

Mesure de qualité d'images couleur par combinaison de classifieurs

Christophe Charrier, Olivier Lezoray, Gilles Lebrun

▶ **To cite this version:**

Christophe Charrier, Olivier Lezoray, Gilles Lebrun. Mesure de qualité d'images couleur par combinaison de classifieurs. Gretsi 2009, 2009, Dijon, France. 4 pp. hal-00403944

HAL Id: hal-00403944
https://hal.science/hal-00403944

Submitted on 21 Jan 2014

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.

Mesure de qualité d'images couleur par combinaison de classifieurs

Christophe CHARRIER, Olivier LÉZORAY, Gilles LEBRUN

Université de Caen-Basse Normandie, GREYC UMR CNRS 6072, ENSICAEN
Équipe Image, 6 Bd. Maréchal Juin, F-14050 Caen, France
christophe.charrier@unicaen.fr,olivier.lezoray@unicaen.fr
gilles.lebrun@unicaen.fr

Résumé – La mesure de la qualité des images est nécessaire dès lors qu'une image subit une dégradation afin de pouvoir quantifier le degré de gène provoquée par cette dernière. Jusqu'à maintenant, toutes les métriques développées fournissent au final une note scalaire de qualité qui doit être en très forte adéquation avec le résultat obtenu en questionnant un être humain. Cependant, le jugement humain ne se résume pas à une simple attribution d'une note de qualité mais opère plutôt par comparaison pour au final aboutir à une classification de la qualité de l'image. La décision est alors prise en fonction de plusieurs échelles internes.

Dans cet article, une méthode de mesure de la qualité d'une image compressée, basée sur un schéma de classification multi-classes est présentée. Au lieu de procéder à un moyennage des critères mesurés, un vecteur multi-dimensionnel est créé et est directement utilisé par le schéma de classification. Les classes finales de qualité utilisées sont celles définies dans la norme ITU 500-11.

Abstract – A crucial step in image compression is the evaluation of its performance, and more precisely the available way to measure the final quality of the compressed image. In this paper, a machine learning expert, providing a final class number is designed. The quality measure is based on a learned classification process in order to respect the one of human observers. Instead of computing a final note, our method classifies the quality using the quality scale recommended by the ITU 500-11. This quality scale contains 5 ranks ordered from 1 (the worst quality) to 5 (the best quality). This was done constructing a vector containing many visual attributes. Finally, the final features vector contains 7 attributes.

1 Introduction

Les schémas de compression avec perte d'information des images couleurs induisent indubitablement des dégradations dont la visibilité est plus ou moins élevée selon leur importance et leur voisinage proche (effet de masquage, ou, son dual, effet de facilitation). Cette évaluation de la qualité peut être basée sur l'observation humaine. Les résultats ainsi obtenus, permettent de déterminer des seuils de sensibilité de l'observateur humain. L'un des inconvénients majeurs de ce type de mesure, est une mise en œuvre particulièrement longue ainsi que l'utilisation d'un protocole contraignant [1]. L'idéal serait alors d'avoir une mesure automatique de qualité, plus rapide, et, *a priori*, sans intervention de l'être humain.

De nombreux travaux ont été réalisés au cours des deux dernières décennies pour développer une mesure de qualité fiable et robuste.

Parmi les plus usités, on retrouve la mesure du PSNR, de l'Erreur quadratique Moyenne (EQM), etc. Dernièrement, à mesure que la connaissance du fonctionnement du Système Visuel Humain s'étoffait, des modélisations de certaines caractéristiques ont été intégrées directement dans les métriques afin de "plagier", le mieux possible le modèle biologique. On trouve ainsi la prise en compte des effets de masquage, de la sensibilité au contraste, etc.

Toutes ces métriques délivrent au final une valeur scalaire permettant ainsi de proceder à un ordonnancement des qualités. En revanche, l'un des inconvénients majeurs est l'utilisation de la métrique de Minkowski pour obtenir le scalaire final, indépendamment de la norme utilisée (distance de Manhattan, distance euclidienne, etc.).

