



**HAL**  
open science

## Compression d'images en norme $L_\infty$

Lamia Karray, Olivier Rioul, Pierre Duhamel

► **To cite this version:**

Lamia Karray, Olivier Rioul, Pierre Duhamel. Compression d'images en norme  $L_\infty$ . 16es Journées Tunisiennes en Electrotechnique et Automatique (JTEA'94), Nov 1994, Hammamet, Tunisie. hal-03330259

**HAL Id: hal-03330259**

**<https://telecom-paris.hal.science/hal-03330259v1>**

Submitted on 10 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Compression d'images en norme $L^\infty$

Lamia Karray\*, Olivier Rioul\* et Pierre Duhamel\*\*

\*France Telecom-CNET/PAB/STC/SGV

38-40, rue du Général Leclerc, 92131 Issy-les-Moulineaux, France

\*\*ENST/SIG, 46, rue Barrault, 75013 Paris, France

Email : lamia.karray@issy.cnet.fr

**Résumé.** Dans ce papier, on développe une nouvelle technique de codage basée sur les propriétés statistiques du bruit de quantification et utilisant un critère de norme  $L^\infty$ . L'image originale est prétraitée, quantifiée, codée et reconstruite à l'intérieur d'un intervalle de confiance donné. Comme prétraitement, on utilise des bancs de filtres itérés et on présente des résultats dans le cas de codage de très bonne qualité (presque sans pertes).

**Mots clés.** Codage d'images, bancs de filtres, erreur de quantification, erreur de reconstruction, distribution statistique, norme  $L^\infty$ .

**Abstract.** A new image coding technique using statistical properties of quantization errors and  $L^\infty$ -norm criterion is investigated. The original image is preprocessed, quantized, encoded and reconstructed within a given confidence interval. We focus on iterated filter banks as a preprocessing technique, and provide results in the case of very good quality (almost lossless) image coding.

**Key words.** Image coding, filter banks, quantization error, reconstruction error, statistical distribution,  $L^\infty$ -norm.

## 1 Introduction

Diverses applications nécessitant le stockage ou la transmission de signaux numériques ont conduit, ces dernières années, à un important travail de recherche dans le domaine de la compression de données aussi bien sans pertes qu'avec pertes. De nombreuses techniques comme les méthodes par prédiction ou les méthodes par transformées orthogonales ont ainsi été développées [6]. Ces transformations sont appliquées sur le signal original afin d'améliorer, par leurs propriétés de décorrélation, les statistiques des données à coder. Elles nécessitent néanmoins des quantificateurs ce qui introduit des erreurs à la reconstruction. Les schémas de compression doivent, par conséquent, être choisis de façon à minimiser ces distorsions [7] et utilisent en général comme critère l'erreur en

moyenne quadratique (norme  $L^2$ ). Cependant, l'évaluation de l'erreur en norme  $L^2$  ne permet qu'un contrôle global des distorsions. Or certaines applications nécessitent un contrôle précis et local de l'erreur ce qui nous mène à introduire le critère de la norme  $L^\infty$  (ou *minmax*) qui consiste à évaluer et surveiller le maximum de l'erreur de reconstruction. Ce critère permet aussi de prendre en compte les éventuelles connaissances *a priori* sur le signal telles que la précision avec laquelle sont données les valeurs de pixels d'une image biomédicale ou le nombre de bits par pixels sur lesquels cette image est codée. L'utilisation du critère de la norme  $L^\infty$  introduit une nouvelle méthode de compression sans pertes qui exploite le fait que le signal original est quantifié en imposant une erreur de reconstruction inférieure au demi pas de quantification initial. Ce critère est aussi efficace en codage avec pertes pour des applications qui nécessitent une certaine précision. On peut ainsi faire une compression assurant un certain pourcentage d'erreurs inférieures à un seuil fixé.

## 2 Schéma de compression

La chaîne de compression, figure 1, comporte les trois blocs habituels :

### 2.1 La décorrélation par transformée

Notre travail se place dans le cas d'une décorrélation du signal effectuée à l'aide d'une transformée en ondelettes discrètes (TOD, ou banc de filtres). Il a été démontré dans plusieurs travaux [8] l'efficacité de ces transformées en compression d'images. Cette transformée permet de décorréler le signal à coder et à améliorer les statistiques de l'entrée du codeur. La TOD peut être calculée à l'aide de bancs de filtres à reconstruction parfaite dont on itère le passe-bas.

