Codage par descriptions multiples pour la transmission vidéo

Christophe TILLIER[1], Teodora PETRISOR[1], Béatrice PESQUET-POPESCU[1], Jean-Christophe PESQUET[2]

[1]ENST, Département du Traitement du Signal et des Images
46, rue Barrault, 75634 Paris Cédex 13, FRANCE

[2]IGM et UMR-CNRS 8049, Université de Marne-la-Vallée,
5, Bd Descartes, Champs sur Marne, 77454 Marne la Vallée Cédex 2, FRANCE
{tillier,petrisor,pesquet}@tsi.enst.fr, pesquet@univ-mlv.fr

Résumé – Dans cet article, nous nous intéressons à l'utilisation du codage par descriptions multiples pour la transmission de séquences vidéo scalables obtenues avec un codeur ondelettes $t+2D$. Ces techniques sont utiles pour des réseaux commutés, à perte de paquets. Contrairement aux approches existantes, la redondance n'est pas introduite dans le domaine spatial mais dans le domaine temporel, grâce à des bancs de filtres dyadiques sur-échantillonnés, compensés en mouvement. Nous fournissons ici un cadre de travail pour construire et étudier des schémas de codage par descriptions multiples temporels, basés sur des décompositions en ondelettes redondantes avec un facteur de redondance réduit, et nous validons cette approche par une application au codage vidéo.

Abstract – In this paper, we address the problem of multiple description coding for the transmission over best-effort networks of scalable video sequences obtained with a $t+2D$ wavelet codec. Unlike previous approaches, which consider redundancy insertion in the spatial domain, in this paper we introduce redundancy in the temporal domain of the video sequence, using motion compensated oversampled dyadic filter banks. We are giving here the framework for building such temporal multiple description schemes based on redundant wavelet decompositions with a reduced redundancy feature. The general approach is applied to video coding.

1 Introduction

En codage vidéo scalable, un intérêt croissant s'est manifesté, au cours de ces dernières années, pour des structures compensées en mouvement utilisant la décomposition de Haar ou des bancs de filtres biorthogonaux 5/3 opérant temporellement [1], [2]. Dans les applications de communication, les représentations scalables sont utiles pour l'adaptation du flux binaire aux variations de bande passante ou aux caractéristiques du récepteur, mais dans le cas de transmission sur des réseaux « best effort », une difficulté nouvelle dans la reconstruction est due aux pertes de paquets.

Le codage par descriptions multiples (MDC) [3], qui est étroitement lié au codage conjoint source-canal, permet de répondre à ce besoin en introduisant de la redondance au niveau de la source pour gérer les pertes de paquets. Cette technique vise à construire des descriptions corrélées qui sont transmises sur des canaux indépendants tout-ou-rien. Dans le cas d'échecs de transmission sur certains canaux, des décodeurs dits « latéraux » doivent être capables de reconstruire la source avec une qualité acceptable. En revanche, la réception de toutes les descriptions doit permettre une reconstruction de qualité supérieure, qui est obtenue en sortie du décodeur dit « central ».

Dans ce travail, nous explorons des schémas de descriptions multiples basés sur une décomposition temporelle en ondelettes redondantes opérant dans un codeur $t+2D$. Notons que la plupart des travaux précédents effectuait la construction des descriptions multiples d'une vidéo à l'aide de plusieurs boucles de prédiction temporelle dans un schéma de codage hybride, [4], [5]. Pour les codeurs hybrides un état de l'art récent des techniques de MDC en vidéo est présenté dans [6].

