

Allocation Optimale de Débit entre le Mouvement et les Coefficients d'Ondelettes pour la Compression Vidéo basée Ondelettes

Marie Andrée AGOSTINI, Marc ANTONINI, Michel BARLAUD

Laboratoire I3S, Université de Nice-Sophia Antipolis, CNRS - UMR 6070
 Bât. Algorithmes/Euclide B, BP 121, 2000, route des Lucioles - 06903 Sophia Antipolis Cedex, France
 Tél. : 33(0)4.92.94.27.85 - Fax : 33(0)4.92.94.28.98
 agostini@i3s.unice.fr, am@i3s.unice.fr, barlaud@i3s.unice.fr

Résumé – Le codage basé ondelettes apparaît comme un schéma efficace pour la compression vidéo scalable. Dans les codeurs vidéo basés ondelettes compensés en mouvement, une estimation précise du mouvement est nécessaire pour minimiser l'énergie des coefficients d'ondelettes. Cependant, un champ de vecteurs mouvement de grande précision coûte cher en ressources binaires par rapport au codage des sous-bandes d'ondelettes, il est donc nécessaire d'optimiser le compromis débit-distorsion entre l'information de mouvement et les coefficients d'ondelettes. Dans ce papier, nous quantifions les vecteurs mouvement en utilisant un codage scalable avec pertes en boucle ouverte. Nous établissons également un modèle théorique de distorsion de l'erreur de codage du mouvement pour évaluer l'impact de ce codage avec pertes des vecteurs mouvement sur la séquence décodée. Ensuite, nous incluons à ce modèle le bruit de quantification des sous-bandes d'ondelettes et nous le généralisons à plusieurs niveaux de décomposition. Grâce à ce modèle, une allocation optimale de débit basée modèle est réalisée entre les vecteurs mouvement et les coefficients d'ondelettes. Les performances de notre codeur sont nettement améliorées sur des séquences CIF, SD et HD.

Abstract – Wavelet-based coding is emerging as a promising framework for efficient and scalable compression of video. In motion-compensated wavelet-based video coders (MCWT), a precise motion estimation is necessary in order to minimize the wavelet coefficients energy. Nevertheless, a motion vectors field of high precision is expensive in binary resources compared to wavelet subbands coding and it is thus necessary to optimize the rate-distortion trade-off between motion information and wavelet coefficients. In this paper, we propose to quantize the motion vectors using a scalable and open-loop lossy coder. We establish a theoretical distortion model of the motion coding error to evaluate the impact of this lossy motion coding on the decoded sequence. Then, we derive a total distortion model of coding error on several decomposition levels, including also the subbands quantization noise. Thanks to the proposed theoretical distortion model, an optimal model-based bitrate allocation is done between motion vectors and wavelet subbands. It increases the coding performances on CIF, SD and HD sequences.

1 Introduction

La compression vidéo connaît un essor important depuis quelques années. Des normes très efficaces ont été établies, aboutissant par exemple aux codeurs hybrides MPEG-4 [1] ou H.264/AVC [2]. Les codeurs basés ondelettes [3, 4], qui font intervenir un schéma lifting $t + 2D$ compensé en mouvement, permettent un meilleur support de la scalabilité [5] et atteignent presque les performances des codeurs hybrides [6]. Toutefois, il est encore nécessaire d'améliorer le traitement des vecteurs mouvement [7] et d'optimiser le compromis débit-distorsion entre l'information de mouvement et les sous-bandes d'ondelettes, dans le but d'améliorer les performances du codeur, surtout à bas débit. Nous avons donc réalisé une approche de codage des vecteurs mouvement (section 2), qui consiste à introduire de la perte sur des vecteurs de grande précision subpixelique [8]. Cette méthode de codage avec pertes en boucle ouverte (conservation de la scalabilité, qualité optimale pour les sous-bandes d'ondelettes) permet de réduire le coût des vecteurs mouvement avec de bonnes performances de codage à bas débit.

L'introduction de pertes sur le mouvement a un impact sur la séquence décodée; pour évaluer cet impact, nous établissons un modèle théorique de distorsion représentant l'erreur de codage du mouvement et des coefficients d'ondelettes (section 3). Ce modèle est ensuite utilisé pour répartir de façon optimale les ressources entre l'information de mouvement et les sous-bandes d'ondelettes pour un débit cible. L'allocation de débit basée modèle s'appuie sur la recherche d'un minimum pour un critère convexe J (section 4). Cette approche améliore les performances de notre codeur (section 5).