### 2.2 La quantification

Le prétraitement est suivi d'une opération de quantification pour redonner au signal des valeurs

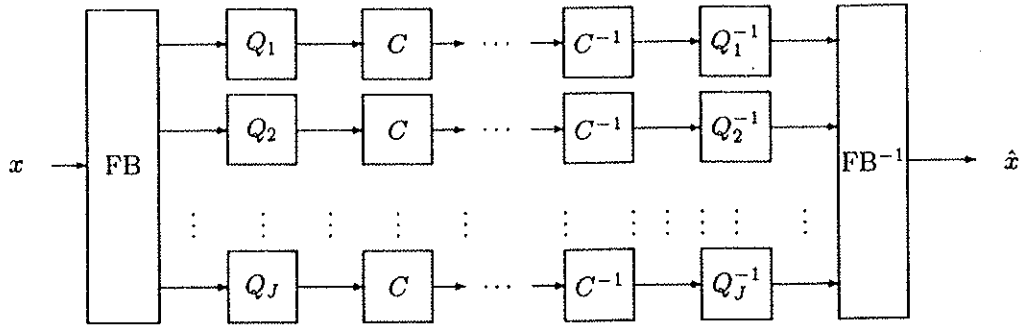


Figure 1: Schéma général de compression. La partie codage utilise un banc de filtres itérés en octaves (TOD), une quantification et un codage sans pertes. La partie décodage fait intervenir les inverses des opérations de codage pour reconstruire le signal  $\hat{x}$ .

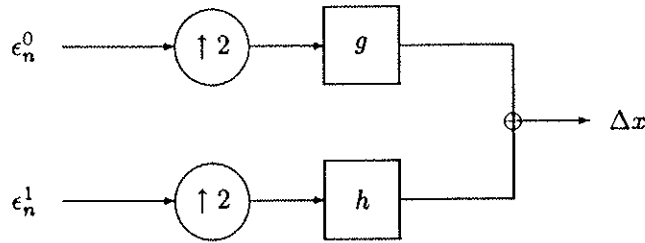


Figure 2: L'erreur de reconstruction. On considère une seule itération d'un signal monodimensionnel.  $\epsilon_n^0$  et  $\epsilon_n^1$  sont les erreurs de quantification.  $\Delta x$  est l'erreur de reconstruction.

entières dans un ensemble fini de niveaux. On choisit une quantification uniforme qui peut être scalaire quand elle considère indépendamment chaque symbole du signal ou vectorielle quand elle fait intervenir des blocs de données.

### 2.3 Le codage

Cette étape consiste à éliminer les redondances statistiques et à affecter des codes aux valeurs quantifiées du signal. On utilise des codeurs sans pertes tels que le codage d'Huffmann global (GHC). Les opérations inverses sont effectuées au décodage.

## 3 Codage en norme $L^\infty$ : point de vue déterministe

En utilisant le schéma de codage décrit précédemment, on cherche à comprimer le signal tout en assurant que l'erreur  $\Delta x$  entre le signal initial,  $x$ , et celui obtenu après compression,  $\hat{x}$ , ne dépasse pas un certain seuil  $s$ . Ceci introduit implicitement le critère du maximum de l'erreur, i.e. critère de la norme infinie ( $L^\infty$ ). Le problème est alors de déterminer les caractéristiques des différents éléments de la chaîne de compression vérifiant la condition

$$\|\Delta x\|_\infty \leq s \quad (1)$$

où  $\|\Delta x\|_\infty = \|\hat{x} - x\|_\infty = \max\{|\hat{x} - x|\}$  est la norme infinie de l'erreur de reconstruction. D'où la nécessité d'évaluer la norme infinie de cette erreur.

### 3.1 Evaluation de la norme infinie de l'erreur

Les filtres étant à reconstruction parfaite et les codeurs sans pertes, les seules distorsions sont générées à la quantification et sont "amplifiées" par filtrage lors de la phase de reconstruction du signal.

