

Influence du codage Jpeg sur des descripteurs d'images

Influence of Jpeg Coding on Image Descriptors

par Jean-Michel JOLION et Stéphane BRES

Laboratoire Reconnaissance de Formes et Vision
Bât 403 INSA
20, Avenue Albert Einstein
69621 Villeurbanne Cedex
France
e.mail : {jolion,sbres}@rfv.insa-lyon.fr
internet : <http://rfv.insa-lyon.fr/~jolion>

résumé et mots clés

Cet article propose une étude expérimentale de la robustesse des principaux descripteurs d'images par rapport aux distorsions introduites par le codage Jpeg. Tout particulièrement, nous présentons les résultats obtenus pour les histogrammes et les éléments différentiels d'une image. Cette étude est préliminaire à nos travaux sur l'indexation d'images et de vidéo¹.

Jpeg, descripteurs d'images, bruit de codage, indexation d'images.

abstract and key words

We present in this paper an experimental study on robustness of main image descriptors regarding to the Jpeg coding noise. We particularly introduce some results concerning histograms and differential characteristics of a given image. This study is a preliminary work of our team in the context of content-based indexation of images and videos.

Jpeg, image descriptors, coding, image indexation.

1. introduction

Dans ce rapport, nous présentons les principaux résultats d'une étude sur la robustesse des descripteurs les plus souvent utilisés en indexation d'images, par rapport au bruit très particulier qu'est le bruit de codage. L'ensemble de cette étude peut être trouvé dans [Jol 98].

La motivation de cette étude réside dans la constatation que, dans le but de pouvoir indexer toutes sortes d'images, il est nécessaire que les outils utilisés ne soient pas sensibles aux différents codages utilisés à l'heure actuelle. Parmi ceux-ci, certains ne sont pas neutres, *i.e.* ils introduisent une perte dans l'information portée par le signal image. Le codage Jpeg est un de ces codages [Wal 91] et a servi d'exemple à notre étude.

Cette perte est bien souvent mineure pour la tâche particulière de visualisation mais elle peut se révéler beaucoup plus importante dès que l'on envisage des traitements automatisés conduisant à des mesures.

En effet, l'ordinateur, et plus particulièrement ses algorithmes, n'ont pas du tout la même sensibilité que l'œil humain. Cette problématique est proche de celle du marquage dans le but de l'authentification, où les techniques utilisées doivent résister aux transformations subies par l'image [Sch 96].

Notre objectif est donc de quantifier les écarts entre les résultats issus de l'application d'un algorithme sur une image originale et sur sa version codée par le système Jpeg en fonction de son paramétrage.

1. Cette recherche est financée conjointement par le projet européen INCO 950363 *TELESUN*, le projet CCETT 96 ME 17 *SESAME*, et le projet Région Rhône-Alpes (mandature 1994-1998) *ACTIV*.

Dans un premier temps, nous allons décrire le protocole de notre étude (images tests, traitements effectués, codage et mesures de performance). Par la suite, nous présenterons les principaux résultats de cette étude.

2. la base de test

2.1. les images

Nous avons constitué une petite base de 50 images qui ne représente pas un choix thématique particulier mais un ensemble très divers d'images de tailles variées représentant des visages, des paysages, des dessins, des textures, ... La figure 1 présente quelques images caractéristiques de cette base. Toutes ces images sont exemptes de bruit de codage, *i.e.* elles sont assimilables à une sortie de capteur. Compte tenu de l'état actuel des recherches sur les descripteurs, nous n'avons pas inclus d'images couleurs dans cette base.

2.2. les traitements

Notre objectif premier était de tester les descripteurs les plus souvent utilisés dans la littérature [DeM 97]. Ceci nécessitait de pouvoir avoir accès aux sources de leurs concepteurs ce qui est le plus souvent impossible et toute tentative de ré-implémentation à partir des articles se heurte à l'absence, dans ceux-ci, de la transcription du savoir-faire. Cependant, ces descripteurs s'appuient tous sur l'image. Il est donc nécessaire de quantifier la dégradation induite sur l'image avant tout test sur le gain introduit par un algorithme. Par la suite, nous proposons de tester des caractéristiques qui s'éloignent de la représentation initiale en incluant une partie de plus en plus prédominante d'algorithmie.

Nous avons ainsi étudié successivement :

L'image : La dégradation de l'image, I , sous sa forme usuelle, *i.e.* matrice de pixels, sera bien sûr le premier test. Ceci permettra de mieux appréhender les dégradations induites sur les descripteurs s'appuyant sur les données contenues dans les images.

L'histogramme : L'histogramme d'une image est un élément de base pour l'indexation (même si il constitue une caractéristique très globale). Il s'agit d'un simple changement de représentation. Il était donc intéressant de voir sa robustesse lorsque l'image est codée en format Jpeg.

La résolution : Avant de rentrer dans le détail des différents descripteurs, nous avons testé la répercussion de la dégradation initiale à différents niveaux de résolution par l'analyse de cinq niveaux de résolutions successifs de la pyramide gaussienne et laplacienne de chaque image [Jol 93].

Les dérivées spatiales partielles : Cette information est la base de nombreux descripteurs, dits descripteurs différentiels,

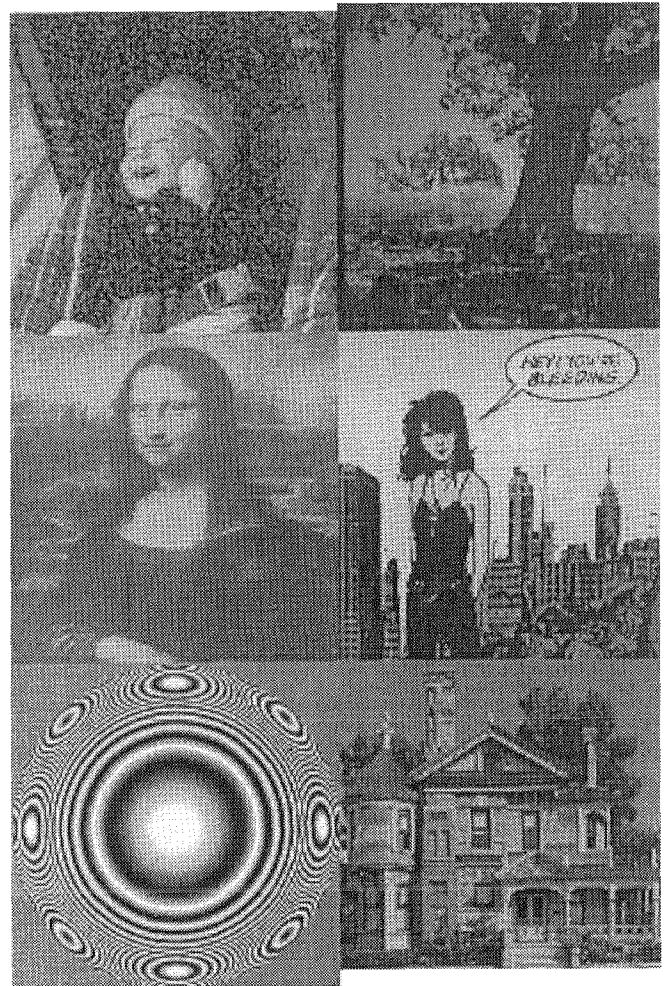


Figure 1. – Extrait de la base d'images test.

comme les invariants que nous aborderons par la suite. Nous nous limitons ici à l'étude des dérivées spatiales partielles d'ordre 1 et 2 : $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, $\frac{\partial^2 I}{\partial x^2}$, et $\frac{\partial^2 I}{\partial y^2}$. Ces dérivées sont estimées en utilisant l'algorithme de Deriche [Der 87] pour une valeur fixe du paramètre α (dans notre cas, nous avons choisi $\alpha = 1$ ce qui correspond à la valeur la plus communément utilisée) pour toute la base.

