

# Mesure de la qualité visuelle des images compressées

## Visual quality measure of the compressed images

par Ahmed TAMTAOUI<sup>1,2</sup>, Driss ABOUTAJDINE<sup>2</sup>

<sup>1</sup>INPT, Avenue Allal Al Fassi, Rabat Instituts, 10100 Rabat, Maroc

<sup>2</sup>GSCM-LEESA, Département de Physique, Avenue Ibn Batouta B.P. 1014, Rabat Maroc

tamtaoui@inpt.ac.ma , aboutaj@fsr.ac.ma

### *résumé et mots clés*

Cet article présente une méthode de mesure objective de la qualité visuelle des images monochromes dégradées par des schémas de codage. La méthode développée se classe dans le cadre des méthodes utilisant une référence. L'originalité de la méthode repose sur la pondération des mesures standards par une densité d'erreur locale calculée sur des fenêtres d'analyse recouvrant l'image. Le calcul de cette densité se base sur trois critères qui sont le contraste, la structure et la quantification. D'autres méthodes, telles que les mesures standards PSNR (Peak Signal to Noise Ratio) et EAM (Erreur Absolue Moyenne), pondérées par le modèle de Daly simplifié [6], et la méthode de Fränti [9], sont décrites dans cet article. Les résultats des différentes méthodes de mesure sont comparés entre eux. La mesure proposée est la plus fiable en terme de coefficient de corrélation avec la mesure subjective pour l'ensemble des images testées.

Qualité objective, qualité subjective, évaluation de la qualité, Mean Opinion Score (MOS), codage d'image, mesure des dégradations, coefficient de corrélation

### *abstract and key words*

This article presents a visual quality measure of the compressed monochrome image. This measure uses a reference image. The originality of this measure is based on the weighting of the Standard measures by a local error density, calculated on the windows overlapping the image. The local error computation is based on the contrast, the structure and the quantification criteria. Actually, this method is compared with standard PSNR (Peak Signal to Noise Ratio) and MAE (Mean of Absolute Error) measures, weighted by the simplified Daly model [6], and Fränti [9] methods. The results of our measure are reliable, compared with the methods mentioned above. The results are then evaluated in terms of the correlation measure with the Mean Opinion Score (MOS).

Quality assessment, objective quality, Mean Opinion Score (MOS), image compression, distortion measures.

## 1. introduction

La mesure de la qualité visuelle des images ou de la vidéo par des méthodes objectives s'avère très utile dans plusieurs applications en imagerie : codage [12][17][23], restauration [2][25], traitement d'images médicales [4]. Toutes les méthodes développées dans ce domaine peuvent être classées en trois grandes catégories : celles qui sont à référence totale (utilisation complète de l'image originale), appelées aussi « Full Reference », celles qui utilisent quelques paramètres de la référence, appelées aussi « Half Reference » ou méthodes à référence réduite [18] et celles qui n'utilisent aucune référence [11]. Les méthodes « Half Reference » sont souvent utilisées dans le cadre de la transmission des images. Les méthodes « Sans Référence » se basent sur la connaissance a priori de la nature des dégradations ou/et des images. Les méthodes sans référence et avec référence réduite ne feront pas l'objet de cet article.

Dans le domaine du codage, plusieurs études ont été faites pour analyser les mesures standards de dégradations basées sur des différences entre pixels des images originale et dégradée [7]. Ces mesures, qui sont à référence complète, ne sont pas commodes pour une évaluation de la qualité à travers différentes dégradations de compression et/ou différents types d'image. En particulier, l'Erreur Quadratique Moyenne (EQM) (ou son corollaire, le PSNR -Peak Signal to Noise Ratio-) est considérée comme une bonne mesure des bruits aléatoires, mais elle est parfois totalement inefficace lorsqu'il s'agit d'erreurs structurées [7][9][22][11] (effet de bloc, erreur de structure, etc.).

Pour résoudre le problème des mesures standards, plusieurs chercheurs ont développé des mesures de dégradations [8][15][19][24][28] en intégrant des modèles de propriétés du Système Visuel Humain (SVH) [29]. Certaines de ces méthodes sont liées à des dégradations très spécifiques telles que les effets de blocs dans JPEG [12][16][23] et le flou dans le codage par les ondelettes [10][27]. D'autres méthodes se sont développées dans le but d'obtenir des mesures qui soient les plus indépendantes possible des types de dégradation [9]. Toutes ces méthodes, dans la plupart des cas, ne sont pas efficaces pour une évaluation de la qualité à travers différents types d'image et masquent les dégradations qui sont localisées et très visibles.

Dans cet article nous proposons une nouvelle méthode de mesure des dégradations de compression en tenant compte de la localité ou de la concentration des erreurs. La mesure proposée se place dans l'hypothèse des méthodes avec référence complète : nous nous plaçons dans le cadre de l'usage des images numériques codées qui peuvent être, par exemple, zoomées, ajustées, affichées sur écran de moniteur ou éditées. Notre préoccupation est d'être le plus indépendant possible de la méthode de codage utilisée, de surmonter le problème de la mesure des dégradations localisées et concentrées et d'être indépendant du type d'image. La méthode proposée est basée sur l'information de concentration d'erreurs locales. L'originalité de cette méthode réside dans

la prise en compte d'une densité d'erreur locale calculée à l'intérieur des fenêtres d'analyse [21][22]. Les types d'erreur considérés sont les erreurs de contraste, de structure spatiale et de quantification (ou de résolution) [9]. Notre méthode est testée sur quatre types de codeur et trois types d'image. Elle est comparée à deux mesures standards (PSNR et EAM) pondérées par le modèle de la Fonction de Sensibilité au Contraste (CSF) de Daly simplifié [6] et à la mesure de Fränti [9]. Nous avons testé le modèle de sensibilité au contraste de Watson [28], qui est initialement développé pour le codage JPEG, pour pondérer plusieurs mesures standards. Les résultats obtenus par la pondération de ce modèle ne se corrélaient pas bien avec la mesure subjective, il ne sera pas traité ici.

