

# Codage vidéo scalable par maillage et ondelettes 3D

Nathalie Cammas<sup>1</sup>, Stéphane Pateux<sup>2</sup>

<sup>1</sup>France Telecom RD, 4 rue du Clos Courtel, 35512 Cesson-Sévigné

<sup>1</sup> et <sup>2</sup> IRISA-INRIA, Campus de Beaulieu, avenue du Général Leclerc, 35042 Rennes

{ncammas, spateux}@irisa.fr

## Résumé

*Cet article présente un schéma de codage vidéo scalable basé sur une approche de type analyse-synthèse. Le mouvement est représenté par un maillage déformable qui permet à l'aide d'une transformée ondelette temporelle appliquée le long de la trajectoire du mouvement d'exploiter au mieux la redondance temporelle dans une séquence vidéo. L'utilisation d'ondelettes 3d dans la stratégie de codage offre une scalabilité naturelle au codeur. De plus, le schéma d'analyse-synthèse permet de découpler le mouvement et la texture et de coder ces deux informations séparément. Le mouvement peut alors être codé avec pertes, le gain en débit pouvant ainsi être mis à profit pour le codage de la texture. Parce que le codage du mouvement est avec pertes, un critère de mesure de qualité dans le domaine texture est proposé.*

*Enfin, le schéma de codage proposé permet de dépasser certaines limites des codeurs vidéo existants liées principalement à l'utilisation de mouvements basés blocs. Il offre des performances comparables au codeur H26L, tout en fournissant un flux scalable.*

## Mots clefs

Analyse-synthèse, scalabilité, maillages déformables, ondelettes 3D, codage vidéo

## 1 INTRODUCTION

Le codage vidéo est utilisé dans de nombreuses applications qui nécessitent des ressources et des bandes passantes variées et variables. Pour répondre à ces différents besoins, il est alors intéressant de disposer d'un flux vidéo possédant des propriétés de scalabilité, c'est-à-dire pouvant s'adapter aux ressources et débits disponibles. La scalabilité peut s'obtenir par l'utilisation de transformées ondelette dans le schéma de codage vidéo, en effet, ces deux notions, ondelette et scalabilité, permettent chacune de représenter un signal de manière hiérarchique. L'idée est alors d'étendre le cas du codage d'images par ondelettes 2D, type JPEG-2000, EZW ou SPIHT, au cas 3D des séquences vidéos. Une transformée ondelette est ainsi appliquée sur chaque axe du signal vidéo: horizontal, vertical et temporel. Mais les ondelettes seront d'autant plus efficaces pour décomposer un signal vidéo que l'on pourra exploiter la corrélation temporelle. Comme le signal vidéo

varie peu le long des trajectoires du mouvement, pour exploiter au mieux la redondance temporelle, il est alors nécessaire d'appliquer la transformée ondelette temporelle le long de ces trajectoires.

Plusieurs schémas de codage vidéo exploitant l'ondelette temporelle et le mouvement ont déjà été proposés. Une première méthode [7] utilise le mouvement global de la scène pour préalablement redresser toutes les images afin d'effectuer une transformée ondelette temporelle. D'autres schémas basés mouvement par blocs ont été proposés par Ohm [1], puis par Choi et Woods [2]. Dans ces schémas, la transformée temporelle est appliquée sur les blocs déplacés des images, mais l'utilisation d'un mouvement par blocs fait apparaître des pixels déconnectés ou doublement connectés qui doivent être gérés différemment. Une autre approche de codage vidéo par ondelettes est celle introduisant la transformation ondelette sous sa forme lifting. Dans [4], la transformée ondelette et la compensation en mouvement sont appliquées dans le même temps. Cette technique permet d'effectuer les calculs sur les grilles d'échantillonnage de chaque image permettant ainsi la réversibilité du schéma. Mais il est alors nécessaire de disposer des champs de mouvement avant et arrière, impliquant alors un coût de codage important pour le mouvement. Dans [5], les auteurs utilisent le même principe avec cependant un nombre de champs de mouvement moindre que dans la méthode précédente grâce à l'utilisation d'un filtre de transformée 5/3 tronqué.

Afin de pallier aux différents inconvénients des méthodes présentées, nous proposons une nouvelle approche de codage vidéo scalable basé sur un schéma de type analyse-synthèse. L'analyse-synthèse nous permet de découpler la texture et le mouvement par projection des images de la séquence vidéo sur des grilles de références. Pour obtenir une bonne représentation du mouvement apparent, nous utilisons un maillage déformable. Les maillages sont de bons outils de compensation en mouvement, ils permettent un suivi continu de la texture qui justifie l'utilisation des ondelettes le long de ces trajectoires. Enfin, les ondelettes 3D offrent une scalabilité naturelle à notre schéma de codage vidéo.