Les cartes de contours : L'utilisation la plus naturelle des descripteurs différentiels est bien sûr la fabrication de la carte des contours. Cette opération fait intervenir l'estimation des dérivées partielles par un filtrage (étape purement calculatoire sans intervention de tests/décisions), et plusieurs étapes plus algorithmiques (filtrage des non maxima locaux dans la direction du gradient, seuillage par double hystérésis, reconstruction de chaînes de contours fragmentées). La robustesse que nous testons est donc celle de l'algorithme entier et non pas seulement celle du calcul du gradient déjà testée par ailleurs. Nous avons utilisé l'algorithme de Canny-Deriche fourni dans la bibliothèque SIMPA du GDR-PRC ISIS.

Points d'intérêt : Une approche classique de l'indexation d'images consiste à résumer celle-ci par un ensemble de points particuliers dits points d'intérêt. Les détecteurs proposés dans la littérature sont très nombreux depuis les travaux initiaux de Moravec [Mor 77] et Beaudet [Bea 78]. Une analyse comparative de ces détecteurs peut être trouvée dans [Sch 97]. Nous nous sommes dans cette étude inspirés des travaux de Förstner [For 94] qui repose sur l'analyse de la distribution locale de l'énergie du gradient (*average squared gradient image*) donnée par une matrice 2×2 , Ψ , calculée en chaque point (x, y) :

$$\Psi(x, y) = \{G_\sigma \otimes \nabla^t g \nabla g [I]\}(x, y)$$

où G_σ est la fonction gaussienne d'écart type σ , \otimes l'opérateur produit de convolution, ∇g l'opérateur gradient $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ et I l'image initiale.

Plusieurs approches ont été proposées pour résumer cette matrice sous la forme d'un indicateur scalaire. Pour notre part, nous avons choisi une mesure simple formée par le produit des valeurs propres. Les points d'intérêt retenus sont les maxima locaux de cette mesure.

Nous testerons la robustesse de cet indicateur par rapport au bruit de codage en prenant en compte la variation du nombre de points d'intérêt détectés et leurs localisations.

Invariants différentiels : Etant donné une carte de points d'intérêt associée à une image initiale, nous testerons les valeurs des invariants différentiels (descripteurs robustes aux transformations appartenant au groupe des similitudes d'images) jusqu'au 3^e ordre (soit 9 invariants) tels que proposés dans [Sch 97] (les indices représentent les dérivées partielles).

- $V_0 = I$
- $V_1 = I_x I_x + I_y I_y$
- $V_2 = I_{xx} I_x I_x + 2I_{xy} I_x I_y + I_{yy} I_y I_y$
- $V_3 = I_{xx} + I_{yy}$
- $V_4 = I_{xx} I_{xx} + 2I_{xy} I_{xy} + I_{yy} I_{yy}$
- $V_5 = I_{xxx} I_y I_y I_y + 3I_{xyy} I_x I_x I_y - 3I_{xxy} I_x I_y I_y - I_{yyy} I_x I_x I_x$
- $V_6 = I_{xxx} I_x I_y I_y + I_{xxy} (-2I_x I_x I_y + I_y I_y I_y) + I_{xyy} (-2I_x I_y I_y + I_x I_x I_x) + I_{yyy} I_x I_x I_y$
- $V_7 = I_{xxy} (-I_x I_x I_x + 2I_x I_y I_y) + I_{xyy} (-2I_x I_x I_y + I_y I_y I_y) - I_{yyy} I_x I_y I_y + I_{xxx} I_x I_x I_y$
- $V_8 = I_{xxx} I_x I_x I_x + 3I_{xxy} I_x I_x I_y + 3I_{xyy} I_x I_y I_y + I_{yyy} I_y I_y I_y$

De nouveau, nous avons utilisé une estimation des dérivées partielles fournie par l'opérateur de Canny-Deriche avec $\alpha = 1$. Dans ce test, c'est la variation sur chaque invariant qui nous intéresse, nous utiliserons donc pour l'image dégradée les mêmes points d'intérêt que ceux de l'image initiale.

Autres caractéristiques : Ce panel n'est pas exhaustif. D'autres traitements ont été testés :

- le filtrage spatial (filtre *SNN* : *Symmetric Nearest Neighbor*);

- le rehaussement de contraste;
- la segmentation par analyse en régions (algorithme *split and merge*);
- la caractérisation par filtre de Gabor.

Cependant, dans cette étude, nous nous focaliserons sur les traitements liés à la composante différentielle de l'image. Pour les autres caractéristiques, le lecteur est invité à se reporter à [Jol 98, Meg 98].

Nous avons également omis la couleur et la texture. Concernant la texture, cette notion repose le plus souvent sur une approche différentielle. Elle n'est donc pas complètement exclue de cette étude.

2.3. le codage Jpeg

Dans cette première étude, notre choix s'est porté sur Jpeg car ce codage présente toutes les caractéristiques recherchées.

Codage répandu : Des codeurs-décodeurs Jpeg sont disponibles sous tous les systèmes d'exploitation. Ce codage est connu du grand public et a vocation à se répandre du fait de son association avec la norme Mpeg utilisée pour les séquences vidéo.

codage avec perte : Le principe du codage Jpeg est simple. L'image initiale est décomposée en blocs 8×8 pixels. Chaque bloc est ensuite modélisé par une somme de fonctions trigonométriques par un algorithme très rapide, la transformée discrète en cosinus (DCT). On obtient alors 64 coefficients par bloc. Cette phase de modélisation du signal est suivie par une phase de quantification (seule partie irréversible du processus de compression et qui introduit donc une perte) et un codage de Huffman des coefficients retenus pour chaque DCT (codage entropique n'introduisant pas de perte d'information). Pour plus de détails, on peut se référer à [Wal 91].

Le codage Jpeg est donc une technique de codage/compression non réversible. Plus le taux de compression souhaité est fort, plus la perte sera importante. Dans notre étude, nous n'avons pas utilisé d'optimisation du codage entropique (la taille du fichier obtenu pourrait donc être légèrement réduite mais notre étude ne porte pas sur la compression et ce paramètre n'a pas d'influence sur la qualité du rendu visuel). Le seul paramètre influent est donc celui que nous nommerons la *qualité de codage*², compris entre 0 et 100% et qui traduit la taille de la table de quantification. Par exemple, une qualité de 25% se traduit par une table de quantification sur 2 bits. Bien sûr, plus cette qualité sera haute, plus le taux de compression sera faible.

Codage perceptuel : Compte tenu de la nature de la base d'approximation, il est clair que ce codage est de type fréquentiel. Il a été conçu pour être adapté à la visualisation [Eud 96]. On constate en effet que dans le cas général (image de scène extérieure par exemple), une qualité de codage de 75% donne une perception

2. Ce paramètre est celui disponible sur le logiciel *cjpeg* sous Unix.

quasiment inchangée. La figure 2 présente, pour une des images de la base, la plus célèbre sans doute, les versions obtenues pour des qualités de codage allant de 10% à 100%. Les effets néfastes du codage (flou et effet de bloc) ne sont pas perceptibles pour des qualités supérieures à 70%³.