## 2. mesures standards pondérées par le modèle de Daly

Nous avons testé deux mesures standards de dégradations (PSNR « Peak Signal to Noise Ratio » et l'EAM « Erreur Absolue Moyenne ») pondérées par le modèle CSF de Daly simplifié [5][6][26]. Ces mesures serviront, par la suite, de comparaison avec la méthode proposée [paragraphe 4].

### 2.1. Le modèle de la fonction de sensibilité au contraste de Daly simplifié

Le modèle de Daly simplifié a été utilisé efficacement dans un schéma d'optimisation de codage JPEG [26]. Il consiste simplement à modéliser le système visuel humain par une fonction de sensibilité au contraste, appelée aussi fonction de transfert de modulation (MTF) :

$$W(u, v) = \begin{cases} 2.2(0.192 + 0.114\tilde{f}(u, v))\exp(-0.114\tilde{f}(u, v))^{1.1} & \text{si } \tilde{f}(u, v) > f_{max} \\ 1.0 & \text{sinon} \end{cases} \quad (1)$$

où  $\tilde{f}(u, v)$  est la fréquence spatiale radiale exprimée en cycles/degré et  $f_{max}$  (égale à 8 cycles/degré) est la fréquence à laquelle l'exponentielle est maximale.

Ce modèle dépend, par le biais de la fréquence spatiale radiale, de l'orientation, de la résolution de l'image et de la distance d'observation. On suppose qu'une image de  $256 \times 256$  pixels est affichée sur une zone de taille de 6,5 cm  $\times$  6,5 cm et que la distance d'observation est égale à quatre fois la hauteur de l'image, on obtient la matrice  $W$  de pondération :

$$W = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.9599 & 0.8746 & 0.7684 & 0.6571 \\ 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.9283 & 0.8404 & 0.7371 & 0.6306 \\ 1.0000 & 1.0000 & 0.9571 & 0.8898 & 0.8192 & 0.7371 & 0.6471 & 0.5558 \\ 1.0000 & 1.0000 & 0.8898 & 0.7617 & 0.6669 & 0.5912 & 0.5196 & 0.4495 \\ 0.9599 & 0.9283 & 0.8192 & 0.6669 & 0.5419 & 0.4564 & 0.3930 & 0.3393 \\ 0.8746 & 0.8404 & 0.7371 & 0.5912 & 0.4564 & 0.3598 & 0.2948 & 0.2480 \\ 0.7684 & 0.7371 & 0.6471 & 0.5196 & 0.3930 & 0.2348 & 0.2278 & 0.1828 \\ 0.6571 & 0.6306 & 0.5558 & 0.4495 & 0.3393 & 0.2480 & 0.1828 & 0.1391 \end{bmatrix} \quad (2)$$

Les éléments de la matrice  $W$  pondèrent les coefficients de la DCT des blocs (de taille  $8 \times 8$  pixels) de l'image et indiquent l'importance visuelle des fréquences spatiales. Ces valeurs diminuent en passant des basses fréquences aux hautes fréquences (parcours en zigzag).

## 2.2. les méthodes PSNR et EAM pondérées par le modèle de Daly simplifié

La pondération (2) des mesures PSNR et EAM est appliquée localement sur des blocs de la matrice erreur entre les images (originale et dégradée) transformées. De ce fait, les valeurs des erreurs dans les hautes fréquences vont diminuer car elles sont supposées être moins visibles par le système visuel humain.

*PSNR pondérée (PSNRP) (ou « weighted PSNR ») :*

$$PSNRP = 10 \cdot \log_{10}(Max^2 / EQMP) \quad (3)$$

où  $Max$  est le maximum pic à pic du signal original.

$EQMP$  est la moyenne des valeurs de  $EQM_p$  (erreur quadratique moyenne pondérée localement sur des blocs de  $8 \times 8$  pixels).  $EQM_p$  est donnée par :

$$EQM_p(I, \hat{I}) = \frac{1}{N^2} \sum_{u=1}^N \sum_{v=1}^N W_{u,v} (I(u,v) - \hat{I}(u,v))^2 \quad (4)$$

$I$  et  $\hat{I}$  représentent les images originale et dégradée respectivement dans le domaine DCT.

$W_{u,v}$  (2) représente le poids de la pondération à la fréquence spatiale  $(u, v)$ .

$N$  (égale à 8) représente la taille de la matrice sur laquelle est appliquée la pondération  $W$ .

De la même manière, la mesure  $EAM$  pondérée est la moyenne des valeurs  $EAM$  pondérées localement ( $EAM_p$ ) par la matrice  $W$  :

*EAM pondérée (EAMP) (ou « weighted EAM ») :*

L' $EAM$  pondérée est la moyenne des valeurs de  $EAM_p$  sur toute l'image :

$$EAM_p(I, \hat{I}) = \frac{1}{N^2} \sum_{u=1}^N \sum_{v=1}^N W_{u,v} |I(u,v) - \hat{I}(u,v)| \quad (5)$$

Prenons un exemple de la mesure PSNRP : l'image Mandrill dégradée de la figure 8 a une valeur de PSNRP de 20.98dB et une valeur de PSNR de 18.54dB. Certaines dégradations se trouvent dans les hautes fréquences, mais ne sont pas très visibles. La mesure PSNRP reflète mieux la qualité de cette image par rapport à la mesure PSNR (PSNRP – PSNR = 2.4 dB).

En même temps, parfois la mesure PSNRP ne reflète pas bien la qualité entre images dégradées. En effet, l'image de la figure 3 est visuellement dégradée par rapport à celle de la figure 6, mais les mesures PSNRP donnent 26.41 dB (figure 3) et 25.53 dB (figure 6). Ceci est dû au fait que le flou apparaît très gênant dans les hautes fréquences, alors qu'il est masqué par la pondération du modèle de Daly.

