, maintenant bien établi, d'une décomposition multibande du signal d'image par le SVH n'est pas pris en compte.

Un second exemple type qui inclue une décomposition multibande est celui proposé par C. Van der Branden

Lambrecht pour l'évaluation de la qualité des images vidéo [2]. La décomposition est très régulière, sûrement trop régulière. Une normalisation des erreurs avec seuillage incorporé permet de prendre en compte l'effet de masquage (purement intra canal). Un premier cumul des erreurs normalisées sur des blocs spatio-temporels approchant la taille des zones de fixation est effectué. Elle est réalisée par une sommation de Minkovski avec un exposant $\beta = 4$. Un dernier cumul par sommation de Minkovski avec un autre exposant β' engendre, après une dernière transformation non linéaire, une mesure Q de la qualité.

Les limitations de ces critères sont multiples. Pour simplifier, les trois plus importantes sont :

- une décomposition en sous-bandes trop régulière, souvent celle de Watson est utilisée [3],
- une interaction entre sous-bandes soit nulle, soit très limitée,
- un cumul spatial des erreurs basé sur des normes usuelles : la norme de Minkovski.

Les travaux présentés dans ce papier concernent la conception d'une méthode de mesure objective de la qualité subjective d'images de documents ayant subi des dégradations dues à un codage avec compression d'informations, la méthode de compression étant de type JPEG. Dans cette étude, la nature des images testées est particulière : celle d'images de documents contenant du texte informationnel. D'autres constituant peuvent être trouvés : graphiques, logos, etc. La qualité observée par l'utilisateur s'exprime essentiellement en terme de lisibilité du texte. Celle-ci fait intervenir les déformations de rendu du couple fond de l'image-corps du tracé des caractères.

Dans une seconde partie, nous présenterons le modèle de qualité d'images de documents tant concernant l'aspect

modèle de perception que celui de modélisation de la fusion des dégradations perçues.

2 Présentation du modèle de qualité d'images de documents

Classiquement un modèle de qualité d'image est constitué de deux parties : une partie A de perception qui engendre pour une image (fixe) I_D observée, un champ de répartition spatiale des dégradations perçues (normalisées souvent), une seconde partie B qui rassemble ces dégradations en une mesure unique scalaire : la qualité. Une échelle de dégradation peut être également utilisée.

2.1 Modèle de perception

Nous avons modélisé le système de perception par une décomposition du signal d'image en sous-bandes visuelles (SBV) élaborée précédemment dans notre équipe [4]. Ce modèle est donné figure 1. Il est constitué pour la décomposition des images vidéo de 5 bandes de fréquence radiales (indice i en chiffre romain), chacune des bandes étant découpée (sauf la bande basses-fréquences) en sous-bandes directionnelles (second indice j en chiffre arabe). Il faut noter que la sélectivité angulaire n'est pas constante. Un sous-échantillonnage adapté à chaque sous-bande est ensuite effectué. On appelle $L_{ij}(s)$ le signal de sortie d'image de la sous-bande (i, j) en un site donné s .

La seconde partie du système de perception de dégradations élabore pour chaque sous-bande (i, j) une répartition spatiale des distorsions perceptuelles normalisées en s'appuyant sur une grandeur perceptuelle fondamentale : le contraste local à bande limitée. Il est défini en un site donné s d'image par :

$$C_{ij}(s) = \frac{L_{ij}(s)}{\bar{L}_{ij}(s)}$$

$\bar{L}_{ij}(s)$ est la valeur moyenne locale (en s) du signal d'image, reconstruite à partir de toutes les sous-bandes de fréquences radiales inférieures à la sous-bande radiale de numéro i . Nous avons montré que celui-ci définissait quasiment une métrique perceptuelle locale : son pas de quantification ΔC_{ij} est constant sans effet de masquage inter sous-bandes, mais sa valeur est propre à chaque sous-bande (i, j). Ceci permet de seuiller et de normaliser les distorsions par le pas de quantification ΔL_{ij} adapté au site considéré s . Ces distorsions perceptuellement normalisées et quantifiées, appelées $e_{ij}^N(s)$, sont données par :

$$e_{ij}^N(s) = 0 \quad \text{si} \quad |e_{ij}(s)| < \Delta L_{ij}(s) \\ = E[e_{ij}(s)/\Delta L_{ij}(s)] \quad \text{sinon}$$

avec $E(y)$ = entier le plus proche de y .

En fait, il existe un masquage inter sous-bandes. Mais dans la représentation du contraste local à bande limitée, seules les sous-bandes (i, j') où j' correspond à des orientations adjacentes à celle de (i, j) participent à l'interaction dans l'effet de masquage. Celui-ci a été modélisé et incorporé dans le calcul de ΔC_{ij} [4].