Notre méthode permet de dépasser certaines limitations des schémas de codage existants, et les résultats montrent qu'en terme de qualité visuelle notre approche par analyse-synthèse est aussi performante qu'un codeur par prédiction,

comme H26L, offrant en plus un flux vidéo scalable.

## 2 SCHEMA PROPOSE

### 2.1 Schéma général

Le schéma 1 présente une vue globale du schéma d'analyse-synthèse. L'analyse permet de découpler les informations de texture et mouvement, qui sont ensuite codées à l'aide d'une transformée ondelette temporelle, et de codeurs offrant des flux progressifs scalables. La phase de synthèse permet de reconstruire la séquence à partir des informations de mouvement et de texture.

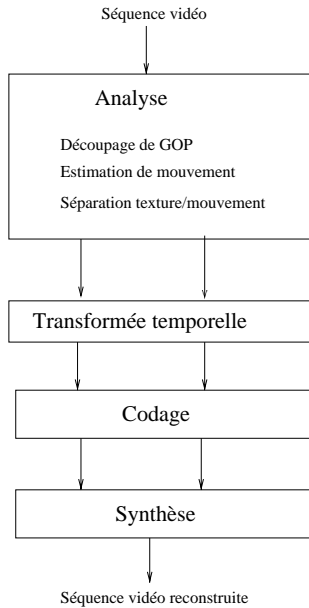


FIG. 1 – Schéma global du codeur vidéo

### 2.2 Estimation de mouvement

Le schéma proposé est basé sur une approche analyse-synthèse. La phase d'analyse extrait des groupes de  $N$  images et effectue une estimation de mouvement à l'aide d'un maillage déformable [6]. Les images sont ensuite redressées par rapport à des grilles de référence, comme dans [7], cette étape permet de séparer le mouvement de la texture.

### 2.3 Transformée temporelle

Une fois plaquée, la texture est transformée en sous-bandes temporelles à l'aide d'une transformée ondelette lifting. La transformée lifting est une alternative au schéma de transformée convolutif, il permet une inversion du schéma plus rapide et le contrôle des erreurs d'arrondis. Lors de la transformée temporelle des images de texture, une compensation en mouvement n'est nécessaire que si les images n'ont pas la même image de référence. Dans notre cas, les images de références peuvent être la première et/ou la dernière image d'un GOP.

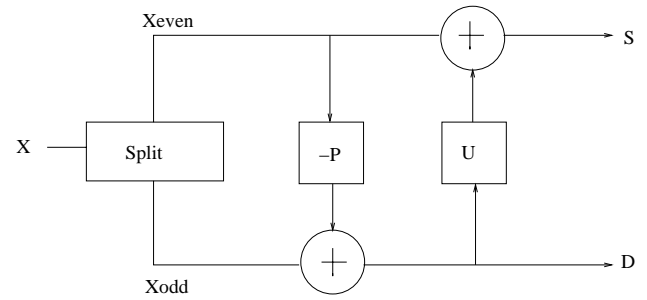


FIG. 2 – Schéma lifting d'une transformée ondelette

### 2.4 Codage

Les sous-bandes temporelles sont ensuite transformées par une transformée ondelette spatiale. Enfin, les sous-bandes spatio-temporelles de texture obtenues sont codées par un codeur scalable, type EBCOT. Une représentation hiérarchique du mouvement est utilisée, [8], les informations de raffinement sont codées avec pertes par plans de bits de façon progressive à l'aide d'un codeur arithmétique. Le gain en mouvement est répercuté sur la texture. Au final, il est possible de décoder le flux à différentes qualités associées à différents débits. Une couche de qualité de base minimale restant à un débit très bas (typiquement 60kbs pour la séquence Foreman, 15Hz).

### 2.5 Synthèse

La phase de synthèse reconstruit la séquence vidéo en re-projetant les images de texture sur leur grille originale. Le codage du mouvement avec pertes se base sur l'hypothèse que l'oeil humain est moins sensible aux erreurs de mouvement qu'aux erreurs de texture, ainsi la vidéo reconstruite en utilisant le mouvement codé avec pertes sera décalée par rapport à la vidéo originale, mais ce décalage ne sera pas perceptible par l'oeil. Cependant, la mesure de qualité de la séquence vidéo reconstruite par le calcul du PSNR se trouve biaisée. Afin de s'affranchir de ce biais, nous avons choisi de mesurer la qualité restituée de la séquence vidéo reconstruite par un calcul de PSNR dans le domaine des images de texture. Le PSNR texture est calculé entre les images de texture initiales issues de la phase d'analyse et les images de texture décodées, cette mesure traduit bien la qualité visuelle de la séquence reconstruite. Ce problème de mesure de qualité a été également abordé dans [9], l'auteur traite de la reconstruction 3D de séquences vidéo à l'aide d'une modélisation géométrique de la scène par un maillage et d'images de texture. Dans notre cas et dans le cas de [9], l'optimisation du critère sur le mouvement et du critère sur la texture est faite séparément, contrairement à un codeur classique type MPEG ou H26x. Le gain sur le mouvement peut alors être répercuté sur la texture, en effet le coût de mouvement des approches classiques est de l'ordre de 45% du débit total pour 13% avec notre approche.