## 3. méthode de Fränti

La méthode de Fränti intègre trois aspects différents : les propriétés du SVH, les facteurs de qualité (erreur de contraste, erreur de structure et erreur de quantification en niveaux de gris) et la pondération des différents facteurs. Les images sont découpées en blocs de  $3 \times 3$  pixels auxquels on applique chaque critère :

$$\text{Critère de contraste : } d_1 = \frac{(\sigma_A - \sigma_B)^2}{\max(1, \sigma_A)} \quad (6)$$

où  $A$  désigne l'image originale,  $B$  l'image distordue et  $\sigma$  la déviation standard.

$d_1$  permet de mesurer l'erreur de contraste entre  $A$  et  $B$ .

$$\text{Critère de structure : } d_2 = \frac{|Gx_A - Gx_B| + |Gy_A - Gy_B|}{2\max(1, \sigma_A)} \quad (7)$$

où  $Gx$  et  $Gy$  sont les réponses horizontales et verticales aux détecteurs de contours.

La normalisation par la valeur de  $\sigma$  du bloc original exprime le masquage des erreurs par le contraste.

$$\text{Critère de quantification : } d_3 = (Q_A - Q_B)^2 \quad (8)$$

où  $Q_A$  et  $Q_B$  sont les nombres de niveaux de gris distincts.

La mesure finale des dégradations est :

$$D = w_1 \cdot f(D_1) + w_2 \cdot f(D_2) + w_3 \cdot f(D_3) \quad (9)$$

avec la fonction  $f$  définie par :

$$f(D_i) = 1 - \min\left(1, \frac{D_i}{k_i}\right) \quad (10)$$

où  $D_i$  est la moyenne des  $d_i$  et ( $k_1 = 3, k_2 = 32, k_3 = 32$ ) sont des paramètres obtenus par Fränti en maximisant la corrélation de la mesure objective ( $D$ ) avec la mesure subjective visuelle sur l'image Lena à travers plusieurs dégradations.

Prenons un exemple de cette mesure sur les images des figures 7 et 8. Les structures de l'image de la figure 8 sont mieux sauvegardées par rapport à celles de la figure 7. Les notes de la mesure reflètent bien cette différence (0.41 pour la figure 8 et 0.28 pour la figure 7), alors que les valeurs PSNR sont proches (18,00 dB et 18,54 dB).

## 4. mesure de dégradations basée sur la densité de pertinence

Il est très fréquent de rencontrer des erreurs concentrées, dans une région de l'image, qui apparaissent pertinentes et gênantes. Nous avons constaté que si des erreurs ponctuelles, même peu importantes, se concentrent, elles deviennent de plus en plus perceptibles pour un observateur. Ces mêmes erreurs apparaîtront non gênantes si elles ne sont pas concentrées. A partir d'un certain degré de concentration elles sont perçues comme une seule masse erronée. Nous avons donc été amenés à nous intéresser à la localité de l'erreur.

Nous introduisons ici une nouvelle approche de mesure des dégradations basée sur une mesure standard pondérée par un critère de densité d'erreur [21][22] que nous appellerons « *critère de densité de pertinence* ». Ce critère mesure la concentration des erreurs dans des zones de l'image et combine trois types d'erreur : l'erreur de contraste, l'erreur de structure et l'erreur de quantification [9]. Le paragraphe ci-dessous montrera comment cette nouvelle mesure a été construite.

### 4.1. construction de la mesure

Puisque dans [9] l'erreur est calculée sur des blocs de taille  $3 \times 3$  pixels, le regroupement des dégradations se fera en accord avec ces blocs. La question qui se pose est de savoir quel serait le degré de ce regroupement. En principe, l'image dégradée devrait être décomposée en régions qui reflètent la même impression sur les erreurs pour l'œil humain. Pour une simplicité de calcul, nous avons choisi un regroupement simple sur des régions carrées que nous appelons « fenêtres d'analyse ».

Les fenêtres d'analyse, dont la taille peut varier de  $3 \times 3$  blocs à la taille de l'image, nous permettent de jouer sur le niveau de détail que nous voulons obtenir. En effet, un détail trop fin n'est pas toujours bon à retenir parce que s'il se situe dans une zone

de forte incertitude ; l'œil humain le percevra mal. Nous obtiendrons donc une mauvaise note alors que l'image globale ne serait pas forcément mauvaise. Nous avons donc choisi d'effectuer les tests avec différentes tailles afin de déterminer plus tard celle qui apparaîtra la plus appropriée.

Dans notre approche de mesure, nous avons choisi une stratégie qui remédie aux défauts des mesures standards en les pondérant par un facteur lié à la densité de pertinence.

*Description de la nouvelle mesure [Schéma 1] :*

Dans une fenêtre d'analyse, une valeur de la dégradation locale est mesurée à l'aide d'une méthode standard choisie. Cette valeur est pondérée par un facteur lié à la densité de pertinence calculée dans cette fenêtre. Les fenêtres analysantes doivent recouvrir toute l'image de manière glissante ou non. Pour obtenir une seule note finale de mesure de dégradations, nous cumule les valeurs pondérées en une seule valeur en utilisant une fonction de cumul.

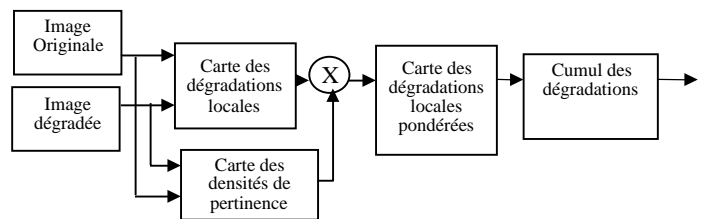


Schéma 1. – Schéma général de la mesure des dégradations basée sur le critère de la densité de pertinence locale.

### 4.2. le critère de la densité de pertinence locale

Le critère de la densité de pertinence locale se base sur le degré de pertinence d'erreur sur les blocs de  $3 \times 3$  pixels à l'intérieur d'une fenêtre d'analyse.

La densité de pertinence, dans la fenêtre d'analyse  $i$ , est définie comme suit :

$$D^i = \frac{N_p^i}{N} \quad (11)$$

$N_p^i$  est le nombre de blocs pertinents dans la fenêtre  $i$ .