Il existe aussi des schémas MDC qui emploient des codeurs en ondelettes 3D. Néanmoins, contrairement à d'autres approches, où le banc de filtres en ondelettes était appliqué indépendamment dans le domaine spatial pour chaque image [7], [8], ou sur les trames résultant d'une décomposition $t+2D$ [9], notre approche concerne l'utilisation de bancs de filtres sur-échantillonnés dans la décomposition temporelle. Des travaux précédents [10], [11] ont introduit de la redondance temporelle dans un banc de filtres 3-bandes. Ici nous proposons plusieurs possibilités de construction des descriptions corrélées à partir d'un banc de filtres dyadique sur-échantillonné, appliqué dans la direction temporelle d'une séquence vidéo traitée avec un codeur $t+2D$. Les propriétés de tels banc de filtres agissant comme des codeurs de canal sur des canaux à effacements ont récemment été explorées dans [12], [13], [14].

Nous nous différencions de ces méthodes en imposant une particularité supplémentaire à nos schémas, qui est un facteur de redondance réduit. Nous obtenons cette propriété en effectuant un sous-échantillonnage d'un facteur 2 des trames de détails temporels après la décomposition de base. Cela est équivalent à dire que nos schémas ont une redondance (en nombre de coefficients d'ondelettes) égale à la taille d'une sous-bande d'approximation temporelle. Pourtant, dans ce cas la reconstruction parfaite n'étant plus garantie, nous sommes amenés à étudier l'inversibilité du décodeur central. De plus, ceci permet aussi d'analyser les effets du bruit de quantification sur la reconstruction.

Dans le paragraphe suivant, nous présentons le banc de filtres considéré ainsi que les schémas MDC proposés. Dans le paragraphe 3 nous considérerons l'application de ces schémas au codage vidéo, puis nous fournissons des résultats de simulation dans le paragraphe 4. Enfin, nous concluons dans le para-

graphe 5.

2 Représentations par bancs de filtres redondants

En considérant un banc de filtres dyadique associé à une décomposition en ondelettes nous allons générer une structure redondante, à partir de laquelle plusieurs configurations de descriptions multiples sont possibles. Nous allons analyser et comparer ces configurations et proposer une solution qui conserve la propriété de reconstruction parfaite, tout en réduisant la redondance.

Soit $(x_n)_{n\in\mathbb{Z}}$, un signal d'entrée temporel et notons $(h_n)_{n\in\mathbb{Z}}$ (resp. $(g_n)_{n\in\mathbb{Z}}$) le filtre passe-bas (resp. passe-haut) d'un banc de filtres d'analyse dyadique à reconstruction parfaite. Les expressions des coefficients d'approximation et de détails sont alors

$$a_n^{\mathrm{I}} = \sum_k h_{2n-k}\, x_k \tag{1}$$

$$d_n^{\mathrm{I}} = \sum_k g_{2n-k}\, x_k. \tag{2}$$

En sous-échantillonnant aux instants impairs plutôt qu'aux instants pairs, les coefficients d'approximation et de détails deviennent

$$a_n^{\mathrm{II}} = \sum_k h_{2n-1-k}\, x_k \tag{3}$$

$$d_n^{\mathrm{II}} = \sum_k g_{2n-1-k}\, x_k. \tag{4}$$

En considérant ces 4 ensembles de coefficients, nous avons naturellement une décomposition redondante, le nombre des coefficients étant multiplié par 2.

Il est toutefois possible de construire des représentations plus « économiques » tout en gardant la propriété souhaitée de reconstruction parfaite. Pour ce faire, nous décimons les séquences précédentes d'un facteur 2, et nous introduisons alors les notations $\hat{a}_n^{\mathrm{I}} = a_{2n}^{\mathrm{I}}$ et $\check{a}_n^{\mathrm{I}} = a_{2n-1}^{\mathrm{I}}$, des notations similaires étant utilisées pour les séquences de coefficients de détails vues précédemment.

