Un algorithme de compression d'images basé sur une représentation mixte par lignes de niveau et paquets d'ondelettes

Jacques FROMENT

PRISME, UFR Math. et Info., Université Paris 5 R. Descartes
45 rue des Saints-Pères, 75270 Paris cedex 06, France
et
CMLA, Ecole Normale Supérieure de Cachan,
61 avenue du Président Wilson, 94235 Cachan cedex, France.
Jacques.Froment@math-info.univ-paris5.fr, Jacques.Froment@cmla.ens-cachan.fr

Résumé – Nous présentons un algorithme de compression d'images basé sur un modèle du type $u = v+w$, où l'image originale u est décomposée en un *sketch* v et en un résidu w. Le *sketch* contient les principales lignes de contour tandis que l'information portée par le résidu est essentiellement du domaine des microtextures. Cette séparation de l'information pertinente permet d'utiliser une représentation adaptée à chaque structure. La géométrie des lignes de contours est détectée et codée en utilisant les lignes de niveau, tandis que les microtextures sont compressées dans une base de paquets d'ondelettes. On remédie ainsi au principal défaut des représentations harmoniques classiques, la génération d'oscillations au voisinage des contours.

Abstract – We introduce a structured image compression scheme based on a $u = v + w$ model, where the orignal image u is decomposed between a sketch v and a residue w. The sketch contains the main edge curves while the information carried by the residue is mostly composed by microtextures. This split of the meaningful information allows the use of a representation adapted to each structure. The edge curve geometry is detected and coded using level lines, while microtextures are compressed by means of a wavelet packet basis. In this way, one can avoid the main drawback of classical harmonic representations, the occurrence of ringing artefacts near the edges.

1 Introduction

Cette dernière décennie a connu une explosion des méthodes de compression d'images utilisant des représentations harmoniques par base d'ondelettes ou de paquets d'ondelettes. Le nouveau standard international JPEG - 2000 [13] entérine ces années de recherche en remplaçant l'ancienne transformée en cosinus par des transformées en ondelettes. L'avantage décisif d'une représentation en ondelettes sur une représentation purement fréquencielle tient à la localisation des atomes dans le plan espace-fréquence : une image présente de nombreuses singularités et, au voisinage de celles-ci, la décroissance des coefficients en ondelettes est plus rapide que celle des coefficients de Fourier. Au prix d'un accroissement de complexité, les paquets d'ondelettes améliorent le pavage de ce plan espace-fréquence en l'adaptant aux caractéristiques de l'image. Obtenons-nous alors une représentation atomique optimale pour la compression de l'information contenue dans les images? Non, car toute représentation espace-fréquence reste assujettie au principe d'incertitude d'Heisenberg qui interdit une localisation parfaite en espace des atomes conservant une localisation fréquencielle. Un contour idéal, défini comme la trace sur l'image de l'arête d'un objet, est une structure purement géométrique qui échappe ainsi à l'analyse harmonique classique. Alors que sur une image correctement échantillonnée un contour serait légèrement flou, dans la pratique on observe qu'un contour est suffisamment singulier pour engendrer sur son voisinage des oscillations, lors de l'opération de seuillage et de quantification des coefficients. Bien que d'intensité moindre ce phénomène est semblable à celui de Gibbs, lequel apparaît lors de l'opération de seuillage des coefficients de Fourier d'une fonction non régulière. Aux taux de compression élevés, les structures oscillantes forment des artefacts très génants pour la perception visuelle. Aux taux plus faibles, bien qu'apparemment imperceptibles, elles peuvent apparaître lors de post-traitements tels qu'une déconvolution, une détection de contour ou un rehaussement de contraste.

Afin de pallier ces défaillances, on peut définir de nouvelles représentations inspirées des ondelettes classiques mais telles que les atomes possèdent une meilleure localisation géométrique [2, 7, 16]. L'approche que nous présentons ici est différente au sens où elle propose une représentation purement géométrique des lignes de contour. Elle consiste d'abord à remarquer que dans une image naturelle l'information pertinente peut être grossièrement séparée entre un *sketch*, c'est-à-dire une esquisse semblable à une image de bandes dessinées, régulière à l'intérieur de chaque région dont les bords correspondent aux contours perceptuels, et entre les microtextures qui sont définies comme la différence sur chaque région de l'image initiale

