

# Nouvelle technique d'adaptation dynamique de la taille du GOP dans le codage Wyner-Ziv des séquences vidéo

JOUMANA FARAH<sup>1</sup>, CHARLES YAACOUB<sup>1,2</sup>, BEATRICE PESQUET-POPESCU<sup>2</sup>

<sup>1</sup>Faculté des Sciences et de Génie Informatique, Université St-Esprit de Kaslik, Liban

<sup>2</sup>Département Traitement du Signal et des Images, Télécom ParisTech, Paris, France

<sup>1</sup>joumanafarah@usek.edu.lb, charlesyaacoub@usek.edu.lb,

<sup>2</sup>beatrice.pesquet@telecom-paristech.fr

**Résumé** – Dans cet article, nous abordons le problème de la détermination de la taille optimale du GOP (Group Of Pictures) en codage vidéo distribué (DVC). Pour cette raison, nous proposons un algorithme permettant d'adapter la taille du GOP, de façon dynamique, en fonction du contenu (niveau de mouvement) de la séquence vidéo. La technique introduite permet de choisir, à tout instant, entre le mode de codage Intra H264 et le mode de codage Wyner-Ziv (WZ), de façon à optimiser la qualité moyenne de décodage. Par ailleurs, l'algorithme permet de déterminer le taux de compression nécessaire des trames WZ sans besoin d'un canal de retour. Les résultats de simulation montrent une amélioration notable dans les performances du système, par rapport au codage Intra H264 seul ou au codage WZ avec une taille de GOP fixe.

**Abstract** – The aim of this paper is to present a novel algorithm for content adaptive GOP size control in distributed video coding. The GOP size is dynamically varied along the sequence, depending on motion activity. Automatic mode selection allows the system to switch between H.264 intra-coding and Wyner-Ziv coding modes to optimize the overall performance. Furthermore, the encoder determines a suitable compression ratio for the Wyner-Ziv frames without the need for a feedback channel. Simulation results show significant improvement in the average system performance, compared to fixed GOP Wyner-Ziv and H.264 intra-coding.

## 1 Introduction

Le codage distribué [1-10] a pris dernièrement beaucoup d'importance en compression vidéo, particulièrement dans le contexte d'applications nécessitant des codeurs à faible complexité. Le paradigme DVC est né du théorème de Slepian et Wolf [11]: étant donné deux sources corrélées  $X$  et  $Y$ , avec  $Y$ , l'information adjacente, comprimée à sa limite entropique  $H(Y)$ ,  $X$  peut être transmise à un débit très proche de l'entropie conditionnelle  $H(X|Y)$ . Le codage Wyner-Ziv [12] est une application de ce concept au codage source avec pertes.

Dans la majorité des systèmes DVC actuels, un sous-ensemble des trames vidéo (trames clés) sont comprimées avec un codeur intra-trames traditionnel [13]. Une ou plusieurs trames succédant aux trames clés, dites trames Wyner-Ziv (WZ), sont comprimées indépendamment les unes des autres (et des trames clés) par un poinçonnage adéquat des bits de parité à la sortie d'un turbo-codeur. Les informations systématiques ne sont pas transmises mais remplacées, dans le décodeur inter-trames, par l'information adjacente (Side Information ou SI) générée grâce aux trames clés (et aux trames WZ précédemment décodées).

L'un des problèmes majeurs de la plupart des systèmes DVC actuels est la nécessité d'un canal de retour pour permettre une allocation de débit flexible des trames WZ générées par le codeur [14]. Ceci rend l'application de ces systèmes difficile dans un environnement temps-réel et multi-utilisateurs [2]. Puisque le décodage des trames WZ dans le récepteur

nécessite l'interpolation de trames clés ou WZ pour la génération de la SI, le décodage correct des trames WZ, à des débits de transmission abordables, nécessite une SI de qualité élevée. Cette qualité se dégrade rapidement lorsque les trames clés sont éloignées, c.à.d. lorsque la taille du GOP (le nombre de trames WZ séparant deux trames clés) augmente. D'un autre côté, le fait de diminuer le nombre des trames clés permet de réduire le débit moyen de transmission d'un GOP, d'où la nécessité d'un compromis.