Cependant, si l'on s'interesse d'un peu plus près à ce que fait un être humain lorsqu'on lui demande de juger la qualité d'une image (avec ou sans connaissance de l'image originale), force est de constater que ce dernier ne donne pas directement de note à l'image, mais opère plutôt par comparaison pour au final aboutir à une classification de la qualité de l'image. En effet, la décision est prise en fonction de plusieurs échelles internes, ce qui permet de qualifier la démarche globale comme une approche multi-critères qui ne semble pas opérer par un simple moyennage des défauts visualisés.

Dans cet article, une méthode originale de la mesure de la qualité d'une image compressée est présentée. Elle s'appuie sur un schéma de classification multi-classes. Ainsi, au lieu de procéder à un moyennage des critères mesurés, un vecteur multi-dimensionnel est créé et est directement utilisé par le schéma de classification. Les classes de qualité utilisées sont celles définies dans la norme ITU 500-11 [1] préconisées lors de la mesure de qualité par observations humaines.

2 Création et manipulation du vecteur de qualité

2.1 Création du vecteur de qualité

Afin de réaliser l'apprentissage de la classification de la qualité des images, il est nécessaire de construire un vecteur d'attributs. Parmi tous les attibuts que l'on peut recenser dans la littérature, nous utiliserons trois critères liés à la structure de l'image [2] (critère de distortion de luminance, de contraste et un critère de comparaison de structure), deux critères colorimétriques permettant de mesurer la distortion de chrominance, ainsi que la dispersion spatio colorimétrique [3], une

mesure des effets de blocs, basée sur une décomposition de Fourier [4], une mesure d'effet de flou basée sur une méthode de détection de contours, ainsi que trois attibuts (énergie, entropie et coefficient d'homogénéité) issus de l'analyse de correlogramme couleur [5].

2.2 Classification du vecteur de qualité

La théorie de l'Apprentissage Statistique de Vapnik et de Chervonenkis [6] a conduit au développement d'une classe d'algorithme connu sous le nom de SVM (Support Vector Machines). Ils permettent de réaliser des estimations en classification (et en régression). Une des originalités de la méthode est de produire une fonction de décision qui n'utilise qu'un sous-ensemble de la base d'apprentissage. Les éléments de ce sous-ensemble sont nommés Vecteurs de Support ($\mathbf{SV}$).

Soit une base d'apprentissage $A = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_k, y_k)\}$ composé de k couples (vecteur d'attributs, label) avec $x_i \in \mathcal{R}^n$ et $y_i \in \{-1, +1\}$. L'algorithme des SVM projette les vecteurs x_i dans un espace de travail $\mathbf{H}$ à partir d'une fonction non linéaire $\phi : \mathcal{R}^n \rightarrow \mathbf{H}$. L'hyperplan optimal de séparation des deux classes dans l'espace $\mathbf{H}$ est ensuite recherché. Cet hyperplan $(\mathbf{w}, b)$ matérialise la frontière de séparation entre les deux classes. La classe y d'un nouvel exemple $\mathbf{x}$ est définie par:

$$y = sign\left(\mathbf{w} \cdot \phi(\mathbf{x}) + b\right) \tag{1}$$

L'hyperplan est optimal s'il maximise la distance qui le sépare des exemples dont il est le plus proche. Cette distance est usuellement appelée marge du classifieur. Il a été démontré [6] que maximiser cette marge correspond à maximiser le "pouvoir" généralisateur du classifieur.

Les SVM étant des classifieurs binaires, la résolution d'un problème multi-classes est effectuée en le transformant en une combinaison de problèmes binaires [7]. Ainsi, dans notre cadre applicatif, cinq fonctions de décision sont utilisées pour réaliser la classification en cinq classes {Excellent, Bonne, MOyenne, MAuvaise, Très Mauvaise} correspondant à leurs discriminations en un contre tous.