La figure 3 montre le facteur de compression obtenu sur cette image pour les différentes qualités testées. On constatera que le taux de compression est de 1.43 pour une qualité imposée de 100%. Ceci est tout à fait possible car la partie sans perte du codage peut en elle-même introduire un gain de place, indépendamment du taux de compression. Cependant, même pour cette qualité (*i.e.* 100%), le codage n'est pas parfait et la reconstruction de l'image fait apparaître 9.5% des points avec un écart non nul entre l'image

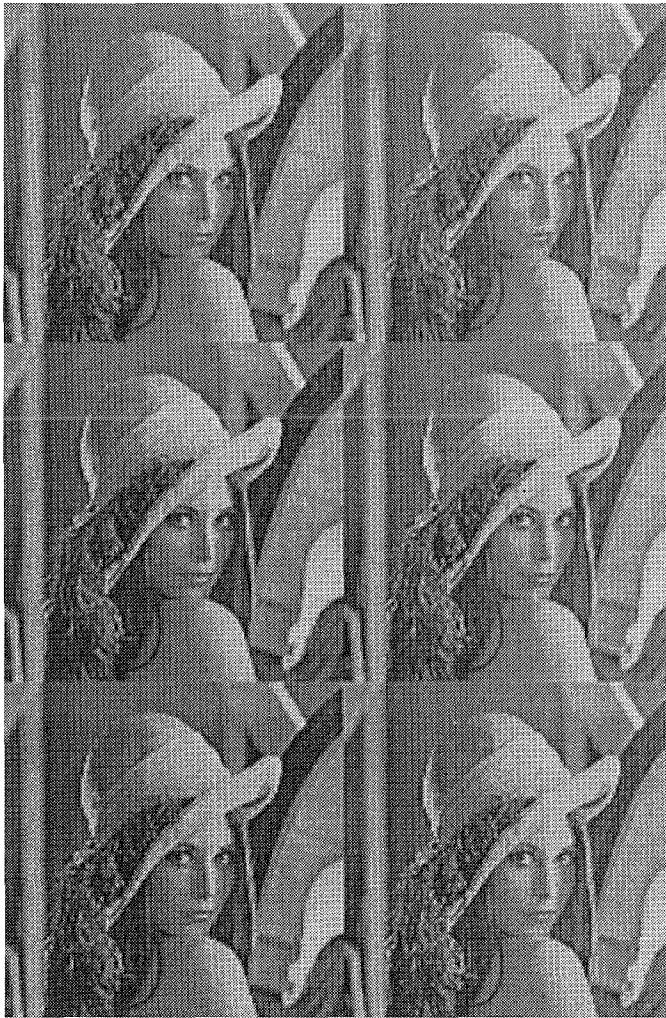


Figure 2. – De gauche à droite et de haut en bas : l'image initiale, les versions codées pour des qualités de 10%, 30%, 50%, 70% et 95%.

3. Il faut cependant signaler ici que cette remarque est surtout valide pour une sortie de type écran car sur papier, le résultat obtenu intègre un facteur de réduction de taille et un effet de filtrage passe-bas dû à l'impression laser, qui atténuent très fortement la dégradation Jpeg.

initiale et l'image Jpeg (cet écart a pour valeur 1, ce qui indique une erreur d'arrondi dans le processus de codage/décodage).

La qualité du rendu perceptif a largement contribué au succès de ce type de codage dans le domaine de l'image. Cependant, ce codage n'étant pas parfait, au sens de la conservation de l'information, il est nécessaire d'étudier la nature de la perte. On sait déjà que ce type de codage est très peu performant pour coder les images trop régulières (images binaires ou images de synthèse). Ceci est dû à la difficulté d'approximer des fonctions continues par des signaux trop discrets.

Dans les tests, chaque image de la base a été codée en Jpeg avec une qualité de 10%, 20%, 30%, 40%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% et 100% (soit 14 codages différents sur chaque image de la base).

2.4. les mesures de performances

Mesurer la performance d'un descripteur, *i.e.* la robustesse dans notre étude, est en soit un problème difficile car cela se ramène à un problème de définition d'une distance entre deux cartes de caractéristiques. Cette question n'est pas nouvelle mais elle a retrouvé toute sa pertinence avec le développement du domaine de l'indexation. On pourra par exemple consulter la très belle étude proposée sur le thème de la similarité entre images (se traduisant par des champs de caractéristiques) par Santini dans [San 97].

Dans notre étude, nous sommes cependant dans un contexte très particulier. En effet, il s'agit de comparer des résultats d'algorithmes appliqués sur des images très similaires, *i.e.* même taille et organisation spatiale (tous les résultats des algorithmes que nous allons tester se présentent sous la forme d'images), et même contenu sémantique. Il est donc possible de construire une mesure globale reposant sur des comparaisons locales (*i.e.* pixel

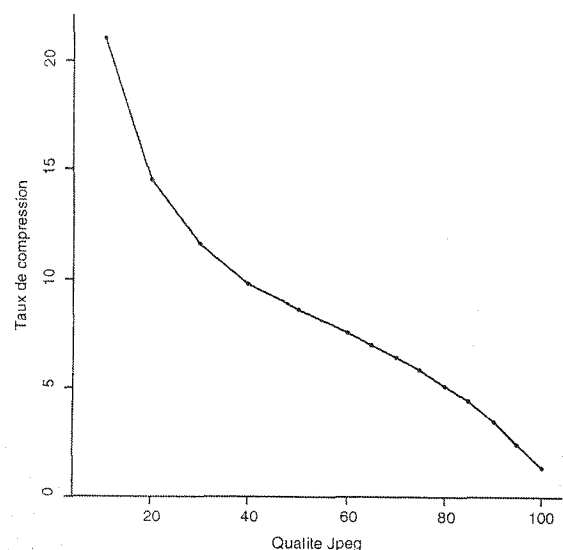


Figure 3. – Taux de compression dû au codage Jpeg pour l'image de la figure 2 en fonction de la qualité du codage.

à pixel). Nous avons retenu trois mesures adaptées à la nature des résultats : le rapport signal à bruit pour les champs denses, la figure de mérite pour les champs éparés (tels que les cartes de contours) et une mesure spécifique aux histogrammes.

Pour deux images I et J , le rapport signal à bruit est défini par :

$$snr(I, J) = 10 \log_{10} \left(\frac{\text{Var}[I]}{\text{Var}[I - J]} \right)$$

Cette mesure n'est pas symétrique, ce n'est donc pas une distance car elle privilégie une des deux images. Cette mesure est cependant intéressante car d'une part nous avons une image de référence dans toutes nos comparaisons, et d'autre part, cette mesure est très utilisée en compression d'images (elle est donc reconnue par notre communauté) [Jai 89]. En revanche, cette mesure est, comme la précédente, adaptée aux perturbations globales plus que locales (un snr crête serait dans ce cas plus approprié mais il s'agit d'un opérateur trop peu robuste). Dans le cas d'une absence totale d'erreur, on doit obtenir un snr infini. Dans ce cas, une valeur arbitrairement fixée à 100 apparaîtra sur les schémas.

Nous avons utilisé cette mesure pour comparer les sorties des différents descripteurs. Il n'est cependant pas possible de l'appliquer pour des champs peu denses comme les cartes de contours ou de points d'intérêt. Dans ces cas particuliers, nous utilisons l'indicateur, ou figure de mérite, introduit par Pratt [Pra 78]

$$F(I, J) = \frac{1}{\max(N_I, N_J)} \sum_{i=1}^{N_J} \frac{1}{1 + d_i^2}$$

où N_I (resp. N_J) est le nombre de points contours dans la carte de contours de I (resp. J), et d_i la distance minimale entre le i ème point contour de la carte J à son homologue (le point le plus proche) dans la carte I (d_i est majorée par 5 pour réduire le temps d'évaluation de cette mesure).

Par définition, cet indicateur prend ses valeurs dans l'intervalle $[0, 1]$ (la valeur 1, resp. 0, correspondant à la ressemblance maximale, resp. minimale, entre les deux cartes contours).