Une étude a été proposée dans [8] qui montre les performances du DVC pour une taille de GOP fixe allant de 2 à 5. Dans [9], un algorithme a été proposé pour permettre de faire varier la taille du GOP de façon dynamique, en fonction du contenu de la séquence. Cependant, les deux études se basent sur l'utilisation d'un canal de retour. D'ailleurs, l'algorithme proposé dans [9] se base sur le calcul de quatre métriques différentes, ce qui augmente considérablement la complexité du codeur. Quant au problème du canal de retour, il a été considéré dans [3], où des tables de performances préétablies sont utilisées pour prédire le niveau de compression nécessaire pour chaque trame, avec des tailles de GOP fixes.

Le but de ce travail est de développer un nouvel algorithme qui permet de faire varier dynamiquement la taille du GOP dans un DVC à basse complexité. En se basant sur notre technique d'estimation de débit présentée dans [2], l'algorithme permet de se débarrasser du canal de retour et, contrairement aux précédentes approches, peut être facilement étendu pour

prendre en compte les imperfections du canal et les contextes multi-utilisateurs. Dans ce but, nous commencerons dans la section 2 par une description détaillée de l'algorithme de contrôle dynamique de la taille du GOP. Ensuite, les résultats de simulation seront présentés et analysés dans la section 3.

## 2 Technique de variation de la taille du GOP dans la séquence vidéo

Lorsque le mouvement est lent dans une séquence vidéo, la corrélation entre les trames successives augmente. Le fait de diminuer le nombre de trames clés (codées en Intra H264 [13]) peut permettre de mieux exploiter cette propriété, puisque dans ces régions le codage WZ aboutit à de meilleures performances en termes de Débit-Distorsion (RD). Dans les zones de fort mouvement, la taille du GOP sera réduite à 1, puisque le codage Intra devient plus performant que le WZ.

Le codec WZ considéré dans cette étude est celui que nous avons précédemment développé dans [2] et auquel nous ajoutons la technique de contrôle automatique de la taille du GOP. A ce but, soit  $S_{max}$  la taille maximale permise d'un GOP. Pour une certaine taille  $S$  de GOP, on note par  $F_0$  la trame clé,  $F_1, F_2, \dots, F_{S-1}$  les trames WZ, et  $F_S$  la trame clé du GOP suivant. On note respectivement par  $R_i$  et  $PSNR_i$  le débit moyen et le PSNR (Peak Signal-to-Noise Ratio) de la  $i^{ème}$  trame du GOP.

L'algorithme de décision de la taille d'un GOP opère de la façon suivante :

*Initialisation* :  $S = 1$ .

*Tant que*  $S \leq S_{max}$ :

*Si*  $S=1$ , *aller à l'étape 5, autrement:*

**Etape 1:** *Interpolation des trames  $F_0$  et  $F_S$*

Puisque l'étape d'estimation de mouvement n'est pas permise au codeur à cause de sa complexité, une interpolation moyenne [1] est utilisée. La trame interpolée sert comme estimation grossière de la SI qui sera générée au récepteur (par compensation de mouvement [1]), pour le décodage de la trame WZ  $F_{\lfloor S/2 \rfloor}$ , localisée à mi-distance entre  $F_0$  et  $F_S$ .

**Etape 2:** *Estimation du débit nécessaire  $R_{\lfloor S/2 \rfloor}$*

En se basant sur la trame WZ  $F_{\lfloor S/2 \rfloor}$  et sa SI estimée, le codeur détermine sa limite de compression inférieure d'après la méthode de calcul entropique introduite dans [2], qui permet de prendre en compte le niveau de mouvement dans la trame. Le taux de compression de la trame WZ est ensuite obtenu en multipliant la limite inférieure par une constante  $T_M$  dont la valeur dépend du nombre  $M$  de bits de quantification par pixel [2]. Ce taux permet de déterminer  $R_{\lfloor S/2 \rfloor}$ .