$N$  est le nombre total de blocs dans cette fenêtre.

Nous définissons un bloc comme pertinent, resp. non pertinent, si la moyenne de ses erreurs de contraste (6), de structure (7) et de quantification (8) est supérieure, resp. inférieure, à un seuil  $S$  préfixé.

De ce fait, on a :  $0 \leq D^i \leq 1$  et la densité de non-pertinence est  $1 - D^i$ .

### 4.3. mesure de dégradations basée sur la densité de pertinence

La mesure des dégradations dans la fenêtre d'analyse  $i$  est donnée par :

$$DP_i = M_i(1 - \alpha D^i) \quad (12)$$

où  $\alpha$  est un paramètre d'adaptation constant identique pour toutes les fenêtres d'analyse.

$M_i$  est la valeur de la mesure standard de dégradation sur la fenêtre d'analyse  $i$ .

Nous avons choisi pour les tests la mesure EAM comme mesure standard de dégradations locales ( $M_i = EAM_i$ ). Ce choix est motivé par le fait que cette mesure est montrée performante dans notre analyse (en terme de la corrélation avec la mesure subjective) en la comparant aux autres mesures standards. Nous appellerons la mesure obtenue : EAM – DEL.

En analysant cette mesure EAM, deux cas se sont présentés :

- 1) Parfois, la mesure EAM augmente rapidement en passant d'une image peu dégradée à une autre plus dégradée (ce qui implique l'augmentation de la densité de pertinence), alors que la qualité subjective diminue moins rapidement. Dans cette situation, le facteur de pondération devrait réduire la vitesse de l'augmentation de la mesure EAM en diminuant sa valeur.
- 2) Parfois, la mesure EAM augmente légèrement en passant d'une image peu dégradée à une autre plus dégradée, tandis que la qualité subjective augmente rapidement. Dans ce cas, le facteur de pondération devrait accélérer la vitesse de l'augmentation de la mesure EAM en augmentant sa valeur.

Le premier cas correspond à une valeur de  $\alpha$  positive et le deuxième cas à une valeur de  $\alpha$  négative.

### 4.4. fonction de cumul

Plusieurs fonctions de cumul d'erreurs locales sont possibles : la moyenne des carrées, la moyenne des valeurs absolues, le minimum ou le maximum, et en général la métrique de Minkowski. Une étude sur plusieurs fonctions de cumul des erreurs est faite dans [14]. Considérant qu'un défaut, même minime, gâche toute l'image, nous avons convenu de retenir comme mesure finale la mesure de la zone la plus mauvaise :

$$Note\_finale = \text{Max}_i(DP_i) \quad (13)$$

### 4.5. optimisation des paramètres de la méthode

Notre méthode utilise trois paramètres, à savoir : le seuil  $S$  de détection des blocs pertinents, le paramètre d'adaptation  $\alpha$  et la

taille  $T$  des fenêtres d'analyse. Soit  $\rho(S, \alpha, T)$  le coefficient de corrélation entre les mesures objective et subjective pour les paramètres  $S$ ,  $\alpha$ , et  $T$  fixés. La valeur de  $\rho$  est toujours négative puisque notre méthode donne des valeurs élevées quand le MOS est petit. L'algorithme d'optimisation de ces paramètres est le suivant :

*Algorithme d'optimisation :*

**Pour**  $T = 3 \times 3$  **jusqu'à** *Taille maximale*

**Recherche du**  $[S_T, \alpha_T] = \underset{S, \alpha}{\text{argmax}} |\rho(S, \alpha, T)|$  sous la

contrainte de  $S > 0$

**Calcul de**  $\rho_T = |\rho(S_T, \alpha_T, T)|$

**Test du** maximum de  $\rho_T$

**Fin**

Soit  $T = T_0$  la valeur de la taille qui réalise le maximum de  $\rho_T$ , alors les paramètres optimisés sont  $(S_{T_0}, \alpha_{T_0}, T_0)$ .

La maximisation de  $|\rho(S, \alpha, T)|$ , sous la contrainte de  $S > 0$ , est faite par la méthode d'optimisation de Lagrange. Pour réaliser cette dernière, nous avons utilisé une fonction numérique sous Matlab (« constr.m ») qui donne la solution optimale en  $S$  et  $\alpha$  pour une taille  $T$  fixée. La formulation de l'optimisation est de la forme :

$$[S_T, \alpha_T] = \underset{\text{sous } -S \leq 0}{\text{argmin}} [\rho(S, \alpha, T)]$$

## 5. évaluation de la qualité subjective des images test

Les évaluations de la qualité subjective sont faites avec trois images de test : Lena, Barbara et Mandrill [Figure 2]. Ces images sont dégradées en utilisant quatre codeurs différents. Sur chaque image nous appliquons les quatre codeurs pour obtenir 10 versions d'images dégradées. Les codeurs utilisés sont JPEG2000 (Joint Picture Expert Group 2000, [3][31]), Fractal [30][32], BTC (Block Truncate Coding, [13]) et QT (QuadTree, [20]). Les 10 versions d'images sont obtenues en faisant quatre dégradations par JPEG2000, quatre par Fractal, une par BTC et une par QT. Nous obtenons au total 30 différentes images dégradées pour les évaluations subjectives.

En ce qui concerne le nombre d'images de test, nous nous sommes alignés sur les travaux de Fränti [9] qui a utilisé 42 images et 5 codeurs. Dans nos travaux antérieurs [21], nous avons testé notre méthode sur une autre base de 19 images avec deux autres types de dégradations (dégradations ponctuelles et dégradations de JPEG).

La figure 3 montre l'image Barbara dégradée par le codage JPEG 2000 avec un taux de compression égal à 40. La figure 4 montre l'image Barbara dégradée par le codage fractal avec un taux de compression égal à 20. Les figures 5 et 6 montrent l'image Barbara dégradée par les codeurs BTC et QT.