avec celle de la bande dessinée [4, 8, 12]. La justification d'un tel découpage n'est pas seulement visuelle, elle procède aussi de propriétés mathématiques : de nombreux éléments portent à croire qu'une image naturelle n'est pas à variation bornée suite à la présence de régions microtexturées [1, 8]. D'autre part, si les microtextures ne possèdent pas (à une échelle suffisamment grossière) de structures géométriques, elles sont bien localisées en fréquence par la répétition de motifs élémentaires. Ces remarques nous ont conduit à proposer dans [8] un modèle pour toute image naturelle u de la forme $u = v + w$, où le *sketch* v est à variation bornée mais où l'image w des microtextures ne l'est pas. Nous avons montré qu'un choix de *sketch* v consistant à la fois pour le modèle mathématique et pour la perception visuelle était donné par une segmentation morphologique de l'image, où les bords des régions étaient les lignes de niveau perceptuellement les plus significatives [5], et où les valeurs de v à l'intérieur des régions étaient obtenues par interpolation Lipschitz des valeurs aux bords [3, 6]. L'objet de notre présente communication est d'exposer une nouvelle méthode de compression basée sur le modèle $u = v + w$, où la composante géométrique du sketch est codée sans perte et où l'image de microtextures est compressée par quantification dans une base de paquets d'ondelettes. Le lecteur pourra se reporter à [10] pour une présentation plus détaillée.

2 Segmentation par lignes de niveau

Nous recherchons une segmentation $\mathcal{P} = (P_i)_{i=1,\dots,n}$ telle que les lignes de contour perceptuel correspondent le mieux possible au bord des régions P_i. Des travaux récents montrent que ces lignes de contour peuvent être représentées à l'aide de portions de lignes de niveau, lesquelles joueraient le rôle d'atomes de la perception [5]. Afin d'obtenir des régions, nous conservons la totalité d'une ligne de niveau contenant des atomes suffisamment significatifs : l'ensemble correspondant que nous appellerons "forme" ("*shape*" suivant la terminologie introduite dans [15]), est défini comme la région à l'intérieur de la frontière donnée par la ligne de niveau (qui est une courbe de Jordan fermée ou rencontrant le bord de l'image). C'est une composante connexe d'un ensemble de niveau inférieur $[u < \lambda] = \{x \in \Omega / u(x) < \lambda\}$ ou supérieur $[u > \mu] = \{x \in \Omega / u(x) > \mu\}$, pour $\Omega \subset \mathbb{R}^2$ le domaine de l'image u et λ le niveau considéré, la composante étant "remplie" pour "boucher les trous" dans le cas où elle n'est pas simplement connexe. Nous noterons une telle forme $[u < \lambda]^k$ ou $[u > \mu]^k$ selon qu'elle est issue d'un ensemble de niveau inférieur ou supérieur. Nous avons ainsi

$$\mathcal{P} = (P_i)_i = \{[u < \lambda]^k\}_{k,\lambda} \cup \{[u > \mu]^k\}_{k,\mu}, \qquad (1)$$

les valeurs λ, μ et les indices k devant être correctement choisis de manière à ce que les bords ∂P_i coïncident au mieux avec les contours perceptuels.

La sélection des formes les plus significatives utilise deux critères complémentaires. Le premier considère le rapport isopérimétrique de la forme S, soit la valeur

$$i(S) = \frac{\mathrm{per}^2(S)}{a(S)}, \qquad (2)$$

pour $a(S)$ l'aire de S et $\mathrm{per}(S)$ son périmètre. Dans [9] nous avons montré que le rapport isopérimétrique, en fournissant une mesure de l'oscillation des lignes de niveau, permettait de discriminer celles qui participent à une ligne de contour (peu oscillantes) de celles qui ne portent pas d'information géométrique significative (très oscillantes). Le principal défaut de ce critère est sa non-localité, une ligne de niveau pouvant être associée à un contour sur une partie seulement. Le second critère est de nature locale. Il considère les singularités de type jonction en T entre lignes de niveau. La présence de jonctions en T révèle des évènements physiques dans la génération de l'image, tels des phénomènes d'occlusion ou de transparence [5]. Au voisinage du bord d'un objet, de multiples jonctions entre le fond et la face de l'objet se produisent, ce qui explique que le nombre de jonctions contenu dans une ligne de niveau puisse être considéré comme un indice de pertinence géométrique [8].

