

AMELIORATION PROGRESSIVE DE L'INFORMATION ADJACENTE POUR LE CODAGE VIDEO DISTRIBUE

Abdalbassir ABOU-ELAILAH¹, Frederic DUFAUX¹, Joumana FARAH²
Marco CAGNAZZO¹, Beatrice PESQUET-POPOSCU¹

¹Laboratoire de Traitement et Communication de l'Information - UMR 5141, Telecom ParisTech
46 rue Barrault, 75634 Paris Cedex 13, France

²Faculté d'ingénierie, Université Saint-Esprit de Kaslik
B.P. 446 Jounieh, Liban

¹{elailah, dufaux, cagnazzo, pesquet}@telecom-paristech.fr

²joumanafarah@usek.edu.lb

Résumé – Ce travail vise à améliorer les performances de compression dans le cadre du codage vidéo distribué. Nous proposons une nouvelle approche, tendant à améliorer itérativement l'information adjacente après le décodage de chaque sous-bande DCT, contrairement à [8] où elle n'est révisée qu'après le décodage de toutes les sous-bandes. Les résultats des simulations indiquent que le schéma proposé mène à des gains importants en termes de débit-distorsion, par rapport au codeur DISCOVER, jusqu'à 0.75 dB pour une taille de GOP = 2 et 2.25 dB pour une taille de GOP = 8, pour des séquences vidéos avec des mouvements complexes.

Abstract – This work aims at improving the compression performance of distributed video coding. We propose a new approach which consists in successively improving the side information after the decoding of each DCT subband, unlike [8] where it is revised after the decoding of all subbands. The experimental results indicate that the proposed scheme can achieve a significant gain in rate-distortion performance, compared to the DISCOVER codec, up to 0.75 dB for a GOP size of 2 and 2.25 dB for a GOP size of 8, for video sequences containing high motion.

1 Introduction

Les activités de recherches antérieures dans le domaine du codage vidéo sont essentiellement concentrées sur les deux principes du codage prédictif et du codage par transformée. En particulier, toutes les normes de codage vidéo conçues par MPEG et ITU-T à ce jour sont basées sur un schéma hybride combinant une compensation de mouvement et une transformée en cosinus discrète (DCT).

Le codage distribué est un nouveau principe, apparu il y a quelques années. Il s'appuie sur deux théorèmes de la théorie de l'information : Slepian-Wolf et Wyner-Ziv. Le premier considère le codage sans perte de deux sources corrélées et montre qu'un codage distribué (c'est-à-dire encodage séparé mais décodage conjoint) peut atteindre le même débit qu'un codage classique (c'est-à-dire encodage et décodage conjoints), avec une probabilité d'erreur résiduelle qui converge vers 0 pour de longues séquences. Le théorème de Wyner-Ziv étend ce résultat au codage avec perte.

Grâce à ses propriétés, le codage vidéo distribué (DVC) offre un certain nombre d'avantages et de fonctionnalités. Il repose sur une formulation statistique, contrairement à l'approche déterministe des schémas classiques. La complexité de calcul peut être allouée de manière flexible entre l'encodeur et le dé-

codeur. En particulier, DVC permet un encodage avec une très faible complexité. L'absence de boucle de prédiction empêche la propagation des erreurs de transmission, propriété particulièrement adaptée à la transmission sur des réseaux sujets à des pertes. DVC permet également des schémas innovants pour la scalabilité et le codage vidéo multi-vue.

Pour ces raisons, DVC est particulièrement intéressant pour une nouvelle classe d'applications caractérisées par une liaison montante, telles que les réseaux de capteurs de faible puissance, les caméras de surveillance vidéo sans fil ou encore les appareils de communication mobile. DVC a suscité un fort intérêt au cours des dernières années et a fait l'objet de nombreux travaux de recherche. Une vue d'ensemble des développements récents est présentée en [1][2].