Cependant, les SVM ne permettent pas d'aboutir directement à des probabilités a posteriori de classification.

Au lieu d'affecter une probabilité a posteriori au classifieur (ce qui peut se traduire par une confiance que l'on accorde à la décision prise par un classifieur), une mixture de gaussienne est calculée pour les 5 classifieurs. Ainsi pour un classifieur $(C_i)_{i\in[1,\ldots,5]}$ et pour un exemple $\mathbf{x_i}$ à classer, l'exentricité à la gaussienne est calculée, ce qui se traduit par la plausibilité que l'exemple $\mathbf{x_i}$ soit correctement classifié. De plus, il est utile de prendre en compte la confiance $p(C_i)$ que l'on accorde au classifieur C_i, via le calcul d'une probabilité a posteriori à l'aide d'une fonction paramétrique de type fonction logistique dans laquelle les paramètres sont estimés selon un maximum de vraisemblance. Ainsi, la confiance que l'on accorde à la classification d'un exemple $\mathbf{x_i}$ par le classifieur C_i est définie par la probabilité conditionnelle $p(\mathbf{x_i}|C_i)$.

Reste alors à combiner les résultats ainsi obtenus afin de prendre la décision finale d'appartenance à l'une des cinq classes de qualité.

Parmi tous les outils qui s'offrent à nous dans ce domaine, nous décidons d'utiliser la théorie de l'évidence (aussi appelée théorie de Dempster-Shafer ou théorie des fonctions de croyance) [8], qui permet d'une part de traiter des informations incertaines, et d'autre part de combiner des informations provenant de plusieurs sources. Dans le cadre de cette théorie, plusieurs règles de décision sont définies pour nous permettre de sélectionner la classe à laquelle le vecteur de caractéristiques représentant la qualité d'une image sera finalement affecté.

2.3 Combinaison des classifieurs

Soit $\Omega = \{\omega_1, \ldots, \omega_N\}$, l'ensemble des N classes finales possibles pour un pixel, appelé cadre de discernement. Dans notre cadre applicatif $N = 5$ et Ω correspond aux cinq classes $\{\omega_E, \omega_B, \omega_{MO}, \omega_{MA}, \omega_{TM}\}$, respectivement pour la classe de qualité excellent, bonne, moyenne, mauvaise et très mauvaise. Au lieu de restreindre ses mesures à cet ensemble (comme le ferait la théorie des probabilités contrainte par son axiome d'additivité), la théorie de l'évidence s'étend sur l'ensemble puissance de Ω, noté 2^Ω, l'ensemble des 2^N sous-ensembles de Ω. On définit alors une fonction de masse m qui représente la croyance que l'on accorde aux différents états du système, à un instant donné. Cette fonction est également connue en tant que masse de croyance élémentaire. Elle est définie de 2^Ω dans $[0, 1]$ et vérifie $\sum_{A\subseteq\Omega} m(A) = 1$ et $m(\emptyset) = 0$ où $m(A)$ représente la croyance que l'on place dans la proposition A. Il est à noter que A représente soit un singleton ω_n, soit une disjonction d'hypothèses. Ici, $m(A)$ quantifie la croyance que l'on accorde au fait que la classe cherchée appartienne au sous-ensemble A de Ω (et à aucun autre sous-ensemble de A). Les propositions A telles que $m(A) > 0$ sont appelées éléments focaux. De plus, dans le cas où l'ensemble des hypothèses est exhaustif et exclusif, la masse de l'ensemble vide $\emptyset$ est nulle.

Deux fonctions de masse initiales m_1 et m_2 représentant les informations respectives de deux sources différentes, peuvent être combinées selon la règle de Dempster [9].

$$m(A) = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K}, \quad \begin{array}{l}\forall A \in \Omega \\ A \neq \emptyset.\end{array} \tag{2}$$

K est appelé le facteur de conflit et représente le désaccord entre les deux sources. Il correspond à la masse de l'ensemble vide $K = \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$.