Une remarque d'importance doit être faite. Fondamentalement cette approche nécessite, comme pour la majorité des autres, que l'on dispose de l'image de référence I_0 (non dégradée) pour pouvoir construire les champs de distorsions perceptuelles pour chaque sous-bandes (figure 1). Des approches plus complexes cherchent à s'affranchir de cette connaissance.

2.2 Modèle de fusion des dégradations perceptuelles.

Cette fonction de fusion des dégradations est appelée pooling dans la littérature anglaise. C'est la partie du modèle de qualité objective la plus difficile à concevoir via des tests psychovisuels car elle fait appel au système neuronal central du SVH. Nous avons par des ensembles de tests psychovisuels cherché à mettre en évidence les fonctions principales, et estimé les paramètres qui les décrivaient. Les tests psychovisuels ont consisté à mesurer la qualité subjective d'images pour des configurations des distorsion contrôlées en sous-bandes visuelles et en distribution d'amplitude, en utilisant comme type de distorsions celles résultant d'une quantification uniforme dans la ou les sous-bandes considérée(s), le pas (ou les pas) étant réglable(s), multiple(s) du pas correspondant au seuil de perception.

Une première série de tests, les plus simples, ont limité les dégradations à une seule sous-bande (i, j). Un second ensemble a successivement impliqué l'interaction des dégradations entre sous-bandes directionnelles adjacentes, puis sur toutes les orientations d'une même bande (i) en fréquences radiales, puis sur deux bandes radiales (les plus importantes) III et IV. Finalement, l'ensemble a été validé sur des dégradations engendrées par un système de vidéo codage d'images de documents (images de lettres grand formats et couverture de magazine envoyés par la poste) utilisant la norme de codage JPEG et pour des taux de compression très variés.

De ces tests, nous avons déduit un modèle de fusion en une mesure de qualité des distributions spatiales des distorsions perçues filtrées (sous-bandes) et normalisées. Il est constitué (figure 2) d'une succession de quatre transformations :

- 1) La première porte sur le cumul directionnel à bande de fréquences radiales donnée i . Un modèle simplifié consiste à prendre le maximum de la valeur absolue de l'erreur selon les différentes orientations : $e_i^N(s) = \sup_j |e_{ij}^N(s)|$
- 2) La seconde de type filtrage spatial local engendre une erreur perceptuelle filtrée en chaque site s de la bande radiale i . Nous avons testé plusieurs formes de filtrage linéaire et non-linéaire. Un filtrage d'ordre de type quartice supérieur a fourni les meilleurs résultats après optimisation. Le support spatial du filtre est de type 3×3 , ou de type en croix à 5 éléments (suivant la bande radiale). La sortie est $\bar{e}_i(s) = \text{filtrage d'ordre}_{3/4}(\{e_i^N(s')\} \quad s' \in V(s))$.
- 3) La troisième fournit pour chaque bande radiale i une mesure F_i globale pour l'image considérée. C'est à ce

niveau que se situe le cumul spatial que nous effectuons sur tout le domaine spatial. La loi de probabilité de la distribution de l'amplitude de \bar{e}_i intervient (cela a été noté par d'autres chercheurs antérieurement) : $F_i = \sum_m x_m f(Pr[\bar{e}_i(s) = x_m])$

où $f(p)$ est une fonction polynomiale monotone croissante, de degrés $1/\alpha$ pour $p \in [0, P_0]$, de degrés α pour $p \in [P_0, 1]$, continue en P_0 et à valeur sur $[0, 1]$, avec $0 < \alpha < 1$.

On peut considérer que F_i qualifie la notion de gêne.

- 4) Enfin la dernière transformation rassemble les gênes F_i en une mesure unique. Pour une mesure de qualité, nous construisons d'abord une note Q_i de qualité propre à chaque bande radiale i (sur une échelle à 5 notes) par :

$$Q_i = 5 - a_i \left[\log_{10} \left(1 + \frac{F_i}{F_{ref_i}} \right) \right]^3$$

La fusion des notes partielles de qualité Q_i s'opère par l'addition des unités Imp qui sont définies par :

$$Imp = \frac{Q_{max} - Q}{Q} \quad (Q_{max} = 5 \text{ dans notre cas})$$

$$\text{avec } Imp_i = \sum_i Imp_i$$

$$\text{et } Imp_i = \frac{Q_{max} - Q_i}{Q_i}$$

3 Résultats

Les différents paramètres modélisant $f(p)$ (2 paramètres) et les Q_i (2 paramètres par bande i) ont été optimisés avec les résultats des tests subjectifs à dégradations contrôlées en sous-bandes. Les résultats obtenus par ce critère objectif issus des tests sur images de documents codées-décodées par la norme JPEG sont en très bonnes concordances avec les notes subjectives. La figure 3 présente cette comparaison, pour 5 types d'images différentes de documents, et à différents taux

de compression (en fait fixés par le coefficient q de qualité de JPEG).