2 Codage du mouvement dans un codeur vidéo basé ondelettes

Totalement scalable, notre codeur vidéo est basé sur un schéma lifting d'ondelettes compensé en mouvement (figure 1). Pour réduire le coût des vecteurs mouvement qui peut être très important par rapport aux coefficients d'ondelettes, nous quantifions avec pertes des vecteurs mouvement estimés avec une grande précision (quart de pixel par

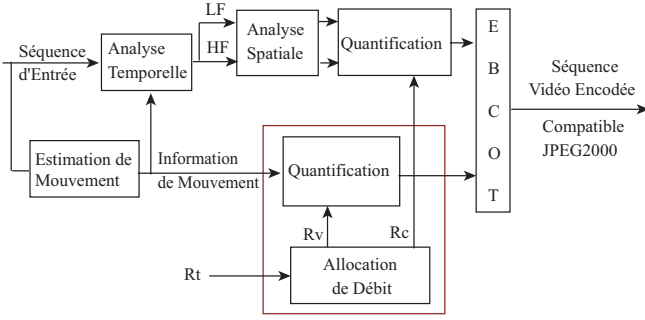


FIG. 1 – Structure générale du codeur proposé (R_c est le débit des coefficients d'ondelettes, R_v celui des vecteurs mouvement, et R_t le débit cible).

exemple), tout en contrôlant le compromis débit-distorsion entre les séquences originales et décodées.

Les vecteurs sont codés en boucle ouverte [8], ce qui permet de rester scalable et d'avoir une bonne qualité pour l'analyse temporelle. Les vecteurs mouvement sont quantifiés par une quantification scalaire uniforme, et codés par un codeur EBCOT [9] et intégrés dans un train binaire compatible-JPEG2000 (ainsi que les coefficients d'ondelettes).

3 Modèle théorique de l'erreur de quantification

Nous avons d'abord établi un modèle pour l'erreur de quantification des vecteurs mouvement, présenté à la section 3.1, puis nous y avons inclus l'erreur de quantification des sous-bandes d'ondelettes (section 3.2) et, enfin, nous l'avons généralisé à plusieurs niveaux de décomposition temporelle (section 3.3).

3.1 Modèle théorique de l'erreur de codage des vecteurs mouvement

Notre modèle est basé sur le calcul de la distorsion entrée-sortie D_t (EQM entre le signal d'entrée x et le signal de sortie \tilde{x}) exprimée par (avec \mathbf{Pn} la puissance du signal et K la taille de la séquence) :

$$D_t = \frac{1}{K} \sum_{k=0}^{\frac{K}{2}-1} (\mathbf{Pn}(x_{2k+1} - \tilde{x}_{2k+1}) + \mathbf{Pn}(x_{2k} - \tilde{x}_{2k})). \quad (1)$$

Puisque nous utilisons le schéma lifting (2,0), un sous-échantillonnage direct est réalisé sur les images de coefficient pair sans l'utilisation de vecteurs mouvement (voir figure 2). Par conséquent, la deuxième partie de cette équation est nulle quand est uniquement considérée la distorsion D_v due à la quantification des vecteurs mouvement.

Finalement, sous des hypothèses asymptotiques (haut débit, faibles erreurs de quantification), la distorsion D_v modélisant l'erreur de reconstruction due à la quantifica-

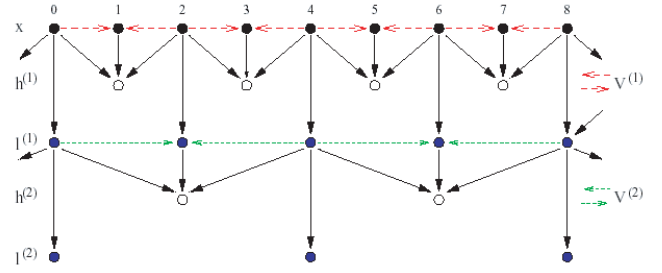


FIG. 2 – Schéma lifting (2,0) pour deux niveaux de décomposition temporelle en ondelettes.

tion des vecteurs mouvement peut s'écrire :

$$D_v = \frac{1}{2K} \sum_{k=0}^{\frac{K}{2}-1} [\mathbf{Pn}(x_{2k}) - \Gamma_{x_{2k}}(\eta_{B_{2k+1}}) + \mathbf{Pn}(x_{2k+2}) - \Gamma_{x_{2k+2}}(\eta_{F_{2k+1}})],$$

où $\mathbf{Pn}(x_{2k})$ et $\mathbf{Pn}(x_{2k+2})$ sont les puissances des images x_{2k} et x_{2k+2} ; $\Gamma_{x_{2k}}(\eta_{B_{2k+1}})$ et $\Gamma_{x_{2k+2}}(\eta_{F_{2k+1}})$ les fonctions d'autocorrélation des images x_{2k} et x_{2k+2} , qui dépendent uniquement de $\eta_{B_{2k+1}}$ et $\eta_{F_{2k+1}}$, les erreurs de quantification sur les vecteurs mouvement "Backward" et "Forward".