Enfin, pour les histogrammes, nous avons utilisé la différence d'entropie :

$$\Delta E(I, J) = \sum_i H_I(i) \log_2(H_I(i)) - \sum_i H_J(i) \log_2(H_J(i))$$

où H_I (resp. H_J) est l'histogramme normalisé (son intégrale est égale à 1) de l'image I (resp. J).

Ces différentes mesures sont moyennées sur toute la base et chaque courbe représentera l'évolution de la moyenne en fonction de la qualité du codage. On peut donc comparer la robustesse (moyenne) des descripteurs mais aussi son évolution en fonction de la qualité de codage. Cependant, nous souhaitons aussi pouvoir associer des appréciations qualitatives du style « bon », « mauvais », etc. Pour cela, nous avons besoin d'une référence. Celle-ci nous sera donnée indirectement par l'application de ces

indicateurs à la mesure de ressemblance entre une image et des versions bruitées de celle-ci et cela pour chaque image de la base.

$$J_{xy} = [I_{xy} + \varepsilon_{xy}]$$

où ε_{xy} est la réalisation d'une variable aléatoire uniforme prenant ses valeurs sur $[-k, +k]$ et $[\]$ l'opérateur partie entière.

Si on néglige les effets de troncature et de saturation (inévitables si on conserve une dynamique fixe de $[0, 255]$), on obtient

$$snr(I, J) = 10 \cdot \log_{10} \left(\frac{3 \text{Var}[I]}{k^2} \right)$$

La figure 4 donne les valeurs obtenues pour le snr à partir de l'ensemble des valeurs de la base. On constate une évolution régulière des indicateurs (moyenne, extréma, ...). Les valeurs moyennes du snr sont présentées dans le tableau 1.

Tableau 1. – SNR moyens en fonction de l'amplitude du bruit.

Bruit (k)	SNR (db)
1	45.212
3	37.227
5	32.965
7	30.101
8	28.956
10	27.043
12	25.477
15	23.556
18	21.987
20	21.083

Nous avons arbitrairement fixé un *seuil de robustesse*⁴ à une valeur de **22db** (donc pour un bruit fort). En ce qui concerne l'indicateur F , figure de mérite, nous avons déterminé le seuil de robustesse par rapport à la valeur 22db retenue pour le snr qui correspond en moyenne à un bruit additif d'amplitude $k = 18$. Nous avons comparé les résultats de l'opérateur de Deriche appliqué à une image originale et à sa version bruitée (avec $k = 18$). La valeur moyenne obtenue pour la figure de mérite, que nous retiendrons donc comme seuil de robustesse, est **0.69**.

Définition : Un descripteur est robuste au bruit de codage Jpeg si la mesure de performance entre le résultat de son application sur

4. La valeur du seuil de robustesse n'est pas en soit une valeur stricte. Elle servira pour établir une comparaison rapide entre descripteurs. C'est donc plus son ordre de grandeur qui est à retenir.

l'image originale d'une part et sur sa version Jpeg d'autre part, est supérieure au seuil de robustesse.

Plus exactement, nous chercherons la qualité minimale du codage Jpeg nécessaire pour atteindre ce seuil.

3. résultats

3.1. dégradation de l'image initiale

La figure 5 présente les indicateurs de performance pour la comparaison directe entre l'image originale et sa version Jpeg. On peut constater que le seuil de robustesse est tout de suite atteint. La détérioration de la performance est lente (par exemple, on note peu de différence entre les valeurs obtenues pour les qualités 40% et 80%). Ceci confirme bien sûr la qualité de codage Jpeg.

3.2. histogramme

La figure 6 représente les valeurs de la différence entre les distributions des niveaux de gris des images comparées (écart mesuré en terme d'écart absolu entre les entropies). Le codage Jpeg induit une différence moyenne de 1.22 bit par image avec un maximum de 5.2 pour une qualité très faible. Cette différence s'estompe au fur et à mesure que la qualité du codage augmente pour finir avec un écart moyen de 0.18 bit. Si l'on considère cet écart avec son signe (conformément à la définition de ΔE), on constate que la valeur moyenne est remarquablement stable autour

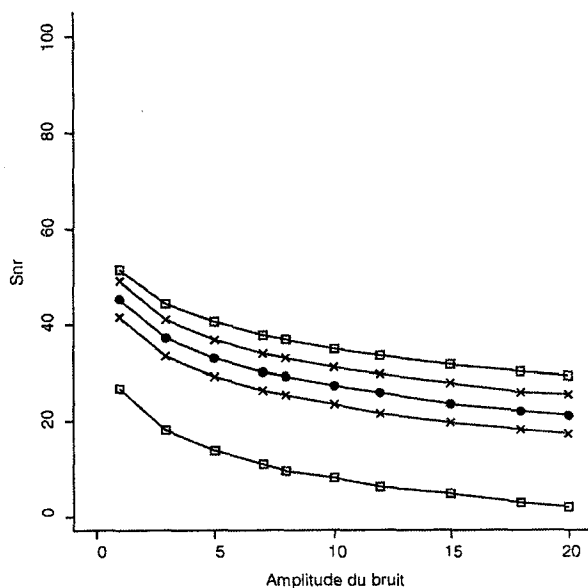


Figure 4. – Mesure moyenne entre image initiale et version bruitée en fonction de la quantité de bruit par *snr*. ● = moyenne, × : moyenne ± écart-type, □ : extréma (min, max).

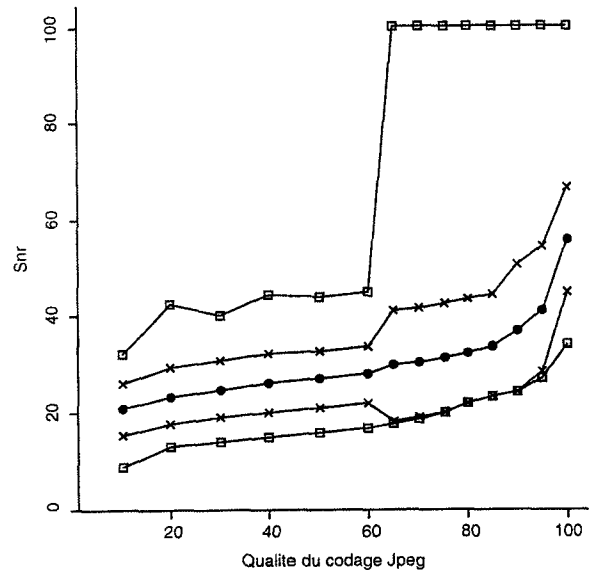


Figure 5. – Mesure moyenne entre image initiale et version jpeg en fonction de la qualité du codage. ● : moyenne, × : moyenne ± écart-type, □ : extréma (min, max).