On considère d'abord un signal monodimensionnel et une itération du banc de filtres. On s'intéresse à la phase de synthèse schématisée par figure 2. Un échantillon de l'erreur en sortie s'écrit

$$\Delta x_{2n} = \sum_{k=0}^{\frac{N_g}{2}-1} \epsilon_{n-k}^0 g_{2k} + \sum_{k=0}^{\frac{N_h}{2}-1} \epsilon_{n-k}^1 h_{2k} \quad (2)$$

$$\Delta x_{2n+1} = \sum_{k=0}^{\frac{N_g}{2}-1} \epsilon_{n-k}^0 g_{2k+1} + \sum_{k=0}^{\frac{N_h}{2}-1} \epsilon_{n-k}^1 h_{2k+1}$$

où  $N_g$  (resp.  $N_h$ ) est la longueur du filtre  $g$  (resp.  $h$ ). La distinction entre les échantillons pairs et impairs de l'erreur provient de l'interpolation qui précède le filtrage. On a alors

$$\|\Delta x\|_\infty \leq \max \left\{ \begin{array}{l} (\Delta x_{2n})_{\max} \\ (\Delta x_{2n+1})_{\max} \end{array} \right.$$

A partir de l'expression (2) on détermine une estimation de la borne supérieure de l'erreur. Une borne supérieure de l'erreur est obtenue quand les

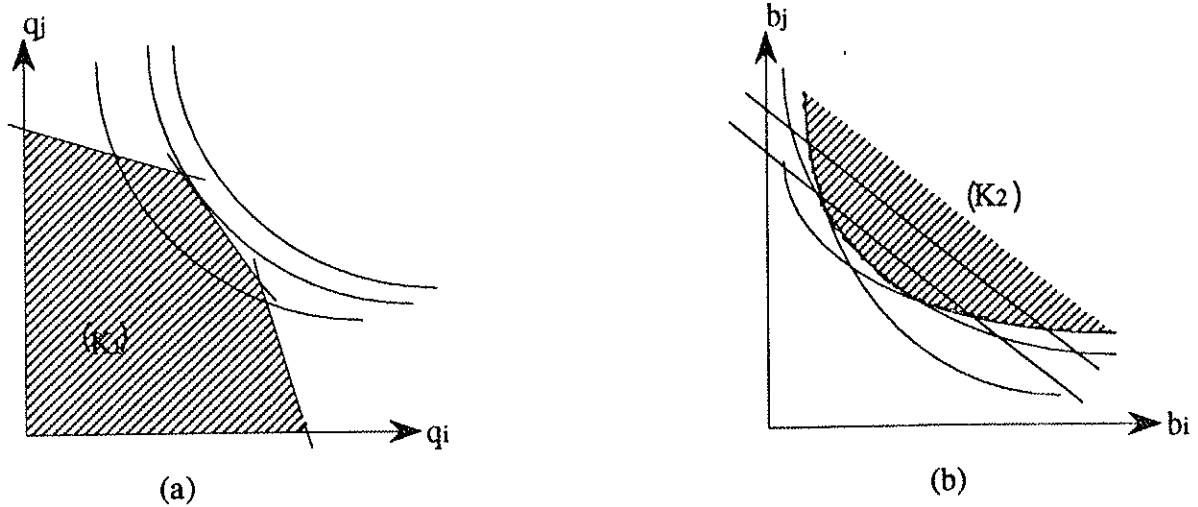


Figure 3: Problèmes équivalents d'optimisation sous contraintes . (a) Le problème  $(P_1)$  a une fonction coût convexe et des contraintes linéaires qui délimitent le domaine permis  $(K_1)$ . (b) Le problème  $(P_2)$  a une fonction coût linéaire et des contraintes convexes, le domaine permis est  $(K_2)$ . On voit facilement qu'il existe une solution optimale située sur les frontières de  $(K_1)$  ou  $(K_2)$ .

erreurs de quantification ( $\epsilon^0$  et  $\epsilon^1$ ) sont extrémales et de même signe que les coefficients des filtres. Or, si  $q^i$  est le pas de quantification l'erreur de quantification  $\epsilon^i$  vérifie  $|\epsilon_k^i| \leq \frac{q^i}{2}$ . Ce qui donne

$$\|\Delta x\|_\infty \leq \max \begin{cases} \frac{q^0}{2} \sum_k |g_{2k}| + \frac{q^1}{2} \sum_k |h_{2k}| \\ \frac{q^0}{2} \sum_k |g_{2k+1}| + \frac{q^1}{2} \sum_k |h_{2k+1}| \end{cases}$$

Sans aucune connaissance *a priori* du signal original, cette estimation est optimale (elle peut être atteinte) bien que pessimiste pour des signaux réels. Par ailleurs, on constate que la borne supérieure de l'erreur est fonction linéaire des pas de quantification. Pour un ensemble donné de filtres, le contrôle de l'erreur est assuré par un choix adéquat des quantificateurs.

