

Codage Basé Région à très bas Débit par la Méthode LAR

O. DEFORGES

L. BEDAT

J. RONSIN

CNRS 6075 URER, INSA de Rennes, 20, Avenue des Buttes de Coësmes, CS 14315, 35043 Rennes
odeforge@insa-rennes.fr, lbedat@insa-rennes.fr, ronsin@insa-rennes.fr

Résumé

Cet article présente un nouveau schéma de codage graduel basé région, performant depuis les très bas jusqu'aux hauts débits. Il est fondé sur une technique de compression basée contenu appelée LAR. Nous introduisons ici une couche d'entrée destinée à représenter et transmettre une segmentation en régions, permettant à la fois d'obtenir une image à très basse résolution, et la possibilité de réaliser ensuite un codage graduel dans les régions d'intérêt. La particularité majeure de notre technique est l'adéquation parfaite entre représentation des régions et structure de données pour coder leur contenu. L'ensemble du codeur allie également une faible complexité avec de très bonnes performances.

Mots clefs

Codage graduel – région d'intérêt – segmentation en régions – grille de sous-échantillonnage non uniforme

1. Introduction

La première génération des codeurs d'image et de vidéo repose essentiellement sur la théorie de l'information. La principale limitation en termes de performances de compression est due aux non-stationarités présentes dans les images naturelles, introduisant des distributions variables de l'information dans l'espace et dans le temps. De plus, les structures de données fixes telles que la grille d'échantillonnage classique cartésienne ou les blocs de données (JPEG ou MPEG) ne permettent pas de traiter de telles non-stationarités.

La seconde génération de codeurs [1] tend à palier ce défaut en représentant les données visuelles en termes de régions, et donc à obtenir une description liée au contenu de l'image. Une région est définie comme une partie convexe de l'image qui partage des caractéristiques communes, alors qu'un objet correspond à un ensemble de régions qui représente une entité sémantiquement cohérente dans l'image [2]. Cette approche tend à réduire le fossé existant entre les systèmes numériques et le système visuel humain (SVH) pour la perception et le traitement des images. Elle apporte également de nouvelles fonctionnalités comme l'interaction entre régions ou objets, la composition de scènes, etc. Une autre possibilité majeure est l'opportunité en compression d'améliorer visuellement les régions d'intérêt de

l'image (ROI pour *Regions Of Interest*), tout en laissant le fond à basse résolution. Cette fonctionnalité suscite beaucoup d'intérêts avec le développement de nouvelles applications à travers les réseaux telles que la consultation de bases de données, la vidéo-surveillance ou encore la vidéo-conférence [3].

Deux types d'information sont nécessaires pour des représentations basées région : la forme et la texture. On rajoute le mouvement pour la vidéo.

Le défaut majeur des approches classiques est que la description de la forme des objets n'est pas corrélée à la structure de représentation de la texture. Dans MPEG-4 par exemple, la technique SA-DCT a été développée pour résoudre ce problème, mais il en résulte un lissage sur les bords des objets, ainsi qu'un coût de calcul supplémentaire.

