

DECODAGE OPTIMAL ORIENTE OBJET POUR LA SUPPRESSION DES EFFETS DE BLOC DANS LES SEQUENCES DV ET MPEG-2

Joël JUNG, Marc ANTONINI, Michel BARLAUD

Laboratoire I3S, CNRS UPRES-A 6070

Université de Nice-Sophia Antipolis,

Les Algorithmes / Bâtiment Euclide B

2000 route des Lucioles - BP 121

06903 Sophia Antipolis Cedex

jung@i3s.unice.fr, am@i3s.unice.fr, Michel.Barlaud@i3s.unice.fr

Résumé – Cet article propose une méthode de décodage permettant d'améliorer la qualité visuelle des séquences vidéos. Deux objectifs sont visés : la réduction des effets de bloc dans les séquences de type DV, M-JPEG et MPEG d'une part, la restitution des blocs perdus (dropout) par le système d'acquisition ou par le réseau d'autre part. La minimisation de critères appropriés permet une estimation et un traitement simultané du fond de la séquence et une segmentation spatio-temporelle des objets en mouvement. Un suivi temporel des objets permet ensuite de traiter chaque objet de manière indépendante et adaptée. La qualité de la séquence ainsi restituée par le décodeur est largement améliorée.

1. Introduction

L'arrivée des dernières technologies numériques engendre un besoin sans cesse croissant de vitesse et d'espace de stockage. C'est pourquoi la compression est devenue incontournable pour les applications grand public. Citons à titre d'exemples les appareils-photos numériques, qui codent les images en JPEG, les caméscopes numériques, qui compressent les séquences au format DV, un dérivé de M-JPEG, ou encore la télévision numérique et le DVD, qui ont adopté le format de compression MPEG-2, sans oublier bien sûr le réseau Internet, où images et séquences transitent compressées.

Parallèlement aux artefacts traditionnels de la compression par DCT se sont ajoutés de nouveaux artefacts : les dropout. Les codeurs temps-réel des caméscopes numériques récents génèrent des pertes sur les images : pertes de blocs, décalages de bandes. La télévision numérique, ainsi que le DVD subissent également des artefacts de ce type. Il est donc nécessaire d'améliorer le décodage des séquences pour permettre, soit une meilleure qualité à débit équivalent, soit un débit plus faible, à qualité égale ou supérieure.

De nombreuses approches ont été proposées pour réduire les effets de bloc dans les images fixes [5][7][9]. De nombreux travaux ont concerné la segmentation dans les séquences [4][8], mais peu ont réellement combiné les deux approches, avec pour objectif final l'amélioration de la qualité de la séquence, en réduisant les effets de bloc.

Nous proposons donc dans cet article une nouvelle méthode de décodage adaptée aux formats de compression utilisant la transformée DCT. Cette méthode traite le problème

spatialement et temporellement, pour obtenir une amélioration significative de qualité de la séquence, et repose sur deux idées :

- traiter simultanément les problèmes de suppression de blocs et de segmentation du mouvement,
- intégrer la notion d'objet, afin de permettre une approche différente pour le traitement du fond et de chaque objet.

De plus, cette méthode présente la particularité de traiter efficacement le problème des dropout dans les séquences vidéos. Les blocs perdus sont parfaitement restitués lorsque la séquence est décodée par la méthode proposée. Ce principe s'applique quel que soit le format de la séquence à décoder.

Dans le deuxième paragraphe, nous décrivons les deux parties de l'algorithme : l'estimation simultanée du fond et la segmentation du mouvement d'une part, le traitement des objets d'autre part. Dans le troisième paragraphe, nous traitons le problème des dropout. Enfin, nous présentons quelques résultats expérimentaux dans le paragraphe 4. Ceux-ci montrent que les effets de bloc et les dropout ont été fortement atténués par rapport à la méthode de décodage standard, ce qui permet d'obtenir une séquence de meilleure qualité.

2. Méthode de décodage

Trois étapes sont réalisées lors du décodage de la séquence : tout d'abord, la restauration du fond de la séquence et l'extraction des objets en mouvement, ensuite, le traitement de chaque objet, et finalement, la restitution de la séquence.

2.1 Restauration du fond

Cette première étape réalise la segmentation temporelle : elle extrait le fond en y supprimant les effets de blocs, et fournit simultanément une carte des objets en mouvement.