### 3.2 Extension au cas bidimensionnel

On considère des bancs de filtres séparables, i.e. la décomposition se fait séparément sur les lignes et les colonnes du signal bidimensionnel. Un raisonnement analogue au précédent nous donne

$$\|\Delta x\|_\infty \leq \max \begin{cases} (\Delta x_{2m,2n})_{\max} \\ (\Delta x_{2m,2n+1})_{\max} \\ (\Delta x_{2m+1,2n})_{\max} \\ (\Delta x_{2m+1,2n+1})_{\max} \end{cases}$$

### 3.3 Cas de plusieurs itérations

C'est le filtre passe-bas qui est itéré. Au bout de  $J$  itérations, on obtient  $3J+1$  sous-bandes. De même

que précédemment, on obtient

$$\|\Delta x\|_\infty \leq \max_{p,q} \left\{ \max_{m,n} (\Delta x_{2^J m+p, 2^J n+q}) \right\}$$

où les  $p$  et  $q$  varient de 0 à  $2^J - 1$  et proviennent des  $J$  interpolations lors de la synthèse. On a alors, pour  $J$  itérations,  $2^J \times 2^J = 2^{2J}$  types d'échantillons de l'erreur.

### 3.4 Choix des quantificateurs

Les quantificateurs doivent être choisis de façon à satisfaire la condition (1). En prenant arbitrairement des pas de quantification suffisamment petits, cette condition sera vérifiée puisque les filtres sont à reconstruction parfaite en l'absence de quantification. Cependant, le pas de quantification  $q_i$  dans la sous-bande numéro  $i$  de dynamique  $d_i$  détermine le débit binaire dans cette sous-bande selon la relation

$$q_i = d_i 2^{-b_i} \quad (3)$$

Or, le but initial est de comprimer le signal, donc de réduire son débit binaire qui s'exprime en fonction des différents débits dans les sous-bandes par

$$b = \sum_i n_i b_i \quad (4)$$

où  $n_i$  est le poids de la sous-bande numéro  $i$  dans l'ensemble du signal (c'est  $\frac{1}{2^{2J}}$  pour les sous-bandes de l'étage  $j$ ). On a donc intérêt à choisir les pas grands pour minimiser le débit mais qui, en même temps, vérifient (1). Or d'après ce qui précède, chaque type d'échantillon de  $\Delta x$  doit satisfaire (1), les pas de quantification  $q_i$  doivent

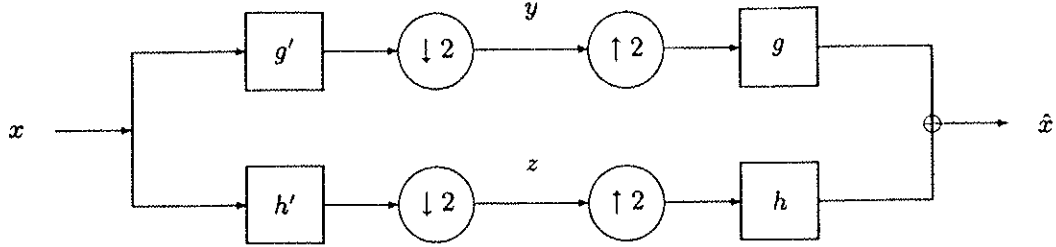


Figure 4: Banc de filtres en une dimension. Les filtres passe-bas et passe-haut d'analyse sont, respectivement,  $g'$  et  $h'$ ; les filtres passe-bas et passe-haut de synthèse sont, respectivement,  $g$  et  $h$ . La branche passe-bas est itérée en utilisant les mêmes filtres (une itération est représentée ici). En deux dimensions, les lignes et les colonnes de l'image sont filtrées séparément.

donc vérifier  $2^{2J}$  contraintes. On a finalement un problème d'optimisation sous contraintes, que l'on exprime de deux façons différentes moyennant la relation (3) :

1. en utilisant les pas de quantification dans les sous-bandes  $q_i$  :

$$(P_1) : \begin{cases} \min(\sum_i -n_i \log_2 q_i) \\ \sum_i \alpha_{ij} q_i \leq s_j \end{cases} \quad (5)$$

avec un critère convexe et des contraintes linéaires (figure 3.a), les  $\alpha_{ij}$  sont exprimés en fonction des coefficients des filtres intervenant dans la sous-bande  $i$ ,

2. en utilisant les débits binaires dans les sous-bandes  $b_i$  :

$$(P_2) : \begin{cases} \min(\sum_i n_i b_i) \\ \sum_i \alpha_{ij} 2^{-b_i} \leq s_j \end{cases} \quad (6)$$

avec un critère linéaire et des contraintes convexes (figure 3.b), les  $\alpha_{ij}$  sont exprimés en fonction des coefficients des filtres intervenant dans la sous bande  $i$  et de la dynamique dans cette sous-bande.