Malgré des progrès considérables ces dernières années, les performances débit-distorsion restent en deçà des attentes. Dans cet article, nous nous concentrons sur l'amélioration de l'information adjacente (SI), qui a un très fort impact sur la performance débit-distorsion. De nombreux travaux se sont d'ailleurs penchés sur cet aspect. En [3], la SI est générée par une interpolation temporelle compensée en mouvement (MCTI). Une approche proposée en [4] consiste à envoyer une information de hachage dans le but d'améliorer l'estimation de mouvement et donc la SI. Similairement, le hachage est com-

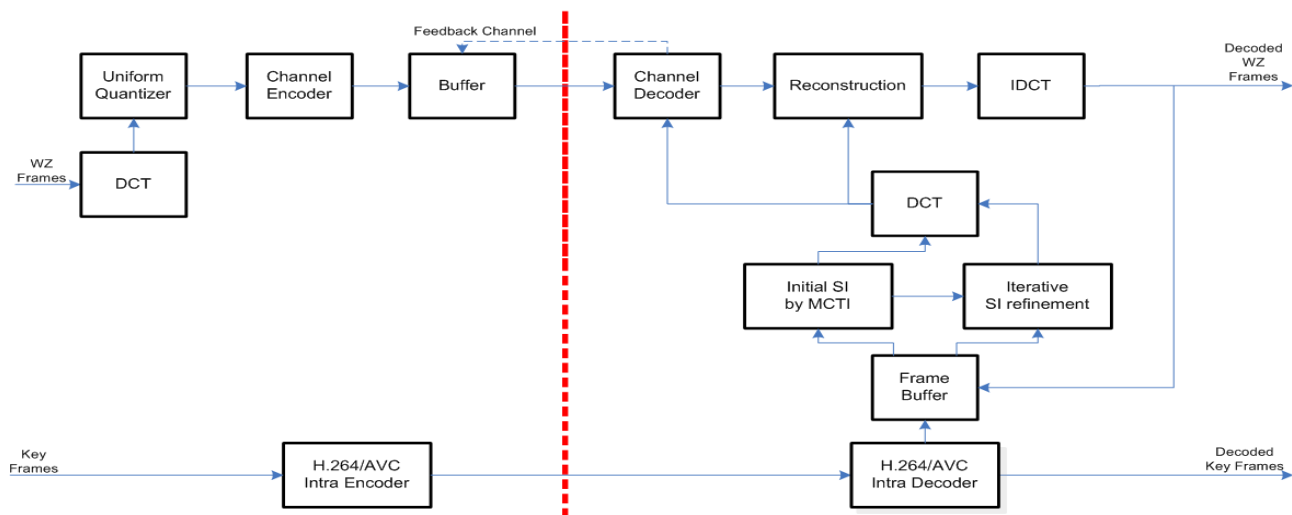


FIGURE 1 – Schéma proposé de codage DVC.

biné avec un algorithme génétique en [5]. En [6], le mouvement est modélisé par un polynôme d'ordre élevé. Un cadre théorique est présenté en [7] pour le codage Wyner-Ziv lorsque la qualité de la SI est incertaine. En [8], la SI est révisée après le décodage de toutes les sous-bandes afin d'améliorer l'étape de reconstruction.

Dans cet article, nous proposons une nouvelle approche en vue d'améliorer la SI dans le cadre d'un système de codage vidéo distribué dans le domaine transformée. La solution proposée consiste à améliorer progressivement la SI après le décodage de chaque sous-bande DCT. Plus spécifiquement, à chaque itération, le décodeur se sert de la trame WZ partiellement décodée (PDWZF), conjointement avec les trames de référence adjacentes, afin d'améliorer la SI. Pour cette dernière opération, la fiabilité des vecteurs de mouvement est premièrement vérifiée. Ensuite, les vecteurs suspects sont à nouveau calculés par une estimation de mouvement bidirectionnelle. Finalement, le mode de compensation optimal est sélectionné.

Les méthodes introduites en [9][10] proposent également une amélioration progressive de la SI après le décodage de chaque sous-bande. Cependant, en [9], l'amélioration de la SI n'exploite que la PDWZF et la SI, bien que les trames de référence adjacentes contiennent des informations utiles. D'autre part, la méthode en [10] repose sur une technique d'amélioration de la SI beaucoup plus simpliste que celle proposée ici.

2 Architecture de codage vidéo distribué proposée

Le schéma de codage vidéo distribué est exposé dans la Figure 1. L'architecture est en partie basée sur le codec DISCOVER [11]. La séquence vidéo est premièrement divisée en trames clés (KF) et Wyner-Ziv (WZF). Le groupe d'images (GOP) est défini comme la distance entre deux KF consécutives. Les KF sont encodées avec H.264/AVC intra. Pour les WZF, une

transformation DCT est préalablement appliquée, suivie d'une quantification uniforme. Les valeurs quantifiées sont ensuite séparées en plans de bits qui sont encodés avec un code Low-Density Parity Check (LDPC). Au décodeur, la SI est générée par une MCTI des trames précédemment décodées. La SI est exploitée dans le décodeur LDPC, conjointement avec les bits de parité des WZF obtenus via des requêtes par un canal de retour, afin de reconstruire les plans de bits, et par conséquent la séquence vidéo décodée.