L'estimation du fond est considérée comme un problème inverse. Soit p_k N images de la séquence MPEG ou M-JPEG contenant les effets de bloc. Comme cela est décrit dans [6], nous trouvons simultanément le fond estimé et la séquence des objets en mouvement, appelée c_k . On souhaite $c_k = 0$ si le point appartient à un objet en mouvement, sinon $c_k = 1$. Le critère correspondant s'écrit :

$$J_1(f, c_1, \dots, c_N) = J_1(f, c_1, \dots, c_N) + J_2(f) \quad (1)$$

avec

$$J_1(f, c_1, \dots, c_N) = \sum_{k=1}^N \int_{\Omega} c_k^2 (f - p_k)^2 + a_c \sum_{k=1}^N \int_{\Omega} (c_k - 1)^2 \quad (2)$$

qui réalise une segmentation spatio-temporelle en utilisant N images consécutives de la séquence, et

$$J_2(f) = I_1 \int_{\Omega} j_1(\|\nabla f\|)$$

le terme de régularisation contenant les a-priori sur la solution. j_1 est une fonction de potentiel [2] qui préserve les discontinuités dans l'image. Le paramètre a_c détermine l'importance que l'on accorde au fond : plus a_c est petit, plus il y a d'objets en mouvement détectés.

Des minimisations alternées en f et en c_k permettent d'obtenir un fond régularisé, ainsi qu'une carte des objets en mouvement. Concernant $J_1(f)$, si p_k est éloigné de l'estimée courante f , c_k est petit : on se trouve sur un objet qui se déplace.

Le critère (1) est insuffisant car il ne traite pas les effets de bloc. Il faut y ajouter

$$J_3(f) = h^2 \int_{\Omega} j_2\left(\frac{|Rf|}{d}\right) \quad (3)$$

où R représente la transformée en ondelettes, et j_2 une fonction de potentiel. d est un seuil dépendant de l'amplitude des effets de bloc, et la valeur de h précise quels sont les coefficients d'ondelettes à seuiller. Le comportement de (3) est détaillé dans [5] : un seuillage doux dans le domaine spatio-fréquentiel des ondelettes est réalisé pour supprimer les effets de bloc.

Ce nouveau critère supprime donc les effets de bloc sur le fond, et réalise simultanément la segmentation des objets en mouvement.

2.2 Restauration des objets

Cette deuxième étape concerne les traitements effectués sur les objets. Les objets ont été séparés du fond par la segmentation temporelle. Il reste donc à les différencier spatialement les uns par rapport aux autres, pour pouvoir les traiter indépendamment.

2.2.1 Localisation spatiale et suivi temporel

A partir de la séquence des c_k , il est possible d'isoler chaque objet les uns des autres, en utilisant par exemple une méthode fondée sur la morphologie mathématique, ou encore sur les contours géodésiques actifs [8], [3]. Pour chaque objet, une estimation de mouvement donne le vecteur $d^t = (d_x^t, d_y^t)$ correspondant au déplacement du pixel (x, y) entre les images $t-1$ et t . La connaissance du mouvement des objets permet, par une méthode classique d'appariement, de réaliser le suivi temporel de chaque objet. Il est donc désormais possible de les traiter séparément.

2.2.2 Suppression des blocs

L'effet de bloc est fixe sur les images de la séquence, par contre il se déplace virtuellement sur l'objet en mouvement : un moyennage temporel avec compensation de mouvement sur $n < N$ images permet donc de réduire ces effets de bloc. n permet de prendre en compte les éventuelles déformations de l'objet au cours du temps : plus l'objet évolue vite, plus n est choisi petit. Chaque objet est obtenu en appliquant l'équation :

$$O(x, y) = \frac{1}{n} \sum_{k=1}^n O_k \left(x + \sum_{j=1}^k d_x^j, y + \sum_{j=1}^k d_y^j \right) \quad (4)$$

Après ce moyennage, une régularisation avec prise en compte des discontinuités contribue à la réduction des effets de bloc.

2.3 Reconstruction de la séquence

La séquence décodée \tilde{p} est reconstruite en utilisant une image de fond f^* sur une durée de N images, et en y projetant les l objets traités :

$$\tilde{p}_k = c_k^{*2} f^* + (c_k^* - 1)^2 O_l^* \quad (5)$$

Si pour un pixel donné on se trouve sur un objet, $c_k^* = 0$, alors on projette le pixel de l'objet O_l^* à l'image k , sinon, $c_k^* = 1$ et l'on projette le pixel du fond f^* .

3. Prise en compte des dropout

La méthode de décodage proposée gère les pertes isolées de blocs sur les images. Ces pertes, qui ne correspondent pas forcément à des blocs DCT, peuvent résulter de :

- Une erreur de transmission sur un réseau en temps réel : le protocole de transmission n'a pas le temps de demander une nouvelle transmission pour les blocs perdus.