La mise en œuvre de notre méthode passe par la définition d'une fonction coût $\mathcal{C}$, telle que si la forme S_1 est plus pertinente que la forme S_2 nous avons $\mathcal{C}(S_1) < \mathcal{C}(S_2)$. Une fonction coût donnant de bons résultats est

$$\mathcal{C}(\mathcal{S}) = \begin{cases} +\infty & \text{si } i(S) > 100 \text{ ou si } \mathrm{per}(S) < 20; \\ -\mathcal{N}(\mathcal{S}) & \text{sinon,} \end{cases} \qquad (3)$$

$\mathcal{N}(\mathcal{S})$ étant le nombre de jonctions en T sur le bord de S. Le test $\mathrm{per}(S) < 20$, en implémentant un filtre de grain [15], permet de réduire de manière très importante le nombre de formes à considérer, sans enlever aucune information importante. Les valeurs 100 et 20 donnent des résultats acceptables sur la plupart des images, bien que des valeurs adaptées puisse améliorer la qualité de la segmentation. L'algorithme de segmentation consiste simplement à décomposer l'image u en ses différentes formes et à sélectionner les plus significatives à l'aide de la fonction coût, la quantité de formes enregistrées étant donnée par le débit allouée pour le codage du *sketch*. Le lecteur désirant expérimenter ces outils trouvera une implémentation rapide de la décomposition en formes (FLST [15]) et de la détection des jonctions en T dans [11].

3 Représentation du *sketch*

Le *sketch* est défini à partir de la segmentation (données géométriques), à laquelle on ajoute des informations sur les niveaux de gris de l'image u afin de reconstruire une image.

De manière à préserver la géométrie des contours, le bord des formes sélectionnées est codé sans perte, en utilisant un codage de chaîne classique (Freeman). La structure particulière des ensembles de niveau dont sont issus ces bords permet cependant quelques optimisations, notamment dans le cas de courbes se recouvrant partiellement [10].

Afin de minimiser la quantité de niveaux de gris à coder, seules deux valeurs de gris sont conservées pour chaque forme : la valeur moyenne sur le côté intérieur de la forme et celle sur le côté extérieur. Ces valeurs sont quantifiées puis codées en utilisant un codage arithmétique prédictif.

Un point $x \in \Omega$ peut appartenir au côté (intérieur ou extérieur) de plusieurs formes : il peut donc y avoir, durant la phase de décodage, ambiguïté quant au niveau à attribuer à x. Grâce à la propriété d'inclusion des ensembles de niveau, il est cependant toujours possible de retrouver la valeur la plus proche de $u(x)$, que nous noterons $g(x)$. Lorsque x n'appartient à aucun côté d'aucune forme, on pose $g(x) = 0$.

La reconstruction d'une image v, à partir des bords des formes $(P_i)_i$ et des valeurs en ces bords données par la graine g, est un problème d'interpolation. Nous utilisons le modèle AMLE (*Absolute Minimizing Lipschitz Extension)* qui permet une interpolation régulière à l'intérieur des formes, quelque soient les conditions aux bords [3, 6]. L'interpolation est réalisée en résolvant une équation d'évolution par un schéma d'Euler implicite.

4 Représentation du résidu

Le résidu w est défini par $w = u - v$. Dans le cas où un nombre suffisant de lignes de niveau peut être considéré comme significatif, le résidu est uniquement composé de microtextures. Cette situation n'est cependant pas réaliste, car le débit du *sketch* serait trop important. Dans la pratique on observe donc une information géométrique, due à la représentation imparfaite des lignes de contours, ainsi qu'une information basse fréquence, due à la différence moyenne de niveaux de gris entre les régions originales et celles reconstruites. Notre approche nous permet cependant de considérer ces informations comme non significatives. Comme les microtextures consituent la composante essentielle du résidu, nous devons utiliser une représentation prioritairement bien adaptée aux motifs très oscillants. Les bases de paquets d'ondelettes offrent une localisation dans les hautes fréquences bien meilleures que les bases d'ondelettes, puisque dans ces dernières la hauteur des rectangles d'Heisenberg est nécessairement croissante avec la fréquence. Nous avons ainsi choisi de représenter le résidu dans une base de paquets d'ondelettes qui est optimale dans le contexte de la compression. L'algorithme de choix de la meilleure base et le codage des coefficients correspondent au codeur défini par F. Meyer, A. Averbuch et J-O. Strömberg dans [14].

5 Illustration de la méthode

Notre méthode de compression structurée est illustrée à la figure 1. Nous l'avons comparée avec l'algorithme de compression par paquets d'ondelettes seul (image G). Bien que la compression structurée ne permette pas d'atteindre le même rapport signal à bruit, nous notons que la qualité de l'image reconstruite est supérieure (comparer F à G). Pour l'essentiel, cela tient à l'absence d'oscillations au voisinage des contours due à la représentation géométrique par lignes de niveau. Cet absence d'artefact est rendu encore plus visible si on applique à l'image décodée certains post-traitement comme un filtre passe-haut, ici un filtre "plus net" (comparer H à I).