**Etape 3:** *Calcul de  $PSNR_{\lfloor S/2 \rfloor}$*

En utilisant la trame WZ  $F_{\lfloor S/2 \rfloor}$  et sa SI estimée, le codeur peut déterminer une estimation  $F'_{\lfloor S/2 \rfloor}$  de la trame qui sera obtenue au récepteur après turbo-décodage, en quantifiant tout d'abord la trame WZ et en réalisant sa reconstruction

[1] en se basant sur la SI estimée. Le PSNR est alors calculé entre  $F_{\lfloor S/2 \rfloor}$  et  $F'_{\lfloor S/2 \rfloor}$ .

**Etape 4:** *Répétition des étapes 1 à 3 jusqu'à ce que les estimations de débit et de PSNR soient obtenues pour toutes les trames du GOP.*

La Figure 1 montre un exemple de toutes les interpolations nécessaires pour une taille de GOP  $S = 4$ . A chaque estimation, si le processus d'interpolation fait intervenir une trame clé à l'instant  $k$  (Ex : trames  $F_0$  et  $F_4$  de l'interpolation 1 dans la Figure 1), la vraie trame  $F_k$  (ou sa version décodée H264) est utilisée puisqu'elle sera disponible au décodeur. Par contre, s'il fait intervenir une trame WZ précédemment décodée (Ex : trame  $F_2$  pour les interpolations 2 et 3 dans la Figure 1), l'estimation  $F'_k$  de cette dernière (obtenue par l'étape 3 de l'itération précédente de l'algorithme) est utilisée au lieu de  $F_k$ , car  $F'_k$  est une estimation plus fidèle de la trame qui sera disponible au décodeur ( $F_k$  n'étant pas connue de ce dernier).

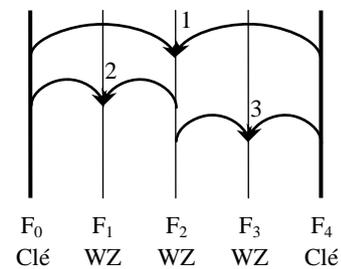


Figure 1 : Etapes d'interpolation pour une taille de GOP  $S = 4$ .

**Etape 5:** *Estimation du débit et du PSNR moyens obtenus avec un GOP de taille  $S$ , respectivement par:*

$$R_{av}^S = \frac{1}{S} \sum_{j=0}^{S-1} R_j \quad \text{et} \quad PSNR_{av}^S = \frac{1}{S} \sum_{j=0}^{S-1} PSNR_j.$$

**Etape 6:** *Calcul du rapport  $\lambda_{av}^S = PSNR_{av}^S / R_{av}^S$*

**Etape 7:** *Incrémentation de la taille du GOP:  $S = S + 1$ .*

Les meilleures performances RD sont obtenues en maximisant le PSNR par unité de débit. Par conséquent, le système décide de la taille optimale  $L$  du GOP en prenant :  $L = \arg \max_{k=1,2,\dots,S_{max}} (\lambda_{av}^k)$ .

Ainsi, si  $L=1$ , une seule trame codée en Intra est transmise. Sinon, une trame clé (Intra) est transmise, suivie de  $L-1$  trames WZ. Cette procédure est répétée à chaque début de GOP.

## 3 Résultats pratiques et discussion

Dans nos simulations, nous considérons 400 trames des deux séquences QCIF Foreman et Grandmother. Dans les Figures 2 et 3, nous représentons les résultats obtenus en termes des courbes RD obtenues en moyennant le débit et le PSNR des trames clés et WZ. Nous avons choisi  $S_{max}=5$  pour l'algorithme de variation dynamique du GOP (appelé VDGOP). Nous faisons

varier le débit total par le choix de la valeur du paramètre de quantification  $M=1, 2, \text{ ou } 4$ .

Nous remarquons que, pour la séquence Foreman, lorsque la taille  $S$  du GOP est fixe, les performances diminuent lorsque  $S$  augmente. Les meilleures performances moyennes sont obtenues lorsque toutes les trames sont codées Intra. Ceci est dû au niveau de mouvement important dans cette scène, qui conduit à une SI de mauvaise qualité au décodeur lorsque les trames clés sont éloignées les unes des autres. Cependant, on remarque que le VDGOP, qui présente des performances similaires au cas du codage Intra à bas débit, permet un gain de 35 kbps pour un PSNR de 39.5 dB.