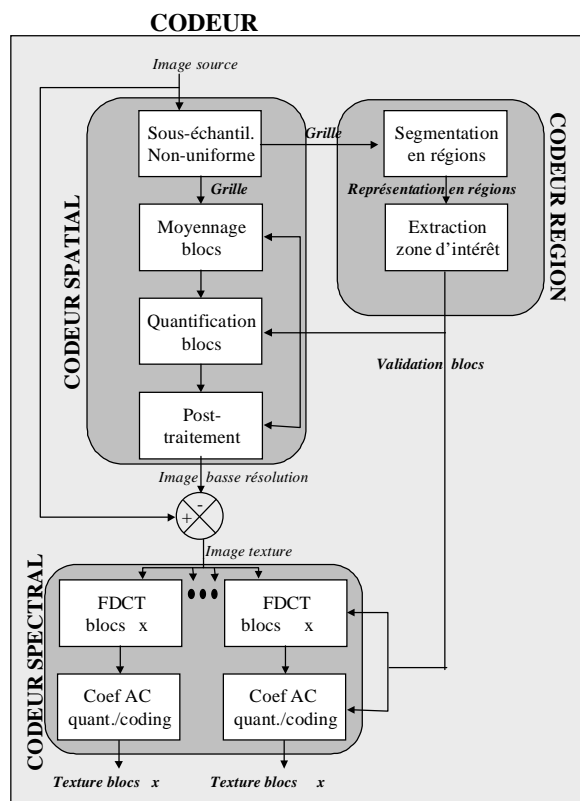
Dans [4] nous avons introduit la méthode LAR comme technique efficace de compression avec ROI évitant ces défauts. A partir d'une image initiale basse résolution issue du codeur spatial LAR, une segmentation en régions est effectuée à la fois au codeur et au décodeur, ce qui élimine tout coût de description des régions. Toutefois le taux de compression est limité par le codage spatial initial. Afin d'atteindre de très bas débit, nous élaborons ici une autre stratégie consistant à coder et transmettre la représentation en régions, puis à utiliser le codeur LAR dans les ROI.

2. La méthode LAR

La méthode LAR (Locally Adaptive Resolution) a été initialement introduite pour le codage d'images multi-niveaux de gris [5]. Le codeur est composé de deux couches (voir figure 1) : un codeur de type spatial permettant les forts taux de compression, et un codeur de type spectral pour coder l'image d'erreur. Le codeur spatial est basé sur une représentation en blocs de l'image en fonction de l'activité (estimée par un gradient morphologique). Typiquement, les tailles de blocs vont de 2x2 à 16x16 par croissance dyadique. Une image basse résolution peut-être obtenue en associant à chaque bloc sa luminance moyenne dans l'image d'origine. Deux informations sont donc à transmettre : d'une part la grille d'échantillonnage, d'autre part les valeurs de luminance encodée par une approche de type MICD adaptative. Une des particularités principale de la méthode est que la quantification tient compte du SVH, en appliquant une forte quantification dans les zones visuellement moins

sensibles (les carrés 2x2 situés sur les frontières), tout en adoptant une quantification fine dans les zones fortement sensibles aux variations de contraste (blocs 8x8 ou 16x16 à l'intérieur des objets).

Grâce notamment à un post-traitement rapide mais efficace, ce codeur spatial fournit une image en sortie qui globalement respecte bien les contours des objets tout en lissant la texture locale (voir figure 2). Cette dernière peut être compressée efficacement pour de meilleures qualités d'image par un codeur spectral basé sur une approche de type DCT à taille de blocs variable [6], dont la taille des blocs est issue du premier codeur. Le critère d'activité initial implique que la taille des blocs est une information sur le contenu de l'image (petits blocs situés sur les contours, grands blocs dans les zones homogènes des objets). Dès lors le codage est rendu sémantiquement progressif en transmettant d'abord la texture des blocs 2x2 (amélioration des contours nets des objets), puis celle des blocs 4x4 (améliorations des zones proches contours ou à faible contour), etc.



Le codeur LAR a ensuite été étendu aux images couleur. Des travaux récents ont pu montrer à travers des tests qualité perceptuelle de qu'un codage d'images couleur basé région par la méthode LAR aboutit à de meilleurs résultats que du JPEG2000 actuel [7].

Dans [4] nous avons étendu la méthode LAR au codage de régions d'intérêt. A partir de l'image

basse résolution issue du codeur spatial, une segmentation en régions était effectuée à la fois au codeur et au décodeur pour une représentation commune sans coût de description supplémentaire. Une région est construite comme union de blocs. La région d'intérêt est alors désignée comme un ensemble de régions élémentaires, et le codeur spectral n'est activé que pour les blocs contenus dans ces régions.