### 3.5 Résolution

Comme le montre la figure 3, le problème a une solution unique qui sera sur la frontière du domaine permis ( $K1$  ou  $K2$ ). Pour trouver numériquement cette solution, on peut appliquer un algorithme d'optimisation général. Mais compte-tenu du nombre élevé de contraintes ( $> 1000$  dans le cas de  $J$  itérations d'un signal bidimensionnel), l'utilisation de ces algorithmes est coûteuse. De plus, ces algorithmes nécessitent un bon point de départ et risquent de diverger si ce point n'est pas suffisamment proche de l'optimum. Par conséquent, on résout le problème en deux étapes : la première consiste à trouver une solution approchée et la deuxième à appliquer l'algorithme d'optimisation général avec comme point de départ cette solution approchée.

#### 3.5.1 Recherche d'une solution approchée

Les contraintes linéaires du problème ( $P1$ ) délimitent un domaine permis ( $K1$ ) qui est un polyèdre convexe dont chacun de ses sommets sature  $n$  contraintes,  $n$  étant le nombre d'inconnues ( $n = 3J+1$  dans le cas de signal bidimensionnel itéré  $J$  fois). Mais, le problème ( $P1$ ) est équivalent à ( $P2$ ), le passage de l'un à l'autre se fait moyennant une transformation strictement convexe ( $x \mapsto -\log_2 x$  ou  $x \mapsto 2^{-x}$ ).

En outre, cette transformation est bijective, donc à un sommet  $Q_i$  de ( $K1$ ) est associé un unique point  $B_i$  de ( $K2$ ) qui sature les mêmes contraintes (par convexité stricte de  $x \mapsto -\log_2 x$ ),  $B_i$  est un sommet "virtuel" de ( $K2$ ). On démontre facilement que l'enveloppe convexe ( $D$ ) de ces sommets "virtuels" de ( $K2$ ) est un polyèdre convexe de sommets les  $B_i$ . La minimisation de coût sur ( $D$ ) donnerait le meilleur sommet du problème ( $P2$ ) (la solution optimale étant sur la frontière de ( $K2$ )). Or, ( $D$ ) est un polyèdre convexe et le critère est linéaire, on résout alors le problème par la méthode du simplexe [10].

Cependant, on ne dispose pas d'une forme explicite simple du domaine ( $D$ ) (on l'obtient par paramétrisation des différentes arêtes reliant les sommets "virtuels" de ( $K2$ ) images de deux sommets adjacents de ( $K1$ )). Mais comme on a l'équivalence entre ( $K1$ ) et ( $K2$ ), on peut ramener le problème à la recherche du meilleur sommet de ( $K1$ ), réalisant le coût minimum. Or dans ce cas le critère est convexe, on résout alors le problème en adaptant l'algorithme du simplexe à ce cas précis.

#### 3.5.2 Recherche de la solution optimale

La première étape de résolution fournit une solution approchée du problème. Cette solution sera le point de départ d'un algorithme général d'optimisation sous contraintes, qui travaillera alors dans de bonnes conditions. La convergence est par conséquent assurée et accélérée. On obtient finalement les pas de quantification optimaux vérifiant

GHC (bpp)	LENA (512 × 512)	Medical (256 × 256) (Coronair)	Aerial (512 × 512) (Aerial view of Corse)
Original	7.47	6.27	3.95
Preliminary Results	6.96 (0.19)	5.09 (0.18)	3.56 (0.19)
Final Results	5.61 (0.5)	3.76 (0.5)	2.42 (0.5)

Table 1: Exemple de codage sans pertes en norme  $L^\infty$ . On affiche les débits binaires globaux obtenus après un Codage d'Huffmann Global (GHC) appliqué 1) sur les images originales, 2) sur les différentes sous-images après transformée (utilisant 5 itérations d'un filtre de Daubechies de longueur 12 [2]). Les erreurs maximales à la reconstruction sont données entre parenthèses. Les résultats préliminaires sont obtenus par application directe de la méthode, il sont ensuite améliorés par la procédure de remise de l'erreur maximale = 0.5 (cf texte).

(1) et assurant un débit minimum.