La séquence Grandmother, caractérisée par un niveau de mouvement plus faible que Foreman, présente un comportement différent. Lorsque la taille du GOP est fixe, les meilleures performances sont obtenues pour  $S=3$ , en particulier à haut débit. Quant au VDGOP, il est plus performant que tous les autres schémas de codage. A un débit de 500 kbps (resp. 750 kbps), un gain de 0.8 dB (resp. 0.9 dB) est observé par rapport au cas  $S=1$  fixe, et de 0.3 dB (resp. 0.2 dB) par rapport à  $S=3$  fixe.

Etant donné que, dans un système pratique à taille de GOP fixe, la valeur optimale de  $S$  ne peut être déterminée à l'avance, notre système VDGOP permet, d'un côté, de résoudre ce problème et, de l'autre, d'améliorer les performances grâce à la variation dynamique de  $S$ .

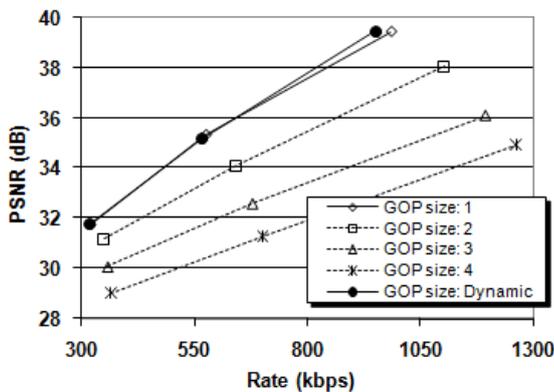


Figure 2 : Performances RD de la séquence Foreman obtenues par un codec WZ à taille de GOP fixe ou variable.

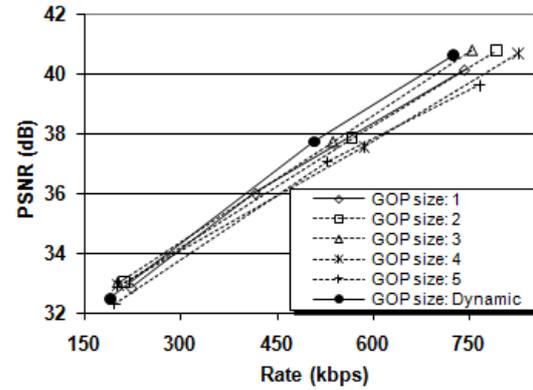


Figure 3 : Performances RD de la séquence Grandmother obtenues par un codec WZ à taille de GOP fixe ou variable

Figure 4 permet de comparer les valeurs du rapport PSNR/R du VDGOP par rapport au codage avec  $S = 1$  ou 3 fixe, pour les 80 premières trames de la séquence Grandmother avec  $M=2$ . Les étiquettes « I » indiquent les positions des trames Intra dans le VDGOP. On remarque que la plupart du temps le VDGOP présente le rapport le plus élevé, sauf pour quelques cas isolés. Ces derniers s'expliquent par le fait que le VDGOP maximise le rapport entre le PSNR et le débit moyennés sur le GOP, alors que la courbe représente la valeur de ce rapport par trame.

Par ailleurs, nous avons observé les pourcentages des tailles de GOP optimales ( $L$ ) obtenues par le VDGOP, pour chaque séquence et pour les différentes valeurs de  $M$  (Figure 5). Pour la séquence Foreman, nous avons remarqué que près de 80 % des GOP ont une taille 1, ce qui indique que le système sélectionne le codage Intra H264 la plupart du temps. Ceci explique la raison de la similarité entre les performances de notre codec WZ et de celui du codage H264 pour la séquence Foreman, comme nous l'avons remarqué dans le graphe de la figure 2. Plus de variations de  $L$  sont observées avec Grandmother: Pour  $M=1$ , aucun GOP n'a pour taille 1, alors que  $L=5$  pour environ 50 % des GOPs. Pour  $M=2$  ou 4, la taille de GOP la plus fréquente (40%) est  $L=2$ .