Ainsi, le vecteur $\overline{\mathbf{c}}_n$ qui contient toutes les séquences sous-échantillonnées possibles est :

$$\overline{\mathbf{c}}_n = \left(\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{a}_n^{\mathrm{II}}\ \check{a}_n^{\mathrm{II}}\ \hat{d}_n^{\mathrm{I}}\ \check{d}_n^{\mathrm{I}}\ \hat{d}_n^{\mathrm{II}}\ \check{d}_n^{\mathrm{II}}\right)^{\intercal}.$$

Le système proposé est facilement décrit en utilisant une décomposition polyphase. Introduisons les composantes polyphases 4-bandes des filtres d'analyse :

$$h_i(n) = h_{4n-i}$$
$$g_i(n) = g_{4n-i}$$

où $i \in \{0,\ldots,3\}$, ainsi que les transformées en z correspondantes $H_i(z)$ et $G_i(z)$.

De la même manière, nous définissons les quatre composantes polyphases du signal d'entrée $x_n^{(i)} = x_{4n+i}$, $i \in \{0,\ldots,3\}$, et le vecteur des composantes polyphases correspondant :

$$\mathbf{x}_n = \left(x_n^{(0)}\ x_n^{(1)}\ x_n^{(2)}\ x_n^{(3)}\right)^{\intercal}$$

Les équations (1) - (4) peuvent se ré-écrire à l'aide de la représentation polyphase comme suit :

$$\overline{\mathbf{C}}(z) = \overline{\mathbf{M}}(z)\ \mathbf{X}(z) \tag{5}$$

où $\overline{\mathbf{C}}(z)$ et $\mathbf{X}(z)$ sont les transformées en z de $(\overline{\mathbf{c}}_n)_{n\in\mathbb{Z}}$ et de $(\mathbf{x}_n)_{n\in\mathbb{Z}}$. $\overline{\mathbf{M}}(z)$ est la matrice de transfert polyphase globale.

$$\overline{\mathbf{M}}(z) = \begin{bmatrix} H_0(z) & H_1(z) & H_2(z) & H_3(z) \\ H_2(z) & H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} \\ H_1(z) & H_2(z) & H_3(z) & H_0(z)z^{-1} \\ H_3(z) & H_0(z)z^{-1} & H_1(z)z^{-1} & H_2(z)z^{-1} \\ G_0(z) & G_1(z) & G_2(z) & G_3(z) \\ G_2(z) & G_3(z) & G_0(z)z^{-1} & G_1(z)z^{-1} \\ G_1(z) & G_2(z) & G_3(z) & G_0(z)z^{-1} \\ G_3(z) & G_0(z)z^{-1} & G_1(z)z^{-1} & G_2(z)z^{-1} \end{bmatrix}.$$

Des schémas de faible redondance s'obtiennent, par exemple, en éliminant 2 des 8 composantes de $\overline{\mathbf{c}}_n$. Néanmoins, un choix judicieux des séquences omises est nécessaire, pour garder une distorsion latérale acceptable. Pour ce faire, tous les schémas que nous construisons conservent les sous-bandes d'approximations issues des deux décompositions, et éliminent seulement la moitié des composantes polyphases des sous-bandes de détails.

Le vecteur $\mathbf{c}_n$ résultant correspond à un facteur de redondance égal à $3/2$. Le banc de filtres sur-échantillonné résultant est tel que $\mathbf{C}(z) = \mathbf{M}(z)\ \mathbf{X}(z)$, où $\mathbf{C}(z)$ est la transformée en z de $(\mathbf{c}_n)_{n\in\mathbb{Z}}$ et $\mathbf{M}(z)$ est la matrice de transfert polyphase du schéma considéré.

Plusieurs décompositions possibles peuvent malgré tout être obtenues suivant le choix de la sous-matrice $\mathbf{M}(z)$. Nous nous sommes intéressés plus particulièrement aux 4 solutions suivantes, pour lesquelles nous spécifions également la manière de construire les 2 descriptions.