3.2 Ajout de l'erreur de quantification des sous-bandes d'ondelettes

Quand les sous-bandes d'ondelettes Haute et Basse Fréquence de chaque GOP ("Group Of Picture") du premier niveau de décomposition sont quantifiées, l'équation (1) doit aussi prendre en compte les bruits de quantification correspondants. Nous pouvons donc montrer que la distorsion totale D_t sur un niveau de décomposition prenant en compte aussi bien la quantification des vecteurs mouvement que celle des coefficients d'ondelettes peut être exprimée comme :

$$D_t \approx \frac{1}{2K} \sum_{k=0}^{\frac{K}{2}-1} \left[\mathbf{Pn}(x_{2k}) - \Gamma_{x_{2k}}(\eta_{B_{2k+1}}) + \mathbf{Pn}(x_{2k+2}) - \Gamma_{x_{2k+2}}(\eta_{F_{2k+1}}) + \frac{5}{2}\mathbf{Pn}(\epsilon_{2k}) + \frac{1}{2}\mathbf{Pn}(\epsilon_{2k+2}) + 2\mathbf{Pn}(\epsilon_{h_{2k+1}}) \right],$$

avec $\epsilon_{h_{2k+1}}$ le bruit de quantification de la sous-bande Haute Fréquence h_{2k+1} , et ϵ_{2k} et ϵ_{2k+2} les erreurs de quantification des sous-bandes Basse Fréquence des images x_{2k} et x_{2k+2} .

3.3 Généralisation à N niveaux de décomposition

L'expression générale de la distorsion totale entrée/sortie D_t , qui représente l'erreur de reconstruction due à la quantification du mouvement et des coefficients d'ondelettes, sur N niveaux de décomposition pour un schéma lifting

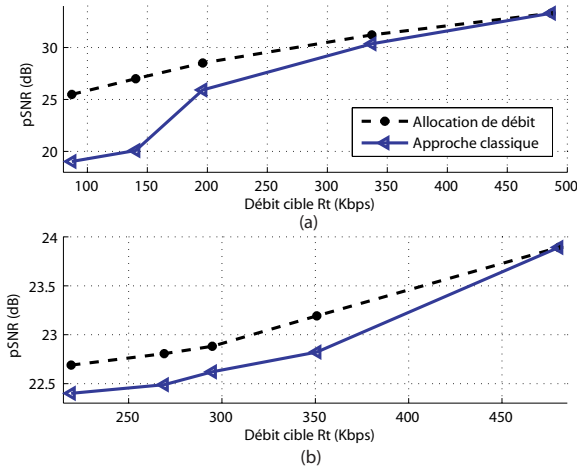


FIG. 3 – Approche classique comparée à l’approche avec allocation de débit optimale, (a) Séquence “Foreman”, au quart de pixel; (b) Séquence “City”, pixel; les 2 avec trois niveaux de décomposition temporelle, en fonction du débit total R_t .

(2,0), peut être exprimée par :

$$D_t = \frac{1}{2K} \sum_{n=0}^{N-1} \sum_{k=0}^{\frac{K}{2^{N-n}}-1} \left[\mathbf{Pn}(x_{2^{N-n}k}) + \mathbf{Pn}(x_{2^{N-n}k+2^{N-n}}) - \Gamma_{x_{2^{N-n}k}}(\eta_{B_{2^{N-n}k+2^{N-n}-1}}) - \Gamma_{x_{2^{N-n}k+2^{N-n}}}(\eta_{F_{2^{N-n}k+2^{N-n}-1}}) \right] + \frac{1}{K} \left(\frac{1}{2^N} \mathbf{Pn}(\epsilon_{l_k}^{(N)}) + \sum_{i=1}^N \frac{1}{2^i} \mathbf{Pn}(\epsilon_{h_k}^{(i)}) \right),$$

avec $l_k^{(N)}$ la sous-bande basse-fréquence et $h_k^{(i)}$ la sous-bande haute-fréquence au i^{ieme} niveau de décomposition. Le détail des calculs et la validation expérimentale du modèle sont présentés dans [10, 11].