- Un défaut du codeur : dans les caméscopes numériques de dernière technologie, les séquences sont comprimées en temps réel au format DV. Ces codeurs câblés génèrent des pertes de blocs, ou des décalages d'images, différents selon les constructeurs.
- Une altération de la bande vidéo.
- Une perte de synchronisation lors du décodage.

L'altération par le temps ou par la fréquence des passages pour les films anciens, ainsi que les dégradations chimiques, engendrent des tâches sur le film, assimilables à des dropout.

3.1 Dropout sur le fond

Une perte sur le fond est prise en charge par l'équation (2), lors du moyennage sur N images. Elle n'apparaît pas sur le fond restauré. Le dropout est un « faux objet » qui est supprimé de la liste des objets. Le critère de suppression est temporel : un objet qui apparaît sur une seule image est considéré comme un dropout. Il est également spatial : la suppression peut se faire selon la taille ou la forme de l'objet.

3.2 Dropout sur les objets

Sur un objet en mouvement, le dropout est supprimé grâce au suivi temporel. La connaissance précise de l'objet sur n images consécutives permet de savoir si un dropout se trouve sur un objet, pour une image donnée. Le moyennage sur l'objet réalisé par l'équation (4) est modifié : si après compensation de mouvement, l'un des n pixels est très différent des $n-1$ autres, il est considéré comme appartenant à un dropout, et n'est pas pris en compte dans le moyennage. Ce moyennage, suivi de la régularisation sur l'objet, permet de supprimer efficacement le dropout.

4. Résultats Expérimentaux

La figure (1) est un exemple de décodage de la séquence « Van », codée en M-JPEG à un taux de 15.6 :1. Deux zooms permettent d'observer la qualité de la reconstruction, à la fois sur le fond et sur un objet en mouvement. Un rehaussement de contraste a été appliqué pour accentuer l'effet de bloc.

La séquence de la figure (2) présente une simulation de dropout. Les décalages de bandes, affectant le fond et l'objet, ainsi que les pertes aléatoires ont été avantageusement supprimés lors du décodage.

La figure (3) est un extrait de la séquence des « Frères Lumières » de 1895 fortement altérée. On observe la suppression de la plupart des tâches présentes sur le film original.

5. Conclusion

Cet article propose un schéma de décodage pour les séquences vidéos codées par blocs. Cette nouvelle méthode diffère des techniques existantes car elle réalise en parallèle deux tâches : l'estimation du fond avec la suppression des effets de bloc, et la détection des objets en mouvement. Une deuxième étape consiste à traiter chaque objet indépendamment et de manière adaptée, grâce à un suivi temporel des objets. Les résultats expérimentaux montrent

l'efficacité du décodage obtenue en substituant le décodeur traditionnel par le décodeur proposé : la qualité visuelle de la séquence a été améliorée, les effets de bloc, amplifiés par l'aspect temporel, ont été largement réduits, qu'ils soient sur le fond ou sur les objets en mouvement. De plus, les dropout sont supprimés de manière efficace, qu'ils soient sur le fond ou sur les objets en mouvement.

Références

- [1] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, *Image Coding Using Wavelet Transform*, IEEE Transactions on Image Processing, vol. 1 no. 2, pp. 205-220, Avril 1992.
- [2] P. Charbonnier, L. Blanc-Féraud, G. Aubert, M. Barlaud, *Deterministic Edge-Preserving Regularization in Computed Imaging*, IEEE Transaction on Image Processing, vol. 5, No. 12, Février 1997.
- [3] E. Debreuve, M. Barlaud, G. Aubert, J. Darcourt, *Attenuation Map Segmentation Without Reconstruction Using a Level Set Method in Nuclear Medicine Imaging*, International Conference on Image Processing, Chicago, Octobre 1998.
- [4] C. Gu, T. Ebrahimi, M. Kunt, *Morphological Object Segmentation and Tracking for Content Based Video Coding*, Multimedia Communication and Video Coding, New-York, Novembre 1995.
- [5] J. Jung, M. Antonini, M. Barlaud, *Optimal JPEG Decoding*, International Conference on Image Processing, Chicago, Octobre 1998.
- [6] P. Kornprobst, R. Deriche, G. Aubert, *Image sequence restoration : A PDE based coupled method for image restoration and motion segmentation*, European Conference on Computer Vision, p. 548—562, Freiburg, Juin 1998.
- [7] S. Minami, A. Zakhor, *An Optimization Approach for Removing Blocking Effects in Transform Coding*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 5, No.2, Avril 1995.
- [8] P. Salembier, F. Marqués, M. Pardàs, R. Morros, I. Corset, S. Jeannin, L. Bouchard, F. Meyer, and B. Marcotegui. *Segmentation-based video coding system allowing the manipulation of objects*. IEEE Trans. On Circuits and Systems for Video Technology, 7(1) : 60-74, Février 1997.
- [9] Y. Yang, N. Galatsanos, A. Katsaggelos, *Regularized Reconstruction to Reduce Blocking Artifacts of Block Discrete Cosine Transform Compressed Images*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 3, No. 6, Décembre 1993.
- [10] J. Jung, M. Antonini, M. Barlaud, Brevet n° 99/07443, 11 Juin 1999.