- **Schéma R :** ce schéma consiste à séparer les coefficients de détails issus d'une analyse classique monodimensionnelle (à sous-échantillonnage critique) en deux groupes : les coefficients d'indice pair et les coefficients d'indice impair, chacun des groupes appartenant à l'une des descriptions. Les coefficients d'approximation sont simplement dupliqués, d'où

$$\mathbf{c}_n = (\underbrace{\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{d}_n^{\mathrm{I}}}_{1^e \text{ description}}\ \underbrace{\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \check{d}_n^{\mathrm{I}}}_{2^e \text{ description}})^{\intercal}.$$

- **Schéma D1 :** on distribue les coefficients de détail de la même manière que dans le schéma précédent, mais dans la seconde description, au lieu de répéter les coefficients d'approximation de la première description, on utilise les coefficients d'approximation de la seconde base. Cela correspond à

$$\mathbf{c}_n = (\underbrace{\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{d}_n^{\mathrm{I}}}_{1^e \text{ description}}\ \underbrace{\hat{a}_n^{\mathrm{II}}\ \check{a}_n^{\mathrm{II}}\ \check{d}_n^{\mathrm{I}}}_{2^e \text{ description}})^{\intercal}.$$

- **Schéma D2 :** on ajoute plus de « diversité » dans le choix des coefficients de détail en prenant en compte les coefficients d'indices pairs de la seconde base :

$$\mathbf{c}_n = (\underbrace{\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{d}_n^{\mathrm{I}}}_{1^e \text{ description}}\ \underbrace{\hat{a}_n^{\mathrm{II}}\ \check{a}_n^{\mathrm{II}}\ \hat{d}_n^{\mathrm{II}}}_{2^e \text{ description}})^{\intercal}.$$

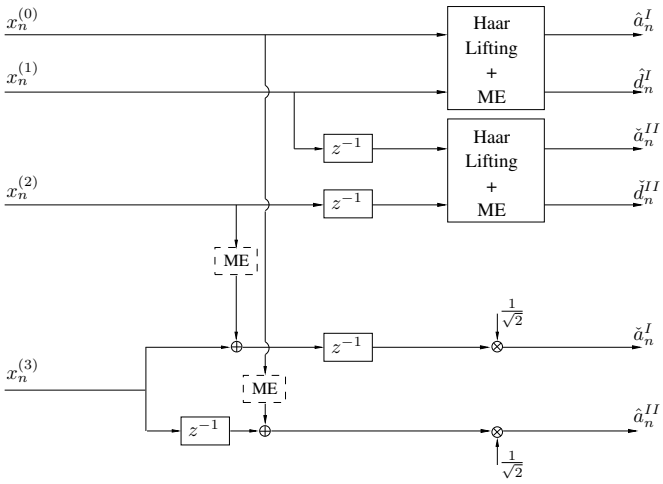

FIG. 1 – Schéma en lifting 4-bandes redondant (l'estimation de mouvement optionnelle est indiquée en pointillés).

– **Schéma D3 :** en sélectionnant dans la seconde base les coefficients de détails d'indices impairs plutôt que ceux d'indices pairs, on obtient alors

$$\mathbf{c}_n = (\underbrace{\hat{a}_n^{\mathrm{I}}\ \check{a}_n^{\mathrm{I}}\ \hat{d}_n^{\mathrm{I}}}_{1^e \text{ description}}\ \underbrace{\hat{a}_n^{\mathrm{II}}\ \check{a}_n^{\mathrm{II}}\ \check{d}_n^{\mathrm{II}}}_{2^e \text{ description}})^{\intercal}.$$

En l'absence de quantification, les deux premiers schémas sont évidemment inversibles puisqu'ils contiennent tous les coefficients obtenus à partir d'une analyse à sous-échantillonnage critique classique. Pour les 2 autres schémas, l'inversibilité n'est pas garantie *a priori*.