4 Allocation de débit basée modèle

Le modèle de distorsion précédemment défini va nous permettre de trouver les débits optimaux pour l’information de mouvement et pour les sous-bandes d’ondelettes dans le but de minimiser la distorsion totale en sortie. Le problème (P) d’allocation de débit peut être formulé ainsi :

$$(P) \begin{cases} \min_{R_v, R_c} D_t(R_v, R_c) \\ \text{avec } R_v + R_c = R_t \end{cases}$$

Pour cela, nous utilisons le critère convexe suivant :

$$J_\lambda(R_v, R_c) = D_t(R_v, R_c) + \lambda(R_v + R_c - R_t).$$

La solution optimale (R_v^* , R_c^*) est donc donnée par l’algorithme d’allocation de débit suivant :

1. $\lambda = \lambda_{init}$
2. Pour chaque valeur de R_c , trouver la valeur optimale $R_v^*(\lambda, R_c)$ qui minimise le critère $J_\lambda(R_v, R_c)$
3. Trouver la valeur optimale R_c^* qui minimise le critère $J_\lambda(R_v^*(\lambda, R_c), R_c)$

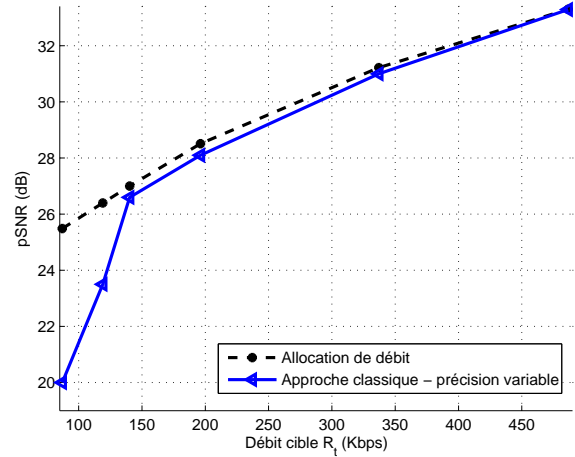


FIG. 4 – Approche classique avec la précision des vecteurs mouvement variable comparée à l’approche avec allocation de débit optimale au quart de pixel, séquence “Foreman”, trois niveaux de décomposition temporelle.

4. Si $R_v^* + R_c^* = R_t$ alors arrêter, sinon changer la valeur de λ et retourner à l’étape 2.

5 Résultats expérimentaux

Nous présentons les résultats de l’allocation de débit, sur trois niveaux de décomposition avec un schéma lifting (2,0), pour la séquence CIF “Foreman” (vecteurs mouvement au quart de pixel), sur la séquence SD “City” (demi-pixel), et sur la séquence HD 720p “Jets” (pixel). La figure 3 présente, pour les deux premières séquences, les courbes du pSNR en fonction du débit cible obtenues en appliquant l’algorithme d’allocation de débit et celles obtenues par l’approche classique (marqueurs triangulaires : pour les mêmes débits cible R_t , $R_v = 143.4$ Kbps pour “Foreman” et $R_v = 160.4$ Kbps pour “City”, codage sans perte des vecteurs mouvement). La figure 4 montre la comparaison pour “Foreman” entre les résultats de l’allocation optimale de débit et les points de l’enveloppe convexe obtenus avec des vecteurs codés sans perte mais de précision subpixelique variable. Il apparaît qu’utiliser les débits optimaux calculés par l’allocation de débit permet d’améliorer significativement la qualité des séquences décodées, surtout à bas débit.

La figure 5 présente une comparaison visuelle de 2 images extraites de “Foreman” codée à 150 Kbps : à gauche, codage sans perte des vecteurs mouvement ; à droite, allocation de débit. Les images extraites de la séquence obtenue par l’approche classique sont très dégradées. En revanche, pour la séquence obtenue après allocation de débit, la qualité est bien meilleure. La figure 6 présente des résultats visuels pour deux images extraites de la séquence “Jets”, avec l’approche classique utilisée à gauche pour un débit total de 1.3 Mbps, et l’allocation de débit optimale utilisée à droite avec un débit total de 1 Mbps, les deux avec un pSNR en sortie de 35.9 dB. La qualité des images est donc la même, mais 300 Kbps ont été préservés pour le débit total.