### 3.6 Codage en sous-bandes

Ayant déterminé les quantificateurs optimaux, il ne reste plus qu'à calculer le débit binaire par application des codeurs sans pertes. Or, ces codeurs nécessitent une évaluation du débit des données à coder. En utilisant (3), on calcule les débits dans les sous-bandes à partir des pas de quantification optimaux. On fait ainsi intervenir les dynamiques dans les sous-bandes que l'on doit, par conséquent, évaluer. Comme pour l'évaluation de l'erreur, on est amené à estimer une borne supérieure de la dynamique. On prend, par exemple, la sous-bande passe-bas dans le cas d'une itération d'un signal monodimensionnel. On a alors à la sortie de la phase d'analyse, figure 4,  $y_n = \sum_{k=0}^{N_{g'}-1} x_{2n-k} g'_k$ , et  $d = y_{max} - y_{min}$ . Or,  $y_{max} = x_{max} \sum_{g'>0} g'_k$  et  $y_{min} = x_{max} \sum_{g'<0} g'_k$ . D'où

$$d \leq x_{max} \sum_k |g'_k| \quad (7)$$

La généralisation au cas bidimensionnel et avec plusieurs itérations se fait par un raisonnement analogue.

### 3.7 Résultats : Application au codage sans pertes

On se place dans le cadre d'images fixes initialement codées sur 8 bpp. On associe les quantificateurs optimaux à une TOD et un codage global d'Huffmann. Les pas de quantification ont été déterminés de façon à avoir une erreur maximale de reconstruction inférieure au demi-pas de quantification initiale, i.e. 0.5, afin de récupérer exactement l'image initiale par requantification sur 8 bpp. Les résultats sont résumés dans la table 1.

La deuxième ligne de cette table nous donne les débits obtenus après application directe de la méthode décrite précédemment. On constate alors une faible réduction du débit binaire mais aussi que l'erreur calculée sur l'image reconstruite est largement inférieure au seuil (0.5). Les estimations

(des bornes supérieures de l'erreur et de la dynamique), effectuées lors de la détermination des quantificateurs optimaux, sont trop pessimistes. Pour cela, on remplace les estimations par les vraies valeurs des normes de l'erreur et de la dynamique calculées directement sur l'image.

Mais, en prenant les dynamiques effectives calculées dans les sous-bandes, on n'améliore pas les résultats. Ceci s'explique par le fait que le codeur sans pertes ne prend en compte que les niveaux effectivement atteints dans les sous-images. Par conséquent, la seule voie d'amélioration est dans l'estimation de l'erreur.

Comme  $\|\Delta x\|_\infty \ll 0.5$ , on peut augmenter le seuil imposé afin d'atteindre effectivement 0.5. Par ailleurs, comme la borne supérieure de l'erreur est une fonction linéaire des pas de quantification, une augmentation uniforme du seuil entraîne la même augmentation sur la borne supérieure de l'erreur, donc sur les pas de quantification. Ainsi, en augmentant uniformément les pas de quantification, on atteint une erreur maximale de reconstruction égale à 0.5, ceci à l'aide d'une procédure de recherche de zéro appliquant la méthode de Brent [5]. Bien que coûteuse en temps de calcul et dépendante du signal, cette méthode assure une erreur de reconstruction de norme infinie égale au seuil toléré (0.5), et on constate, par ailleurs, une baisse importante des débits binaires (cf la troisième ligne de table 1).

## 4 Point de vue statistique

Dans certaines applications, comme les images biomédicales, les valeurs de pixels sont obtenues avec une précision donnée. Ceci permet de définir un intervalle de confiance. Cette information supplémentaire sur le signal peut être exploitée pour la compression sans pertes dans la mesure où l'on peut permettre une erreur de reconstruction tout en restant à l'intérieur de l'intervalle de confiance initial. Ceci nécessite un contrôle local de l'erreur de reconstruction. On a vu qu'avec le critère de la norme  $L^\infty$  on a un contrôle local des distorsions, échantillon par échantillon, qui permet de tenir

compte de l'intervalle de confiance pour comprimer un signal donné avec une erreur à 100% inférieure à un seuil fixé. Par ailleurs, certaines applications autorisent un certain pourcentage d'erreurs supérieures au seuil fixé. Les quantificateurs seront dans ce cas déterminés à partir d'une étude statistique de façon à vérifier la condition

$$\text{prob}\{\|\Delta x\|_\infty \leq s\} \geq p\% \quad (8)$$

où  $p$  est le pourcentage d'erreurs inférieures au seuil  $s$ .