### 5.1. conditions d'évaluation

Les conditions d'évaluation de la qualité subjective sont normalisées par les recommandations du CCIR [1]. Cependant, ces recommandations sont, à l'origine, désignées pour l'image télévisuelle sans tenir compte de la mesure des dégradations avec l'image de référence (image originale). Notre but est de mesurer les dégradations entre deux images au lieu de la qualité d'une seule image : nous supposons de plus que les images peuvent être éditées, zoomées et observées à la plus faible distance possible. De ce fait, nous nous sommes référés aux conditions d'évaluation qui ont été utilisées par Fränti [9] où les recommandations du CCIR sont respectées partiellement [Table1] :

Table 1. – Conditions d'évaluation subjective de la qualité.

Qualité d'image	Images photographiques
Conditions d'observation	Environnement de bureau normal
Distance d'observation	A la libre appréciation de l'observateur
Temps d'observation	Illimité
Nombre d'observateurs	Entre 15 et 39
Echelle	L'échelle du MOS étendu (entre 0 et 10)

Pour notre part, les images de test sont affichées, sur un écran de PC, avec des dimensions de 6.5 cm × 6.5 cm (≈ 256×256 pixels). L'échelle de note retenue est le MOS (Mean Opinion Score) étendu de 0 à 10 (2 : dégradation très gênante, 4 : gênante, 6 : un peu gênante, 8 : perceptible, mais pas gênante, 10 : imperceptible) et est normalisée entre 0 et 1. Nous avons aussi accordé la possibilité aux observateurs de donner des notes par demi-valeurs. Pour calculer la mesure subjective finale d'une image dégradée, nous faisons la moyenne des notes données par 17 observateurs.

La mesure subjective sert seulement, dans cet article, pour évaluer la mesure objective développée en terme du coefficient de corrélation qui sera décrit dans le paragraphe suivant.

### 5.2. le coefficient de corrélation de la mesure objective

Pour mesurer la corrélation d'une mesure objective (notée par le vecteur  $X$ ) avec la mesure subjective (notée par le vecteur

observation  $Y$ ), nous utilisons le coefficient de corrélation qui est défini par :

$$\rho(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (14)$$

où  $(X_i)$  et  $(Y_i)$ ,  $i = 1$  à  $n$ , sont les composantes des vecteurs  $X$  et  $Y$  respectivement.  $n$  représente le nombre de valeurs utilisées dans la mesure. Par la suite  $n$  prendra la valeur 10 ou 30 suivant que la corrélation est utilisée à travers un seul type d'image ( $n = 10$ ) ou toutes les images ( $n = 30$ ).  $\bar{X}$  et  $\bar{Y}$  représentent, respectivement, les valeurs moyennes des vecteurs  $X$  et  $Y$ .

## 6. remise à l'échelle de la mesure objective

Lors de notre analyse des résultats, nous avons remarqué que le coefficient de corrélation à travers toutes les images ( $n = 30$  dans (14)) est généralement faible pour une mesure donnée, alors qu'il est élevé à travers chaque image ( $n = 10$ ). Ce problème est essentiellement dû à la mauvaise adaptation de la mesure objective aux différents types d'image à la fois. Dans la figure 1 nous montrons un exemple de ce problème : la mesure  $M$  est très bien corrélée avec la mesure MOS pour chaque image, mais elle ne l'est pas à travers les deux images. Pour y remédier, une simple translation est faite sur la mesure  $M$  pour obtenir la mesure  $R$  [Figure 1, (b)].

En général, pour résoudre ce problème, nous procédons manuellement à une remise à l'échelle du MOS des mesures objectives. Elle est décrite comme suit :

Soient  $M_I$  et  $MOS_I$ , respectivement, les vecteurs mesures objective et subjective d'une image  $I$ . Remettre à l'échelle la mesure  $M_I$ , c'est trouver un vecteur  $R_I = aM_I + b$  ( $a$  et  $b$  sont des constantes à déterminer) qui minimise la norme Euclidienne  $\|R_I - MOS_I\|$ . On a :  $|\rho(R_I, MOS_I)| = |\rho(M_I, MOS_I)|$ .

Dans l'exemple de la figure 1, on a :

$$|\rho(R_{I1}, MOS_{I1})| = |\rho(M_{I1}, MOS_{I1})|,$$

$$|\rho(R_{I2}, MOS_{I2})| = |\rho(M_{I2}, MOS_{I2})|,$$

et

$$\begin{aligned} |\rho([R_{I1} \ R_{I2}], [MOS_{I1} \ MOS_{I2}])| \\ \gg |\rho([M_{I1} \ M_{I2}], [MOS_{I1} \ MOS_{I2}])| \end{aligned}$$

Pour chaque image, on identifie les paramètres  $a$  et  $b$  de la remise à l'échelle. Ces paramètres dépendent de l'image et sont calculés pour chaque méthode de mesure.

Comme exemple de remise à l'échelle, nous prenons les images Barbara et Mandrill des figures 3 et 8. La figure 3 (PSNR = 25.36 dB) est légèrement dégradée par rapport à la figure 8 (PSNR = 18.54 dB) et le décalage des valeurs des PSNR est de 6.82dB. Après remise à l'échelle de la mesure PSNR sur les deux images, nous obtenons la valeur 0.44 pour la figure 3 et 0.40 pour la figure 8, ce qui montre bien l'importance de cette remise à l'échelle.

Par la suite, la comparaison des mesures de dégradations se fera en terme des coefficients de corrélation des mesures objectives remises à l'échelle du MOS.

## 7. résultats

Nous avons procédé à la mesure de 30 images dégradées en utilisant une fenêtre d'analyse non glissante.

Après avoir optimisé le coefficient de corrélation sur chaque image, nous obtenons les valeurs suivantes des paramètres  $S$ ,  $\alpha$  et **Taille** :

Lena :  $\alpha = 0$  et **Taille = 8x8 blocs** (i.e.  $24 \times 24$  pixels)

Barbara :  $S = 1.04$ ,  $\alpha = 0.57$  et **Taille = 8x8 blocs**

Mandrill :  $S = 1.00$ ,  $\alpha = -0.48$  et **Taille = Taille de toute l'image**

$\alpha = 0$  correspond à la mesure EAM sans pondération.