Reconstruire le signal à partir des représentations redondantes proposées, revient en termes de MDC à définir le schéma de synthèse utilisé par le décodeur central. Avec les notations précédentes, le problème est de trouver une matrice de transfert $\mathbf{W}(z)$ polynomiale de dimension $N \times K$ telle que :

$$\mathbf{W}(z)\mathbf{M}(z) = \mathbf{I}_{N\times N} \tag{6}$$

où

$$\mathbf{W}(z) = [W_{i,j}(z)]_{1\le i\le N, 1\le j\le K}$$

et

$$\mathbf{M}(z) = [M_{i,j}(z)]_{1\le i\le K, 1\le j\le N}$$

avec $K = 6$ et $N = 4 < K$. Ce problème n'ayant pas de solution unique, nous avons mis en évidence une solution optimale (minimisant les effets dûs aux bruits de quantification) par une approche de type pseudo-inverse [15].

3 Application au codage vidéo

Une difficulté supplémentaire inhérente au codage vidéo est la prise en compte de l'estimation/compensation de mouvement.

Dans chaque description une paire de coefficients d'approximation et de détail est obtenue par un étage lifting classique. L'application d'une transformée compensée en mouvement pour cette paire se fait naturellement [16]. En revanche, le coefficient d'approximation d'indice complémentaire apparaît seul, suite à l'élimination du coefficient de détail correspondant. Ceci est illustré dans la figure 1.

Pour le calcul de ce coefficient, deux stratégies différentes peuvent être envisagées, suivant que l'on fasse intervenir ou non une compensation de mouvement. En introduisant cette compensation de mouvement (en pointillés sur la figure 1), on augmente un peu la redondance car on a à coder un champ de vecteurs de mouvement supplémentaire. Pourtant, ce champ permet d'obtenir une sous-bande d'approximation très proche d'une image naturelle, ce qui facilite son codage. En plus, lors du décodage latéral l'efficacité de reconstruction est aussi augmentée.

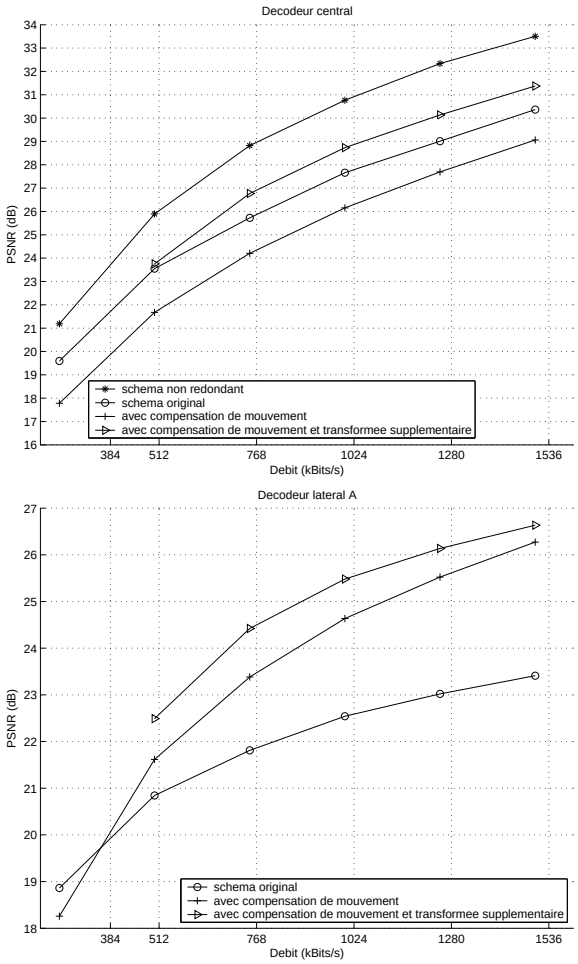

FIG. 2 – Courbes débit-distorsion comparant les différentes stratégies d'introduction de compensation de mouvement dans le schéma D3 pour le décodeur central (gauche) et un décodeur latéral (droite) pour la séquence « MOBILE » (CIF à 30fps)

Après une étude sur la corrélation temporelle des deux sous-bandes d'approximation d'une description, une troisième stratégie s'est avèrée intéressante. Elle consiste à appliquer une nouvelle analyse temporelle de Haar sur ces approximations pour augmenter l'efficacité du codage.