La résolution de ce problème nécessite une connaissance de la distribution statistique de l'erreur de reconstruction.

#### 4.1 Statistique de l'erreur de reconstruction

Comme les pas de quantification sont très faibles ( $q \ll 1$ ), le bruit de quantification dans les sous-bandes est blanc et décorrélié avec le signal. Par conséquent, l'erreur de quantification suit une loi uniforme entre  $-\frac{q}{2}$  et  $\frac{q}{2}$ . L'erreur totale de reconstruction est la somme des erreurs de chaque sous-bande à l'issue de la phase de synthèse. Sa distribution est alors celle d'une combinaison linéaire de lois uniformes. On vérifie en pratique que cette distribution tend vers une gaussienne de moyenne nulle et de variance  $\sigma^2$  (fonction des pas de quantification et des coefficients de filtres) [3, 4]. Sa densité de probabilité s'écrit :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

#### 4.2 Résolution du problème statistique

Il s'agit de déterminer les quantificateurs qui vérifient (8) tout en minimisant le débit binaire. On a, de nouveau, un problème d'optimisation sous contraintes :

$$(P) : \begin{cases} \min(\sum_i n_i b_i) \\ \text{prob}\{\|\Delta x\|_\infty \leq s\} \geq p\% \end{cases} \quad (9)$$

Or, on a vu que  $\Delta x$  est la contribution de plusieurs types d'erreurs (cf section 3.3), par exemple pour  $J$  itérations d'un signal bidimensionnel, il y a  $2^{2J}$  types d'échantillons  $(\Delta x_{2^J m+p, 2^J n+q})$  où  $p, q = 0, \dots, 2^J - 1$ , qui ont chacun une distribution gaussienne de variance  $\sigma_{p,q}^2$ . On peut considérer que ces différents types d'erreurs ont des contributions équiprobables dans le signal global d'erreur. Alors,

$$\text{prob}\{\|\Delta x\|_\infty \leq s\} =$$

$$\frac{1}{2^{2J}} \sum_{p,q} \text{prob}\{\max_{m,n}(|\Delta x_{2^J m+p, 2^J n+q}|) \leq s\}$$

D'autre part, puisque les distributions sont gaussiennes, pour  $p$  et  $q$  donnés on a

$$\text{prob}\{\max_{m,n}(|\Delta x_{2^J m+p, 2^J n+q}|) \leq s\} = \frac{1}{\sigma_{p,q}\sqrt{2\pi}} \int_{-s}^s e^{-\frac{x^2}{2\sigma_{p,q}^2}} dx \quad (10)$$

En supposant que les erreurs de quantification sont indépendantes et que les différentes réalisations d'une erreur sont aussi indépendantes, on détermine facilement  $\sigma_{p,q}$  [3] sachant que

$$\sigma_{p,q}^2 = E[(\Delta x_{2^J m+p, 2^J n+q})^2]$$

Par exemple, dans le cas d'une itération d'un signal monodimensionnel on obtient :

$$\sigma_{p=0,1}^2 = \frac{q_0^2}{12} \sum_k g_{2k+p}^2 + \frac{q_1^2}{12} \sum_k h_{2k+p}^2$$

où  $\frac{q_i^2}{12}$  est la variance de l'erreur de quantification puisqu'elle est uniforme.

La variance est donc fonction des pas de quantification, la résolution du problème statistique revient alors à l'optimisation des quantificateurs.

#### 4.3 Détermination des pas de quantification

D'après l'équation (8), certaines erreurs dépassent le seuil,  $s$ , donné. Les pas de quantification vont donc être choisis avec un seuil plus grand tout en vérifiant (8). On rappelle par ailleurs que dans l'estimation de la borne supérieure de l'erreur, les pas de quantification dépendent linéairement du seuil. Autrement dit, si l'on multiplie le seuil par une certaine constante  $\alpha$ , les pas de quantification sont multipliés par  $\alpha$  et la variance de l'erreur par  $\alpha^2$ . Ainsi, pour résoudre le problème (9), on calcule d'abord les pas optimaux obtenus dans le cas déterministe (cf 3.4). Les nouveaux pas de quantification seront alors  $q'_i = \alpha q_i$ , et la nouvelle variance  $\sigma'^2 = \alpha^2 \sigma^2$ . En utilisant (10), le facteur d'échelle  $\alpha$  est déterminé à partir de la relation  $\text{erf}(\frac{s}{\alpha \sigma \sqrt{2}}) \geq \frac{p}{100}$ . En tenant compte des différents types d'erreurs  $(\Delta x_{2^J m+p, 2^J n+q})$  et de l'hypothèse de leur équiprobabilité, on obtient  $\alpha$ , donc  $q'_i$ , en résolvant, à l'aide d'algorithmes de recherche de zéro, l'équation

$$\frac{p}{100} = \frac{1}{2^{2J}} \sum_{p,q} \text{erf}\left(\frac{s}{\alpha \sigma_{p,q} \sqrt{2}}\right)$$