L'approche précédente ne permet toutefois pas d'opérer à très bas débit dès lors que la segmentation s'appuie sur l'image du codeur spatial. Nous proposons ici l'introduction de la phase de segmentation dès le début du processus, et la transmission de cette description au décodeur

3. Codage basé région

3.1. Segmentation en régions

La structure de données de base reste la représentation en blocs de l'image. A partir de ce maillage, le codeur effectue une segmentation en régions comme union de blocs. C'est la seule contrainte à imposer pour le choix de l'algorithme de segmentation. Dans cet article, nous réutilisons le procédé de segmentation en régions de [4] basé sur un critère de taille et de luminance des blocs qui a l'avantage d'être très rapide, mais d'autres techniques plus élaborées peuvent être considérées.

3.2. Compression de la représentation en régions

La représentation en régions doit être transmise pour le décodeur. Elle constituera une image à très basse résolution en associant à chaque région sa luminance moyenne. Nous avons développé ici une nouvelle technique de compression de la forme des régions, non pas basée contour mais contenu. A partir du maillage en blocs, l'étiquette des blocs est obtenue en fournissant la position du premier bloc possédant la même étiquette, suivant le voisinage défini figure 3.a. Noter que comme les blocs sont de taille différente, la connectivité n'est pas limitée à 4 en voisinage causal.

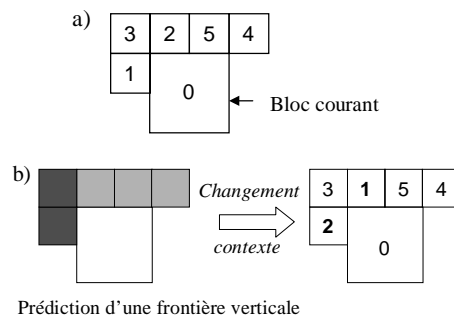


Figure 3 : Voisinage et changement de contexte

Nous définissons ainsi la fonction connectivité C :

$$\begin{cases} C(p_0) = 0 \text{ si } E(p_0) \neq E(p_x), \forall x > 0 \\ C(p_0) = x \text{ si } E(p_0) = E(p_x) \\ \quad \text{et } x < y, \forall y \neq x \text{ et } / E(p_0) = E(p_y) \end{cases}$$

avec p_x la position en x selon le voisinage fig. 3.a et $E(p_x)$ la valeur d'étiquette en p_x .

Les positions étant ensuite encodées par un codeur arithmétique, il est intéressant pour une meilleure efficacité de posséder des probabilités différentes. Ainsi nous avons défini un contexte prédictif de positions pour favoriser la position 1 (voir figure 3.b). Quatre contextes de position ont alors été définis (hypothèse de ligne horizontale, verticale, diagonales à $+45^\circ$ et -45°).

Ce codage des régions a été comparé à celui obtenu par une technique markovienne basée contour proposée dans [7] et considérée très performante. Sur l'image 4.a présentant 130 régions, notre coût en codage est de 0.07 bpp pour 0.068 par l'approche Markovienne. Cette méthode ne peut s'appliquer bien sûr qu'en connaissant le pavage en bloc (0.03 bpp), mais celui-ci fait partie intégrante du codeur LAR. L'intérêt majeur de notre approche ici est qu'elle ne requiert qu'un seul balayage régulier de l'image, et par conséquent est tout à fait à propos dans un contexte temps-réel.

3.3. Codage graduel de régions d'intérêt

La figure 4 montre un exemple de codage graduel de régions d'intérêt, avec un fond de l'image restant à très basse résolution. La raffinement de l'image possède six niveaux graduels de qualité:

- premier niveau avec la description en régions, et la luminance associée à chaque région (figure 4.a),
- second niveau avec le codeur spatial, la luminance associée à chaque bloc (figure 4.c),
- quatre autres niveaux avec le codeur spectral pour coder la texture locale, d'abord pour les blocs 2×2 afin de raffiner les contours (figure 4.d), puis les 4×4 , 8×8 , et enfin 16×16 pour raffiner l'intérieur des objets (figure 4.e).