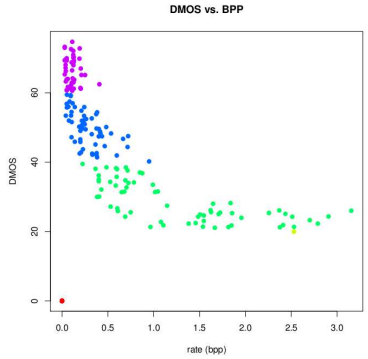
Après combinaison, il reste à prendre une décision quant à l'élément le plus "probable" de Ω. Plusieurs règles de décision sont alors possibles, mais une des plus utilisées reste celle du maximum de probabilité "pignistique" BetP(ω). Cette règle de décision présentée par SMETS [10] utilise la transformation pignistique, qui permet de répartir uniformément la masse associée à un sous-ensemble de Ω sur chacun de ses éléments :

$$\text{BetP}(\omega) = \sum_{\omega\in A\subseteq\Omega} \frac{m(A)}{|A|}, \forall\omega \in \Omega \tag{3}$$

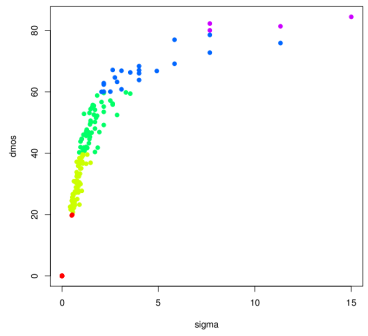
$|A|$ étant le cardinal de A. La décision va alors vers l'élément de Ω dont cette valeur est la plus grande :

$$\omega^* = \text{Arg}\left\{\max_{\omega\in\Omega}\left[\text{BetP}(\omega)\right]\right\}. \tag{4}$$

L'une des difficultés réside dans le choix de la modélisation des fonctions de masse. Parmi les modélisations existantes, celle proposée par DENŒUX [11] a été utilisé car elle est basée sur la notion de distance aux voisins et permet d'intégrer divers critères de voisinage. Ainsi la masse $m(\{\omega_n\})$ est définie comme



(a) DMOS vs taux de compression (bpp) pour la base d'images JPEG2000



(b) DMOS vs σ pour la base d'images ayant subies des dégradations liées à un bruit gaussien

FIG. 1: *les couleurs sont associées aux différentes classes de qualité comme suit: violet–5 (excellente), bleu–4 (bonne), vert–3 (moyenne), jaune–2 (mauvaise) and rouge–1 (très mauvaise)*

une fonction décroissante de la distance d entre le vecteur $\mathbf{x_i}$ à classifier et le barycentre de la classe ω_n:

$$\begin{cases} m(\omega_n) = \alpha \exp\left(-\gamma_n d^2\right) \\ m(\Omega) = 1 - m(\omega_n) \end{cases} \quad (5)$$

où $\alpha = 0.5$ est une constante. $\gamma_n > 0$ dépend de la classe ω_n et correspond à la probabilité conditionnelle $p(\mathbf{x_i}|\omega_n)$. Deux formulations de la distance d ont été étudiées. Il s'agit 1) de la distance euclidienne d_E, 2) de la distance euclidienne d_{EP} pondérée par un critère de corrélation permettant de comparer la dispersion spatio-colorimétrique de deux nuages couleurs.

3 Mesure des performances

3.1 Bases et outils

La méthode proposée a été testée sur deux bases d'images : la première représente des images compressées avec le standard JPEG2000, et la seconde représente des images ayant subies des dégradations liées à un bruit gaussien, toutes deux présentes dans la base LIVE [12]. Pour chacune de ces deux bases d'images nous disposons de la mesure de qualité DMOS $\bar{d}_j$ obtenue en faisant intervenir des observateurs humains. Cette mesure est obtenue selon

$$\bar{d}_j = \frac{1}{N} \sum_{i=1} N(r_{i\text{ref}(j)} - r_{ij}) \quad (6)$$

où N représente le nombre d'observateurs, $r_{i\text{ref}(j)}$ la note donnée par le i-ème sujet pour l'image de référence j et r_{ij} la note donnée par le i-ème sujet pour l'image de dégradée j. Ainsi en fonction de la note DMOS obtenue pour l'image j, il est possible de fournir une classe de qualité.