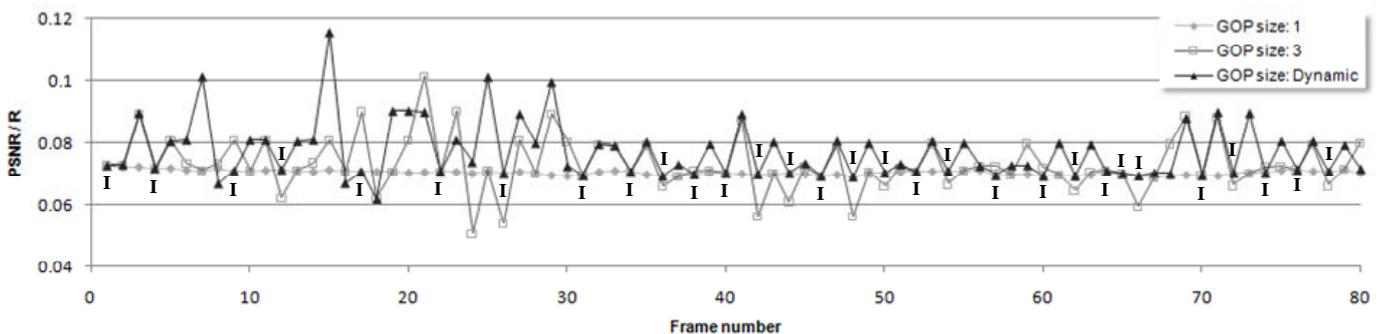


Figure 4 : Rapport PSNR/R en fonction du numéro de la trame pour la séquence Grandmother avec  $M=2$

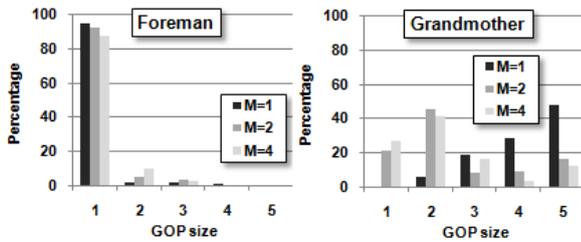


Figure 5 : Pourcentage des tailles de GOP pour les différentes séquences

Dans les graphes de la figure 6, nous représentons les variations des tailles de GOP tout au long de la séquence Foreman, pour  $M = 2$  et  $4$  respectivement. Les régions contenant de longues successions de trames Intra H264 (c.à.d. des tailles de GOP égales à 1) correspondent aux parties de la vidéo caractérisées par les niveaux de mouvement les plus élevés. Dans de telles régions, le codage Intra H264 est plus performant que le codage WZ. Cependant, dans certaines régions, une taille de GOP de 2 ou 3 permet d'obtenir de meilleures performances par rapport au codage Intra. Il en résulte que les performances RD, obtenues par notre système pour la séquence Foreman, sont légèrement meilleures que celles du codage Intra pur, comme nous l'avons constaté dans la figure 2.

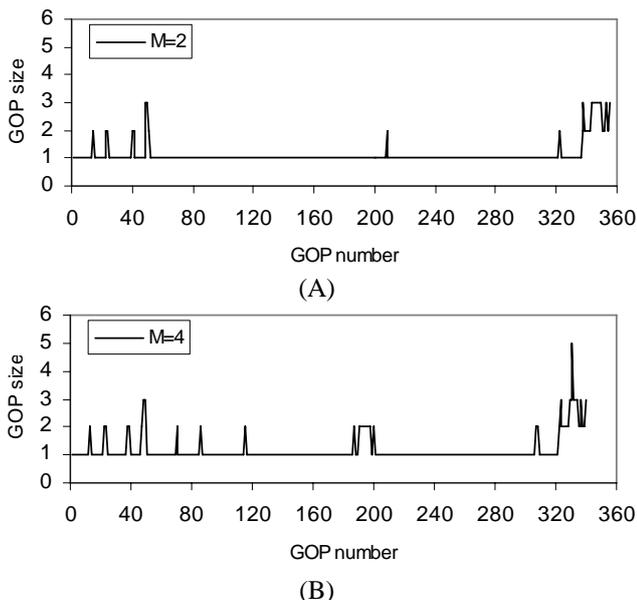


Figure 6 : Variations de la taille du GOP le long de la séquence Foreman pour  $M = 2$  (A) et  $4$  (B).