Références

[1] L. Alvarez, Y. Gousseau, and J-M. Morel. Scales in natural images and a consequence on their bounded variation norm. In *Scale-Space Theories in Computer Vision*, pages 247–258. Lecture Notes in Computer Science 1682, 1999.

[2] E. Candès and D. Donoho. Curvelets: a surprisingly effective nonadaptive representation of objects with edges. In L. Schumaker A. Cohen, C. Rabut, editor, *Curve and Surface Fitting : Saint-Malo 1999*. Vanderbilt Univ. Press, 2000.

[3] F. Cao. Absolutely minimizing lipschitz extension with discontinuous boundary data. *Note aux C.R. Acad. Sci. Paris*, t.327(I):563–568, 1998.

[4] S. Carlsson. Sketch based coding of grey level images. *Signal Processing North-Holland*, 15(1):57–83, 1988.

[5] V. Caselles, B. Coll, and J-M. Morel. Topographic maps and local contrast changes in natural images. *Int. J. Comp. Vision*, 33(1):5–27, 1999.

[6] V. Caselles, J-M. Morel, and C. Sbert. An axiomatic approach to image interpolation. *IEEE Trans. on Image Proc.*, 7(3):376–386, 1998.

[7] M. N. Do and M. Vetterli. Orthonormal finite ridgelet transform for image compression. In *Proc. of ICIP'2000*, volume 2, pages 367–370, 2000.

[8] J. Froment. A functional analysis model for natural images permitting structured compression. *ESAIM:COCV Control, Opt. and Cal. of Var.*, 4:473–495, 1999.

[9] J. Froment. Perceptible level lines and isoperimetric ratio. In *Proc. of ICIP'2000*, volume 2, pages 112–115, 2000.

[10] J. Froment. Image compression through level lines and wavelets packets. Manuscript. `http://www.math-info.univ-paris5.fr/~jf/publi.html`, 2001.

[11] J. Froment and L. Moisan (ed.). Megawave2 v.2.01. A free and open-source Unix image processing software for reproducible research, available at `http://www.cmla.ens-cachan.fr`, 2001.

[12] J. Froment and S. Mallat. Second generation compact image coding with wavelets. In C.K. Chui, editor, *Wavelets - A Tutorial in Theory and Applications*, pages 655–678. Academic Press, 1992.

[13] M.J. Gormish, D. Lee, and M.W. Marcellin. Jpeg 2000: overview, architecture and applications. In *Proc. of ICIP'2000*, volume 2, pages 29–32, 2000.

[14] F. Meyer, A. Averbuch, and J-O. Stromberg. Fast adaptive wavelet packet image compression. *IEEE Trans. on Image Proc.*, 9(5), 2000.

[15] P. Monasse and F. Guichard. Fast computation of a contrast-invariant image representation. *IEEE Trans. on Image Proc.*, 9(5):860–872, 2000.

[16] E. Le Pennec and S. Mallat. Image compression with geometrical wavelets. In *Proc. of ICIP'2000*, volume 1, pages 661–664, 2000.