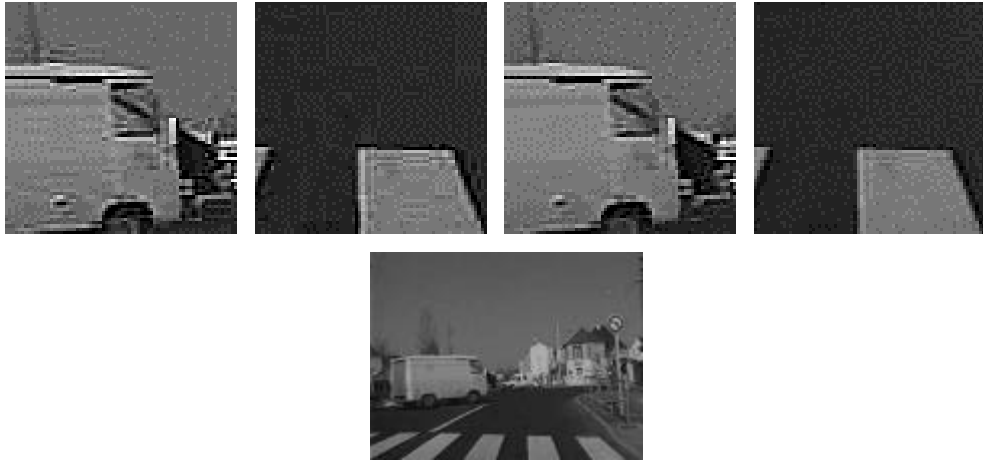


Figure 1 : Réduction des effets de bloc : séquence « Van », codée par M-JPEG (TC=15.6:1). A gauche, décodeur M-JPEG standard, à droite, décodeur proposé (zoom + rehaussement de contrastes).

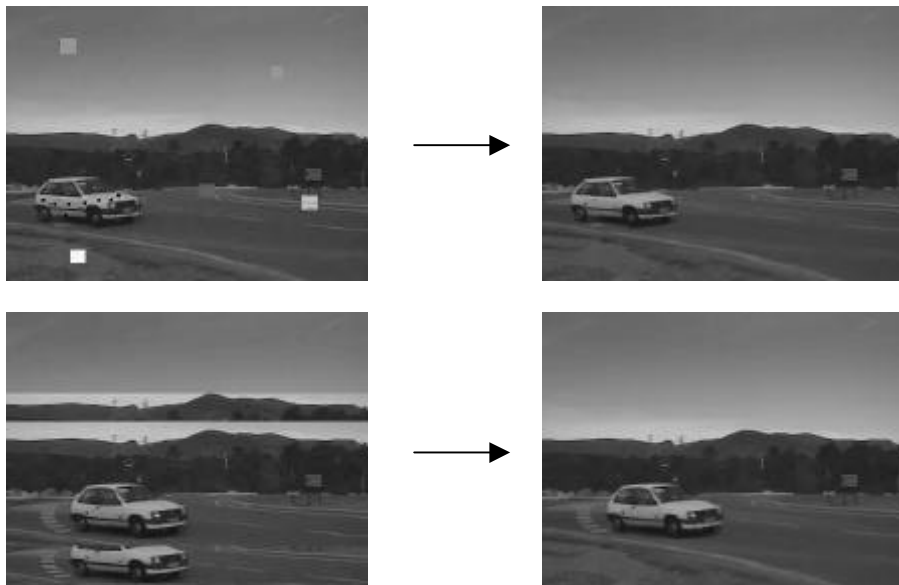


Figure 2 : Suppression de dropout : simulation de décalages de bandes et de pertes aléatoires. A gauche, décodeur M-JPEG standard, à droite, décodeur proposé.



Figure 3 : Suppression de dropout : extrait de la séquence des Frères Lumière (1895). A gauche, séquence originale numérisée, à droite, séquence traitée.