La figure 1 représente les valeurs DMOS obtenues pour les deux bases d'images testées.

Une sélection d'un ensemble d'apprentissage et d'un ensemble de test est ensuite opérée afin de pouvoir entraîner notre modèle. Chacun de ces deux ensembles est construit en respectant les distributions de qualité présentes initialement dans la base globale.

La méthode introduite a ensuite été comparée avec les méthodes de mesures de qualité existantes telles que MS-SSIM [2], VIF [13], et VSNR [14]. En effet dans [15], les auteurs ont démontré que les deux premières métriques ont un taux de corrélation avec les notes moyennes obtenues lors d'expériences psychovisuelles supérieures aux métriques existantes et testées.

Afin de réaliser une comparaison des résultats obtenus selon la méthode proposée avec les métriques précédentes, deux techniques ont été utilisées. La première consiste à réaliser une classification basique pour chacune des trois métriques utilisées. Ainsi une normalisation sur l'intervalle [0,1] est tout d'abord opérée pour les valeurs obtenues par les trois métriques. Puis cet intervalle est ensuite subdivisé en 5 sous-intervalles $\{[0; 0.2[, [0.2; 0.4[, [0.$ qui sont respectivement associés aux cinq classes de qualité. Ainsi il est possible d'obtenir les taux de reconnaissance pour chacune des métriques utilisées.

La seconde technique de comparaison repose sur le calcul d'une note de qualité en fonction de la probabilité pignistique BetP(ω_i) obtenue par le vecteur candidat x (représentant l'image I) et l'intervalle numérique associée à la classe ω_i. La note finale $N_f(x)$ est alors obtenue à l'aide d'une fonction logisitique comme suit

$$N_f(x) = K_i . \frac{1}{1 + exp(-\sqrt{\sigma} BetP(\omega_i)^2)} \quad (7)$$

avec K_i une constante correspondant à la note maximale que l'on puisse attribuer à une classe ω_i et σ représente la longueur d'un intervalle lié à une classe ω_i. Ainsi en se référant aux préconisation décrites dans la norme ITU 500-11, un intervalle numérique est de longueur 20. Par exemple, la 5$^{\text{e}}$ classe ω_E est représentée par l'intervalle des notes $]80, 100]$.

Les métriques MS-SSIM, VIF et VSNR sont comparées avec l'approche proposée selon trois critères : 1) le coefficient de corrélation de Spearman (SROCC), 2) le coefficient de corrélation linéaire de Pearson (CC) et 3) l'erreur des moindres carrées (RMSE); ces deux dernières mesures étant réalisées après l'application d'une régression non linéraire [1].

3.2 Résultats

Le tableau **??** présente les taux de reconnaissances pour chacune des métriques comparées ainsi que pour les deux bases de données testées. On constate que l'utilisation de la distance euclidienne d_{EP} pondérée par un critère de corrélation founit des résultats meilleurs que ceux obtenus en utilisant uniquement la distance euclidienne classique d_E. De plus le taux de reconnaissance obtenu par la méthode proposée est légèrement supérieur à celui obtenu par la meilleurs métrique pour les deux bases testées.