FIG. 5 – Séquences reconstruites “Foreman”, format CIF à 150 Kbps, trois niveaux de décomposition ; images 3 et 57 ; à gauche : vecteurs au quart de pixel codés sans perte ($R_v = 143.4$ Kbps) ; à droite : vecteurs quart-pixéliques quantifiés à $R_v^* = 48.5$ Kbps et allocation optimale de débit.

6 Conclusion

Il est nécessaire d’optimiser le compromis débit - distorsion entre l’information de mouvement et les sous-bandes d’ondelettes dans les codeurs vidéo basés ondelettes compensés en mouvement. Pour cela, nous proposons de quantifier avec pertes en boucle ouverte les vecteurs mouvement estimés avec une grande précision subpixélique et de réaliser une allocation de débit pour distribuer de façon optimale les ressources entre les vecteurs mouvement et les coefficients d’ondelettes. Pour évaluer l’impact des pertes introduites lors des quantifications de l’information de mouvement et des sous-bandes d’ondelettes, nous avons établi sur plusieurs niveaux de décomposition un modèle théorique de distorsion.

En outre, nous avons réalisé un algorithme d’allocation de débit optimale basé modèle pour allouer les ressources binaires entre les vecteurs mouvement et les coefficients d’ondelettes. Cette approche permet de trouver analytiquement pour un débit cible donné, les débits optimaux des vecteurs mouvement et des coefficients d’ondelettes pour avoir une distorsion minimale au décodage. Les résultats obtenus sur des séquences CIF, SD et HD montrent une amélioration significative des performances. Des travaux futurs devraient améliorer encore ces résultats en utilisant une estimation de mouvement plus efficace.

Références

[1] MPEG4 Video Group, *Coding of audio-visual objects : Video, ISO/IEC JTC1/SC29/WG11 N2202*, Mar. 1998.

[2] Gary J. Sullivan, Pankaj Topiwala, and Ajay Luthra, “The H.264/AVC advanced video coding standard : Overview and introduction to the fidelity range extensions,” in *SPIE Conference on Applications of Digital Image Processing*



FIG. 6 – Séquences reconstruites “Jets”, format HD 720p ; images 14 et 59 ; à gauche : vecteurs pixéliques codés sans perte, $R_t = 1.3$ Mbps ($R_v = 766.5$ Kbps) ; à droite : vecteurs pixéliques quantifiés à $R_v^* = 357.5$ Kbps et allocation optimale de débit utilisée avec $R_t = 1$ Mbps ; pSNR = 35.9 dB dans les deux cas.

XXVII, special Session on Advances in the New Emerging Standard : H264 / AVC, Aug. 2004.

- [3] S. J. Choi and J.W. Woods, “Motion-compensated 3-D subband coding of video,” *IEEE Transaction on Image Processing*, vol. 8, no. 2, pp. 155–167, Feb. 1999.
- [4] S. Cho and W.A. Pearlman, “A full-featured, error-resilient ; scalable wavelet video codec based on the set partitioning in hierarchical trees (SPIHT) algorithm,” *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 12, pp. 170–171, Mar. 2002.
- [5] G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans, “Motion compensation and scalability in lifting-based video coding,” *EURASIP Signal Processing : Image Communication, special issue on Wavelet Video Coding*, pp. 577–600, Aug. 2004.
- [6] J-R. Ohm, “Three dimensional subband coding with motion compensation,” *IEEE Trans. On Image Processing*, vol. 3, no. 5, November 1994.
- [7] L. A. Da Silva Cruz and J. W. Woods, “Adaptive motion vector quantization for video coding,” in *IEEE Intern. Conf. on Image Processing*, Vancouver, Canada, Sept. 2000, pp. 867–870, vol. 2.
- [8] M. A. Agostini, T. André, M. Antonini, and M. Barlaud, “Scalable motion coding for video coders with lifted MCWT,” in *Proc. of International Workshop on Very Low Bit-rate Video-coding (VLBV)*, Costa Rei, Sardinia - Italy, Sept. 2005.
- [9] D. Taubman, “High performance scalable image compression with ebcot,” *IEEE Transaction on Image Processing*, vol. 9, no. 7, pp. 1158–1170, July 2000.
- [10] M. A. Agostini, T. André, M. Antonini, and M. Barlaud, “Modeling the motion coding error for MCWT video coders,” in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 2006.
- [11] M. A. Agostini and M. Antonini, “Theoretical model of the coding error in mcwt video coders,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, Atlanta, USA, Oct. 2006.