4 Conclusion et perspectives

Nous avons présenté un schéma complet de construction d'une norme objective de qualité d'images de documents. Sa partie complètement novatrice concerne la fusion des distributions spatiales d'erreurs perceptuelles à bandes limitées en une sortie unique de qualité. La construction de cette fusion a été faite à partir de la conception et la menée d'une suite logique de tests psychovisuels spécifiques. Les résultats obtenus sont très bons sur images de documents monochromes. L'extension d'une telle recherche sera naturellement le développement de critères objectifs pour des images quelconques monochromes et couleur puis pour les séquences d'images animées.

Remerciements : Ce travail a été fait dans le cadre d'un contrat de recherche avec le Service de Recherche Technique de la Poste (SRTP), que nous remercions pour son aide.

Références

[1] M. MIYAHARA, K. KOTANI, V.R. ALGAZI « Objective Picture Quality Scale For Image Coding » - IEEE Transactions on communications, Vol XX, N° Y, 1998

[2] C. J. VAN DEN BRANDEN LAMBRECHT, O. VERSCHURE « Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System ». Proceedings of the SPIE, Vol 2668, pp. 450-461, Sans Jose, CA, January 28-February 2, 1996. (available on http://ltswww.epfl.ch/pub_files/vdb/.)

[3] A.B WATSON, « The cortex transform : rapid computation of simulated neural images », C.V.G.I.P, Vol 39, 1987, pp. 311-327

[4] A. SAADANE, H. SENANE, D. BARBA, « Visual coding : Design of psychovisual quantizers », Journal of Visual Communication and Image Representation, Vol. 9, N°4, December. 1998, pp.381-391.

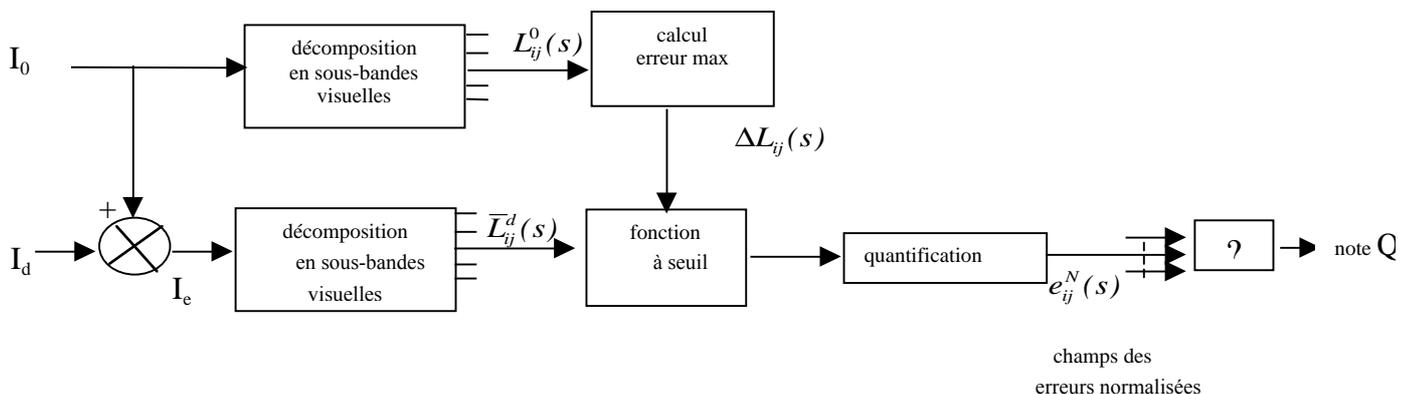


Figure 1 : Décomposition en sous-bandes et construction des champs d'erreurs perceptuelles normalisé

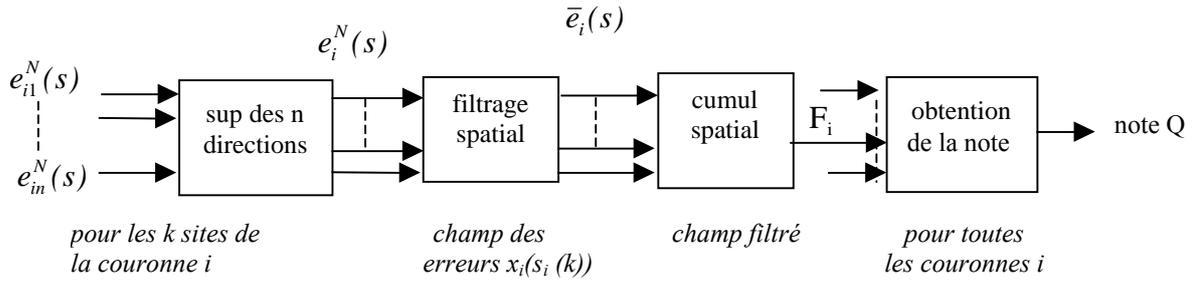


Figure 2. Bloc-diagramme de la méthode de fusion des champs de distorsions perceptuelles normalisées