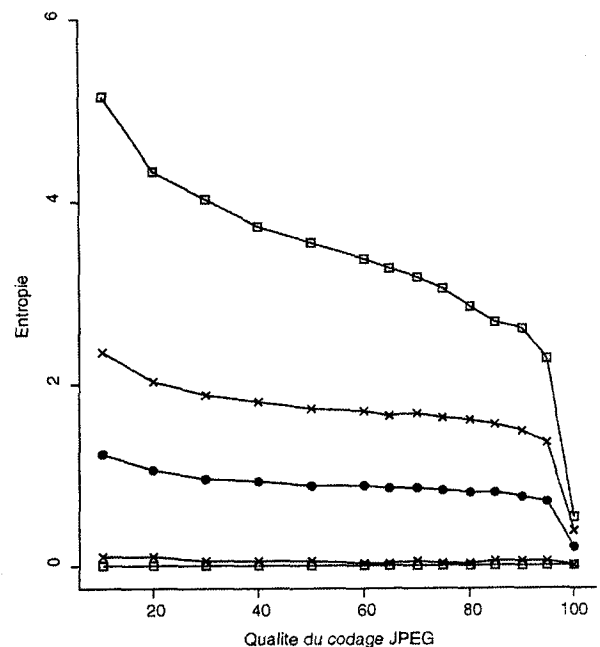


Figure 6. – Différence d'entropie (en valeur absolue) entre les histogrammes des images initiales et de leurs versions jpeg en fonction de la qualité du codage. ● = moyenne, × : moyenne ± écart-type, □ : extréma (min, max).

de la valeur -0.75 bit (cf. figure 7), la dispersion se réduisant progressivement autour de cette valeur. En moyenne, le codage Jpeg augmente la valeur de l'entropie d'une image (ceci est dû à la présence de notre base d'images simples, *i.e.* quasi binaires, dont le codage Jpeg provoque l'apparition de niveaux de gris intermédiaires et donc augmente l'entropie du signal).

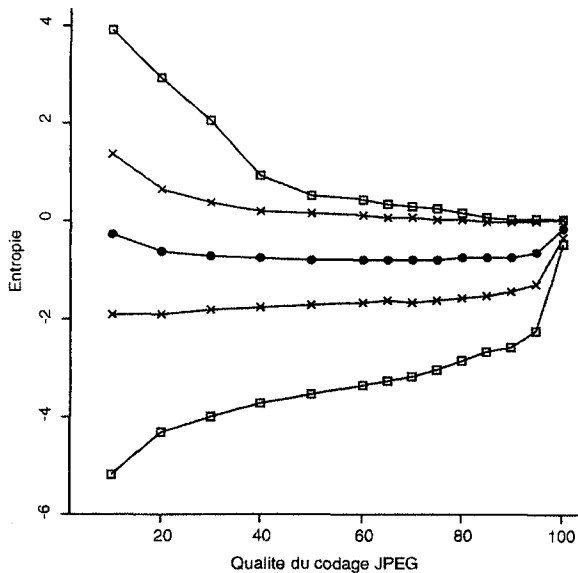


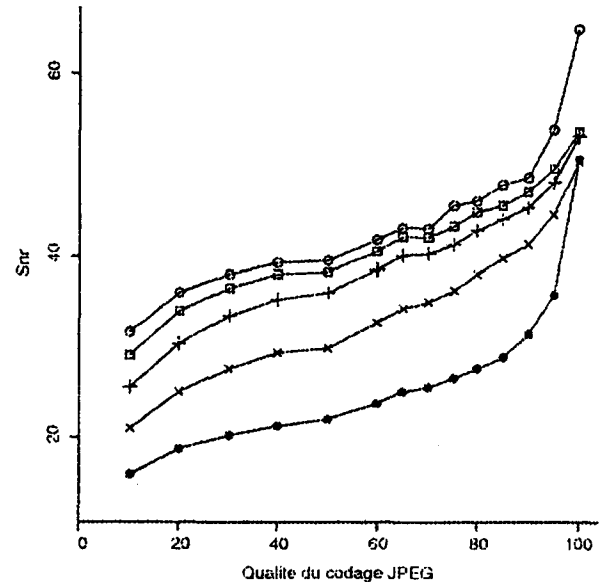
Figure 7. – Différence d'entropie entre les histogrammes des images initiales et de leurs versions jpeg en fonction de la qualité du codage. ● = moyenne, × : moyenne \pm écart-type, □ : extréma (min, max).

3.3. la multirésolution

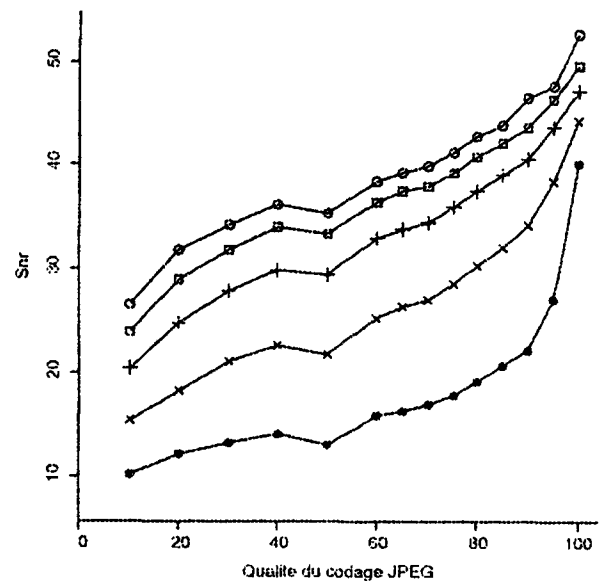
Notre premier test porte sur les cinq premiers niveaux de la pyramide gaussienne de chaque image (le facteur de réduction de la résolution est de 2 dans chaque dimension entre chaque couple de niveaux de résolution). La figure 8a montre l'évolution de la moyenne de l'indicateur *snr*. Il apparaît clairement que la réduction de résolution atténue l'effet du bruit de codage. Cependant, le seuil de robustesse est rapidement atteint (*i.e.* dès 20%). La répercussion de la réduction de la résolution est donc peu importante.

Pour confirmer ce premier résultat, on peut étudier le même indicateur moyen pour les cinq premiers niveaux de la pyramide laplacienne. La figure 8b montre cette évolution. Si le comportement est de même nature (amélioration quand la résolution diminue), les valeurs sont significativement plus faibles indiquant une détérioration dans le domaine fréquentiel. En effet, si l'on considère le seuil de robustesse, il est atteint beaucoup plus tard, successivement 90%, 40%, 20%, 10% et 10%. La décomposition laplacienne est une décomposition en bandes de fréquences. Le premier niveau correspond aux plus hautes fréquences. On vérifie donc que ce sont ces fréquences qui sont les moins bien conservées par le codage Jpeg. Le gain lors d'une réduction de résolution est en moyenne 2db plus fort sur la pyramide laplacienne que sur la pyramide gaussienne.

Par ailleurs, l'amélioration est de plus en plus faible quand on réduit la résolution. Il n'y a pas de discontinuité comme on pourrait s'y attendre compte tenu de l'importance de la taille du bloc dans le codage Jpeg.



(a)



(b)

Figure 8. – Indicateur de robustesse pour les cinq premières résolutions de la pyramide gaussienne (a) et de la pyramide laplacienne (b). ● : 0, × : 1, + : 2 □ : 3, ○ : 4.

3.4. mesure des dérivées et de leurs combinaisons

La figure 9 présente l'indicateur *snr* en moyenne pour les dérivées spatiales partielles. La valeur de l'indicateur pour l'image est ajoutée pour comparaison.