Dans cet article, notre approche vise à améliorer itérativement la SI après le décodage de chaque sous-bande DCT. Plus spécifiquement, la SI est initialement générée par MCTI comme pour DISCOVER. Une PDWZF est obtenue en exploitant les bits de parité de la première sous-bande DCT. Le décodeur se sert de cette PDWZF, conjointement avec les trames de référence adjacentes, afin de réviser la SI. Cette dernière est ultérieurement utilisée pour le décodage de la prochaine sous-bande et ce processus est itéré pour toutes les sous-bandes.

La technique proposée pour l'amélioration de la SI est composée de trois étapes : la détection des vecteurs de mouvement erronés, la correction de ces vecteurs de mouvement, et la sélection du mode optimal de compensation de mouvement. L'approche proposée est similaire à celle exposée en [8]. Toutefois, dans le schéma proposé, la SI est progressivement améliorée après le décodage de chaque sous-bande, contrairement à [8], où elle n'est révisée qu'après le décodage de toutes les sous-bandes.

Les trois étapes principales de l'algorithme sont plus amplement détaillées comme suit :

- Détection de vecteurs de mouvement erronés : Les vecteurs de mouvement ne sont pas toujours fiables en présence de mouvements complexes ou rapides, ou d'occlusions. Afin d'identifier ces vecteurs suspects, la différence absolue moyenne (MAD) est estimée pour un bloc donné entre la PDWZF et la SI actuelle, et comparée à un seuil T_1 :

Foreman - GOP size = 2					
	$T_2 = 4$				
	$T_1 = 0$	$T_1 = 2$	$T_1 = 4$	$T_1 = 6$	$T_1 = 8$
Δ_R (%)	-11.48	-13.22	-13.40	-11.00	-8.98
Δ_{PSNR} [dB]	0.60	0.71	0.73	0.60	0.47
	$T_1 = 4$				
	$T_2 = 0$	$T_2 = 2$	$T_2 = 4$	$T_2 = 6$	$T_2 = 8$
Δ_R (%)	-9.88	-13.07	-13.40	-13.41	-13.43
Δ_{PSNR} [dB]	0.54	0.70	0.73	0.73	0.73

TABLE 1 – PSNR et débit de notre méthode par rapport à DISCOVER, pour différentes valeurs de T_1 et T_2 , en utilisant la métrique de Bjontegaard.

$$\text{MAD}(\text{PDWZF}, \text{SI}(\mathbf{MV})) < T_1 \quad (1)$$

où $\mathbf{MV}=(\text{MV}_x, \text{MV}_y)$ est le vecteur de mouvement candidat. Si la condition requise par l'équation (1) est satisfaite, le vecteur de mouvement est considéré comme fiable et sera conservé. A défaut, il est considéré comme peu fiable et sera corrigé à l'étape suivante.

- Correction des vecteurs de mouvement erronés : Pour améliorer les vecteurs de mouvement suspects, ces derniers sont à nouveau calculés par une estimation de mouvement bidirectionnelle et un critère MAD. Plus précisément, pour le bloc considéré dans la PDWZF, le vecteur de mouvement qui minimise le MAD avec la trame de référence précédente est calculé. De manière similaire, un vecteur de mouvement entre le bloc considéré dans la PDWZF et la trame de référence suivante est estimé.
- Sélection du mode optimal de compensation de mouvement : L'objectif de cette étape est de générer une compensation de mouvement optimale en sélectionnant le bloc le plus semblable au bloc courant parmi trois sources : la trame de référence précédente (mode BACKWARD), la trame de référence suivante (mode FORWARD), et la moyenne bidirectionnelle compensée en mouvement des trames de référence précédente et suivante (BIDIRECTIONAL). La sélection est effectuée selon les équations suivantes :

$$\left\{ \begin{array}{l} \text{if } |\text{MAD}_n - \text{MAD}_p| < T_2 \\ \quad \{\text{mode} = \text{BIDIRECTIONAL}\} \\ \text{else if } \text{MAD}_n < \text{MAD}_p \\ \quad \{\text{mode} = \text{FORWARD}\} \\ \text{else} \\ \quad \{\text{mode} = \text{BACKWARD}\} \end{array} \right.$$

où MAD_p et MAD_n sont les différences absolues moyennes entre le bloc considéré dans la PDWZF et les blocs correspondants dans les trames de référence précédente et suivante, respectivement, et T_2 constitue un seuil.