Avant de faire les comparaisons, les valeurs de la mesure objective de chaque image sont remises à l'échelle du MOS dans l'intervalle [0, 1].

La performance de la mesure proposée est testée en comparant ses valeurs avec les évaluations subjectives. Les résultats des autres méthodes (EAM, « EAM pondérée », « PSNR pondérée », et la méthode de Fränti) sont aussi inclus dans les comparaisons. La méthode EAM-DEL est bien corrélée avec la qualité subjective à travers chaque image et à travers toutes les images à la fois (voir [Tableau 2]). Les autres méthodes présentent des coefficients de corrélation moins bons pour les images Barbara et Mandrill (sauf la mesure de Fränti sur Mandrill où le coefficient de corrélation est élevé (0.98)).

Nous remarquons que la pondération des mesures PSNR et EAM par le modèle de Daly (1) améliore les coefficients de corrélation de manière significative par rapport au cas sans pondération [Tableau 2].

Nous remarquons aussi que la mesure PSNR pondérée ou non par le modèle de Daly représente les plus bas coefficients de corrélation à travers chaque image et à travers toutes les images. La figure 9 présente les valeurs de la mesure proposée de l'image Barbara en comparaison avec celles de la mesure « PSNR pon-

Tableau 2. – Les coefficients de corrélation des méthodes de mesure des dégradations.

Mesure	Lena	Barbara	Mandrill	Coefficient de corrélation à travers toutes les images
PSNR	0.95	0.87	0.91	0.91
PSNR pondérée (Daly)	0.97	0.90	0.94	0.94
EAM	<b>0.99</b>	0.91	0.94	0.94
EAM pondérée (Daly)	<b>0.99</b>	0.94	0.94	0.96
Fränti	0.98	0.91	<b>0.98</b>	0.95
<b>EAM-DEL</b>	<b>0.99</b>	<b>0.98</b>	0.96	<b>0.97</b>

dérée ». Les valeurs de la nouvelle mesure se corrélaient bien aux valeurs du MOS, tandis que les valeurs de la mesure « PSNR pondérée » le sont moins. En d'autres termes, cette dernière mesure donne d'une part, des valeurs élevées (par rapport au MOS) pour les images dégradées par JPEG 2000 à forts taux de compression (images 3 et 4) et d'autre part une valeur faible pour l'image dégradée par QT (image 10). Spécialement, les effets de blocs dans le codage QT et le flou dans JPEG2000 semblent être problématiques. La mesure proposée donne aussi des valeurs légèrement élevées pour les images 3 et 4, mais moins élevées par rapport à la mesure « PSNR pondérées ».

Les paramètres de la méthode proposée sont optimisés pour une image en maximisant le coefficient de corrélation de la mesure objective avec la mesure subjective. Pour déterminer le degré de fiabilité de réutilisation des paramètres, nous avons fait des tests de calcul des coefficients de corrélation en réutilisant les paramètres d'une image sur une image différente. Les tests sont réalisés seulement sur la base d'images {Lena, Barbara et Mandrill}. Les résultats obtenus sont indiqués dans le Tableau 3. Le degré de fiabilité de réutilisation des paramètres est élevé entre les images Lena et Mandrill (décalage du coefficient de corrélation de  $0.99 - 0.98 = 0.01$  ou de  $0.96 - 0.94 = 0.02$ ). Nous constatons que le degré de fiabilité baisse dans le cas de réutilisation des paramètres de Mandrill pour Barbara (décalage du coefficient de corrélation de  $0.97 - 0.85 = 0.12$ ) et de Lena pour Barbara (décalage du coefficient de corrélation de  $0.97 - 0.91 = 0.06$ ). Ces résultats préliminaires indiquent que le type de l'image Barbara est différent de celui des autres images et que les types des images Lena et Mandrill sont proches en terme de réutilisation des paramètres de la méthode proposée.

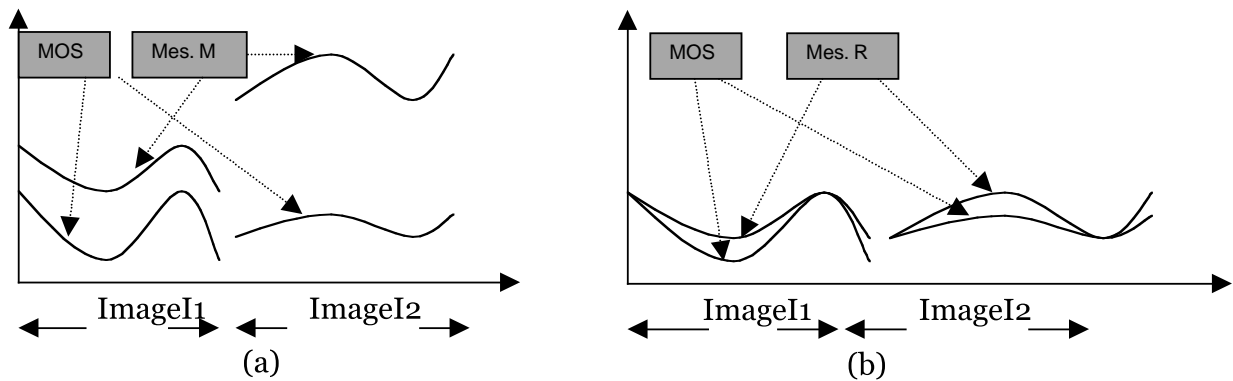
Nous n'avons pas testé la fiabilité des paramètres de la méthode pour d'autres types d'image telles que les images satellitaires ou

médicales. Nous pensons que dans ce cas les informations pertinentes sont différentes et que les dégradations sont perçues différemment.

**Tableau 3.** – Les coefficients de corrélation de la mesure proposée EAM-DEL dans le cas de réutilisation des paramètres d’une image pour les autres images.