On peut constater que le passage de l'image à ses dérivées conduit à une perte d'un ordre de grandeur sur l'indicateur. Ceci s'explique

Influence du codage Jpeg

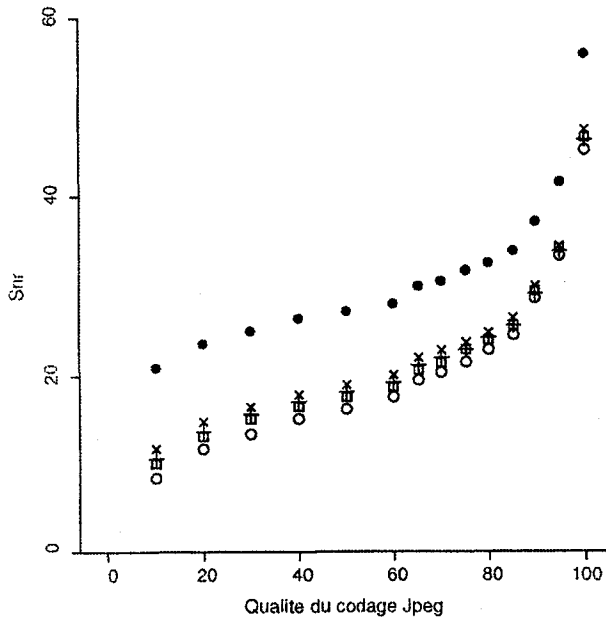


Figure 9. – Mesure moyenne entre les dérivées des images initiales et leurs versions jpeg en fonction de la qualité du codage. ● : I_x , × : $I_x + I_y$, □ : I_x^2 , ○ : I_y^2 .

aisément par le fait bien connu que toute mesure de variation amplifie le bruit. De même, le passage des dérivées premières aux dérivées secondes s'accompagne d'une nouvelle dégradation mais de faible amplitude.

Pour valider en moyenne le seuil de robustesse, il est nécessaire d'avoir une qualité de codage jpeg au moins égale à 65% pour les dérivées premières et 70% pour les dérivées secondes.

Pour analyser plus finement ce comportement, il serait nécessaire de regarder si il y a des inversions d'ordre dans les valeurs des dérivées, *i.e.* déplacement des maxima (qui sont souvent les points recherchés). Nous n'avons pas fait cette étude de manière générale mais nous verrons plus loin des éléments de réponse dans le cas de la détection des points d'intérêt.

L'utilisation la plus courante de ces dérivées est la localisation des contours (qui entre autre s'appuient sur des maxima locaux du gradient, donc des dérivées). La figure 10 nous montre l'évolution de l'indicateur de performance figure de mérite. On constate que les fluctuations peuvent être très importantes et surtout non régulières comme le montre la courbe de la valeur minimale. Ceci est dû à la détérioration importante que subissent certaines images particulières comme les images synthétiques. En moyenne, le comportement est cependant régulier. Il est difficile de prévoir pour cette mesure un seuil de robustesse. Cependant, le seuil de référence (0.69) est atteint pour une qualité moyenne de codage très faible (20%).

Cette plus grande robustesse, par rapport aux dérivées, s'explique par le fait qu'un détecteur de contour prend en compte toute une partie algorithmique qui augmente la robustesse du processus.

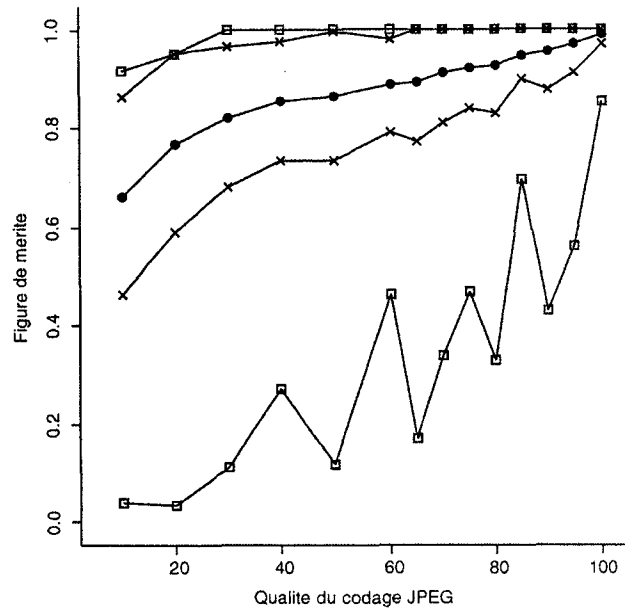


Figure 10. –Indicateur de robustesse entre les contours des images initiales et ceux de leurs versions jpeg en fonction de la qualité du codage. ● : moyenne, × : moyenne ± écart-type, □ : extréma (min, max).

3.5. détection de points d'intérêt

La figure 11 propose un exemple de points d'intérêt sur deux images de la base (un visage et une forme binaire simple – le *whale*).

Schématiquement, ces points d'intérêt sont localisés en des lieux où les dérivées partielles de 1^{er} ordre sont fortes sur les deux directions principales du support de l'image. La densité de ces points dépend donc du contenu de l'image. Ils sont plus nombreux si celle-ci contient des textures. L'indicateur de performance utilisé, la figure de mérite, prend en compte à la fois la variation de cette densité et les écarts de localisation pour chaque point d'intérêt détecté.

La figure 12 résume l'évolution de l'indicateur. Celle-ci est rapide pour les valeurs faibles de la qualité (< 40%). La progression est plus lente par la suite (40%-80%) pour réaccélérer pour les fortes valeurs de la qualité du codage (> 80%).

Si l'on compare avec le même indicateur mesuré pour les cartes de contours (*cf.* Figure 10), on constate que les valeurs obtenues sont nettement plus faibles (bien que l'écart se réduise pour les fortes valeurs de la qualité de codage). Ceci traduit la moindre robustesse, de ce descripteur. L'avantage de celui-ci, un résumé très concis et limité à quelques points, se retourne contre sa robustesse car le moindre point perdu ou rajouté par erreur fait fortement réduire l'indicateur de performance.

Cette mesure intégrant à la fois la variation du nombre de points et leur localisation, nous avons procédé à une mesure spécifique de la variation du nombre de points d'intérêt (en pourcentage).

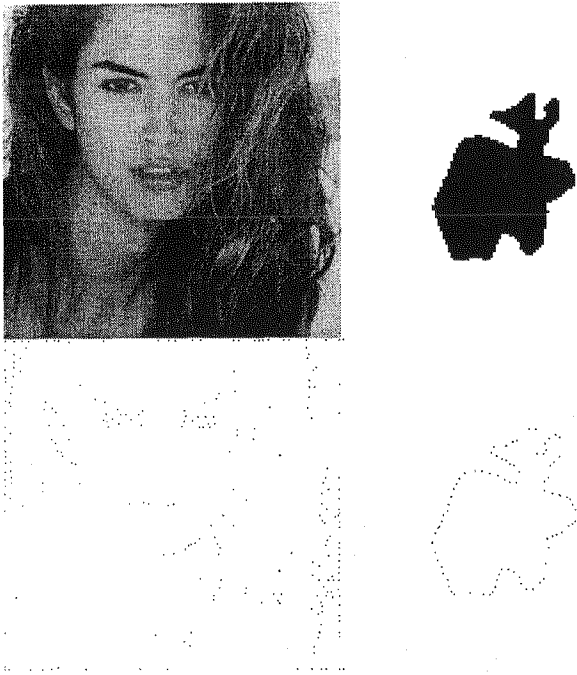


Figure 11. – Exemple de points d'intérêt détectés sur deux images de la base.

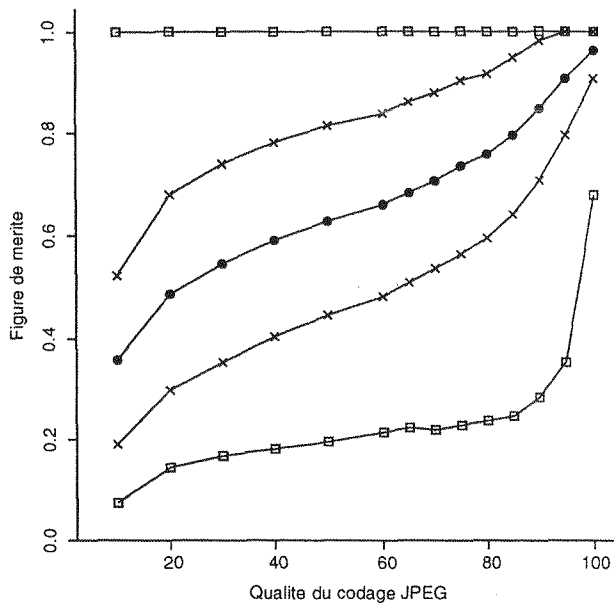


Figure 12. – Indicateur de robustesse pour la détection des points d'intérêt en fonction de la qualité du codage jpeg. ● : moyenne, × : moyenne ± écart-type, □ : extréma (min, max).