4 Simulations et résultats

Parmi les 4 schémas étudiés, R et D2 donnent de moins bonnes performances que D1 et D3 en terme de puissance du bruit de quantification dans la séquence reconstruite, et nous ne nous attarderons donc pas à présenter leurs résultats. Dans les simulations, nous effectuons $J = 3$ niveaux de décomposition temporelle avec la transformée de Haar compensée en mouvement [17]. Au dernier niveau, nous utilisons l'un des 2 schémas

analysés D1 ou D3. La redondance globale est $1 + 2^{-J}$, les trames de détails pour les résolutions $j < J$ étant distribuées de manière équitable entre les 2 descriptions. Enfin, pour les décodeurs latéraux, une approche par pseudo-inverse a aussi été développée pour la reconstruction optimale des descriptions.

Les schémas proposés ont été testés sur plusieurs séquences CIF à 30fps. Dans un premier temps, on compare les schémas D1 et D3 en faisant une décomposition temporelle sur trois niveaux, compensée en pel entier, seulement sur les deux premiers et, pour l'instant, sans aucune estimation de mouvement sur le niveau ou nous avons introduit la redondance. Les sous-bandes obtenues sont décomposées spatialement avec des ondelettes biorthogonales 9/7 puis codées en utilisant l'algorithme MC-EZBC [18].

Cette comparaison en termes de performances débit-distorsion est présentée dans le tableau 1 pour les décodeurs centraux et latéraux. Comme préconisé par la théorie, le décodeur central du schéma D1 donne de meilleurs résultats que celui de D3. Le décodeur latéral noté A (correspondant à la première description) dans le tableau 1 est identique pour les 2 schémas, alors que pour le décodeur noté B (description 2), une dissymétrie de construction dans le schéma D1 explique l'écart de performances.

TAB. 1 – Comparaison débit-distorsion : PSNR (dB) à différents débits (Kbs), pour les séquences « FOREMAN » et « MOBILE » (CIF à 30fps) sur 3 niveaux de décomposition en ondelettes.

« FOREMAN » schéma D1						
débit	250	500	750	1000	1500	3000
central	29.48	32.19	33.85	34.98	36.85	40.53
latéral A	26.05	27.20	27.78	28.13	28.66	29.51
latéral B	24.32	24.84	25.06	25.16	25.29	25.43
« FOREMAN » schéma D3						
débit	250	500	750	1000	1500	3000
central	29.27	32.01	33.68	34.79	36.68	40.39
latéral A	26.05	27.20	27.78	28.13	28.66	29.51
latéral B	25.26	26.16	26.62	26.88	27.28	27.96
« MOBILE » schéma D1						
débit	250	500	750	1000	1500	3000
central	19.89	22.18	23.54	24.88	26.55	30.61
latéral A	18.96	20.15	20.78	21.24	21.91	23.07
latéral B	18.36	19.24	19.70	19.93	20.37	20.90
« MOBILE » schéma D3						
débit	250	500	750	1000	1500	3000
central	19.70	21.95	23.33	24.61	26.33	30.43
latéral A	18.96	20.15	20.78	21.24	21.91	23.07
latéral B	18.81	19.83	20.35	20.72	21.28	22.23

Dans la suite on augmente l'efficacité du codage en passant à une compensation au $8^{ème}$ de pel sur tous les niveaux temporels et nous comparons les résultats des trois stratégies décrites dans la section 3 sur le schéma D3, qui est plus équilibré. Ces résultats se trouvent dans la figure 2. Nous pouvons remarquer que l'introduction de compensation de mouvement pour le $2^{ème}$ coefficient d'approximation améliore considérablement les performances des décodeurs latéraux, mais diminue les performances du décodeur central. Cet inconvénient est résolu avec la nouvelle transformée de Haar appliquée sur les coefficients d'approximation.