Contrairement aux méthodes standard (JPEG2000 ou MPEG4), la description de la région d'intérêt partage la même topologie que la structure de données utilisée pour coder le contenu. De ce fait, une ROI agit directement comme un masque des blocs à prendre en compte dans la suite des codeurs spatiaux et spectraux. Un effet direct est qu'il n'y a pas d'effet de bord sur les frontières des ROI qui correspondent complètement aux bords réels de l'image.

La ROI elle-même est désignée à travers les étiquettes des régions qui la composent, et ne constitue donc qu'un très faible surcoût.

A noter que le PSNR dans la ROI reste faible comparé à la qualité observée. Ceci est une particularité du codeur LAR qui tend à placer les

erreurs de quantification dans les zones moins sensibles du système visuel humain.

Par ailleurs, cette technique s'avère globalement très rapide: une image 512×512 peut être traitée en une seconde environ, sur un PC 450 MHz, cela sans optimisation du code.

4. Conclusion

Nous avons présenté dans cet article une technique originale basée région, appelée LAR, de représentation et de compression d'image. Le codeur LAR est à l'origine une méthode basée contenu, mettant en œuvre un sous-échantillonnage non uniforme sur une grille carrée. La segmentation en régions est réalisée sur cette grille, permettant d'obtenir une adéquation complète entre supports des régions et codage de la texture associée. Des travaux précédents avaient montré qu'une stratégie possible était de transmettre une image basse résolution, puis d'effectuer la segmentation à la fois au codeur et au décodeur, n'entraînant aucun surcoût pour la description des régions. Afin d'atteindre de plus forts taux de compression, nous proposons ici la transmission de la représentation en régions en premier lieu. Une technique très rapide de codage de cette représentation a été développée, en tenant compte des spécificités du codeur LAR. Une application directe en est la représentation graduelle avec région d'intérêt dont la mise en œuvre est immédiate avec notre approche.

Références

- [1] M. Kunt, A. Ikonopoulos, M. Kocher, "Second generation Image coding techniques", *Proceedings of the IEEE*, Vo. 73, No. 4, April 1985, pp 549-575.
- [2] Ebrahimi T., Kunt M., "Visual data compression for multimedia applications", *Proceedings of the IEEE*, Vo. 86, No. 6, June 1998, pp 1109-1125.
- [3] N. Peterfreund, Y. Zeevi, "Nonuniform image representation in area-of-interest systems", *IEEE Trans. Image Processing*, Vol. 4, No 9, sept. 1995, pp 1202-1212.
- [4] O. Déforges, J. Ronsin, "Codage de région d'intérêt par la méthode LAR", CORESA'00, Poitiers, 19-20 octobre 2000
- [5] O. Déforges, L. Bédard, J. Ronsin, "La méthode LAR pour le codage basé région des images couleur", *Gretsi'01*, Toulouse, septembre 2001.
- [6] C. Chen, "Adaptive transform coding via quad-tree based variable block-size DCT", *ICASSP'89*, Glasgow, 1989, pp 1854-1856.
- [7] S. Pateux, C. Labit, "Efficient way of coding arbitrary geometric partition for region-based video sequence coding", *VCIP'97*, San-José, USA, feb.1997, pp 742-753.