Cette mesure est représentée sur la figure 13. Sur notre base d'images, les variations s'équilibrent, la variation moyenne est très proche de 0 (si l'on tient compte de signe de la variation) et il n'y a pas de comportement spécifique en fonction de la qualité de codage. En revanche, prises en valeur absolue, les valeurs ont un comportement moyen qui traduit une amélioration nette avec

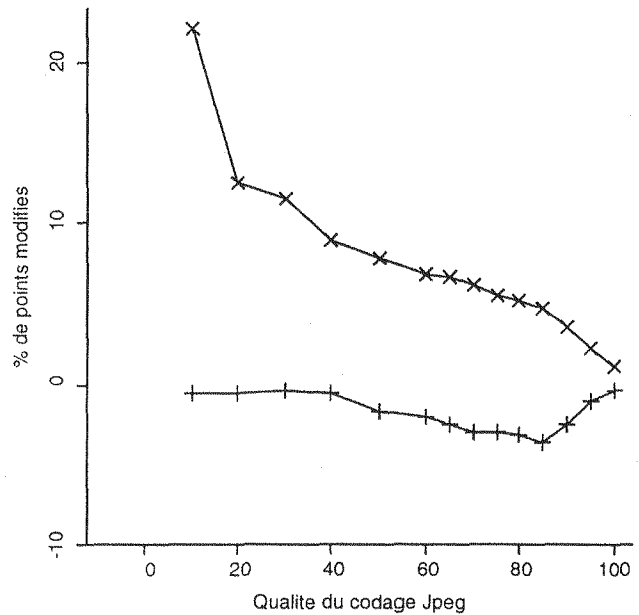


Figure 13. – Variation moyenne (pourcentage) du nombre de points d'intérêt en fonction de la qualité du codage jpeg. + : moyenne des valeurs signées, × : moyenne des valeurs prises en valeur absolue.

l'accroissement de la qualité de codage. Ces valeurs ne sont pas du tout négligeables puisqu'il faut une qualité moyenne de codage d'au moins 85% pour avoir une variation moyenne du nombre de points inférieure à 5%.

Le seuil de robustesse est atteint pour une qualité moyenne de codage au moins égale à 70% (à comparer avec les 20% des contours et lieux de contraste).

3.6. invariants différentiels

Les valeurs des invariants différentiels sont mesurées uniquement aux points d'intérêt détectés dans l'image de référence avec le codage Jpeg. Nous ne tiendrons donc pas compte ici du phénomène de délocalisation.

Les tableaux 2 et 3 contiennent l'évolution des moyennes des indicateurs pour les neuf invariants. Ces deux tableaux sont regroupés sur la figure 14.

Bien entendu, V_0 , constitué par la valeur portée par l'image, *i.e.* le niveau de gris, est l'invariant le plus robuste. Cependant, la mesure indique une perte de 2db en moyenne par rapport à la mesure correspondant à l'image prise dans sa totalité (*cf.* figure 5). Le sous-ensemble de l'image constitué par les points d'intérêt est donc moins robuste pour ce descripteur (ce qui est normal car les points d'intérêt sont caractérisés par des configurations locales riches en variation). Il y a ici une piste à développer sur la prise en compte du critère de codage dans la détermination des points d'intérêt.

Influence du codage Jpeg

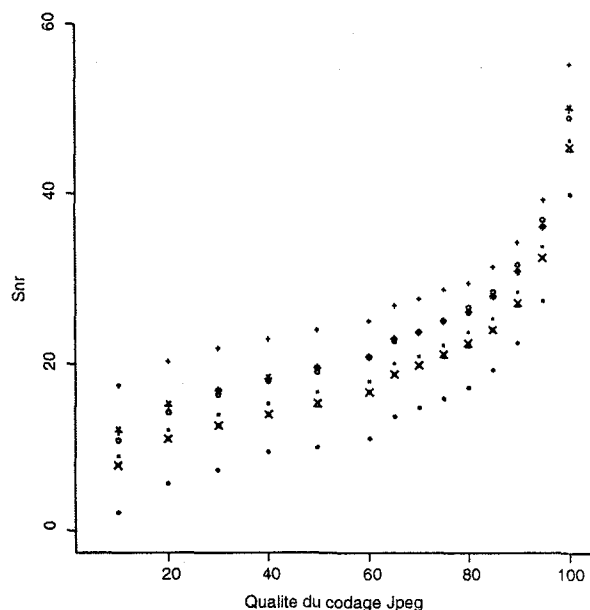


Figure 14. – Indicateur de robustesse pour les neuf invariants différentiels. En fonction de la qualité du codage jpeg. + : V_0 , × : V_1 , □ : V_2 , □ : V_3 , ○ : V_3 , ○ : V_4 , ● : V_5 , · : V_6 , + : V_7 , X : V_8 .

En allant du plus robuste vers le moins robuste, on trouve en deuxième position V_1 , qui est le seul invariant utilisant uniquement les dérivées partielles du premier ordre. Ce résultat est conforme avec l'analyse globale des dérivées présentée précédemment. La valeur de l'indicateur est similaire, même très légèrement supérieure, à celle obtenue pour l'image entière. Caractérisés par de fortes dérivées premières, il est normal que les points d'intérêt soient robustes sur cet invariant différentiel (V_0 n'est qu'un cas extrême peu caractéristique d'une notion différentielle).

Le classement de la robustesse des invariants se poursuit ainsi : $V_4, V_3, V_8, V_2, V_6, V_5, V_7$. La figure 14 montre que V_4 est très proche de V_1 ; V_3, V_8 et V_2 constituent un groupe et V_6, V_5 et V_7 forment un troisième groupe caractérisé par des valeurs beaucoup plus faibles de l'indicateur.

Le tableau 4 indique la qualité moyenne minimale de codage nécessaire à chaque invariant pour atteindre le seuil de robustesse de 22db. On retrouve une décroissance de la robustesse en accord avec l'utilisation de dérivées partielles d'ordre croissant et la complexification des expressions. C'est ce dernier critère qui permet d'expliquer que V_4 soit plus robuste que V_3 (présence de carrés), et V_8 du même ordre que V_2 .

Très schématiquement, il apparaît que les valeurs brutes des invariants sont de robustesses très différentes et que pour les plus complexes d'entre eux, une grande qualité de codage est nécessaire pour en autoriser l'utilisation.

On retrouve ici un paradoxe bien classique de l'analyse d'images : plus un descripteur est pauvre en information (donc peu caractéristique, par exemple V_0), plus il est robuste; plus il est riche (car traduisant des configurations très spécifiques), plus il est instable.

Tableau 2. – Indicateur de robustesse moyen pour les neuf invariants différentiels.

Qualité	V_0	V_1	V_2	V_3	V_4
10	17.26	11.95	7.72	8.69	10.53
20	20.24	15.09	10.86	11.94	14.15
30	21.72	16.84	12.39	13.81	16.23
40	22.73	18.27	13.73	15.13	17.71
50	23.73	19.35	14.95	16.34	18.92
60	24.89	20.70	16.42	17.82	20.65
65	26.68	22.74	18.51	19.83	22.62
70	27.58	23.7	19.62	20.8	23.54
75	28.51	24.83	20.68	22.04	24.99
80	29.51	26.01	22.05	23.46	26.49
85	31.17	27.85	23.82	25.3	28.42
90	34.06	30.69	26.74	28.32	31.54
95	39.11	36.02	32.19	33.7	36.86
100	55.14	49.97	44.87	46.15	48.77

Tableau 3. – Indicateur de robustesse moyen pour les neuf invariants différentiels.