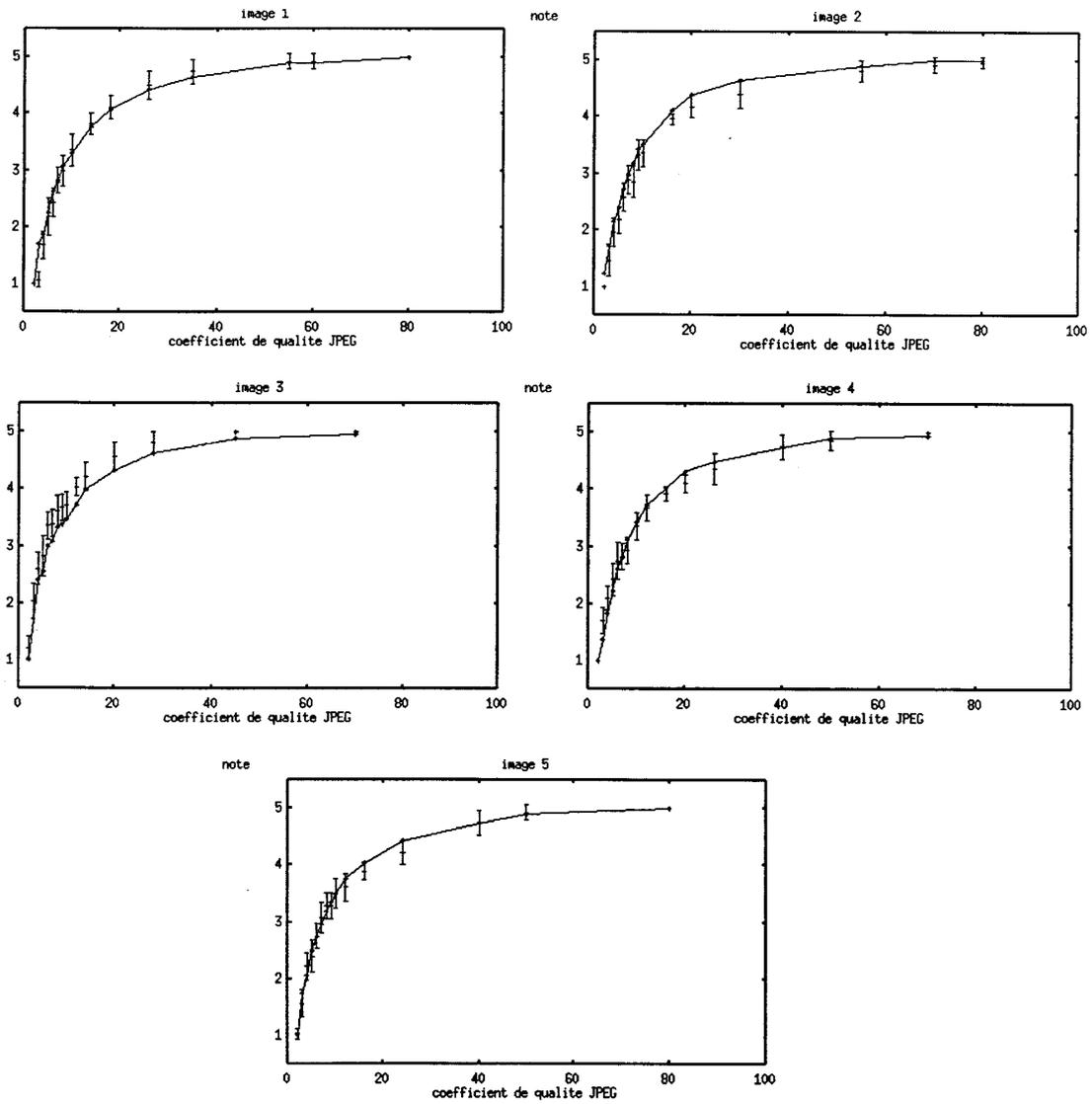


Figure 3 : Notes subjectives moyennes et notes objectives calculées, en fonction du coefficient de qualité JPEG