Image	Paramètres de Lena	Paramètres de Barbara	Paramètres de Mandrill
Lena	<b>0.99</b>	0.96	0.98
Barbara	0.91	<b>0.97</b>	0.85
Mandrill	0.94	0.94	<b>0.96</b>
À travers toutes les images	0.95	<b>0.96</b>	0.93

La méthode proposée est simple à implémenter. Sa complexité de calcul dépend fortement du calcul des trois critères de structure, de contraste et de quantification. Ces critères sont simples à calculer mais ils ne peuvent pas être implémentés en temps réel (exemple pour une optimisation de paramètres de codeurs qui fonctionnent en temps réel). Les méthodes PSNR et EAM pondérées sont rapides, mais elles sont moins fiables par rapport à la méthode proposée.



**Figure 1.** – Mesures des images I1 et I2 avant (a) et après (b) la remise à l'échelle.



**Figure 2.** – Images originales (Lena-Barbara-Mandrill).

## 8. conclusion et perspectives

Dans cet article, nous avons développé une nouvelle méthode de mesure des dégradations basée sur la pondération de la mesure EAM par la densité de pertinence d'erreur locale. En comparaison avec les mesures standards (PSNR et EAM) pondérées par le modèle de Daly et la mesure de Fränti, cette nouvelle méthode offre de meilleurs coefficients de corrélation avec la mesure subjective.

En perspective, nous pensons développer un schéma de codage basé sur le critère de la densité de pertinence et étudier le problème de la réutilisation des paramètres de la méthode développée pour d'autres types d'image.

## 9. remerciements

Nous tenons à remercier l'équipe TEMICS à l'IRISA/INRIA de Rennes, pour nous avoir permis de développer une partie de nos recherches au sein de son laboratoire dans le cadre de l'action intégrée inter-universitaire Franco-Marocaine AI 74/SI/97.

Une partie du travail a bénéficié du support du programme marocain de recherche PROTARS n° P4T2/07.





Figure 3. – Image Barbara dégradée par JPEG2000 (PSNR = 25.63 dB).



Figure 4. – Image Barbara dégradée par Fractal (PSNR = 23.36 dB).



Figure 5. – Image Barbara dégradée par BTC (PSNR = 28.19 dB).



Figure 6. – Image Barbara dégradée par QT (PSNR = 24.03 dB).

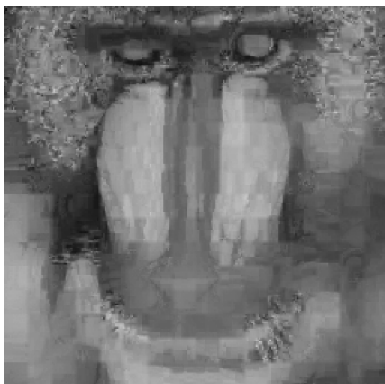


Figure 7. – Image Mandrill dégradée par Fractal (PSNR = 18.00 dB).



Figure 8. – Image Mandrill dégradée par JPEG200 (PSNR = 18,54 dB).

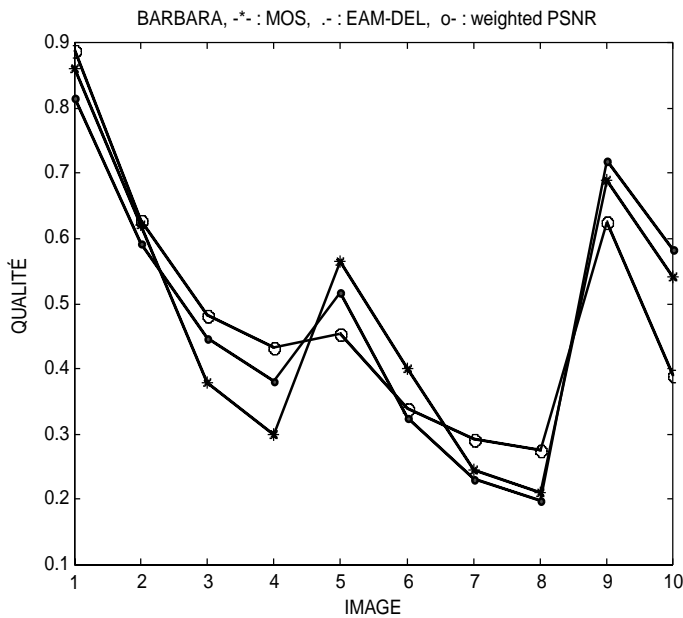


Figure 9. – Les valeurs de la mesure proposée et de la mesure « PSNR pondérée ».

**BIBLIOGRAPHIE**

[1] CCIR, «Rec. 500-2, Method for the subjective assessment of the quality of television picture», Recommendations et Rapports du CCIR, Genève, 1982.

[2] K. BELKACEM-BOUSSAID and A. BEGHADADI, *A new image smoothing method based on a simple model of spatial processing in the early stages of human vision*, IEEE Trans. On Image Processing, Vol. 9, NO. 2, February 2000.

[3] C. CHRISTOPOULOS, A. SKODRAS and T. EBRAHIMI, *The JPEG2000 still image coding system: An overview*, IEEE Transactions on Consumer Electronics, Vol. 46, No 4, pp. 1103-1127, November 2000.

[4] P. C. COSMAN, R. M. GRAY and R. O. OLSHEN, *Evaluating quality of compressed medical images: SNR, Subjective Rating, and Diagnostic Accuracy*, Proceedings of the IEEE, Vol. 82, No. 6, June 1994.

[5] S. DALY, *The visible differences predictor: An algorithm for the assessment of image fidelity*, In « Digital images and human vision », edited by Andrew B. Watson, A Bradford Book, The MIT Press, Cambridge, 1993.

[6] S. DALY, *Subroutine for the generation of a two dimensional human visual contrast sensitivity function*, Technical Report, Eastman Kodak, 1987.

[7] A. M. ESKICIOGLU and P. S. FISHER, *Image quality measures and their performance*, IEEE Trans. on Comm., Vol. 43, No. 12, December 1995.

[8] T. EUDE, *Quality assessment of still monochrome images*, ICISP'01, International Conference on Image and Signal Processing, May 3-5, Agadir, 2001.