FIGURE 2 – Résultats visuels de la SI finale de notre méthode (à droite - PSNR = 29.72 dB) comparés avec DISCOVER (à gauche - PSNR = 22.14 dB) - Trame numéro 92 de la séquence Foreman.

Foreman		
GOP size = 2		
	Ref. [11]	Proposed
Δ_R (%)	-04.26	-13.40
Δ_{PSNR} [dB]	00.22	0.73
GOP size = 4		
Δ_R (%)	-12.09	-32.33
Δ_{PSNR} [dB]	00.54	01.60
GOP size = 8		
Δ_R (%)	-19.40	-48.09
Δ_{PSNR} [dB]	00.80	02.28

TABLE 2 – PSNR et débit de notre méthode pour *Foreman*, par rapport à DISCOVER, en utilisant la métrique de Bjontegaard.

3 Résultats des simulations

Afin d'évaluer la performance du schéma proposé, nous avons réalisé de nombreuses simulations, en utilisant des conditions identiques à celles de DISCOVER [11]. Nous avons utilisé la séquence de test 'Foreman' avec une résolution QCIF et un échantillonnage à 15 images/sec.

Le paramètre T_1 joue un rôle important dans notre système. En particulier, il détermine la complexité de notre méthode. Le Tableau 1 compare les valeurs du PSNR et du débit, obtenus par notre technique, par rapport à DISCOVER, pour différentes valeurs de T_1 et T_2 , en utilisant la métrique de Bjontegaard [12]. On remarque que la performance de notre système est optimale pour $T_1 = 4$. En outre, pour $T_1 = 0$, on ré-estime tous les vecteurs de mouvement pour tous les blocs, même si les vecteurs sont fiables. En conséquence, la complexité augmente. Pour $T_1 = 8$, on a quelques vecteurs erronés qui ne sont pas ré-estimés en raison de la grande valeur de T_1 . Concernant le paramètre T_2 , il est clair que l'utilisation du mode "bidirectionnel - $T_2 > 0$ " est toujours meilleur que le mode "unidirectionnel - $T_2 = 0$ ". Par la suite, les paramètres T_1 et T_2 ont été fixés à $T_1 = T_2 = 4$. Pour référence, nous comparons nos résultats avec ceux obtenus par DISCOVER [11], H.264/AVC Intra, H.264/AVC No motion, ainsi que par la méthode proposée en [8].

La Figure 2 montre les résultats visuels de la SI finale de notre méthode comparée avec DISCOVER. L'amélioration itérative de la SI apporte un gain de 7.5 dB pour la trame 92 de la séquence Foreman.

La Figure 3 montre la performance de notre système en

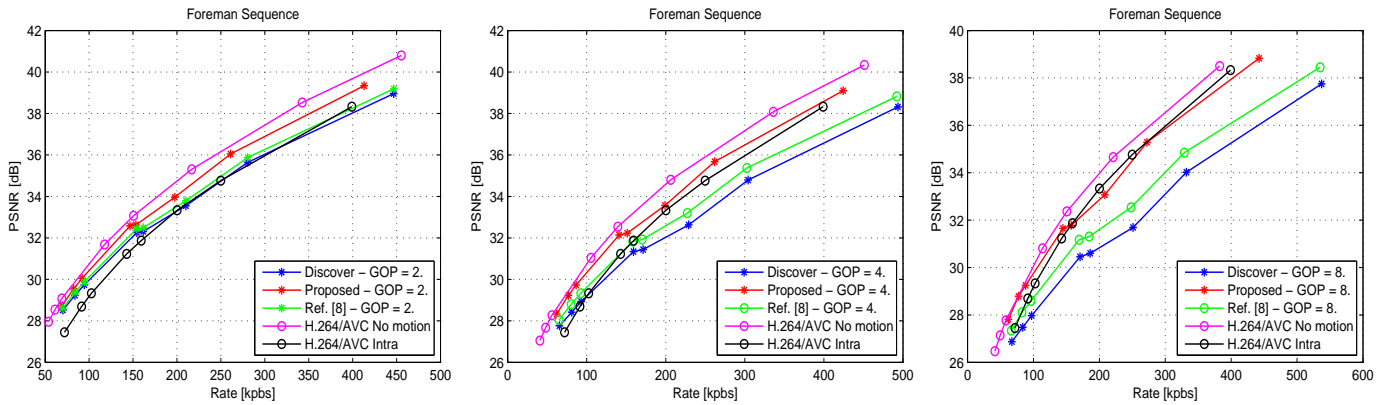


FIGURE 3 – Performance débit-distorsion.