5 Conclusions

Dans cet article, nous avons présenté plusieurs schémas de codage par descriptions multiples temporelles basés sur des décompositions en ondelettes redondantes avec un facteur de redondance réduit. Le facteur de redondance est réglable à l'aide du nombre de niveaux de décomposition temporelle effectué. Nous avons étudié la reconstruction de ces schémas, optimale en termes d'effet du bruit de quantification. Les schémas étudiés ont été appliqués au codage vidéo scalable et robuste et nous avons constaté que, dans ce cas, la compensation de mouvement joue un rôle essentiel.

Références

[1] J. Xu, Z. Xiong, S. Li, and Y. Zhang, "Three-Dimensional Embedded Subband Coding with Optimized Truncation (3D-ESCOT)," *Applied and Computational Harmonic Analysis*, vol. 10, pp. 290–315, 2001.

[2] B.-J. Kim, Z. Xiong, and W.A. Pearlman, "Very low bit-rate embedded video coding with 3-D set partitioning in hierarchical trees (3D-SPIHT)," *IEEE Trans on Circ. and Syst. for Video Tech.*, vol. 8, pp. 1365–1374, 2000.

[3] V.K. Goyal, *Beyond Traditional Transform Coding*, Ph.D. thesis, Univ. California, Berkeley, 1998.

[4] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri, "Multiple-description video coding using motion-compensated temporal prediction," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 12(3), pp. 193–204, March 2002.

[5] Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 12(6), pp. 438–452, June 2002.

[6] Y. Wang, A.R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93(1), pp. 57–70, January 2005.

[7] I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Trans. on Image Proc.*, vol. 12(18), pp. 1211–1225, October 2003.

[8] M. Fumagalli, R. Lancini, and A. Stanzione, "Video transmission over ip by using polyphase downsampling multiple description coding," *Proc. IEEE Int. Conf. on Multimedia and Expo*, August 2001, pp. 1095–1098.

[9] M. Pereira, M. Antonini, and M. Barlaud, "Multiple description coding for internet video streaming," *Proc. IEEE Int. Conf. Image Processing*, 2003.

[10] B. Pesquet-Popescu C. Tillier and M. van der Schaar, "Multiple descriptions scalable video coding," *Proc. of EUSIPCO*, Sept. 2004.

[11] M. van der Schaar and D.S. Turaga, "Multiple description scalable coding using wavelet-based motion compensated temporal filtering," *Proceedings of the IEEE International Conference on Image Processing*, Barcelona, Sept. 2003.

[12] V.K. Goyal, J. Kovačević, and Martin Vetterli, "Quantized frame expansions as source-channel codes for erasure channels," *Proc. IEEE Data Compression Conf.*, pp. 326–335, March 1999.

[13] J. Kovačević, P.L. Dragotti, and V. Goyal, "Filter bank frame expansions with erasures," *IEEE Trans. on Inf. Theory*, vol. 48, no. 6, pp. 1439–1450, June 2002.

[14] R. Motwani and C. Guillemot, "Tree-structured oversampled filterbanks as joint source-channel codes : application to image transmission over erasure channels," *IEEE Transactions on Signal Processing*, vol. 52 (9), pp. 2584 – 2599, Sept. 2004.

[15] T. Petrisor, C. Tillier, B. Pesquet-Popescu, and J.-C. Pesquet, "Temporal multiple description schemes for scalable video using wavelet frames," en préparation.

[16] T. Petrisor, C. Tillier, B. Pesquet-Popescu, and J.-C. Pesquet, "Redundant multiresolution analysis for multiple description video coding," *International Workshop on Multimedia Signal Processing, Sienne, Italie*, 2004.

[17] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Salt Lake City, UT, May 2001.

[18] "3D MC-EZBC software package," disponible sur MPEG CVS.