### 3 RESULTATS

#### 3.1 Conditions expérimentales

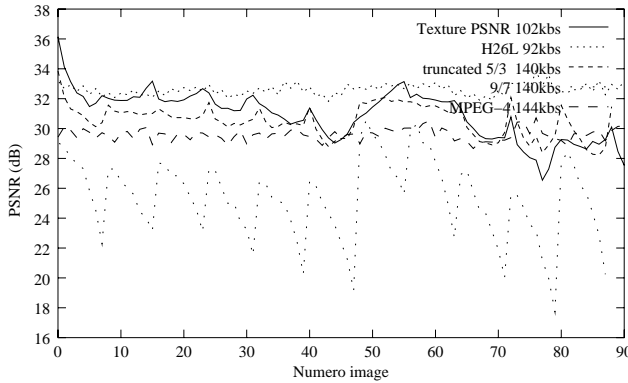


FIG. 3 – Séquence Foreman CIF, YUV420, 15Hz

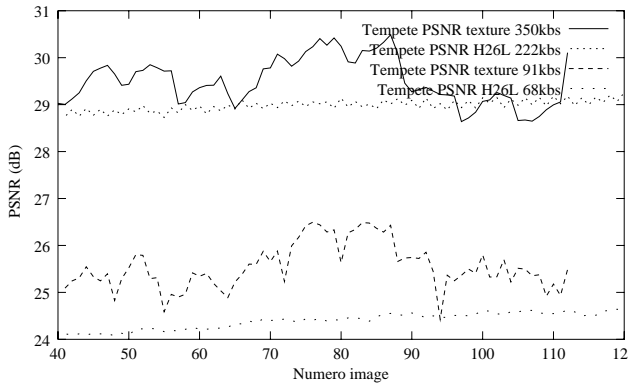


FIG. 4 – Séquence Tempete CIF, YUV420, 30Hz

Le schéma proposé a été testé sur différentes séquences et à différents débits. Nous présentons les résultats obtenus sur la séquence Foreman CIF, YUV420, 15Hz et la séquence Tempête CIF, YUV420, 30Hz. La figure 3 compare les performances obtenues par notre schéma sur la séquence Foreman avec le codeur H26Lv8, le codeur MPEG-4 M-omusys et deux codeurs basés ondelette3D décrits dans [3]. Le premier codeur vidéo ondelette représente le mouvement par blocs et effectue une transformée temporelle avec compensation en mouvement à l'aide d'un filtre lifting 9/7. Les sous-bandes temporelles sont codées par un codeur progressif scalable et le coût du mouvement représente 45% du débit total. Le deuxième codeur par ondelette utilise aussi un mouvement par blocs, mais effectue la transformée temporelle à l'aide d'un filtre lifting 5/3 tronqué. Le codeur H26Lv8 a été utilisé dans une configuration de performances maximales avec l'option d'optimisation débit-distorsion activée, QP=31, profil IPBB et codage arithmétique.



FIG. 5 – Foreman image 117 reconstruite par le schéma maillage+ondelette 3D



FIG. 6 – Foreman image 117 codée-décodée par H26L

#### 3.2 DISCUSSION

Les résultats montrent que les codeurs ondelette utilisant un mouvement par blocs sont limités à des filtres ondelette temporels courts. Ceci s'explique par le fait que le mouvement par blocs contraint en débit produit des discontinuités aux frontières des blocs, ces discontinuités sont inadaptées à la transformation par ondelette temporelle, car les sous-bandes hautes fréquences temporelles résultantes sont de forte énergie et sont difficiles à coder (cf [3] pour plus de détails). L'utilisation de filtres ondelette temporel courts donne alors de meilleurs résultats (5/3 tronqué de la figure 3) que les filtres longs (9/7 de la figure 3). Les maillages qui procurent un suivi continu de la texture au cours du temps paraissent alors plus adaptés à la représentation

du mouvement dans le cadre de transformée ondelette temporelle, ceci est confirmé par les résultats de notre codeur qui offrent les mêmes performances que le codeur par blocs utilisant le filtre 5/3 tronqué. Ces résultats sont d'autant plus encourageants que le codeur par maillage ne gère pas les zones à problème du maillage.