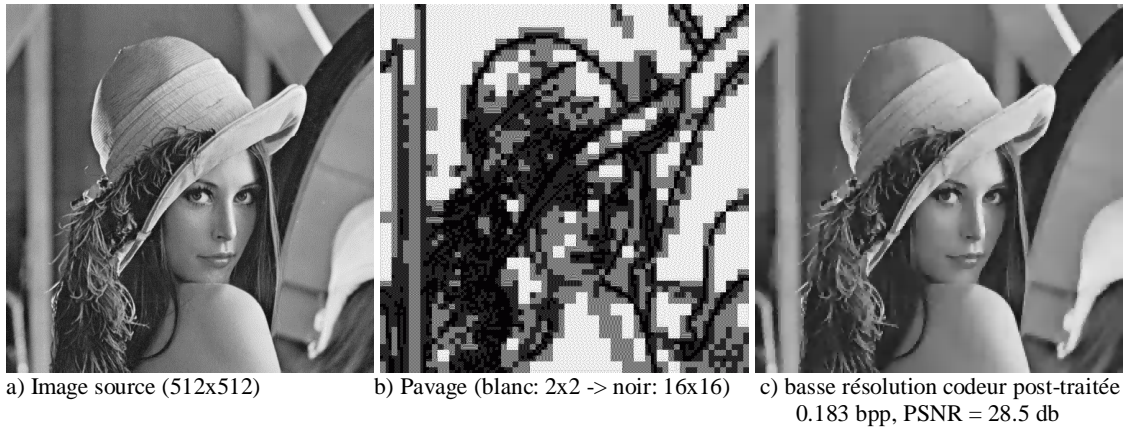


Figure 2 : Images résultantes du codeur spatial

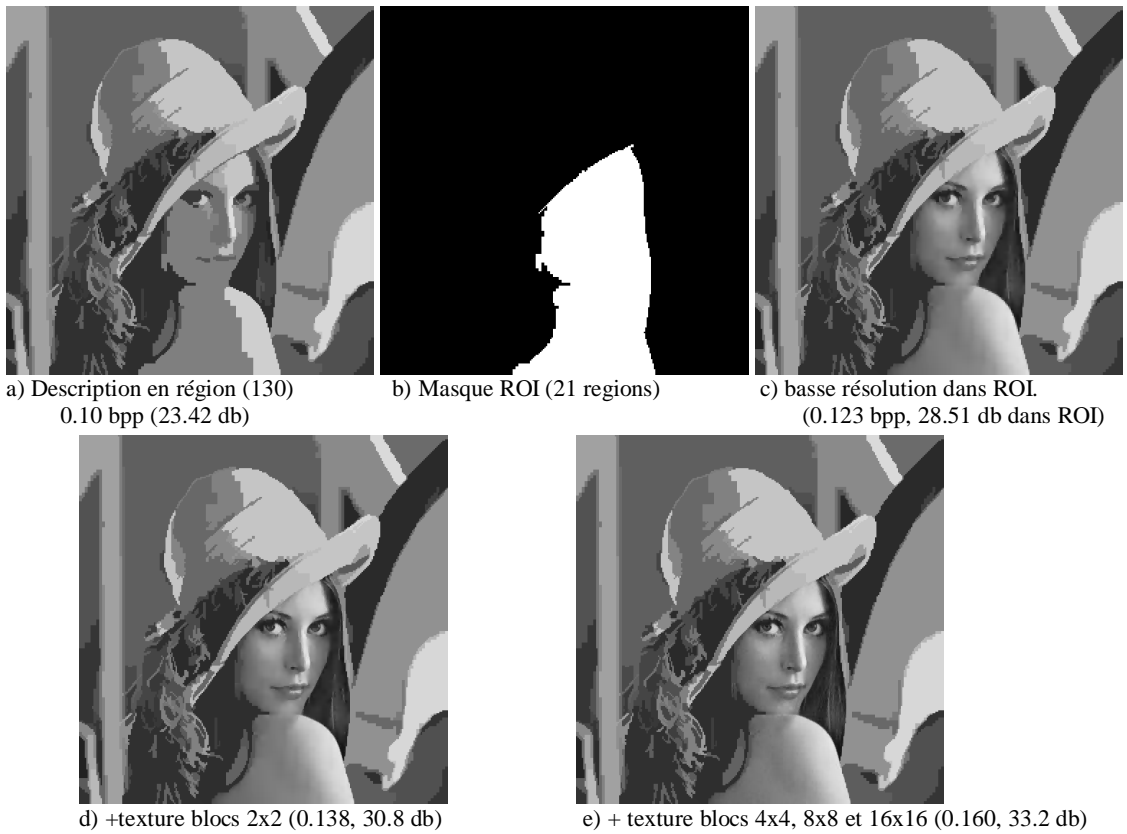


Figure 4 : Exemple de codage graduel dans une région d'intérêt