LENA (512 × 512)							
percentage (%) of errors ≤ 0.5	required	100	99	95	90	85	80
	observed	100	99.30	95.30	89.90	84.63	79.40
GHC (bpp)		5.61	5.01	4.61	4.37	4.18	4.03
Medical image (coronair, 256 × 256)							
percentage (%) of errors ≤ 0.5	required	100	99	95	90	85	80
	observed	100	99.01	95.32	89.89	84.56	79.36
GHC (bpp)		3.76	3.13	2.82	2.61	2.43	2.31

Table 2: Un exemple d'application : codage presque sans pertes. On représente les débits binaires globaux obtenus par un Codage d'Huffmann Global (GHC) appliqué dans les différentes sous-bandes (après 5 itérations du filtre de Daubechies de taille 12 [2]). Le seuil étant fixé à 0.5, on fait varier l'intervalle de confiance. Pour  $p = 100\%$ , c'est le critère déterministe qui est utilisé.

#### 4.4 Résultats

Un exemple d'application de cette approche statistique est donnée dans la table 2. Les résultats sont obtenus dans le contexte d'une "compression presque sans pertes" donc pour un seuil d'erreur maximale tolérée  $s = 0.5$ . La première colonne présente les résultats obtenus dans le cadre de l'approche déterministe où aucune erreur ne dépasse 0.5, le signal original est alors entièrement reconstruit sans erreurs par requantification sur 8 bpp, ceci avec une réduction importante du débit. Les autres colonnes montrent la variation du débit en permettant, de manière contrôlée, de plus en plus d'erreurs au dessus du seuil fixé. Les résultats sont obtenus par application directe du facteur d'échelle  $\alpha$ , calculé par la méthode décrite ci-dessus, aux pas de quantification obtenus par la méthode décrite au paragraphe 3.5. On constate que les intervalles de confiance effectifs sont très peu différents de ceux imposés au départ.

## 5 Conclusion

L'association d'une transformation au critère de la norme  $L^\infty$  dans la compression d'images permet d'avoir une approche simple, efficace et flexible de codage sans pertes. Elle constitue un moyen de contrôle local des distorsions. Ceci permet, pour certaines applications, de tenir compte de la précision de mesure en maintenant l'erreur de reconstruction à l'intérieur de l'intervalle de confiance. Le critère de l'intervalle de confiance permet d'introduire, en outre, la notion de masquage spatial qui consiste à contrôler l'erreur selon les zones d'intérêt dans l'image.

## References

- [1] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image Coding Using Wavelet Transform," IEEE Trans. on Image Processing, Vol. 1, NO. 2, pp. 205-220, 1992.
- [2] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," Comm. Pure Appl. Math, vol XLI, no. 7, pp. 909-294.
- [3] W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley International Edition, NY, 1967, Vol. I, chap. 7.
- [4] M. Loève, *Probability Theory*, The university series, Princeton, NJ, 1960, chap. 6.
- [5] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, 1988, chap. 10.
- [6] M. Rabbani and P.W. Jones, "Digital Image Compression Techniques," SPIE Optical Engineering Press : Bellingham, Washington, USA, 1991.
- [7] K. Ramchandran A. Ortega and M. Vetterli, "Best Wavelet Packet Bases in the Rate Distortion Sense," IEEE Trans. Image Processing, vol. 5, pp. 381-384, Apr. 1993.
- [8] O. Rioul and M. Vetterli, "Wavelets and Signal Processing," IEEE S.P. Magazine, pp. 14-34, Oct. 1991.
- [9] A.B. Sripad and D.L. Snyder, "A Necessary and Sufficient Condition for Quantization Errors to be Uniform and White," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-25, NO. 5, Oct. 1977.
- [10] D.A. Wismer and R. Chattergy, *Introduction to Nonlinear Optimization*, North Holland New York, 1978, (Appendix on linear programming).