En effet, il reste des zones où les maillages ne modélisent pas le mouvement correctement, impliquant des artefacts visuels importants qu'il faut corriger en augmentant le coût de codage de la texture. Ces zones difficiles à modéliser sont les zones d'occlusion et les zones présentant un fort changement de résolution. Le premier cas intervient quand des objets se découvrent ou se recouvrent, la texture est fortement expansée ou contractée dans ces zones. L'aplatissement des mailles conduit alors à des retournements générant une texture reconstruite faite d'aliasing et de fausses textures qui sont difficiles à coder et n'apportent aucune information nécessaire à la bonne reconstruction de la séquence. Une solution permettant de gérer ces problèmes serait l'utilisation de technologie de représentation et de codage basé objet et l'introduction de lignes de rupture, notion présentée dans [8]. Les zones de changement de résolution dans la texture provoquent des artefacts chers à coder dans le domaine texture, une solution serait de pondérer ces zones au moment du codage par un facteur prenant en compte la déformation.

Le schéma de codage par maillage et ondelettes 3D permet de dépasser certaines limitations des codeurs existants type H26L ou MPEG-4. La vidéo reconstruite ne présente pas les artefacts dûs aux effets de blocs des schémas standardisés, ni l'effet lissant des filtrages utilisés pour limiter les effets de blocs, la texture de la vidéo reconstruite par le schéma d'analyse-synthèse est beaucoup plus nette que celle reconstruite par H26L (voir figures 5 et 6). De plus, le schéma par analyse-synthèse offre une scalabilité naturelle au niveau du codage des informations de texture mais aussi au niveau des informations de mouvement. La vidéo peut être reconstruite à différents niveaux de qualité de texture et de mouvement.

Par rapport aux codeurs basés blocs, le codeur basé maillage et ondelette 3D possède l'avantage d'avoir un coût de mouvement moindre. En effet, les maillages sont beaucoup moins chers à coder qu'un mouvement basé blocs et l'inversion possible du mouvement implique qu'un seul champ de mouvement est nécessaire.

## 4 CONCLUSION

L'approche présentée propose un codeur vidéo scalable basé sur un schéma de type analyse-synthèse utilisant les maillages et les ondelettes 3D. Le schéma d'analyse-synthèse offre une nouvelle méthode de représentation des informations en séparant par une phase d'analyse les informations de mouvement et de texture et en codant séparément ces informations. La phase de synthèse permet de reconstruire une vidéo de qualité visuelle aussi proche

que possible de la vidéo originale, tout en offrant la possibilité de reconstruire la vidéo à différentes qualités. L'utilisation des ondelettes permet d'offrir une scalabilité naturelle à notre schéma, couplé à un outil de compensation par maillage offrant une continuité dans le suivi de la texture, ces outils permettent d'exploiter au mieux la redondance temporelle dans la séquence vidéo. L'hypothèse de base du schéma d'analyse-synthèse, qui est que l'oeil humain est moins sensible aux défauts de mouvement qu'aux défauts de texture nous permet de coder le mouvement avec pertes. Toutefois, nous avons vu que le codage du mouvement avec pertes implique l'utilisation d'une mesure de qualité dans le domaine texture.

## Références

- [1] J.R Ohm. Three-dimensional subband coding with motion compensation. *IEEE Transactions on Image Processing*, 3(5):559–571, September 1994.
- [2] S-J. Choi et J.W. Woods. Motion-compensated 3-d subband coding of video. *IEEE Transactions on Image Processing*, 8(2):155–167, february 1999.
- [3] J. Vieron, C. Guillemot, et S. Pateux. Motion compensated 2d+t wavelet analysis for low rate fgs vido compression. Dans *International Thyrrhenian workshop on digital communications 2002 (invited paper)*, 2002.
- [4] A. Secker et D. Taubman. Motion-compensated highly scalable video compression using an adaptative 3d wavelet transform based on lifting. *IEEE*, 2001.
- [5] L.Luo, J.Li, S.Li, Z.Zhuang, et Y-Q.Zhang. Motion-compensated lifting wavelet and its application in video coding. Dans *IEEE International Conference on Multimedia and Expo*, August 2001.
- [6] G.Marquant, S. Pateux, et C.Labit. Mesh and "crack lines": Application to object-based motion estimation and higher scalability. Dans *IEEE International Conference on Image Processing, ICIP'00, vancouver, Canada*, September 2000.
- [7] D. Taubman et A. Zakhor. Multirate 3-d subband coding of video. *IEEE Transactions on Image Processing*, 3(5):572–588, september 1994.
- [8] Gwenaëlle Marquant. *Représentation par maillage adaptatif déformable pour la manipulation et la communication d'objets vidéo*. Thèse de doctorat, Université de rennes 1, 2000.
- [9] F. Galpin. *Représentation 3D de séquences vidéo, Schéma d'extraction automatiquze d'un flux de modèles 3D, applications à la compression et à la réalité virtuelle*. Thèse de doctorat, Université Rennes 1, janvier 2002.