Base JPEG2000					
	méthode avec				
	d_E	d_{EP}	MS-SSIM	VIF	VSNR
TR	0.967	0.979	0.975	0.968	0.973
Base bruit gaussien					
	méthode avec				
	d_E	d_{EP}	MS-SSIM	VIF	VSNR
TR	0.953	0.970	0.965	0.953	0.960

TAB. 1: *taux de reconnaissance pour chacune des méthodes testées et pour les deux bases JPEG2000 et de bruit gaussien.*

Base JPEG2000					
	méthode avec				
	d_E	d_{EP}	MS-SSIM	VIF	VSNR
SROCC	0.9614	0.9649	0.9647	0.9720	0.9406
CC	0.9609	0.9712	0.9708	0.9789	0.9476
RMSE	6.0199	5.9151	5.9899	5.0925	5.6143
Base bruit gaussien					
	méthode avec				
	d_E	d_{EP}	MS-SSIM	VIF	VSNR
SROCC	0.9340	0.9699	0.9519	0.9706	0.9507
CC	0.9401	0.9677	0.9487	0.9762	0.9501
RMSE	5.9776	5.7435	5.8225	4.9905	5.5443

TAB. 2: *Performance des différentes méthodes.*

Le tableau 3.2 contient les résultats de la comparaison des métriques en terme de corrélation aux valeurs DMOS fournies avec la base LIVE. Les résultats obtenus en utilisant le distance euclidienne d_E sont inférieurs à ceux obtenus en utilisant la distance d_{EP}. De même, les résultats obtenus avec la méthode proposée sont quasi identiques à ceux obtenus par les trois autres métriques, et la différence des valeurs n'est pas statistiquement significative.

Ainsi, on peut conclure que la méthode proposée fait jeu égal avec les métriques couramment utilisées et surpasse ces dernières lorsque l'on se réfère aux taux de reconnaissance. Les performances en terme de coefficients de corrélation, qui s'apparentent à ceux obtenus avec les trois autres métriques, sont probablement dûs aux fonctions noyaux utilisées qui ne sont pas optimisées par rapport aux données du problème. Néanmoins cela permet d'ouvir la voie à de nouvelles techniques de mesure de qualité, plus proche de ce qui est réalisé par le système visuel humain, autrement que par une simple notation.

4 Conclusion

Le méthode proposée se veut une méthode originale de mesure de la qualité des images basée sur une approche de classification. Ce processus est suffisament générique pour être apliqué quelque soient les défauts que l'on souhaite mesurer : dans ce cas, il suffit d'adapter le vecteur de caractéristiques utilisé ainsi que les fonctions noyaux utilisés.

Qui plus est, il peut être utilisé dans le cadre d'une mesure de qualité avec ou sans référence, ce qui n'existe pas jusqu'à maintenant.

Références

[1] ITU-R Recommendation BT.500-11, "Méthodologie d'évaluation subjective de la qualité des images de télévision," tech. rep., UIT, Geneva, Switzerland, 2002.

[2] Z. Wang and A. C. Bovik, "A universal quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[3] A. Trémeau, C. Charrier, and E. Favier, "Quantitative description of image distorsions linked to compression schemes," in *Proceedings of The Int. Conf. on the Quantitative Description of Materials Microstructure*, (Warsaw), Apr. 1997. QMAT'97.

[4] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *International Conference on Image Processing*, (Vancouver, BC), pp. 981–984, Sept. 2000.

[5] P. Gastaldo, G. Parodi, J. Redi, and R. Zunino, "No-reference quality assessment of JPEG images by using CBP neural networks," in *ICANN 2007*, vol. LNCS 4669, pp. 564–572, 2007.

[6] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[7] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 415–425, 2002.

[8] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.

[9] A. Dempster, "Upper and Lower Probablilities Induced by Multivalued Mapping," *Ann. Math. Statist.*, vol. 38, pp. 325–339, 1967.

[10] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," *Uncertainty in Artificial Intelligence*, vol. 5, pp. 29–39, 1990. Elsevier Science Publishers.

[11] T. Denoeux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 5, pp. 804–813, 1995.

[12] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2." http://live.ece.utexas.edu/research/quality.

[13] H. R. Sheik and A. C. Bovik, "Image information and visual quality," in *Proc. IEEE ICASP*, 2004.

[14] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[15] H. R. Sheik, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 5, no. 11, pp. 3441–3452, 2006.