terme de débit-distorsion pour des longueurs de GOP=2, 4 et 8. On peut observer que la performance de notre méthode est constamment supérieure à DISCOVER ainsi qu'à [8], pour toutes les tailles de GOP. Les valeurs du PSNR et du débit de la méthode proposée et de référence [8] par rapport à DISCOVER sont montrées dans le Tableau 2 en utilisant la métrique de Bjontegaard [12]. La méthode proposée apporte une amélioration de la performance débit-distorsion jusqu'à 0.73 dB avec une réduction de débit de 13.4 % par rapport à DISCOVER pour GOP=2. Les améliorations sont encore plus significatives pour GOP=4 ou GOP=8 : jusqu'à 1.6 dB avec une réduction de débit de 32.33 % et 2.28 dB avec une réduction de débit de 48 % respectivement.

On peut observer que la performance de la méthode proposée s'approche de celle de H.264/AVC No motion, et qu'elle surpasse celle de H.264/AVC Intra pour toutes les tailles de GOP.

Des gains similaires sont obtenus pour la séquence 'Soccer'. Pour les séquences 'Hall Monitor' et 'Coastguard', dont le mouvement est nettement plus simple, les gains en PSNR sont plus petits et se situent entre 0.2 et 0.8 dB.

Références

- [1] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardini et J. Ostermann. *Distributed Monoview and Multiview Video Coding*. IEEE Signal Processing Magazine, vol. 24, no. 5, pp. 67-76, Sept. 2007.
- [2] F. Dufaux, W. Gao, S. Tubaro et A. Vetro. *Distributed Video Coding : Trends and Perspectives*. EURASIP Journal on Image and Video Processing, vol. 2009, Jan. 2009.
- [3] C. Brites, J. Ascenso et F. Pereira. *Improving transform domain Wyner-Ziv video coding performance*. Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), May 2006.
- [4] A. Aaron, S. Rane et B. Girod. *Wyner-Ziv video coding with hash-based motion compensation at the receiver*. Proc. of IEEE Int. Conf. on Image Processing, Oct. 2004.
- [5] T. Maugey, C. Yaacoub, J. Farah, M. Cagnazzo et B. Pesquet-Popescu. *Side information enhancement using an adaptive hash-based genetic algorithm in a Wyner-Ziv context*. Proc. of IEEE International Workshop on Multimedia Signal Processing (MMSp 2010), Saint-Malo, France, Oct. 2010.
- [6] G. Petrazzuoli, M. Cagnazzo et B. Pesquet-Popescu. *High order motion interpolation for side information improvement in DVC*. Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), March 2010.
- [7] F. Bassi, M. Kieffer et C. Weidmann. *Wyner-Ziv Coding with uncertain side information quality*. Proc. of European Signal Processing Conference (EUSIPCO), August 2010.
- [8] S. Ye, M. Ouaret, F. Dufaux et T. Ebrahimi. *Improved side information generation for distributed video coding by exploiting spatial and temporal correlations*. EURASIP Journal on Image and Video Processing, vol. 2009, Article ID 683510, doi :10.1155/2009/683510, 2009.
- [9] R. Martins, C. Brites, J. Ascenso et F. Pereira. *Refining side information for improved transform domain Wyner-Ziv video coding*. IEEE Transactions on circuits and systems for video technology, vol. 19, no. 9, pp. 1327 - 1341, Sept. 2009.
- [10] M.B. Badem, W.A.C. Fernando, J.L. Martinez et P. Cuenca. *An iterative side information refinement technique for transform domain distributed video coding*. Proc. of IEEE Int. Conf. on Multimedia and Expo (ICME), June 2009.
- [11] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov et M. Ouaret. *The DISCOVER codec : architecture, techniques and evaluation*. Proc. of Picture Coding Symposium, Nov. 2007.
- [12] G. Bjontegaard. *Calculation of average PSNR differences between RD-curves*. In VCEG Meeting, Austin, USA, Apr. 2001.