Qualité	V_5	V_6	V_7	V_8
10	2.34	5.4	1.93	8.19
20	5.51	8.08	4.46	11.27
30	7.34	9.67	6.29	12.90
40	9.40	11.24	7.81	14.48
50	9.90	12.37	8.59	15.54
60	10.98	13.67	10.39	16.98
65	13.56	15.86	12.07	18.86
70	14.52	16.52	12.88	19.89
75	15.56	17.9	13.99	21.28
80	17.06	19.19	15.37	22.42
85	19.07	21.45	17.29	24.47
90	22.16	23.96	20.25	27.44
95	27.41	29.45	25.8	32.8
100	39.59	42.93	39.01	46.06

Tableau 4. – Seuil de robustesse pour chaque invariant différentiel.

Invariant	Qualité
V_0	40
V_1	65
V_2	80
V_3	75
V_4	70
V_5	90
V_6	90
V_7	95
V_8	80

Une estimation statistique du bruit de codage est tout à fait envisageable au regard d'une base d'images. Celle-ci peut ainsi permettre de construire des estimations de l'incertitude qui sont très importantes lors d'une mesure de distance entre descripteurs utilisant les invariants (grâce à l'emploi de distances telles que celles de Mahalanobis ou de Bhattacharyya [Fuk 72] qui prennent en compte des matrices de covariances incluant l'incertitude des mesures, ou encore plus explicitement dans des modèles tel que le filtre de Kalman).

4. conclusions

Nous avons, dans cette étude préliminaire, montré l'importance de la prise en compte du bruit de codage dans la mise en place d'un schéma d'indexation d'images. Nous avons appuyé nos remarques sur le cas du codage Jpeg qui, bien que très performant, n'en est pas moins un codage avec pertes. Nous avons testé les principaux descripteurs d'images que l'on retrouve très majoritairement dans les systèmes actuels d'indexation en insistant plus particulièrement sur les descripteurs différentiels.

La prise en compte de la présence d'un bruit de codage dans la conception d'un descripteur est une nécessité pour la généralité des futurs systèmes d'indexation. Peut-on pour autant envisager la construction de descripteurs invariants à un type de codage particulier, comme on le fait pour les descripteurs que l'on rend invariants à une transformation de type changement d'échelle, translation, ...?

Enfin, comme nous l'avons signalé en introduction, il ne s'agit que d'une première étude sur une base d'images très simple. Nous allons maintenant tester deux autres bases et d'autres descripteurs (en particulier la couleur et la texture). La première base sera thématique et constituée de 220 visages pris dans les

mêmes conditions. Notre souhait est de tester sur cette base des descripteurs spécifiques de la détection et reconnaissance automatique de visages. La deuxième sera formée d'images extraites du corpus vidéo proposé par l'INA (Institut National de l'Audiovisuel). Nous testerons dans ce dernier cas l'influence du codage Mpeg (1 et 2), qui, plus encore qu'un codage comme Jpeg, est devenu un passage obligé pour tout système d'indexation et de recherche à vocation grand public sur des données vidéo numériques.

Remerciements

Les auteurs remercient les rapporteurs pour leurs remarques constructives lors de l'élaboration de cet article.

BIBLIOGRAPHIE

- [Bea 78] P. R. Beaudet, Rotationally invariant image operators, *Proc. of 4th Int. Conf. on Pattern Recognition*, pp. 579-583, 1978.
- [DeM 97] M. De Marsicoi, L. Cinque, S. Levialdi, Indexing pictorial documents by their content : a survey of current techniques, *Image and Vision Computing*, 15, pp. 119-141, 1997.
- [Der 87] R. Deriche, Optimal edge detection using recursive filtering, *Proc. 1st Int. Conf. on Computer Vision*, London, pp.501-505, June 8-11.
- [Eud 96] T. Eude, H. Cherifi, Optimisation perceptuelle du codeur Jpeg, *Proc. of CORESA 96*, pp. 220-227, 1996.
- [For 94] W. Förstner, A Framework for Low Level Feature Extraction, *Proc. of 3rd European Conf. on Computer Vision*, pp. 383-394, 1994.
- [Fuk 72] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, New York, Academic Press, 1972.
- [Jai 89] A. K. Jain, *Fundamentals of digital image processing*, Prentice Hall, Information and System Sciences Series, 1989.
- [Jol 93] J. M. Jolion, A. Rosenfeld, *A Pyramid Framework for Early Vision*, Kluwer, 1993.
- [Jol 98] J.M. Jolion, S. Bres, Indexation d'images et bruit de codage, *Rapport de recherche RR 98.01*, Laboratoire Reconnaissance de Formes et Vision, INSA, Lyon, 1998.
- [Meg 98] R. Megret, Mesure de l'erreur introduite par Jpeg sur le filtrage de Gabor, *Rapport de recherche RR 98.03*, Laboratoire Reconnaissance de Formes et Vision, INSA Lyon, 1998.
- [Mor 77] H. Moravec, Toward automatic visual obstacle avoidance, *Proc. of 5th Int. Joint Conf. on Artificial Intelligence*, 584, 1977.
- [Pra 78] W.K. Pratt, *Digital Image Processing*, New York, Wiley, Interscience, 1978.
- [San 97] S. Santini, R. Jain, Image Database are not Databases with Images, *Proc. of 9th Int. Conf. on Image Analysis and Process*, Florence, Italy, Sept. 1997, Lecture Notes in Computer Sciences 1311, Springer, pp. 38-45, 1997.
- [Sch 96] M. Schneider, S.F. Chang, A Robust Content Based Digital Signature for Image Authentication, *Proc. of ICIP 96*, 1996.
- [Sch 97] C. Schmid, R. Mohr, Local Gray-value Invariants for Image Retrieval, *IEEE trans. on Patt. Anal. Mach. Intell.*, 19(5), pp. 530-535, 1997.
- [Wal 91] G.K. Wallace, The JPEG still image compression standard, *Communication of the ACM*, 34, pp. 30-44, 1991.

Manuscrit reçu le 12 mars 1998.

Influence du codage Jpeg

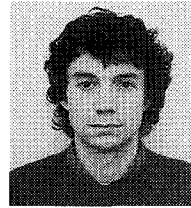
LES AUTEURS

Jean-Michel JOLION



Jean-Michel Jolion est docteur de l'Institut National des Sciences Appliquées (INSA) de Lyon (1987). De 1987 à 1988, il a effectué un séjour au Computer Vision Laboratory (Université du Maryland, USA). De 1988 à 1994, il est Maître de Conférences à l'Université Lyon I. Il est, depuis 1994, Professeur à l'INSA de Lyon. Ses recherches, au sein du laboratoire Reconnaissance de Formes et Vision, portent sur la vision artificielle et plus particulièrement la multirésolution et les pyramides, l'indexation d'images, et l'apport de la théorie des graphes en analyse d'images.

Stéphane BRES



Stéphane Bres est ingénieur de l'Institut National des Télécommunications (INT) d'Evry (1988) et docteur de l'INSA de Lyon (1994). Il est maître de conférences du département informatique de l'INSA de Lyon depuis 1995 et enseigne le traitement du signal et l'analyse numérique. Sa recherche, au sein du laboratoire Reconnaissance de Formes et Vision, porte actuellement sur l'indexation d'images dans le cadre de projets régionaux et européens.