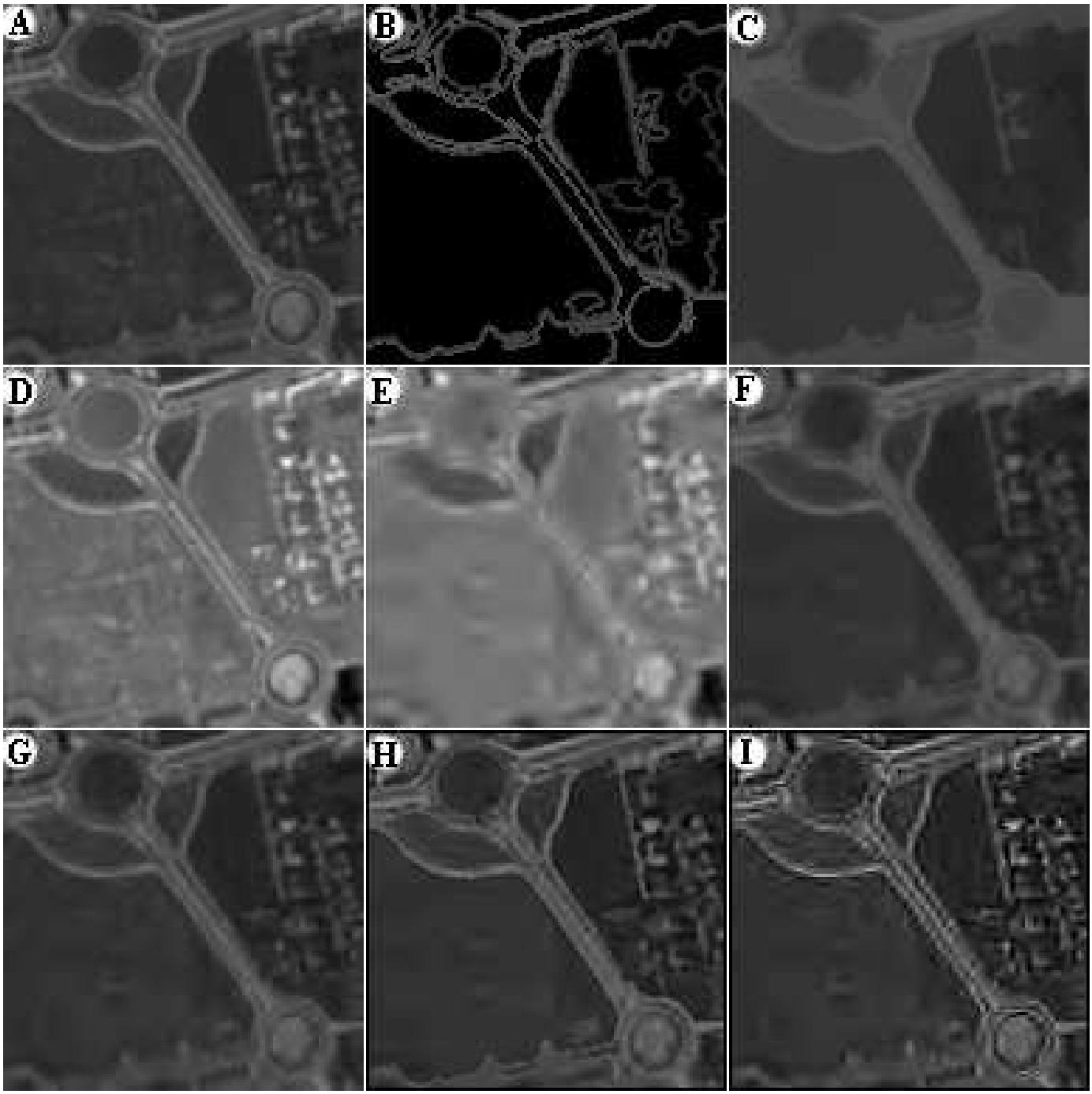

FIG. 1: Illustration de la méthode de compression, pour un taux de compression de 20 (débit de 0.4 bpp). **A** : image originale 128×128, 8 bpp (vue satellite simulée, crédit CNES). **B** : carte des valeurs aux bords utilisées pour reconstruire le *sketch* (c'est la graine g) ; codage de la géométrie des lignes de niveau sur 0.198 bpp et codage des niveaux de gris sur 0.002 bpp, soit un total de 0.2 bpp. **C** : sketch v obtenu à partir de B par AMLE ; les principales structures (réseau routier) sont bien reconstruites tandis que seule une petite partie du lotissement est conservée. **D** : résidu A-C ; les "microtextures" sont ici essentiellement composées du champ, et du lotissement non représenté dans le *sketch*. **E** : résidu décodé après compression par le codeur en paquets d'ondelettes de F. Meyer, A. Averbuch et J-O. Strömberg; débit de 0.2 bpp ; noter la bonne restitution du lotissement. **F** : C+E décodage de l'image A compressée par notre méthode ; débit de $0.2 + 0.2 = 0.4$ bpp, PSNR $= 33.8$ db. **G** : pour comparaison, image A décodée après compression par le codeur de F. Meyer, A. Averbuch et J-O. Strömberg ; débit de 0.4 bpp, PSNR $= 34.8$ db. Si le rapport signal à bruit est supérieur, la qualité visuelle est inférieure à cause des oscillations au voisinage des contours (pseudo-phénomène de Gibbs). **H** : image F filtrée par un filtre passe-haut "plus net", pour mettre en évidence d'éventuels artefacts. **I** : image G filtrée par le même filtre, pour mettre en évidence le pseudo-phénomène de Gibbs. Les chaînes de traitement de nombreux systèmes d'imagerie, comme l'imagerie satellitaire, imposent souvent aux images décodées des post-traitements de ce type.