[9] P. FRÁNTL, *Blockwise distortion measure for statistical and structural errors in digital images*, Signal Processing: Image communication 13, pp89-98, 1998.

[10] S. HEMAMI, *Supra-threshold wavelet coefficient quantization in natural images : analysis and application to compression*, ICISP'01, International Conference on Image and Signal Processing, May 3-5, Agadir, 2001.

[11] M. JUNG, D. LEGER, M. GAZALET, *Méthode d'évaluation univariante*

*de la qualité des images. Application à des images comprimées selon la norme JPEG*, Revue de Traitement du signal, Volume 17 numéro 2, 2000

[12] S. A. KARUNASEKERA and N. G. KINGSBURY, *A distortion measure for blocking artifacts in image based on human visual sensitivity*, IEEE Transactions on Image Processing, Vol. 4, No. 6, June 1995.

[13] Chun-Hung KUO, Chang-Fuu CHEN and Wen-Nan HSIA, *A compression algorithm based on classified Interpolative block truncation coding and vector quantization*, Journal of Information Science and Engineering 15, pp. 1-9, 1999.

[14] P. Le CALLET et D. BARBA, *Critères objectifs de qualité visuelle : le passage d'erreurs locales a une appréciation globale de qualité*, CORE-SA 2000, Octobre 2000.

[15] A. MAYACHE, T. EUDE and H. CHERIFI, *A comparison of image quality models and metrics based on human visual sensitivity*, ICIP'98, October 4-7, 1998 Chicago, Illinois, USA.

[16] M. MIYAHARA, K. KOTANI and V.R. ALGAZI, *Objective picture quality scale (PQS) for image coding*, IEEE Transactions on Communications, 1998

[17] N. B. NILL, *A visual model weighted cosine transform for image compression and quality assessment*, IEEE Transactions on Communications, Vol. COM-33, No. 6, June 1985.

[18] S. OLSSON, M. STROPPIANA, J. BAĪNA, *Objective methods for assessment of video quality: state of the art*, IEEE Transaction on Broadcasting, Vol. 43, No 4, December 1997.

[19] A. SAADANE, H. SENANE and D. BARBA, *Visual coding: design of psychovisual quantizers*, Journal of Visual Communication and Image Representation, Vol. 9, No. 4, December 1998, pp. 381-391.

[20] H. SAMET, *The Quadtree and related hierarchical data structures*, ACM Compt. Surv., Vol. 16, pp. 188-260, June 1984.

[21] A. TAMTAOUI, D. ABOUTAJDINE, *Elaboration de critères de mesures de qualité objective pour la qualité subjective*, GRETSI'99, Septembre 1999, Vannes, France

[22] A. TAMTAOUI, D. ABOUTAJDINE, *A new objective distortion measure for compressed image quality*, ISIVC'00, Rabat, Maroc, 2000

[23] K.T. TAN and GHANBARI, *Measuring blocking artefacts using harmonic analysis*, Electronic letters, Vol.35, No. 16, August 1999.

[24] T. TIFFANY et S. HAKIM, *L'apport d'un bloc de segmentation d'erreur dans l'évaluation de la qualité d'images*, GRETSI'01, 2001, Toulouse, France.

[25] N. DAMERA-VENKATA, *Image Quality assessment based on a degradation model*, IEEE Transaction on Image Processing, Vol. 9 No. 4, April 2000.

[26] Ching-Yang WANG, Shih-Ming LEE, Long-Wen CHANG, *Designing JPEG quantization tables based on human visual system*, Signal Processing: Image Communication 16, pp. 501-506, 2001.

[27] A. B. WATSON, G. Y. YANG, J. A. SOLOMON, and J. VILLASENOR, *Visual thresholds for wavelet quantization error*, Proc. SPIE, pp. 382-392, 1996.

[28] A. B. WATSON, *DCTune: A technique for visual optimization of DCT quantization matrices for individual images*, Society for Information Display Digest of Technical Papers XXIV, pp. 946-949, 1993.

[29] *Digital images and human vision*, edited by Andrew B. Watson, A Bradford Book, The MIT Press, Cambridge, 1993.

[30] *Fractal image compression-theory and application*, Y. Fisher, Springer-Verlag edition, New York, 1994.

[31] <http://www.imagepower.com/products/jp2bugs.htm>

[32] <http://inls.ucsd.edu/y/Fractals/Mars-1.0.tar.gz>

Manuscrit reçu le 8 juillet 2002

### LES AUTEURS

#### Ahmed TAMAOUI



Ahmed TAMAOUI est titulaire d'un doctorat en Traitement du Signal et Télécommunications de l'Université de Rennes I, France, en 1992. Depuis Janvier 1995 il est enseignant chercheur au sein de l'Institut National des Postes et télécommunications (INPT) de Rabat, Maroc. Actuellement, il est responsable du laboratoire Traitement d'Images à l'INPT. Ses principales recherches portent sur les méthodes d'estimation de la disparité pour la reconstruction 3D, l'analyse des mesures de la qualité subjective des images dégradées par les schémas de compression, et le codage conjoint source-canal des images et de la vidéo.

#### Driss ABOUTAJDINE



Driss ABOUTAJDINE a obtenu le diplôme de 3ème cycle et le doctorat d'état en traitement du signal et de l'image de l'université Mohammed V en 1980 et 1985 respectivement. Il est actuellement professeur à la Faculté des Sciences de Rabat où il anime le groupe de recherche GSCM « Signaux, Communications et Multimédia » et l'UFR SYSCOM « Systèmes et communications ». Depuis 2001, il est coordinateur national du pôle de compétences STIC « Sciences et Technologies de l'Information et de la Communication » qui regroupe plus de 30 équipes de recherche universitaires et industrielles. Il est membre d'IEEE depuis 1988 et senior member depuis 1999. Il a organisé le colloque international ISIVC'2000 en avril 2000 à Rabat. Il a contribué à plus de 100 papiers dans des conférences et journaux nationaux et internationaux. Ses domaines d'intérêt comprennent, le traitement de l'information multimédia et ses applications en télécommunications.