## 4 Conclusion

Dans cet article, nous avons présenté un algorithme simple qui permet d'adapter de façon dynamique la taille du GOP dans un codeur vidéo Wyner-Ziv, en fonction du contenu de la scène vidéo codée. Le système se base sur des calculs théoriques des limites de compression pour estimer le débit nécessaire au bon décodage des trames WZ, sans besoin d'un canal de retour. Par suite, cette étude est particulièrement intéressante dans le contexte d'applications de diffusion (Broadcasting). La sélection automatique du mode de codage permet de choisir entre le codage Intra H264 et le codage WZ de façon à optimiser les performances globales du système. Les résultats de simulation

montrent un gain en performances qui peut atteindre 0.9 dB par rapport au codage Intra H264 et 0.3 dB par rapport au codage WZ avec une taille de GOP fixe.

Comme perspectives à ce travail, nous comptons étendre l'étude présentée ici pour prendre en compte les conditions du canal de transmission dans la variation dynamique de la taille du GOP. Nous comptons aussi étudier l'implémentation du système dans le contexte plus réaliste d'un réseau multi-utilisateurs, en se basant sur nos recherches précédentes [2, 10].

## 5 Remerciements

Ce travail a été en partie financé par le CNRS Libanais et a eu lieu dans le cadre du projet franco-libanais CEDRE.

## 6 Bibliographie

- [1] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv Coding of Motion Video", 36th Asilomar Conference on Signals, Systems and Computers, pp. 240-244, November 2002.
- [2] C. Yaacoub, J. Farah, B. Pesquet-Popescu, "Feedback Channel Suppression in Distributed Video Coding with Adaptive Rate Allocation and Quantization for Multiuser Applications", EURASIP Journal on Wireless Communications and Networking (WCN), 2008.
- [3] X. Artigas and L. Torres, "Improved signal reconstruction and return channel suppression in Distributed Video Coding systems", 47th International Symposium ELMAR-2005, Croatia, June 2005.
- [4] M. Morbee, J. Prades-Nebot, A. Pizurica and W. Philips, "Rate Allocation Algorithm For Pixel-Domain Distributed Video Coding Without Feedback Channel", 32nd International Conference on Acoustics, Speech and Signal Processing, Hawaii, USA, April 2007.
- [5] D. Kubasov, K. Lajnef and C. Guillemot, "A hybrid encoder/decoder rate control for a Wyner-Ziv video codec with a feedback channel", IEEE MultiMedia Signal Processing Workshop, Grece, October 2007.
- [6] C. Brites and F. Pereira, "Encoder rate control for transform domain Wyner-Ziv video coding", International Conference on Image Processing, USA, September 2007.
- [7] Y. Tonomura, D. Shirai, T. Nakachi, and Tetsuro Fujii, "Optimal Bit Allocation for Wavelet-Based Distributed Video Coding", 8th IEEE International Symposium on Multimedia, USA, December 2006.
- [8] A. Aaron, E. Setton and B. Girod, "Towards practical Wyner-Ziv coding of video", IEEE International Conference on Image Processing, Barcelona, Spain, Sept. 2003.
- [9] J. Ascenso, C. Brites, F. Pereira, "Content Adaptive Wyner-Ziv Video Coding Driven by Motion Activity", IEEE International Conference on Image Processing, Atlanta, USA, October 2006.
- [10] C. Yaacoub, J. Farah, B. Pesquet-Popescu, "Optimal Rate Allocation in Multi-User Wyner-Ziv Video Coding Systems with Coded Key Frames", 19th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, France, September 2008.
- [11] D. Slepian and J.K. Wolf, "Noiseless Coding of Correlated Information Sources", IEEE Transactions on Information Theory, Vol. IT-19, pp. 471-480, July 1973.
- [12] D. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", IEEE Transactions on Information Theory, Vol. IT-22, pp. 1-10, January 1976.
- [13] ITU-T and ISO/IEC JTC1, "Advanced Video Coding for Generic Audiovisual Services." ITU-T Recommendation H.264 – ISO/IEC 14496-10 AVC, 2003.
- [14] C. Brites, J. Ascenso and F. Pereira, "Feedback Channel in Pixel Domain Wyner-Ziv Video Coding: Myths and Realities", 14th European Signal Processing Conference, Italy, September 2006.